Cancer Modeling: A Personal Perspective

Rick Durrett

Cancer modeling comes in a wide variety of styles. Indeed, it can involve almost any type of applied mathematics. My personal favorite approach is the use of probability models to understand how genetic mutations lead to cancer progression, metastasis, and resistance to therapy. Ordinary differential equations can be used to study the growth of tumor cell populations, often leading to a conclusion of Gompertzian growth [21]. PDE models using cell densities and nutrient concentrations as state variables can be used to analyze various spatiotemporal phenomena; see [13].

Individual and agent-based models that treat cells as discrete objects with predefined rules of interaction can offer an improvement over PDE methods in some situations, such as the study of angiogenesis, the development of new blood vessels to bring nutrients to a growing tumor [1]. For a comparison of individual-based and continuum approaches in one particular example, see [4]. Agent-based systems are one of many computationally intensive methods [24] and are often components of multiscale models (see [16], [6], and [8]).

Rather than spend the entire article in the land of generalities with random pointers to the literature, I will next give a description of a useful, simple, and flexible model: multitype branching processes. The types represent stages in the cancer progression. For example, in colon cancer, type 1 cells have one copy of the gene APC inactivated, type 2 cells have both copies inactivated, type 3 cells have the KRAS turned on, and type 4 cells have a mutation in P53. For more on this example, see the classic paper by Luebeck and Moolgavkar [17] or Steven Frank's book [12].

I realize that many readers of this article have not heard of these three genes. For the purposes of this article, it is enough to know a few simple facts. APC is a tumor-suppressor gene. When both copies are knocked out in a cell, trouble starts. KRAS is an oncogene. A mutation of one copy changes the behavior of the cell. In population genetics these would be called advantageous mutations, but what is advantageous to a subset of your cells is not necessarily good for the whole. Finally, P53, which is named for its 53 Kilo-Dalton size, is a housekeeping gene that helps keep the cell-replication machinery running smoothly.

Exactly what these three genes do is not important. Indeed, in many cases, such as the BRCA genes, there was a long time interval between when they were discovered to have a statistically significant correlation with breast cancer and when the mechanism that caused this association was understood. For the branching process model we are about to describe, the key fact, which is used to give estimates of the mutation rates u_i, is that there is a fairly specific sequence of mutations that leads to the disease. Before turning to the mathematical details of the model, one last thing that should be said is that this "key fact" is an oversimplification. In 20 percent of colon cancers, APC is not mutated, but the oncogene β-catenin, which is in the same metabolic pathway as APC, is turned on. The fact that pathways and not individual genes are the targets of cancer-causing mutations has greatly complicated the tumor genome sequencing approach to understanding the mechanisms of cancer.

Rick Durrett is a James B. Duke Professor of Mathematics at Duke University in Durham, NC. His email address is rtd@math.duke.edu.

In multitype branching processes, cells of type i die at rate b_i, give birth to nonmutant offspring at rate a_i, and produce cells of type $i + 1$ at rate u_{i+1}. Here the rates are the transition rates of a continuous time Markov chain; e.g., the probability a cell gives birth in t units of time is $a_i t + o(t)$ as $t \to 0$. Let $Z_i(t)$ be the number of cells of type i at time t. The behavior of the type 0's has been known for many years; see, e.g., the classic book by Athreya and Ney [2]. If $\lambda_i = a_i - b_i$ is the net growth rate, then the expected number of type 0 at time t is

$$EZ_0(t) = e^{\lambda_0 t} Z_0(0).$$

If $a_0 > b_0$, then the probability that the type 0's do not die out is λ_0/a_0, and when they do not die out

$$e^{-\lambda_0 t} Z_0(t) \to V_0 = \text{exponential}(a_0/\lambda_0).$$

Here $\to$ indicates that with probability one the sequence of numbers converges. $V = \text{exponential}(r)$ is read "exponential with rate r" and means $P(V > v) = e^{-rv}$ for $v \geq 0$.

Since the types progress $0 \to 1 \to 2 \to \dots$, this process at first seems childishly simple. Breaking things down according to the time of the mutation to type 1, the expected number of type 1's satisfies

$$EZ_1(t) = \int_0^t EV_0 e^{\lambda_0 s} \cdot u_1 \cdot e^{\lambda_1(t-s)} \, ds.$$

The first factor gives the size of the type 0 population at time s, the second the mutation rate, and the third the number of offspring the mutant has at time t. However, if $0 < \lambda_0 < \lambda_1$, the expected value $EZ_1(t)$ drastically overestimates the number of type 1's. The main contribution to the integral comes from times near 0, but in cancer u_1 is very small (10^{-5} or less), so, as in the state lottery, the expected value comes from a rare event that produces a very large result.

In analyzing the growth of the multitype process, it is more productive to focus one's attention on times at which $V_0 e^{\lambda_0 s} \approx 1/u_1$ and type 1's are being produced at a positive rate. In terms of the theory of stochastic processes, mutations to type 1 are a nonhomogeneous Poisson process. Using elementary properties of the Poisson process and simple computations with Laplace transforms, one can show that

$$e^{-\lambda_1 t} Z_1(t) \to V_1 \quad \text{with} \quad Ee^{-\theta V_1} = (1 + cu_1 \theta^\alpha)^{-1},$$

where $\alpha = \lambda_0/\lambda_1$. This result is more easily understood (and generalized to types $k > 1$) if one conditions on the value of V_0.

$$E(e^{-\theta V_1}|V_0) = \exp(-cu_1 V_0 \theta^\alpha).$$

See [11] for more details. The right-hand side is the Laplace transform of a one-sided stable law. These distributions come up in a typical graduate course in probability because they are part of the answer to the question, "What are the possible limits of normalized sums of independent random variables $(S_n - b_n)/a_n$?"

The appearance of stable laws in this setting is (at first) somewhat surprising, but as we will see in a moment, it is very useful in quantifying the relative frequency of mutations that make up a tumor. Understanding the amount of tumor heterogeneity is important for several reasons. Higher levels of heterogeneity have been correlated with tumor aggressiveness in a clinical study of Barrett's esophagus [19] and is thought to be predictive of malignant progression in other cancers as well. In addition, tumor heterogeneity poses challenges for the development of successful therapies. For example, increased genetic heterogeneity means a higher probability that a tumor harbors cells resistant to treatment. On the other hand, there are some new therapies that take advantage of the competition between different types of tumor cells. See [20] (which is available for free at PubMedCentral) for an explanation and for more on the causes and consequences of heterogeneity.

One measure of tumor diversity, Simpson's index R_k, can be defined as the probability that two randomly chosen type k cells are descended from the same mutation. In genetics this is called the homozygosity. Using some of the facts about stable laws that have accumulated over the last seventy years, one can prove a remarkably simple result for type k cells:

$$ER_k = 1 - \alpha_k, \quad \text{where } \alpha_k = \lambda_{k-1}/\lambda_k.$$

With more work one can compute moments of R and obtain insights into its distribution. See [10] for more details.

Two things make these results possible: (i) A Poisson process representation of the relative contribution of the different mutations to the cancer cell population, which is the key to identifying the limits V_i as stable laws. See Figure 1 for a graphical
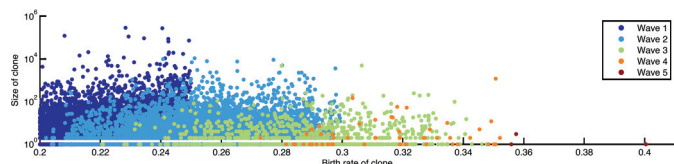

Figure 1. Tumor heterogeneity in a multitype branching process. Generalizing the model discussed in the text, the growth rate of a mutant is that of its parent plus an amount uniformly distributed on [0, .05]. Dots indicate the growth rate of the mutant and the logarithm of the number of its descendants at a fixed time. Note that a half dozen type 1 families are responsible for most of the population. Figure by Jasmine Foo.

display of tumor heterogeneity in a simulation of a multitype branching process. (ii) Mathematicians who have done intensive study of Poisson-Dirichlet distributions, which appear naturally in a wide variety of topics from the infinite alleles model in biology to properties of the cycle structure of randomly chosen permutations. See Pitman's book [23] for more details.

An Application to Ovarian Cancer

While it is nice that our multitype branching processes have a detailed and interesting theory, it is more important that they are useful in applications. One simple example comes from work I did with Duke undergraduate Kaveh Danesh to answer a question asked by two doctors in Duke Medical Center's Department of Obstetrics and Gynecology. Evan Myers and Laura Havriletsky wanted to know, "How often should ovarian cancer screening be done in order to be effective?" To address this question, we built a multitype branching process model in which type 0 cells are in the primary tumor in the ovary or fallopian tube, type 1 cells are floating in the abdominal cavity, and type 2 cells are those that have attached to the peritoneum. Type 2 cells infiltrate the cellular matrix and eventually metastasize to distant organs, so when they are present in significant numbers, the cancer is classified as stage III. Note that in this model, the transitions from type $0 \rightarrow 1$ and $1 \rightarrow 2$ involve migration of cells, not genetic mutation.

One of the problems with ovarian cancer is that many cases are diagnosed in stage III or IV, with a five-year survival rate of less than 30 percent, compared to 90 percent for stage I tumors, so many lives could be saved if they were caught in stage I, when the cancer has not spread. (Contrary to intuition, the clinically defined stage II—tumors in both ovaries/fallopian tubes—often does not come between stages I and III.) We defined the window of opportunity for screening to be $T_2 - T_0$, where T_0 is the first time the primary tumor is 0.5 cm in diameter (and hence visible on a transvaginal ultrasound) and T_2 is the first time there are 10^9 cells of type 2 (which corresponds to one gram). These definitions are somewhat arbitrary, but our formulas easily give results for other cutoffs.) Using what is known about tumor growth and mutation rates, we concluded that the window of opportunity was 2.9 years, with most of the distribution concentrated on $[2.5, 3]$ years. Thus, in order to be effective, screening needs to be done every two years. See [7] for more details. Figure 2 gives a picture of the growth of the three cell populations on a log scale.
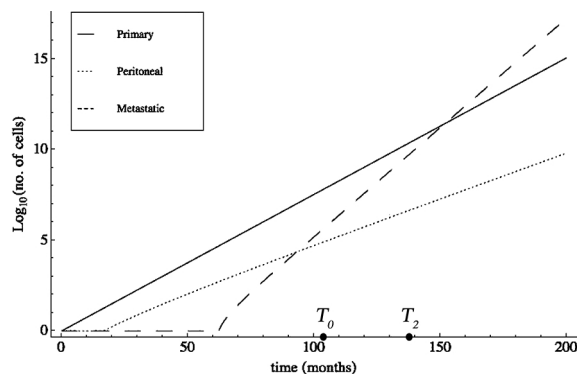

Figure 2. Plot of the sizes of the Primary (solid line), Peritoneal (dotted line), and Metastatic (dashed line) cell subtypes in the ovarian cancer model on a logarithmic scale. The window of opportunity for screening is $[T_0, T_2]$, where T_0 is the first time the primary tumor has diameter 0.5 cm. T_2 is the time at which there are 10^9 metastatic cells (approximately one gram). Figure by Kaveh Danesh.

Models: Simple or Detailed?

I prefer simple models that can be analyzed mathematically. For example, early in my career I studied percolation and the Ising model. In the stochastic Ising model each iron atom has a spin that can be $+1$ (up) and -1 (down), and spins flip at a rate that depends on the number of neighbors of the opposite type. Of course, in a real iron bar, atoms have a spin that points in some direction in three-dimensional space, and spins interact with other than their nearest neighbors. However, despite these simplifications, the Ising model yields insights into qualitative properties of the magnetization phase transition.

On the other hand, cancer biologists often prefer models that include all of the relevant details. Chapter 6 of Cristini and Lowengrub's book [6] describes a model of ductal carcinoma in situ, the most prevalent precursor to invasive breast cancer. The authors use an agent-based modeling framework that takes cell motility and various cell-cell interactions into account, but then one ends up with more than two dozen parameters, simulations that are restricted to a 1 mm duct, and a model that treats a two-dimensional slice instead of the three-dimensional tube. See Figure 3 for a picture of a simulation.

While the model is complex and analysis can only be based on simulation, it has the advantage of being realistic. At the SIAM Life Sciences meeting in San Diego, August 7–10, 2012, I heard Paul Macklin talk about how computations with this model were useful in informing medical treatment decisions [18]. The issue is that the calcified core of dead cells in a breast cancer duct, which is
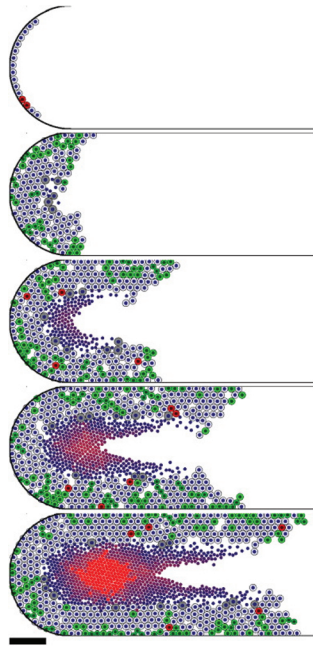
Figure 3. Simulation of ductal carcinoma in situ. Cell nuclei are the small dark blue circles; quiescent (resting) cells are pale blue; proliferating cells are green; apoptosing (dying) cells are red; and necrotic (dead) cells are grey until they lyse, after which their solid fraction remains as debris (dark circles in center of duct). The shade of red in the necrotic debris indicates the level of calcification. Simulated times (from top to bottom): 0, 7, 14, 21, and 28 days. Figure from [18], reprinted with the permission of the *Journal of Theoretical Biology*.

what can be seen in a mammogram, is smaller than the region in which cancer cells are present. Thus one must estimate the size of the region of the breast to be removed. This is an important problem, because 50 percent of women who have a lumpectomy must return for a second surgery because not all the cancer cells were removed the first time.

If you are interested in learning more about detailed models, the first 2012 issue of *Mathematical Medicine and Biology* has the proceedings of a 2009 SIAM three-part minisymposium, "State of the Art in Computational Modeling of Cancer". This style of modeling does not suit my skill set very well, but the approach is valuable, and I am doing my best to educate myself about it. Simple and detailed models complement each other. Simple models give insight into the working of detailed models. Detailed models give mechanistic insights that shape the form of the simple models and rigorously calibrate their parameters.

Breast Cancer Heterogeneity

An article like this would not be complete without some unsolved problems, so I'll mention two. Due to the introduction of microarray analysis and the use of statistical classification techniques [22], a subdivision of breast cancers into a half dozen subtypes has been developed. The grouping of tumors according to the established markers estrogen receptor (ER), progesterone receptor (PR), and HER2 has improved treatment outcomes, but challenges remain. The first and most obvious is that 16 percent of tumors show none of these markers and are classified as triple-negative [27].

During the summer of 2012, one of my postdocs, Marc Ryser, and I learned of another issue in a meeting with Kimberly Blackwell, a researcher in the Duke Cancer Institute. Due to tumor heterogeneity, a small sample of one section of the tumor may not reflect all of the mutations present. Thus, one would like to understand the spatial structure of heterogeneity in breast cancer, with the aim of conducting more informative biopsies. That is, one wants to avoid false negatives due to insufficient sampling and tumor heterogeneity.

Much is known about the heterogeneity of breast cancer between and within patients (see [25]). However, despite a large amount of money spent on tumor-sequencing studies, many of the things we need to know concerning mechanisms of that disease in order to develop a model are not known, so it is hard to know where to begin. Being a mathematician, my approach is to study a simple spatial model and to understand how the observed patterns of heterogeneity depend on the model's parameters. In this way we have results that can be applied to a variety of cancer types.

An Exciting New Development

Announced in the August 23, 2012, issue of *Nature* (see pages 462–463 for the "News and Views" summary) is the fact that there is now conclusive evidence of the existence of cancer stem cells (CSC) that can produce tumor cells in the same way that ordinary stem cells can produce normal tissue. The existence of CSC has long been controversial. For example, transplantation studies have shown that human cancer cells when injected into mice rarely cause cancer, leading to the conclusion that only a small fraction of the tumor cells are responsible for the disease. However, skeptics have pointed out that removing cells from their natural environment may change their behavior. For more see [29].

The three new papers, two in *Nature* [5], [9] and one in *Science* [26], avoid this objection by using a genetic technique called lineage tracing to track cells in an existing tumor. The laboratory techniques are ingenious and for most of us are difficult to read and understand. However, I

believe that mathematics can play an important role in sorting out whether the assumed stem cell dynamics will indeed produce the behavior observed in the laboratory experiments. To see what I have in mind, read [28] or just look at their Figure 1, which compares tumor morphology under the CSC and classical models.

It is remarkable what simple facts are not known about the normal stem cells in our bodies, whose existence is well documented. They sit at the bottom of each of the 10^7 crypts in the colon, but the best estimate is that there are 4–20 of them in each crypt. Stem cells in the bone marrow produce white blood cells, and some of them are quiescent at any one time, but the quiescent fraction may be 10 percent or it may be 90 percent.

For a long time it was thought that stem cells always divide asymmetrically, producing one stem cell and one progenitor cell. However, recent studies have shown that sometimes the result of cell division is two stem cells and sometimes two progenitor cells; see [15]. I found it interesting that the authors' analysis led them to a thirty-year-old paper of Bramson and Griffeath on the voter model [3] and that they applied its result for the one-dimensional system to their data on colon crypts.

Where Are the Math Problems in the Last Two Sections?

Biology is not physics. Statistical mechanics, quantum mechanics, relativity, and string theory have given rise to complex mathematical problems which can be happily studied by mathematicians with no knowledge of physics. However, in my twenty-five-year experience of working on problems that come from ecology, genetics, and now cancer, 80 percent of the problem in a biological application is to figure out what the question is, what mathematical tools to use, and, in my case, how to find a model that is simple enough to study analytically and that can say something useful about the application.

If You Are Excited About Cancer Modeling, Then What's Next?

Well, you just missed the SIAM Life Sciences Meeting in San Diego, August 7–10, 2012, where a lot of cancer modeling was discussed. Presumably the 2014 meeting will have more of the same. An easy entry into the subject will be provided by the 2014–2015 year on cancer modeling at the Mathematical Biosciences Institute. Till then you can read a few of the papers in the references, visit my webpage to see some of my papers, or go to `http://michorlab.dfci.harvard.edu/index.php/publications` for a more extensive set of publications by collaborator Franziska Michor.

There are more papers to read than anyone has time. When I searched MathSciNet recently, it reported 2,423 papers on cancer and 212 on breast cancer since 2000, but many of the latter are statistical analyses. PDE people will probably have more fun with the more than one hundred papers on angiogenesis. The quantitative biology section of the arXiv has cancer modeling papers, but by far the most comprehensive collection is the open-access PubMedCentral, where NIH grantholders are required to put their papers. This is good news and bad news: There are 3,800 papers with "breast cancer" in the title since 2010.

While you can get a lot of information from journals and books, by far the best way to get into math biology is to find a biologist or medical researcher to talk to. Bridging the communication gap and trying to figure out what you can do to help with their research is not easy, but then again, things that are worth doing rarely are.

References

[1] T. ALRÓN, Modeling tumor-induced angiogenesis: A review of individual-based models and multi-scale approaches, pages 45–75 in *Mathematics, Developmental Biology, and Tumor Growth,* Contemporary Math., vol. 492, Amer. Math. Soc., Providence, RI, 2009.

[2] K. ATHREYA and P. NEY, *Branching Processes*, Springer, New York, 1972.

[3] M. BRAMSON and D. GRIFFEATH, Asymptotics for interacting particle systems on $\mathbb{Z}^d$, *Probability Theory and Related Fields* **53** (1980).

[4] H. BYRNE and D. DRASO, Individual-based and continuum models of growing cell populations: A comparison, *J. Math. Biol.* **58** (2009), 657–687.

[5] J. CHEN et al., A restricted cell population propagates glioblastoma growth after chemotherapy, *Nature* **488** (2012), 522–526.

[6] V. CRISTINI and J. LOWENGRUB, *Multi-Scale Modeling of Cancer,* Cambridge Univ. Press, 2010.

[7] K. DANESH, R. DURRETT, L. HAVRILESKY, and E. MYERS, A branching process model of ovarian cancer, *J. Theoret. Biol.*, to appear.

[8] T. S. DEISBOECK, Z. WANG, P. MACKLIN, and V. CRISTINI, Multiscale cancer modeling, *Ann. Rev. Biomed. Engr.* **13** (2011), 127–155.

[9] G. DRIESSENS, B. BECK, A. CAAUWE, B. D. SIMONS, and C. BLANPLAN, Defining the mode of tumour growth by clonal analysis, *Nature* **488** (2012), 527–531.

[10] R. Durrett, J. Foo, K. Leder, J. Mayberry, and F. Michor, Intratumor heterogeneity in evolutionary models of tumor progression, *Genetics* **188** (2011), 461–477.

[11] R. Durrett and S. Moseley, Evolution of resistance and progression to disease during clonal expansion of cancer, *Theor. Pop. Biol.* **77** (2010), 42–48.

[12] Steven A. Frank, *Dynamics of Cancer*, Princeton Univ. Press, 2007.

[13] A. Friedman, *Tutorials in Mathematical Biosciences, II: Cell cycle, proliferation, and cancer,* Springer Lecture Notes in Math., vol. 1872, 2005.

[14] M. Kimmel and D. E. Axelrod, *Branching Processes in Biology*, Springer, New York, 2002.

[15] C. Lopez-Garcia, A. M. Klein, B. D. Simons, and D. J. Winton, Intestinal stem cell replacement follows a pattern of neutral drift, *Science* **330** (2010), 822–825.

[16] J. Lowengrub, H. B. Frieboes, F. Jin, Y. L. Chuang, X. Li, P. Macklin, S. M. Wise, and V. Cristini, Nonlinear modeling of cancer: Bridging the gap between cells and tumors, *Nonlinearity* **23** (2010), R1–R91.

[17] E. G. Luebeck and S. H. Moolgavkar, Multi-stage carcinogenesis and the incidence of colorectal cancer, *Proc. Nat. Acad. Sci. U.S.A.* **99** (2002), 15095–15100.

[18] P. Macklin, M. E. Edgerton, A. M. Thompson, and V. Cristini, Patient-calibrated agent-based modelling of ductal carcinoma in situ (DCIS): From microscopic measurements to macroscopic predictions of clinical progression, *J. Theoret. Biol.* **301** (2012), 122–140.

[19] C. C. Maley et al., Genetic clonal diversity predicts progression to esophageal adenocarcinoma, *Nature Genetics* **18** (2006), 468–473.

[20] A. Marusyk and K. Polyak, Tumor heterogeneity: Causes and consequences, *Biochem. Biophys. Acta.* **1805**, no. 1 (2010), 105–117. PMC2814927

[21] L. Norton, A Gompertzian model of human breast cancer growth, *Cancer Research* **48** (1988), 7067–7071.

[22] C. M. Perou, Molecular portraits of human breast tumors, *Nature* **406** (2000), 747–752.

[23] J. Pitman, *Combinatorial Stochastic Processes,* Springer-Verlag, Berlin. Available at http://works.bepress.com/jim_pitman/1/

[24] V. Quaranta, K. A. Rejniak, P. Gerlee, and A. R. A. Anderson, Invasion emerges from cancer cell adaptation to competitive micro-environments, *Seminars in Cancer Biology* **18** (2008), 338–348.

[25] H. G. Russnes, N. Navin, J. Hicks, and A. L. Borresen-Dale, Insight into the heterogeneity of breast cancer through next-generation sequencing, *Journal of Clinical Investigation* **121** (2011), 3810–3818.

[26] A. G. Schepers et al., Lineage tracing reveals $Lgr5^+$ stem cell activity in mouse intestinal adenomas, *Science* **337** (2012), 730–735.

[27] S. P. Shah et al., The clonal and mutational evolution spectrum of primary triple-negative breast cancers, *Nature* **486**, (2012), 395–399.

[28] A. Sottoriva, L. Vermeulen, and S. Tavaré, Modeling evolutionary dynamics of epigenetic mutations in hierarchically organized tumors, *PLoS Computational Biology* **7** (2011), Issue 5, paper e1001132.

[29] M. H. Tomlinson, Cancer stem cells: A guide for skeptics, *J. Cell. Biochem.* **106** (2009), 745–749.