PLISKA
STUDIA MATHEMATICA
BULGARICA

ПЛИСКА
БЪЛГАРСКИ
МАТЕМАТИЧЕСКИ
СТУДИИ

Pliska Stud. Math. Bulgar. **22** (2013), 5–24

PLISKA
STUDIA MATHEMATICA
BULGARICA

A COMPUTATIONAL APPROACH FOR THE STATISTICAL ESTIMATION OF DISCRETE TIME BRANCHING PROCESSES WITH IMMIGRATION

Dimitar Atanasov, Vessela Stoimenova*

Abstract. It is well known that the estimation of the parameters of branching processes (BP) as an important issue used for studying and predicting their behavior needs lots of energy consuming work and faces many computational difficulties. The task is even more complicated in the presence of outliers – "wrong", "untypical" or "contaminated" data, which require a different statistical approach. The existing asymptotic results for the classical estimators can be combined with a generic method for constructing robust estimators, based on the trimmed likelihood and called weighted least trimmed estimators (WLTE). Despite the computational intensity of the procedure it gives well interpretable results even in the case of minor a priori satisfied asymptotic requirements. In the paper we explain the main outlines of this routine and show some classical estimators and their robust modifications in the important class of discrete-time branching processes with immigration. We present a software package for MATLAB for simulation, plotting and estimation of the process parameters. The package is available on the Internet, under the GNU License.

---

*The research was partially supported by appropriated state funds for research allocated to Sofia University (contract No 125/2012), Bulgaria.

2010 *Mathematics Subject Classification:* 60J80

Key words: branching processes, immigration, estimation, statistical software

1. Introduction. The present paper considers some aspects of the simulation, visualization and statistical estimation of the parameters of two related classes of discrete time branching processes: the classes of the Bienaymé-Galton-Watson processes with an increasing random number of ancestors (BPR processes) and of the branching processes with immigration (BPI processes).

Inspired from real situations requiring the modeling and the study of cell proliferation Yakovlev and Yanev [25] showed the usefulness of the branching processes with a large and often random number of ancestors. They have studied these processes also in applications to nuclear chain reactions. The statistical inference for Bienaymé-Galton-Watson processes with an increasing random number of ancestors (BPR processes) was introduced and developed by Yanev [27] and Dion and Yanev [3], [4], [5], [6]. Robustified versions in the sense of the weighted and trimmed likelihood of the classical estimators are proposed by Stoimenova, Atanasov, Yanev [17], [18]. In the class of the power series offspring distributions some topics of the parametric estimation are considered in [19] and of the robust parametric estimation – in [16].

The first model of branching process with immigration was proposed by Sevastyanov in 1957 [15]. He investigated a Markov branching process with immigration occurring in the moments of a homogeneous Poisson process. Since then there are interesting generalizations of the Sevastyanov's model. In [13] a Markov branching processes is considered where the moments of immigration form a time nonhomogeneous Poisson process.

The statistical inference for branching processes with immigration (BPI process) is considered in a huge number of sources. The reader is referred to the papers [11], [12],[10], [28], [24], [22], [23] and others. In these papers the asymptotic properties of the BPI process are studied as well as the nonparametric maximum likelihood, the conditional least squares and the weighted conditional least squares estimation are introduced. The estimation of the parameters in BPI processes can be considered from another point of view using the relationship between the BPI and the BPR processes [5]. The effectiveness of the estimators of Dion and Yanev of the individual and immigration mean in BPI processes, based on the relationship to the BPR processes, is studied in [1], [2], where also their robust modifications are proposed.

These sources indicate that in addition to the theoretical studies related to the statistical estimation of branching stochastic processes their implementation is deeply dependent on the availability of powerful computational tools for simulation, visualization and estimation of the simulated or real data. As to the robust modifications of the classical estimators, their calculation is impossible without a statistical package that performs the computationally intensive procedures. In this work we present such a system for simulation and estimation, written in Matlab and available free on the Internet under the GNU license.

Within the present paper under robustness we mean weighted and trimmed likelihood (WLT(k)) estimators, defined independently by Vandev and Neykov [21] and Hadi and Luceño [7], based on the principle of the maximum likelihood estimation and applying the basic idea of the Least Trimmed Squares (LTS) estimators of Rousseeuw [14] using appropriate weights. For the study of the breakdown properties of the estimators in the above cited papers the d-fullness technique is used, proposed and developed by Vandev [20] and Vandev and Neykov [21].

2. Two related processes – the process with immigration and the process with a random number of ancestors.

2.1. The Branching process with a random number of ancestors. This process (denoted BPR) may be considered as a Bienayme-Galton-Watson process or as a multitype branching process with one type of particles, which starts with a random number of ancestors $Z(0,n)$, where an additional index n is introduced, allowing for the increase of the initial number of particles as $n \to \infty$. One can think of n as a function of time t or conversely, of the time t – as a function of n. The asymptotic behaviour of the process and its statistical properties depends on the rate of convergence of n/t [5]. Note that usually when the process with one type of particles has to be considered as a BPR process, the initial number of particles $Z_1(0)$ becomes a function of n: $Z_1(0) = N(n)$.

The BPR process can be defined as follows.

Assume there exists on some probability space a set of i.i.d. r.v. $\{\xi(t,l,n)\}$, distributed like the r.v. ξ, with values in the set of nonnegative integers $\{0,1,2,\dots\}$; $\{\xi(t,l,n),\ i \in N\}$ are independent of $Z(0,n)$ such that $Z(t+1,n) = \sum_{l=1}^{Z(t,n)} \xi(t,l,n)$, if $Z(t,n) > 0$, $t = 1,2,\dots$. Then for each $n = 1,2,\dots$ $\mathbf{Z(n)} = \{Z(t,n), t = 0,1,\dots\}$ is a Bienayme-Galton-Watson process having a random number of ancestors $Z(0,n) \geq 1$.

The nonparametric mle (the Harris estimator) [9] of the offspring mean $m = E\xi$ is given by the formula

$$\widehat{m}_t(n) = \frac{\sum_{j=1}^{t} Z(j,n)}{\sum_{j=0}^{t-1} Z(j,n)}. \tag{1}$$

2.2. The discrete-time branching process with immigration. Dion and Yanev [5] show the duality between the BPR processes and the branching processes with immigration (BPI), which allows us to transfer results on the statistical inference of these processes.

We remind that the BPI 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. At time t the l-th native particle gives rise only to a random number $\{X(t,l)\}$ of 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 $\{I(t)\}$ 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, which does not become extinct. Due to its importance for applications it is treated separately and is considered in a different way from the multitype model.

Let $\{Y(t)\}_{t=0}^{\infty}$ be a BPI process defined by the recursive formula

$$Y(t) = \begin{cases} \displaystyle\sum_{l=1}^{Y(t-1)} X(t,l) + I(t), & \text{if} \quad Y(t-1) > 0, \\ I(t), & \text{if} \quad Y(t-1) = 0. \end{cases}$$

Without any loss of generality one can assume that $Y_0 = 0$.

Assume further that the offspring mean $m < \infty$. As usually the process is called subcritical if $m < 1$, critical if $m = 1$ and supercritical if $m > 1$.

In the supercritical situation many of the estimators for the classical Bienayme-Galton-Watson process (BGW process) for the individual characteristics can be applied for the BPI process. Concerning the individual mean m, the Harris estimator $\widehat{m}_t = \sum_{j=1}^{t} Y(j) / \sum_{j=0}^{t-1} Y(j)$ is completely applicable. In the subcritical case, however, the estimator $\widehat{m}_t$ is not consistent, therefore other estimators are needed.

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

$$(2)\quad \begin{aligned} \overline{m}_t &= \frac{t\sum\limits_{k=1}^{t} Y(k)Y(k-1) - \sum\limits_{k=1}^{t} Y(k)\sum\limits_{k=1}^{t} Y(k-1)}{t\sum\limits_{k=1}^{t} Y^2(k-1) - \left(\sum\limits_{k=1}^{t} Y(k-1)\right)^2} \\ \overline{\lambda}_t &= \frac{t\sum\limits_{k=1}^{t} Y(k)\sum\limits_{k=1}^{t} Y^2(k-1) - \sum\limits_{k=1}^{t} Y(k-1)\sum\limits_{k=1}^{t} Y(k)Y(k-1)}{t\sum\limits_{k=1}^{t} Y^2(k-1) - \left(\sum\limits_{k=1}^{t} Y(k-1)\right)^2}. \end{aligned}$$

In [24] it is noted that the conditional least squares estimators are not satisfactory for the reasons that the estimator $\overline{m}_t$ has a larger asymptotic variance than the m.l.e. of Harris in the supercritical case and the estimator $\overline{\lambda}_t$ is not a consistent estimator for λ in the supercritical case. To avoid these disadvantages Wei and Winnicki [22] proposed to use the weighted conditional least squared estimators

$$(3)\quad \begin{aligned} \widetilde{m}_t &= \frac{\sum\limits_{k=1}^{t} Y(k)\sum\limits_{k=1}^{t} \frac{1}{Y(k-1)+1} - t\sum\limits_{k=1}^{t} \frac{Y(k)}{Y(k-1)+1}}{\sum\limits_{k=1}^{t} (Y(k-1)+1)\sum\limits_{k=1}^{t} \frac{1}{Y(k-1)+1} - t^2} \\ \widetilde{\lambda}_t &= \frac{\sum\limits_{k=1}^{t} Y(k-1)\sum\limits_{k=1}^{t} \frac{Y(k)}{Y(k-1)+1} - \sum\limits_{k=1}^{t} Y(k)\sum\limits_{k=1}^{t} \frac{Y(k-1)}{Y(k-1)+1}}{\sum\limits_{k=1}^{t} (Y(k-1)+1)\sum\limits_{k=1}^{t} \frac{1}{Y(k-1)+1} - t^2}. \end{aligned}$$

In [24] it is noted that in the supercritical case $\widetilde{m}_t$ is a more efficient estimator than $\overline{m}_t$ in the sense of achieving a lower asymptotic variance. However the estimator $\widetilde{\lambda}_t$ is not consistent. In fact it is proved in [22] that the parameters λ and b^2 do not have consistent estimators in the supercritical case.

2.3. Transferring results. As noted in [5], the estimation theory for the BPR processes can be transferred to the BPI processes without taking account of the criticality of the processes. This can be achieved in the following way.

One may consider the partial tree 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, $n, t =$

$0, 1, 2, \ldots$. 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 BPR process.

The Dion-Yanev estimator of the offspring mean m, obtained on the basis of the transformed into a BPR process immigration process, is the Harris type estimator $\widehat{\widehat{m}}_t(n) = \frac{\sum_{i=1}^{t} Z(i,n)}{\sum_{i=0}^{t-1} Z(i,n)}$.

The estimators of the immigration component λ of Dion and Yanev when the offspring mean m is known or unknown are correspondingly $\widehat{\widehat{\lambda}}_t(n) = \frac{Z(t,n)}{nm^t}$ and $\widetilde{\widetilde{\lambda}}_t(n) = \frac{Z(t,n)}{n(\widehat{\widehat{m}}_t(n))^t}$.

In [2] a quantitative comparison of the offspring mean estimators is described and is shown that the estimators of Heyde-Seneta overestimate the real parameter values in the subcritical case and underestimate them in the supercritical case. From simulated trajectories the estimator of Wei-Winnicki and its relative difference are calculated (f.e. in the case of the immigration mean, n sample paths of the immigration process $\mathbf{Y_i}$, $i = 1, 2 \ldots, n$, and corresponding immigration mean estimator $\lambda_i(\mathbf{Y_i})$, the relative difference is defined as $D_n = \frac{1}{n-1} \sum_{i=1}^{n} [\lambda_i(\mathbf{Y_t}) - \lambda])$ and is shown that this estimator is not appropriate for the subcritical case and for the supercritical case the estimation is very unstable. 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 BPR process is the most stable among all estimators we have just considered. 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 by the fact, that the transformation BPI – BPR requires more information about the immigration process (not only the generation sizes). It is noted in Dion and Yanev [6] that in general the knowledge of $\{Z(0,n), \ldots, Z(t,n)\}$ would seem to be asymptotically equivalent to $\left\{ [Y(k)]_{k=0}^{t+n}, \sum_{k=0}^{n} I(k) \right\}$ as $n, t \longrightarrow \infty$ on the set of the nonextinction.

3. The problem of the robust modification – from mathematical to computation point of view. First note that if one has a BPR process, starting with a large number of ancestors, it can be split into several independent subprocesses, thus obtaining observations over a set of sample paths of the process (for details see [26]).

In [1] the estimation of the immigration component and in [2] - of the individual mean in the BPI processes in the presence of outliers is considered.

3.1. The mathematical model. The idea can be summarized in the following steps [1] (we explain it in the case of the immigration mean, the case of the offspring mean is analogous):

- Let us suppose that we have a set of r sample paths of the entire family tree of a branching process with immigration. This means that we are able to observe also the equivalent BGW process, starting with a random number of ancestors (BPR process). Let us consider the set $\mathbf{Z} = \{\mathbf{Z^{(1)}}(\mathbf{n}), \cdots, \mathbf{Z^{(r)}}(\mathbf{n})\}$, where $\{\mathbf{Z^{(i)}}(\mathbf{n})\}$ is a single realization of a BPR process (equivalent to a corresponding realization of the process with immigration) with the same parameters n and t.
- Using this set and the estimators of Dion and Yanev over each realization we obtain a number of values for the offspring distribution.

  Let

  $$(4)\quad \widehat{\widehat{\lambda}}_t^{(i)}(n) = \frac{Z^{(i)}(t,n)}{nm^t} \quad \text{and} \quad (5)\quad \widetilde{\widetilde{\lambda}}_t^{(i)}(n) = \frac{Z^{(i)}(t,n)}{n(\widehat{\widehat{m}}_t^{(i)}(n))^t},$$

  $i = 1, 2, \ldots, r$, be the estimators of Dion and Yanev for the immigration mean λ for the sample path $\mathbf{Z^{(i)}}(\mathbf{n})$ when the offspring mean m is correspondingly known or unknown. Here $\widehat{\widehat{m}}_t^{(i)}(n)$ is the Harris estimator (1) for the unknown individual mean based on this sample path.
- Under the conditions given in [5] these values are asymptotically normally distributed: Let the function $Est(\mathbf{Z^{(i)}}(\mathbf{n}), \lambda)$ be a transformation of the estimator (4) or (5) of the parameter λ, which gives the asymptotic normality:

  $$Est(\mathbf{Z^{(i)}}(\mathbf{n}), \lambda) \xrightarrow{d} N(0, a\lambda + c), t \to \infty,$$

  where a and b are appropriate constants. Depending on the criticality of the process $Est(\mathbf{Z^{(i)}}(\mathbf{n}), \lambda)$ is expressed in a different way. For instance, if one considers the estimator (4) in the subcritical case $m < 1$

  $$Est(Z^{(i)}(n), \lambda) = \sqrt{nm^t}(\overline{\lambda}_t^{(i)}(n) - \lambda) \xrightarrow{d} N(0, \lambda).$$

For details see [1].

- If the conditions for the asymptotic normality from the previous point are not satisfied, the estimated value is far from the real value of the immigration mean. The aim is to apply the weighted and trimmed likelihood in order to eliminate the cases, which do not satisfy these conditions, and to obtain estimators of the immigration component, closer to the real value. This is achieved via the WLT(k) estimator

$$\widehat{\Lambda}_t^T(n) = \underset{\lambda>0}{\text{argmin}} \sum_{i=1}^{T} -w_i f(Est(\mathbf{Z}^{(\nu(\mathbf{i}))}(\mathbf{n}), \lambda)), \tag{6}$$

where T is properly chosen trimming factor, $0 < T \geq r/2$, $f(x)$ is the log-density of the asymptotically normal distribution of the used Dion - Yanev estimators (4) and (5) , ν is a permutation of the indices, such that

$$f(Est(\mathbf{Z}^{(\nu(\mathbf{1}))}(\mathbf{n}), \lambda)) \geq f(Est(\mathbf{Z}^{(\nu(\mathbf{2}))}(\mathbf{n}), \lambda)) \geq \cdots \geq f(Est(\mathbf{Z}^{(\nu(\mathbf{T}))}(\mathbf{n}), \lambda))$$

and λ is the unknown immigration mean. The weights w_i are nonnegative and at least T of them are strictly positive.

Thus in the likelihood are included only the "most probable" estimators with the largest value of the density function.

- The robust properties of an estimator can be studied by the measure of robustness, called breakdown point (BP). We adopt the definition of a finite sample breakdown point of Hampel at al. [8]. For a given estimator S it is defined as $BPoint(S) = \frac{1}{n} \max\{m : \sup \|S(X_m)\| < \infty\}$, where X_m is a sample, obtained from the sample X by replacing any m of the observations by arbitrary values. Vandev and Neykov in 1998 [21] and Vandev in 1993 [20] propose a method for determining the value of the $BPoint$ of a statistical estimator based on the index of fullness of the set of log-density functions in the log-likelihood. Applying this concept, in [1] the lower bound for the breakdown point of the estimator $\widehat{\Lambda}_t^T(n)$ is given according to the criticality of the process.

3.2. The algorithmic approach. The algorithm [18] for the case of the BPI process can be described concisely in the following way:

1. Setting the initial value for the unknown parameter $\lambda = \lambda_0 \in (0, \infty)$.

2. Sorting the observations according to the log-density function f at the current value of the unknown parameter :

$$f_{\nu(1)}(Est(\mathbf{Z}^{(\nu(1))}(n),\lambda)) \leq \cdots \leq f_{\nu(r)}(Est(\mathbf{Z}^{(\nu(r))}(n),\lambda))$$

3. Finding the value λ which satisfies (6).

4. If the exit conditions are not satisfied than go back to 2.

The optimization algorithm in 3. is implemented in MATLAB and is based on Golden Section search, parabolic interpolation and Nelder-Mead simplex (direct search) method.

4. The *Immigration processes* part of the package: some illustrative examples and graphics. The presented package contains functions for the calculation of estimators in the BPR, BPI and discrete time multitype processes (BPM processes). Additionally functions for the simulation of these processes are provided. The functionality is based on the information about the entire family trees. The package is available at `http://www.ir-statistics.net/index.cgi/software-bp-tools` licensed under the CC-GNU LGPL. The package covers more than 35 functions for manipulation and estimation of different types of branching processes. The complete list of functions is available at the project site.

In this section we describe the main functionality of the package via several illustrative examples.

The branching process in the package is presented as a vector of parent pointers Z. The i-th column of Z represents the information about the i-th particle of the process. The value $Z(1,i)$ contains a pointer (a column in Z) of the parent particle. For the root particle the parent pointer is 0. The second row $Z(2,:)$ gives us information about the generation of the particle. $Z(3,i)$ represents the type of the particle, for example: 1 – alive; 2 – dead; 3 – immigrant. The first three rows of the process Z are obligatory. If there is some additional information available, it can be included in the next rows of the matrix Z, for instance the lifespan of the particle can be added as a fourth row of Z.

Example 1. The matrix

```
Z1 =[ 0    1    1    2    2    2    3    3    0    9    0   11
      3    4    4    5    5    5    5    5    3    4    4    5
      3    1    1    1    1    1    1    1    3    2    3    1]
```

defines a process with 3 generations, presented on Figure 1. Its ninth column shows that the 9-th particle lives in the third generation, has no parent and is immigrant, while the last column shows a native particle (the 12-th particle in the trajectory), which lives in the 5-th generation and has the immigrant parent from the 11-th row.

Figure 1 is plotted with one of the most useful functions in the package – the *bp_treeplot* function, which produces the tree plot for the family tree of the process with all the information about the particles as type and lifespan.

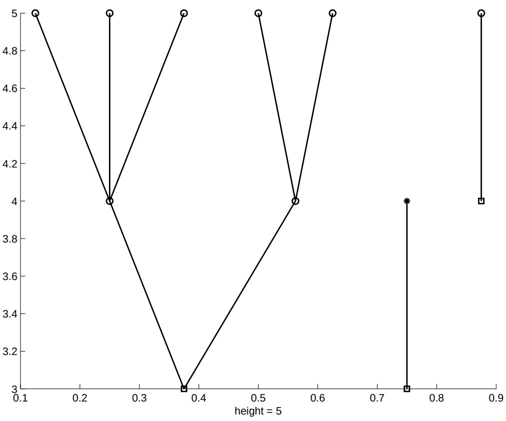

Fig. 1. The BPI process from Example 1. ♢

The functionality of the package allows us to use three types of simulated processes. They can be generated by the functions *gen_bp* in the case of BPR, *gen_bp_immigr* in the case of BPI and *gen_multi_type_bp* for the BPM processes.

Example 2. A simulation of a random process having 12 generations, $Bi(5, 0.2)$ offspring distribution (critical process) and $Po(3)$ distribution of the immigration component can be generated using the *Offspring* and *Immigration* structures responsible for the parameters of the offspring and immigration distribution with the corresponding fields. The result is presented on Figure 2.

```
Immigration = dist: 'poiss'
              par1: 3.000
Offspring   = dist: 'bino'
              par1: 5
              par2: 0.200
Z2 = gen_bp_immigr(12,Offspring,Immigration);
bp_treeplot(Z2,Z2(3,:))
```

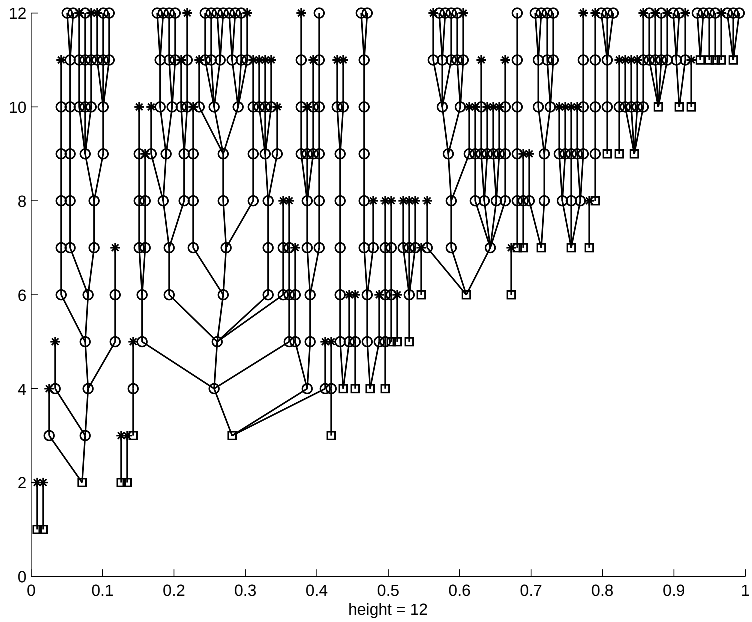

Fig. 2. The simulated BPI process from Example 2. ◊

Example 3. Some estimates of the parameters of the process $Z2$ from *Example 1* are given in Table 1.

Table 1. Estimated process parameters

Description	Command	Result	True value
Harris estimator of the offspring mean	bp_harris_est(Z2)	1.0000	1.1695
MLE of the immigration mean	bp_immigr_est(Z2)	3.0000	2.9167
The individual variance estimator with mean 1	bp_ind_var(Z2,1)	0.8	1.6416

The offspring and immigration mean estimators of Heyde-Seneta (HS) and Wey-Winnicki (WW) for this process are calculated via the functions *bp_heyde_seneta_est(Z2)* and *bp_wei_winnicki_est(Z2)* and are given in Table 2 for $k = 1, 2, \ldots, 7$ consecutive generations of the process.

The aggregate properties such as the number of particles for each generation or the number of particles with a given number of children can be found using the functions *bp_count_popul* and *bp_count_children*.

Example 4. The result of the command line

Table 2. Estimated parameters via the conditional least squares estimators. $\Diamond$

Estimator	based on k generations					
	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
HS offspring mean	0	1.5000	1.6429	2.1923	1.4831	1.1936
HS immigr. mean	0	0	-0.0714	-0.7308	0.7712	1.6396
WW offspring mean	0	0.3333	0.2874	0.1506	-0.0909	0.2675
WW immigr. mean	0	0.3333	0.9310	2.6566	3.3091	3.3130
Estimator	based on k generations					
	$k=7$	$k=8$	$k=9$	$k=10$	$k=11$	$k=12$
HS offspring mean	1.3506	1.0361	1.1324	1.0684	1.0395	0.9803
HS immigr. mean	1.0523	2.8667	2.3382	2.8666	3.1287	3.7215
WW offspring mean	0.4860	0.6062	0.7606	0.8757	0.9666	1.0065
WW immigr. mean	4.6967	3.8051	4.7926	4.5871	4.4559	3.8294

```
bp_count_children(Z2)
```

is given in Table 3:

Table 3. The number of particles with k children, $k = 0, 1, 2, 3, 4$

Number of children	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
Number of particles	76	100	41	15	4

Analogously the result from the command line

```
bp_count_popul(Z2)
```

is summarized in Table 4:

Table 4. The number of particles in the k-th generation, $k = 0, 1, 2, \ldots, 11$. $\Diamond$

Generation number	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Number of particles	2	3	5	11	16	19
Generation number	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	$k=11$
Number of particles	29	28	38	41	44	42

Using *bgwi2bgwr* a BPR process, equivalent to the immigration process can be obtained.

Example 5. The command line

```
Z3 = bgwi2bgwr(Z2); bp_treeplot(Z3)
```

produces the equivalent BPR process on Figure 3 for the family tree on Figure 2.

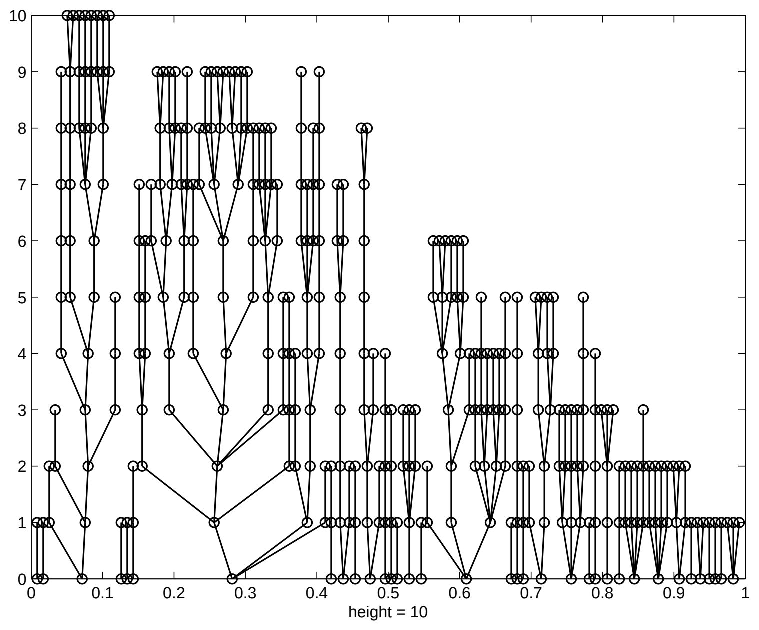

Fig. 3. The equivalent BPR process of the BPI process from Example 1. $\Diamond$

The function *bp_immigr_dy_est* can be used to obtain the Dion-Yanev estimator $\widehat{\widehat{\lambda}}_t(n)$ for a BPI process with a given offspring mean. The result is in a matrix form with rows representing the values of the estimator for different n and the columns - the values of the estimator for different t:

Example 6. The Dion-Yanev estimates for different values of n and t for the BPI process from *Example 2.* are given in the next Table 5, where the zeros mean that the estimators can not be calculated due to lack of observations $(1 \geq n \geq t)$. Note that according to [5] and [6] when $m = 1$, n and t have to be chosen to converge as $\dfrac{t}{n} \to 0$ and that the reasonable estimates are "round" the diagonal above the zeros.

Table 5. The Dion-Yanev estimators for the immigration mean with known offspring mean 1 for different values of n and t. $\Diamond$

n/t	1	2	3	4	5	6	7	8	9
1	35.0000	40.9091	27.2727	22.5394	16.3923	15.5230	12.9829	9.7500	9.3301
2	17.5000	20.4545	13.6364	11.2697	8.1962	7.7615	6.4915	4.8750	4.6651
3	11.6667	13.6364	9.0909	7.5131	5.4641	5.1743	4.3276	3.2500	0
4	8.7500	10.2273	6.8182	5.6349	4.0981	3.8808	3.2457	0	0
5	7.0000	8.1818	5.4545	4.5079	3.2785	3.1046	0	0	0
6	5.8333	6.8182	4.5455	3.7566	2.7321	0	0	0	0
7	5.0000	5.8442	3.8961	3.2199	0	0	0	0	0
8	4.3750	5.1136	3.4091	0	0	0	0	0	0
9	3.8889	4.5455	0	0	0	0	0	0	0
10	3.5000	0	0	0	0	0	0	0	0

As already mentioned, the BPI process is in fact a two-type decomposable BPM process. This fact allows us to use the estimation theory for the BPM process and hence to estimate the individual and immigration distribution of the BPI process.

The generation of multitype processes is based on information about the probabilities for a given set of offspring types. These probabilities are given in a matrix with a special structure. The first column represents the type of the parent particle, the second column is the corresponding probability for the offspring, presented in the rest of the row.

Example 7. Let us consider the BPI process with the probability distribution given in the left hand side (the matrix P) of Table 6. The third row of the matrix P in Table 3 states that the particle of type 1 has 2 children of type 1 and no children of type 2 with probability 0.4. Here the particles of type 1 are the “natives” and of type 2 is the “immigrant”. This matrix should be returned from a user defined function.

Table 6. Generation and estimation of a BPI process as a BPM.

```
P = [1 0.2 1 0
     1 0.2 0 0
     1 0.4 2 0
     1 0.2 3 0
     2 0.7 1 1
     2 0.3 2 1];
```

```
1.0000    0.1899    1.0000         0
1.0000    0.2089         0         0
1.0000    0.4051    2.0000         0
1.0000    0.1962    3.0000         0
2.0000    0.6667    1.0000    1.0000
2.0000    0.3333    2.0000    1.0000
```

A simulation of the process with the probability matrix P, starting with one particle of type 1 and one particle of type 2 is presented on Figure 4.

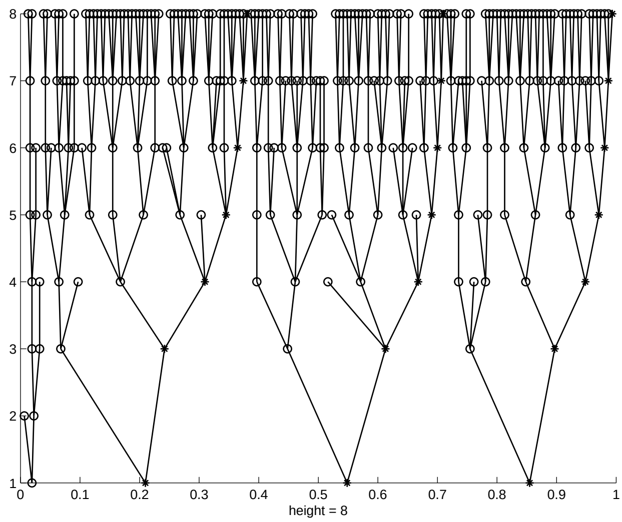

Fig. 4. The BPI process with probability matrix P from Example 6.

Having the family tree one can estimate (*multi_type_est(Z4)*) the individual probabilities via the formula for the relative frequencies $\Delta_i(t,n) = \dfrac{Z_i(t,n)}{\sum_{i=1}^{k} Z_i(t,n)}$ (the fraction of the number of particles of type i and the total number of particles in the moment t in a BPM process with k types of particles: for details see [26]) in the same form as the matrix P. The result is given in the right part of Table 6.

The complete code for generating, plotting and estimation for this example is

```
Z4=gen_multi_type_bp(7,[1 3],@test_multi_type_gen); % test_multi_type_gen returns P
bp_treeplot(Z4,Z4(3,:))
multi_type_est(Z4)
```

4.1. Robust estimation. The robust estimation of the process parameters can be realized using a package for WLT estimation (distributed separately from the same site). The estimation is based on the information from a set of family trees. We introduce the usage of the package with two examples.

Example 8. Let us calculate the robust modification of the Harris estimator with a trimming factor of 80% for a sample path over a BPR process with $Ge(0.2)$ individual distribution, 20 generations and 100 ancestors.

```
Offspring.dist = 'geo';            % Theoretical distribution
Offspring.par1 = 0.2;
Z3 = bp_gen_mt(20,100,Offspring);   % A set of 100 family trees with 10 generations
HarrisEstimator = bp_harris_est(Z3);% The Harris estimator for the trees.
C{1,1} = Z3; % Additional parameters for the normalization of the Harris estimator
             %
             % Calculate the WLTE(80) robust estimator and the std. err. in the
             % interval (0,5) with a starting point of 0.5, weights all equal to
             % one, and level of convergence  1e-4
[RobustEstimator, stdErr] = wlte( 80 , HarrisEstimator , 0, 5, 0.5, 'r_lklh',...
                                  [],'ones', 1e-4 , C)
```

Here the function *r_lklh*, which is included in the package, gives the log-likelihood function of the asymptotic distribution of the Harris estimator:

```
function res=r_lklh(m,x,th_add,C)
Z = C{1,1}(1,C{2,1});
res = Z*(x-m).^2;
```

Example 9. Let us consider the robust estimation of the immigration mean in a single realization of a BPI process using the robust modified Dion-Yanev estimator $\widetilde{\widetilde{\lambda}}_t(n)$ based on the Harris estimator $\widehat{\widehat{m}}_t(n)$, which can be described as follows. First we generate a BPI process, then we transform it into a BPR process, which is split into a set of independent subprocesses with one ancestor. For each realization from this set the estimates are calculated and later are robust modified.

```
O.dist = 'hyge';                % HG(10,5,3) offspring distribution
O.par1 = 10;
O.par2 = 5;
O.par3 = 3;
I.dist = 'bino';                % Bi(4,0.3) immigration distribution
I.par1 = 4;
I.par2 = 0.3;
Z=gen_bp_immigr(30,D,I);        % Generating a sample path with 30 generations
                                % Calculating the Dion-Yanev estimator
DYEstimator = bp_immigr_dy_est(Z, bp_harris_est(Z));
                       % Calculating the  robust estimator and the std. error
                       % with trimming factor 70% of the number of immigrants
[RobustEstimator StdError] = rbs_immigr_dy_est(DYEstimator,0.7,...
                                   O.par1, O.par1, I.par1*I.par2*(1-I.par2));
```

4.2. GUI. Almost all of the functionality of the package is available through a graphical user interface application (distributed separately) called BP_Engine. The main window of the application is presented on Figure 5.

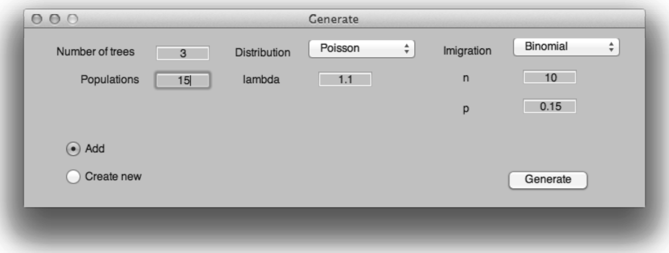

Fig. 5. The main window of BP_Engine

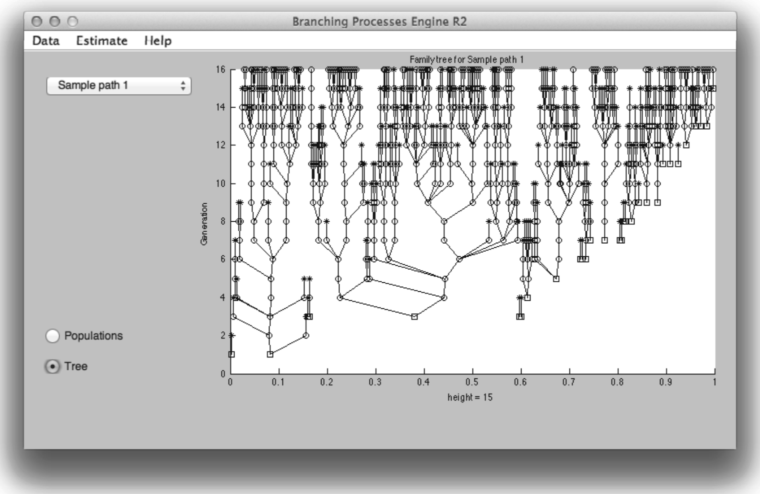

Fig. 6. The visualization window

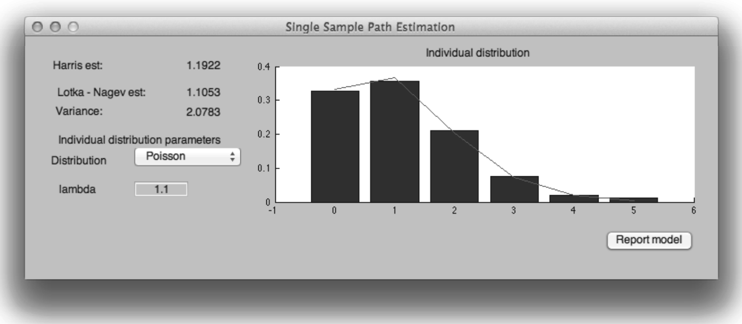

Fig. 7. The offspring characteristics estimation window

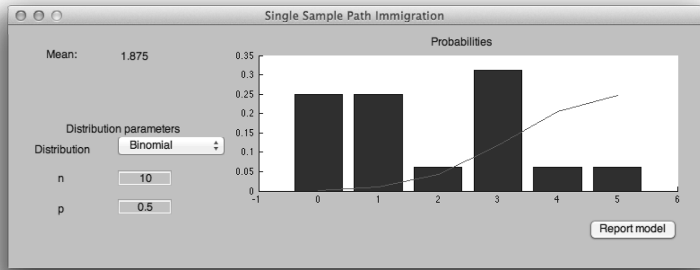

Fig. 8. The immigration characteristics estimation window

The menu of the application gives access to the main functions of the package. The *Data* menu includes different features for generating, loading and manipulating information about the sample paths of the process. The *Estimate* menu item allows access to the different types of estimators, supported by the package. For example, on Figure 6 the estimation of the immigration mean as well as the fit of the immigration distribution is presented.

Acknowledgements. This research is part of the study on the statistical estimation and visualization of the three types of discrete time branching processes which started with the coauthor professor Nikolay Yanev from the Institute of Mathematics and Informatics, Bulgarian Academy of Science, to whom the authors owe the interesting idea and the beneficial advice.

REFERENCES

[1] D. Atanasov, V. Stoimenova, N. M. Estimators in branching processes with immigration. *Pliska Stud. Math. Bulgar.* **18** (2007), 19–40.

[2] D. Atanasov, V. Stoimenova, N. Offspring mean estimators in branching processes with immigration. *Pliska Stud. Math. Bulgar.* **19** (2009), 69–82.

[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** (1991), No 3, 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** (1992), No 11, 27–30.

[5] J. P. Dion, N. M. Yanev. Statistical inference for branching processes with an increasing number of ancestors. *J. Statistical Planning & 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] A. Hadi, A. Luceno. Maximum trimmed likelihood estimators: a unified approach, examples and algorithms. *Comput. Statist. and Data Analysis.* **25** (1997), 251–272.

[8] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel. Robust Statistics: The Approach Based on Influence Functions. New York, John Wiley and Sons, 1986.

[9] T. E. Harris. The Theory of Branching Processes. Berlin , Springer-Verlag, 1963.

[10] C. C. Heyde. On estimating the variance of the offspring distribution in a simple branching process. *Adv. Appl. Prob.* **3** (1974), 421–433.

[11] C. C. Heyde, J. P. Leslie. Improved classical limit analogues for Galton–Watson process with or without immigration. *Bull. Austral. Math. Soc.* **5** (1971), 145–156.

[12] C. C. Heyde, E. Seneta. Estimation theory for growth and immigration rates in a multiplicative process. *J. Appl. Prob.* **9** (1972), 235–256.

[13] O. Hyrien, K. Mitov, N. Yanev. Limit Theorems for Supercritical Markov Branching Processes with Non-Homogeneous Poisson Immigration. *C. R. Acad. Bulg. Sci.* bf 65 (2013), No 4, 485-492.

[14] P. Rousseeuw. Least median of squares. *J. Amer. Statist. Assoc.* **79** (1984), 871–880.

[15] B. A. Sevastyanov. Limit theorems for Branching random Processes of Special Type. *Theory Prob. Appl.* **2** (1957), 339–348 (in Russian).

[16] V. Stoimenova. Robust parametric estimation of branching processes with random number of ancestors. *Serdica Math. J.* **31** (2005), No 3, 243–262.

[17] V. Stoimenova, D. Atanasov, N. Yanev. Robust estimation and simulation of branching processes. *C. R. Acad. Bulg. Sci.* **57** (2004), No 5, 19–22.

[18] V. Stoimenova, D. Atanasov, N. Yanev. Simulation and robust modification of estimates in branching processes. *Pliska Stud. Math. Bulgar.* **16** (2004), 259–271.

[19] V. Stoimenova, N. Yanev. Parametric estimation in branching processes with an increasing random number of ancestors. *Pliska Stud. Math. Bulgar.* **17** (2005), 295–312.

[20] D. L. Vandev. A note on breakdown point of the least median of squares and least trimmed estimators. *Statistics and Probability Letters* **16** (1993), 117–119.

[21] D. L. Vandev, N. M. Neykov. About regression estimators with high breakdown point. *Statistics*, **32** (1998), 111–129.

[22] C. Z. Wei, J. Winnicki. Some asymptotic results of branching processes with immigration. *Ann. Statist.* **31** (1989), 261–282.

[23] C. Z. Wei, J. Winnicki. Estimations of the means in the branching processes with immigration. *Ann. Statist.* **18** (1990), 1757–1773.

[24] J. Winnicki. Estimation theory of the branching processes with immigration. *Contemporary Mathematics* **80** (1988), 301–323.

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

[26] A. Yu. Yakovlev, N. M. Yanev. Relative frequencies in multitype branching processes. *Ann. Appl. Probab.*, **19** (2009), No 1, 1–14.

[27] N. M. Yanev. On the statistics of branching processes. *Theor. Prob. Appl.* **20** (1975), 623–633.

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

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

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