COMPUTATIONAL MODELLING OF METASTASIS DEVELOPMENT IN CANCER EVOLUTION

1

Maroussia Slavtchova–Bojkova Sofia University and Institute of Mathematics and Informatics bojkova@fmi.uni-sofia.bg "Workshop on branching processes and applications (WBPA'2018)" 10–13 April 2018, Badajoz, SPAIN Collaborators for this research Plamen Trayanov, PhD, 2016 Stoyan Dimitrov, PhD student Kaloyan Vitanov, MSc degree student, 2017 1. Slavtchova–Bojkova, M. (2016). On two–type decomposable branching processes in continuous time and time to escape extinction. Branching Processes and their Applications. Lecture Notes in Statistics, Springer, 219, 319–329.

2. Slavtchova–Bojkova, M., Trayanov, P., Dimitrov, S. (2017). Branching processes in continuous time as models of mutations: Computational approaches and algorithms. http://dx.doi.org/10.1016/j.csda.2016.12.013.

3. Vitanov, K., Slavtchova–Bojkova, M. (2017). Multi–type branching processes in continuous time as models of cancer. Annuare de l'Universite de Sofia "St. Kl. Ohridski", (accepted)

4. Slavtchova–Bojkova, M., Vitanov, K. (2018). Computational modelling of the evolution of cancer by multi–type branching processes, (in preparation)

As it is stated in Durrett, 2015:

"Metastasis, the spread of cancer to distant organs, is the most common cause of death for cancer patients. It is very complex process: cell must enter the blood stream (intravasation), survive the trip through the circulatory system, leave the blood stream at its destination (extravasation), and survive in an alien environment, e.g., cells from the breast tissue living in bone." We have two main consequences from the biological nature of cancer:

we can have more than one type of metastasis in the human organism, possibly after local elimination of the initial tumor followed by proper medical treatment;

arriving in a completely different environment the cancerous cell may change its characteristics concerning lifespan and division;

stemming from the multistage theory of cancer, our main idea in this study is to introduce a novel multi-type decomposable branching process model of a cell population, with n types of cells, n > 2.



1 Main results

(1) the basic functional equations for the probability generating function (p.g.f.) of the process itself and of both the number of mutations occurred up to time t and the number of mutations to the escape type in the whole process,

(2) probabilities of ultimate extinction of the process and those of extinction before given moment t,

(3) properties of the time until occurrence of the first "successful" mutant, initiating a non-extincting Bellman- Harris branching process,

(4) Immediate risk of escaping extinction,

(5) an integral equation for the distribution of the event that jointly the first "successful" mutant does not appear and no cells of types $1, 2, 3, \ldots, n$ exist at time t,

(6) numerical schemes for calculating the introduced integral equations,

(7) calculations, using the given numerical algorithms, in order to demonstrate the behaviour of the new model under different setups.

2 Integral equations

2.1 Integral equations for p.g.f. of the process.

We denote p.g.f. of the multi-type BHBP under consideration, starting with one cell of type i, $0 \le i \le n$, with:

 $F_i(t; s_0, s_1, \dots, s_n) = E(s_0^{Z^0(t)} s_1^{Z^1(t)} \dots s_n^{Z^n(t)} | Z^i(0) = 1, Z^j(0) = 0, j \neq i),$

where $|s_m| \le 1, 0 \le m \le n$. Let us note that because of the inability of the offspring of type 0 cells to change its type, it follows that

$$F_0(t;s_0,s_1,...,s_n) = F_0(t;s_0) = E(s_0^{Z^0(t)} | Z^0(0) = 1, Z^j(0) = 0, j \neq 0).$$

Denote (f_i, G_i) is the p.g.f. of offspring and lifespan d.f. of a cell of type $i, i = 0, ..., n, u_{ij}$ are mutation probabilities, $\sum_{j=0}^{\infty} u_{ij} = 1$ for each i = 1, ..., n and $u_{00} = 1$.

1. For type 0:

$$F_0(t; s_0, s_1, ..., s_n) = F_0(t; s_0)$$

= $s_0(1 - G_0(t)) + \int_0^t f_0(F_0(t - y; s_0)) dG_0(y).$

2. For types $1 \leq i \leq n$:

$$F_{i}(t; s_{0}, s_{1}, ..., s_{n}) = s_{i}(1 - G_{i}(t)) + \int_{0}^{t} f_{i}[u_{i0}F_{0}(t - y; s_{0}) + u_{i1}F_{1}(t - y; s_{0}, s_{1}, ..., s_{n}) + ... + u_{in}F_{n}(t - y; s_{0}, s_{1}, ..., s_{n})]dG_{i}(y),$$

where $F_i(0; s_0, s_1, \ldots, s_n) = s_i, |s_i| \le 1, i = 1, 2, \ldots, n$. The proof follows by use of multynomial distribution and could be seen in Vitanov and Slavtchova-Bojkova (2017, accepted).

2.2 Number of occurred mutants

- "mutation" whenever a daughter cell has a different type than the mother cell;
- "mutant" only when a daughter cell is of type 0 and has a mother cell from any of the subcritical types;
- "trivial mutant" whenever a daughter cell with a subcritical type has a mother with different subcritical type.

Let us denote by $I_i(t), 1 \le i \le n$ the random variable (r.v.) being the number of mutants that have so far occurred until moment t in a main model process, starting with a single cell of type i. The p.g.f. of $I_i(t), 1 \le i \le n$ will be denoted by:

$$h_{I_i(t)}(s) = E(s^{I_i(t)}), |s| \leq 1.$$

Let us denote by $I_i, 1 \leq i \leq n$ the r.v. being the number of mutants that have occurred during the whole main model process, in a process starting with a single cell of type *i*. The p.g.f. of $I_i, 1 \leq i \leq n$ will be denoted by:

$$h_{I_i}(s) = E(s^{I_i}), |s| \le 1.$$

Using again the assumption of independence in cell reproduction, in Vitanov and Slavtchova-Bojkova (2017, accepted) the following integral equations are established:

$$h_{I_i(t)}(s) = 1 - G_i(t) + \int_0^t f_i(u_{i0}s) + u_{i1}h_{I_1(t-y)}(s) + \dots + u_{in}h_{I_n(t-y)}(s))dG_i(y),$$

 $h_{I_i}(s) = f_i(u_{i0}s + u_{i1}h_{I_1}(s) + \dots + u_{in}h_{I_n}(s)).$

3 Probabilities of extinction

We define the probabilities of extinction of the main model process before fixed moment t, as follows:

$$q_i(t) = P[Z^m(t) = 0, 0 \le m \le n, t \text{ is fixed} | Z^i(0) = 1, Z^j(0) = 0, j \ne i], 0 \le i \le n.$$

Again, we are able to identify a recurrent relationship between $q_i(t)$. **Theorem 2.1**. The following integral equations hold:

$$q_0(t) = \int_0^t f_0(q_0(t-y)) dG_0(y),$$

$$q_i(t) = \int_0^t f_i(u_{i0}q_0(t-y) + u_{i1}q_1(t-y) + \dots + u_{in}q_n(t-y))dG_i(y), 1 \le i \le n.$$

In what follows we will establish the limit behaviour of the probabilities $q_i(t), 0 \le i \le n$.

Considering the reasoning involved in the derivation of q_i , it follows that q_i correspond to the probabilities of ultimate extinction of the main model process. We can denote these probabilities as:

$$q_i = P[Z^0(t) = Z^1(t) = \dots = Z^n(t) = 0 \text{ for some } t > 0|$$
$$Z^i(0) = 1, Z^j(0) = 0, j \neq i], 0 \le i \le n.$$

Now we are able to obtain relationships between p.g.f. $f_i(s)$ and q_i , $0 \le i \le n$:

Firstly, from Jagers, p. 140, we immediately have $q_0 = f_0(q_0)$.

Secondly, the probability of final extinction of a process starting with a single cell of type i, q_i , $1 \le i \le n$ depends only on q_0 because type 0 cells are the only ones that are supercritical. If the process starts with a single cell of type i, the number of mutants that occur during the whole process is modeled through the r. v. I_i . In order for the main model process to become extinct, it is sufficient that with all occurring mutants start processes which will die out. Using the total probability argument and if denoting by $r_{ik} = P(I_i = k)$ we obtain:

$$q_i = \sum_{k=0}^{\infty} r_{ik} q_0^k = E(q_0^{I_i}) = h_{I_i}(q_0), 1 \le i \le n.$$

In addition, for $s = q_0$ we have

$$q_i = h_{I_i}(q_0) = f_i(u_{i0}q_0 + u_{i1}h_{I_1}(q_0) + \dots + u_{in}h_{I_n}(q_0)).$$

4 Time until occurrence of a mutant, starting a non-extincting BHBP

We introduce r.v. T_i , denoting the time it takes for the occurrence of the first mutant, initiating a non-extincting BHBP main model. Such a mutant, leading to non-extincting processes, is called "successful" and the fact that it starts such a process is often paraphrased as "the process escaping extinction". We define $T_i = \infty$ as the event that no successful mutant has occurred during the process. That way $T_i \in (0, \infty]$.

Note that the proofs for the case in which the process starts with a single cell of type i, $1 \le i \le n$ are analogous to those in Slavtchova–Bojkova (2016, LNS, Springer), with the difference that I_i , $I_i(t)$, and q_i have p.g.f. of rather different form.

Theorem 3.1 Let the process starts with 1 cell of type i, $1 \le i \le n$. Then the distribution of r.v. T_i has the following properties: (*i*) $P(T_i > t | Z^i(0) = 1, Z^m(0) = 0, m \ne i) \equiv Q_{i,t} = h_{I_i(t)}(q_0)$, (*ii*) $P(T_i = \infty | Z^i(0) = 1, Z^m(0) = 0, m \ne i) = q_i = h_{I_i}(q_0)$, (*iii*) $E(T_i | T_i < \infty, Z^i(0) = 1, Z^m(0) = 0, m \ne i) = \frac{1}{1-q_i} \int_0^\infty \left[h_{I_i(t)}(q_0) - h_{I_i}(q_0) \right] dt$. If the process starts with k_i particles of type $i, 1 \le i \le n$, then the distribution of r.v. T has the following properties:

$$\begin{aligned} (iv) \ P(T_i > t | Z^i(0) = k_i, 1 \le i \le n) = \\ h_{I_1(t)}^{k_1}(q_0) \times h_{I_2(t)}^{k_2}(q_0) \times \ldots \times h_{I_n(t)}^{k_n}(q_0), \\ (v) \ P(T_i = \infty | Z^i(0) = k_i, 1 \le i \le n) = q_1^{k_1} \times q_2^{k_2} \times \ldots \times q_n^{k_n} \\ = h_{I_1}^{k_1}(q_0) \times h_{I_2}^{k_2}(q_0) \times \ldots \times h_{I_n}^{k_n}(q_0), \\ (vi) \ E(T_i | T_i < \infty, Z^i(0) = k_i, 1 \le i \le n) = \\ - \end{aligned}$$

$$\frac{1}{1-q_1^{k_1} \times q_2^{k_2} \times \ldots \times q_n^{k_n}} \int_0^\infty [h_{I_1(t)}^{k_1}(q_0) \times h_{I_2(t)}^{k_2}(q_0) \times \ldots \times h_{I_n(t)}^{k_n}(q_0) - h_{I_1}^{k_1}(q_0) \times h_{I_2}^{k_2}(q_0) \times \ldots \times h_{I_n}^{k_n}(q_0)] dt$$

5 Immediate risk of escaping extinction

Another useful characteristic associated with the occurrence of successful mutants in the main BHBP model is the probability of a "successful" mutant occurring within a very small interval dt after moment t. More precisely, this probability will be called "immediate risk of escaping extinction".

If there are no subcritical cells left in the cell population, the probability of occurrence of a "successful" mutant is 0. Therefore we will investigate a modification of the standard formulation of the hazard function. Let us define a modified hazard function $g_i(t)$ for each subcritical type $1 \le i \le n$ in the following way:

$$g_i(t)dt = P(T_i \in (t, t+dt]|T_i > t, \sum_{j=1}^n Z^j(t) > 0, Z^i(0) = 1, Z^m(0) = 0, m \neq i).$$

In other words, we will consider the probability of a "successful" mutant occurring immediately after moment t, under the additional condition that at moment t the population has at least one cell from an arbitrary subcritical type.

From this definition we have:

$$g_{i}(t)dt = \frac{P(T_{i} \in (t, t + dt] | Z^{i}(0) = 1, Z^{m}(0) = 0, m \neq i)}{P(T_{i} > t, \sum_{j=1}^{n} Z^{j}(t) > 0 | Z^{i}(0) = 1, Z^{m}(0) = 0, m \neq i)}$$

=
$$\frac{P(T_{i} \in (t, t + dt] | Z^{i}(0) = 1, Z^{m}(0) = 0, m \neq i)}{Q_{i,t} - P(T_{i} > t, \sum_{j=1}^{n} Z^{j}(t) = 0 | Z^{i}(0) = 1, Z^{m}(0) = 0, m \neq i)},$$

which can be rewritten as:

$$g_i(t) = \frac{F_{T_i}^{(1)}(t|Z^i(0) = 1, Z^m(0) = 0, m \neq i)}{Q_{i,t} - P(T_i > t, \sum_{j=1}^n Z^j(t) = 0|Z^i(0) = 1, Z^m(0) = 0, m \neq i)}$$

where $F_{T_i}^{(1)}$ is the probability density function of T_i and

$$P(T_i > t | Z^i(0) = 1, Z^m(0) = 0, m \neq i) \equiv Q_{i,t} = h_{I_i(t)}(q_0)$$

To simplify the notation, we introduce

$$V_{i,t} = P(T_i > t, \sum_{j=1}^n Z^j(t) = 0 | Z^i(0) = 1, Z^m(0) = 0, m \neq 0, i).$$

Theorem 4.1. The joint probability of the event that jointly the first "successful" mutant does not appear and no cells of types $1, 2, 3, \ldots, n$ exist at time t satisfies the following integral equation:

$$V_{i,t} = \int_0^t f_i \bigg(u_{i0}q_0 + u_{i1}V_{1,t-y} + \ldots + u_{in}V_{n,t-y} \bigg) dG_i(y),$$

for i = 1, 2, ..., n.

Theorem 4.2. There exist $\lim_{t\to\infty} V_{i,t} = V_i$, such that $V_{i,t} \leq V_i$, $\forall t \geq 0, 1 \leq i \leq n$.

6 Schemes for numerical calculations

6.1 Calculation of $h_{I_i(t)}(s)$, $h_{I_i}(s)$ and the time until the occurrence of a mutant, starting a non-extincting main BHBP

We extend the algorithm provided in Slavtchova–Bojkova (CSDA, 2017) in the case of n + 1 types, provided the offspring of the cells with indices $i, 1 \le i \le n$ can be of any type (including type 0), the offspring of type 0 cells may be of type 0 only.

For every $1 \le i \le n$ we have:

$$h_{I_i(t)}(s) = 1 - G_i(t)$$

-
$$\int_0^t f_i(u_{i0}s + u_{i1}h_{I_1(t-y)}(s) + \dots + u_{in}h_{I_n(t-y)}(s))dG_i(y)$$

Below we will assume that s is fixed. I. Let t = 0. For every $1 \le i \le n$ we have

$$h_{I_i(0)}(s) = 1 - G_i(0) = 1, \quad h_{I_i(0)}(s) = 1.$$

Then we have: $h_{I_1(0)}(s) = h_{I_2(0)}(s) = \ldots = h_{I_n(0)}(s) = 1$. II. Let t = h. We approximate the integral

$$\int_0^h f_i(u_{i0}s + u_{i1}h_{I_1(h-y)}(s) + \dots + u_{in}h_{I_n(h-y)}(s))dG_i(y)$$

using the right rectangle rule:

$$h_{I_i(h)}(s) \approx 1 - G_i(h)$$
+ $f_i(u_{i0}s + u_{i1}h_{I_1(h-h)}(s) + u_{i2}h_{I_2(h-h)}(s) + \dots +$
+ $u_{in}h_{I_n(h-h)}(s)) \times (G_i(h) - G_i(0)),$

$$h_{I_i(h)}(s) \approx 1 - G_i(h)$$
+ $f_i(u_{i0}s + u_{i1}h_{I_1(0)}(s) + u_{i2}h_{I_2(0)}(s) + \dots +$
+ $u_{in}h_{I_n(0)}(s)) \times G_i(h).$

Now we have:

$$\begin{split} h_{I_1(0)}(s) &= 1, \\ h_{I_1(h)}(s) &\approx 1 - G_1(h) + f_1(u_{10}s + u_{11} + u_{12} + \dots + u_{1n}) \times G_1(h), \\ h_{I_2(0)}(s) &= 1, \\ h_{I_2(h)}(s) &\approx 1 - G_2(h) + f_2(u_{20}s + u_{21} + u_{22} + \dots + u_{2n}) \times G_2(h), \\ \dots \\ h_{I_n(0)}(s) &= 1, \\ h_{I_n(h)}(s) &\approx 1 - G_n(h) + f_n(u_{n0}s + u_{n1} + u_{n2} + \dots + u_{nn}) \times G_n(h) \end{split}$$

III. Let t = 2h. Consider:

$$\begin{split} &\int_{0}^{2h} f_{i}(u_{i0}s + u_{i1}h_{I_{1}(2h-y)}(s) + \ldots + u_{in}h_{I_{n}(2h-y)}(s))dG_{i}(y) = \\ &= \int_{0}^{h} f_{i}(u_{i0}s + u_{i1}h_{I_{1}(2h-y)}(s) + \ldots + u_{in}h_{I_{n}(2h-y)}(s))dG_{i}(y) + \\ &+ \int_{h}^{2h} f_{i}(u_{i0}s + u_{i1}h_{I_{1}(2h-y)}(s) + \ldots + u_{in}h_{I_{n}(2h-y)}(s))dG_{i}(y). \end{split}$$

Again, we approximate the integrals through the right rectangle rule:

$$h_{I_{i}(2h)}(s) \approx 1 - G_{i}(2h)$$

$$+ f_{i}(u_{i0}s + u_{i1}h_{I_{1}(h)}(s) + u_{i2}h_{I_{2}(h)}(s) + \dots + u_{in}h_{I_{n}(h)}(s))$$

$$\times (G_{i}(h) - G_{i}(0))$$

+
$$f_i(u_{i0}s + u_{i1} + u_{i2} + \dots + u_{in}) \times (G_i(2h) - G_i(h)).$$

6.2 Approximation of $P(T > t, \sum_{j=1}^{n} Z^{j}(t) = 0 | Z^{i}(0) = 1, Z^{m}(0) = 0, m \neq 0, i)$ I. Let t = 0. We have $V_{i,0} = 0, 1 \le i \le n$. II. Let t = kh. For each i, $1 \le i \le n$ we have: $V_{i,kh} \approx \sum^{\kappa} f_i(u_{i0}q_0 + u_{i1}V_{1,(k-j)h} + \dots + u_{in}V_{n,(k-j)h}) \times (G_i(jh) - G_i((j-1)h))$ i=1

6.3 Approximation of modified hazard functions $g_i(t), 1 \le i \le n$

$$g_i(t) = \frac{F_T^{(1)}(t|Z^i(0) = 1, Z^m(0) = 0, m \neq i)}{Q_{i,t} - V_{i,t}}$$

I. Forward difference approximation: we discretize the [0,T] interval with step h, i.e. we use points 0, h, 2h, ..., Nh = T, (n+1)h, thus yielding

$$F^{(1)}(kh) = \frac{F_T((k+1)h) - F_T(kh)}{h} + O(h), k = 0, 1, \dots N.$$

II. Centered difference approximation: we discretize the [0, T] interval with step h, i.e. we use points 0, h, 2h, ..., Nh = T, (n + 1)h. thus yielding

$$F^{(1)}(kh) = \frac{F_T((k+1)h) - F_T((k-1)h)}{2h} + O(h^2), k = 1, 2, ..., N.$$

6.4 Examples –three setups

In this section we will present the results obtained from calculations done in three setups. These setups differ from each other in the type of distributions used for modeling the lifespan of the distinct cell types. In general, setup 1 considers distributions which do not exhibit heavy-tails, i.e. their tails are not exponentially bounded, setup 2 is the same as setup 1 except that a heavy-tailed distribution is used for one of the subcritical types, and setup 3 considers only heavy-tailed distributions. More precisely, we will restrain ourselves with the cases where we model all cell types in setup 1 with exponential distributions, in setup 2 we will change the distribution of type 1 cells from exponential to lognormal, and in setup 3 we will model all cell types with lognormal distributions. In our full research we used numerous different combinations of exponential, truncated normal, gamma, lognormal, Pareto, Weibull and Cauchy distributions, throughout cell types 0, 1, ..., n, all of those combinations yielded similar results as those stated below.

Comparison of $q_i(t)$ and q_i for the first setup						
	$f_i(s)$	u_{i0}	u_{i1}	u_{i2}	u_{i3}	lifespan distr.
Type 0	$0.2s^0 + 0.45s^2 + 0.35s^4$	1				Exp(2)
Type 1	$0.64s^0 + 0.36s^2$	0.05	0.7	0.1	0.15	Exp(3)
Type 2	$0.7s^0 + 0.12s^2 + 0.18s^4$	0.1	0.07	0.8	0.03	Exp(4.5)
Type 3	$0.78s^0 + 0.22s^4$	0.01	0.07	0.02	0.9	Exp(6)
Our coloulations are denough $t = 200$ and $h = 0.01$						

Our calculations are done with t = 300 and h = 0.01.











t = 300 and h = 0.01. The value of q_0 is 0.2233 as was already calculated.



Behaviour of $h_{I_i(t)}(q_0), i = 1, 2, 3$ in setup 1.



Behaviour of $h_{I_i(t)}(q_0), i = 1, 2, 3$ in setup 2.



Behaviour of $h_{I_i(t)}(q_0), i = 1, 2, 3$ in setup 3.

6.6 Comparison of the modified hazard functions $g_i(t)$

h = 0.1 for setups 1-3 and we begin with setup 1, t = 300.



Behaviour of $g_i(t), i = 1, 2, 3$ in setup 1.

t = 500



Behaviour of $g_i(t), i = 1, 2, 3$ in setup 2.





Behaviour of $g_i(t), i = 1, 2, 3$ in setup 3.

7 Conclusions

Once again we state that the numerical results presented in the current paper are a subset of all calculations we made using exponential, truncated normal, gamma, lognormal, Pareto, Weibull and Cauchy distributions with various values for their parameters.

We argue that modelling lifespan of cancer cell types with heavy-tailed distributions is a perspective approach as it yields close to real world behaviour for $g_i(t)$ without the need of inflating any of the parameters of the distributions involved. However, one should be cautious with the presence of oscillations when calculating for large t. Whether possible oscillations are a result of a deficiency in our model and corresponding numerical calculation scheme, or are a manifestation of a real-world behavior of the cancer disease, is open for exploration.

Remark 2. We have used PYTHON 3.5.2. for implementing the numerical methods.

8 Acknowledgments.

The research is supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No DFNI-I02/17 and Ministerio de Economía y Competitividad and the FEDER through the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, grant MTM2015-70522-P, the Junta de Extremadura (grant GR15105), Spain.

9 References.

Durrett, R. (2015). Branching Process Models of Cancer. Springer.

Haccou, P., Jagers, P., Vatutin, V. (2007). Branching processes: variation, growth and extinction of populations. Cambridge University Press.

Iwasa, Y., Michor, F., Nowak, M. (2003). Evolutionary dynamics of escape from biomedical intervention. Proc. Biol. Sci., B 270(1533), 2573–2578.

Iwasa, Y., Michor, F., Nowak, M. (2004). Evolutionary dynamics of invasion and escape. J. Theor. Biol., 226(2), 205–214.

THANK YOU FOR YOUR ATTENTION!

18th International Summer Conference on Probability and Statistics (ISCPS'18) 25-30 June 2018, Pomorie, BULGRIA http://www.math.bas.bg/ statlab/ISCPS2018/