Modeling of the Process of Telomere Shortening: an Overview

Peter Olofsson and Marek Kimmel

Department of Statistics
Rice University, P.O.Box 1892, Houston, TX 77251
USA

1 Introduction

Shortening of telomeres is one of the supposed mechanisms of cellular aging and death. The hypothesis is that each time a cell divides it loses pieces of its chromosome ends. These ends are called telomeres and consist of repeated sequences of nucleotides, telomere units. When a critical number of telomere units is lost, the cell stops dividing. There is an extensive biological literature on the subject; a good reference for the basic facts and an intuitively appealing explanation of the deletion process is Levy *et al.* (1992). This also seems to be the first paper to formulate the problem mathematically and its essentially deterministic model serves as the basis for all subsequent mathematical models in the literature.

Telomeres are assumed to consist of telomere units, repeated sequences of nucleotides. When a chromosome replicates, each newly synthesized strand loses one telomere unit at one of its ends. This means that the pair of daughter chromosomes each has one old, unchanged strand and one new, one unit shorter. Once a critical number of telomere units are lost, a so called Hayflick checkpoint is reached and the cell stops dividing. Under this assumption, only the length of the shortest telomere will matter and thus a chromosome is said to be of type j if its shortest telomere has j remaining units. This leads to a model where a type j chromosome gets two offspring, one of type j and one of type $j-1$. Cells of type 0 do not divide.

In this chapter, the attempts to model and analyze this process with

mathematical methods are reviewed. The emphasis is put on multi-type branching process models where the main features turn out to be reducibility and polynomial growth.

2 Branching Process Model

We now describe the telomere loss process as a multi-type branching process. For generalities of such processes, see for instance Mode (1971). The type of a chromosome is, as described above, the number of remaining telomere units on its shortest end. Fix one particular chromosome and let the type of the cell be the type of that chromosome. A cell of type j divides into two daughter cells of types j and $j-1$ if $j \geq 1$; a cell of type 0 does not divide but stays in the population forever (is immortal). Dividing cells have life-lengths that are i.i.d. random variables. Start with one cell of type k and let $M_j(t)$ denote the expected number of cells of type j at time t. The main feature of the resulting multi-type branching process is reducibility, meaning that types do not communicate. As a result of this, the asymptotics are polynomial rather than exponential. The main results of Arino, Kimmel and Webb (1995) are

Theorem 2.1 *Assume that life-lengths are i.i.d. exponential with mean* $1/\alpha$. *Then*

$$M_j(t) = \frac{(\alpha t)^{k-j}}{(k-j)!}$$

for $j = 0, 1, ..., k$

and

Theorem 2.2 *Assume that life-lengths are bounded and continuous i.i.d. random variables with mean* μ. *Then as* $t \to \infty$,

$$M_j(t) \sim \frac{t^{k-j}}{(k-j)!\mu^{k-j}}.$$

for $j = 0, 1, ..., k$.

In Olofsson and Kimmel (1999) and Olofsson (2000), probabilistic methods were used to prove these and more general results which are presented in the next section.

3 Incorporating Cell Death

In Olofsson and Kimmel (1999), new probabilistic methods of proof are introduced. The model is the same as above but the methods of proof are simpler, clearer and easier to generalize. Let the ancestor be of type k throughout. The basic idea is to relate the real-time process to the process counted generation-wise. Recall $M_j(t)$, the expected number of j-type cells at time t and let $m_j^{(n)}$ be the expected number of j-type cells in the nth generation. The two relate in the following way:

$$M_j(t) = \sum_{n=k-j}^{\infty} m_j^{(n)}(G^{*n}(t) - G^{*(n+1)}(t)),$$

where G is the distribution function of the life-length and G^{*n} its n-fold convolution, the distribution of the sum of n i.i.d. life-lengths. To understand this formula, note that at time t, cells from any generation may be present. Since there are $m_j^{(n)}$ individuals in the nth generation and each of these is alive at time t with probability $G^{*n}(t) - G^{*(n+1)}(t)$ (born before t but not yet dead at t), the expected number of cells from the nth generation present at time t is simply $m_j^{(n)}(G^{*n}(t) - G^{*(n+1)}(t))$. Summing over all the generations then gives us $M_j(t)$. The identity may be formally derived as the unique solution to a certain renewal equation, for a more general version of this, see Jagers (1989).

In particular, if life-lengths are exponential(α), it is well known that

$$G^{*n}(t) = e^{-\alpha t} \sum_{i=0}^{n-1} \frac{(\alpha t)^i}{i!},$$

a gamma distribution with parameter n and α and hence

$$M_j(t) = e^{-\alpha t} \sum_{n=k-j}^{\infty} m_j^{(n)} \frac{(\alpha t)^n}{n!}. \tag{3.1}$$

If there is no cell death, $m_j^{(n)} = \binom{n}{k-j}$ and hence

$$M_j(t) = e^{-\alpha t} \sum_{n=k-j}^{\infty} \binom{n}{k-j} \frac{(\alpha t)^n}{n!} = \frac{(\alpha t)^{k-j}}{(k-j)!} e^{-\alpha t} \sum_{n=k-j}^{\infty} \frac{(\alpha t)^{n-(k-j)}}{(n-(k-j))!} = \frac{(\alpha t)^{k-j}}{(k-j)!}$$

which is Theorem 2.1.

The first generalization is to introduce possible cell death. As before, let life-lengths of proliferating cells be i.i.d. exponential random variables with mean $1/\alpha$. Suppose that cells of type 0 live for a time that is exponential with mean $1/\tau$ where $\tau \leq \alpha$. Another way of thinking of this is that 0-cells live for a time that is exponential with mean $1/\alpha$ just like all other cells and then 'tries to divide' and either dies or lives on. If the probability of survival is p, we have the relation $\tau = \alpha(1-p)$.

The formula for $M_j(t)$ remains unchanged for $1 \leq j \leq k$ and for $j = 0$ we get the following asymptotic result:

Theorem 3.1 *As $t \to \infty$,*

$$M_0(t) \sim \frac{1}{1-p} \frac{(\alpha t)^k}{k!}.$$

In the model with immortal cells, the 0-cells dominate since their growth rate is the fastest. In the model with cell death, the 1-cells have the same growth rate so asymptotically they dominate together with the 0-cells in the proportions 1 to $1/(1-p)$. This is already a qualitative difference between the two models.

Next, we introduce possible death also for the non-zero cells. Thus, let p be the probability of survival for a 0 cells and q the corresponding probability for non-zero cells. The asymptotics now depend on the relation between p and q as the following result shows.

Theorem 3.2 *For $1 \leq j \leq k$,*

$$M_j(t) = \frac{(\alpha q t)^{k-j}}{(k-j)!} e^{-\alpha(1-q)t}.$$

For $j = 0$ there are three cases. If $p = q$, then

$$M_0(t) = \frac{(\alpha q t)^k}{k!} e^{-\alpha(1-q)t}.$$

If $p < q$ then

$$M_0(t) \sim \frac{q}{q-p} \frac{(\alpha q t)^{k-1}}{(k-1)!} e^{-\alpha(1-q)t}.$$

Finally, if $p > q$, then

$$M_0(t) \sim \left(\frac{q}{q-p}\right)^k e^{-\alpha(1-p)t}.$$

Note that the polynomial growth is now asymptotically killed by an exponential decay factor, determined by the larger of the two survival probabilities p and q.

A second important generalization of the model and results from Arino, Kimmel and Webb (1995) is to relax the exponential assumption and let the life-length distribution be arbitrary. The next section is devoted to this.

4 General Life-length Distributions

In Olofsson and Kimmel (1999), the results from Arino, Kimmel and Webb (1995), were also extended to general life-length distributions. We state the theorem with proof.

Theorem 4.1 *Suppose the life-lengths are i.i.d. with common distribution function G and common mean μ. Then, as $t \to \infty$,*

$$M_j(t) \sim \frac{t^{k-j}}{\mu^{k-j}(k-j)!}$$

for $0 \le j \le k$.

The proof is based on the following so called Tauberian theorem (Feller (1971)).

Theorem 4.2 *Let* H *be a non-decreasing function such that the Laplace-Stieltjes transform* $\widehat{H}(s) = \int_0^\infty e^{-su} H(du) < \infty$ *for* $s > 0$. *Then*

$$\widehat{H}(s) \sim \frac{A}{s^r} \quad \text{as } s \to 0$$

if and only if

$$H(t) \sim \frac{At^r}{\Gamma(r+1)} \quad \text{as } t \to \infty.$$

Proof of Theorem 4.1. First let $k = 1$. Then

$$\begin{aligned} M_1(t) &= \sum_{n=1}^{\infty} n(1-G) * G^{*n}(t) \\ &= \sum_{n=1}^{\infty} n(G^{*n}(t) - G^{*(n+1)}(t)) \\ &= \sum_{n=1}^{\infty} \sum_{j=1}^{n} (G^{*n}(t) - G^{*(n+1)}(t)) \\ &= \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} (G^{*n}(t) - G^{*(n+1)}(t)) \\ &= \sum_{j=1}^{\infty} G^{*j}(t) \sim \frac{t}{\mu} \end{aligned}$$

as $t \to \infty$, by the renewal theorem, Asmussen (1987). That this theorem applies can be realized directly: the k-type cells split according to a renewal process and at each renewal an infinite line of $(k-1)$-type cells is initiated since there is no death. Hence the number of $(k-1)$-type cells at time t is

exactly the number of such renewals up to time t.

In the absence of cell death, clearly $M_1(t)$ is non-decreasing and

$$\widehat{M_1}(s) = \int_0^\infty e^{-su} M_1(du) = \sum_{n=1}^\infty \int_0^\infty e^{-su} G^{*n}(du)$$

$$= \sum_{n=1}^\infty \widehat{G}(s)^n = \frac{\widehat{G}(s)}{1-\widehat{G}(s)} < \infty$$

for $s > 0$ unless $G(0) = 1$. Hence, Theorem 4.2 above applies and

$$\widehat{M_1}(s) \sim \frac{1}{\mu s}$$

as $s \to 0$.

Now, consider $M_k(t)$ for an arbitrary k. We obtain

$$\widehat{M_k}(s) = \int_0^\infty e^{-su} M_k(du)$$

$$= \sum_{n=k}^\infty \binom{n}{k} \int_0^\infty e^{-su} (1-G) * G^{*n}(du)$$

$$= (1-\widehat{G}(s)) \sum_{n=k}^\infty \binom{n}{k} \widehat{G}(s)^n$$

$$= (1-\widehat{G}(s)) \frac{\widehat{G}(s)^k}{(1-\widehat{G}(s))^{k+1}} = \widehat{M_1}(s)^k.$$

Hence

$$\widehat{M_k}(s) \sim \frac{1}{\mu^k s^k}$$

as $s \to 0$ and applying Theorem 4.2 again yields

$$M_k(t) \sim \frac{t^k}{\mu^k k!}$$

as $t \to \infty$.

Note that Theorem 2.2 follows as a special case.

The models and main results in Arino, Sanchez and Webb (1998) and Olofsson (2000) are essentially the same and extend the previous model by allowing for loss of a random number of telomere units in each division. The methods in the first paper are deterministic, mainly using differential equations. The second paper uses variants of the probabilistic methods introduced above and involves standard results from renewal theory. We state the main result from Olofsson (2000).

Theorem 4.3 *Assume that a cell of type j gives birth to a cell of type j and a cell of type i with probability $p_{j,i}$ for $i = 0, ..., j-1$. Let a cell of type j have life-length distribution function G_j. Start from a cell of type k and let $M_j^{(a)}(t)$ be the expected number of j-type cells of age less than a at time t. Then, as $t \to \infty$,*

$$M_j^{(a)}(t) \sim \frac{C_{k,j}}{\mu_j} \int_0^a (1 - G_j(y))dy \cdot \frac{t^{k-j}}{(k-j)!}$$

for $1 \leq j \leq k$. For $j = 0$ we have, as $t \to \infty$,

$$M_0^{(a)} \sim C_{k,0} \cdot a \cdot \frac{t^{k-1}}{(k-1)!}.$$

The constant $C_{k,j}$ is

$$C_{k,j} = \prod_{i=j+1}^{k} \frac{p_{i,i-1}}{\mu_i}.$$

5 Discussion

As was pointed out, the type of a cell is the length of the telomere of one single chromosome. This may be reasonable if there is one particular chromosome that is responsible for stopping cell division (or if there only is one chromosome such as is the case for the protozoan *Tetrahymena thermophila*; see Larson *et al.* (1987). If however cell division is stopped whenever any of several chromosomes reaches the critical telomere length, the situation is

different. We then consider a minimum of several processes and the dynamics are not the same.

For a simple illustration of this, assume a cell has two chromosomes and that the type of the cell is (j, k) where these are the remaining numbers of telomere units on the respective chromosomes. As soon as one of the chromosomes hit 0, the cell stops dividing. Now suppose the cell is of type $(1, 1)$. Then the two type 1 chromosomes each produce two daughter chromosomes, one of type 0 and one of type 1. If chromosomes are randomly allocated to daughter cells, the two daughter cells may thus either be of types $(0, 0)$ and $(1, 1)$ or of types $(0, 1)$ and $(1, 0)$ with equal probabilities. In the second case, both cells stop dividing, in the first case the $(1, 1)$-type continues to divide. Clearly there will only be a finite number of divisions in this model (the total number of cells is a geometric random variable) and the population no longer exhibits polynomial growth.

In humans, it could theoretically be any number between 1 and 46 chromosomes that are involved in stopping cell division. There are so far three papers that address these problems, two of which use computer simulation approaches.

In Arino, Kimmel and Webb (1995), the number of chromosomes involved is taken to be 40 and the model is compared to the data from Levy *et el.* (1992). The 40 chromosomes are assumed to evolve according to independent branching processes, an assumption which simplifies analysis and gives a good fit, but does not address the problems mentioned above with observing the minimum.

In Tan (1999), it is investigated what number of chromosomes best fit the data of Jones *et al.* (1985). The model allows for random lengths of the deletions and it is concluded that only a few, most likely only two, chromosomes are responsible for sending the initial signal to stop cell division.

In contrast, Rubelj and Vondraček (1999), find a good fit to the same data using all 46 chromosomes and a model which includes what they call "abrupt telomere shortening". This means that the telomeres decrease in unit steps but once they have reached a certain length n_0, they may lose anything from 1 to n_0 number of units according to some specified probability distribution on $\{1, 2, ..., n_0\}$.

A comparison of the simulations shows that Tan (1999) mimics the cloning of the laboratory by letting cells reproduce to a certain number, then harvesting half the cells at random, letting the reproduce again and so on until all

cells have stopped dividing. The procedure in Rubelj and Vondraček (1999) is to start from one cell, choose one of the two daughter cells at random and continue in this fashion a specified number D_0 of times. This gives a good fit to the bimodal distributions observed in the experiments of Jones *et al.* (1985).

6 References

[1] Arino, O., Kimmel, M. and Webb, G.F. Mathematical modeling of the loss of telomere sequences. *J. Theor. Biol.* **177** (1995), 45-57.

[2] Arino, O., Sanchez, E. and Webb, G.F. Polynomial growth dynamics of telomere loss in a heterogeneous cell population. *Dynam. Contin. Discrete Impuls. Systems* **3** (1997), 262-282.

[3] Asmussen, S. *Applied Probability and Queues.* Wiley (1987), New York.

[4] Feller, W. *An Introduction to Probability Theory and its Applications, Vol 2.* Wiley (1971), New York.

[5] Jagers, P. General branching processes as Markov fields. *Stoch. Proc. Appl.* **32** (1989), 183-212.

[6] Jones, R.B., Whitney, R.G. and Smith, J.R. Intramitotic variation in proliferative potential: stochastic events in cellular aging. *Mech. Ageing. Dev.* **29** (1985), 143-149.

[7] Larson, D.D., Spangler, E.A. and Blackburn, E.H. Dynamics of telomere length variation in tetrahymena thermophila. *Cell* **50** (1987), 477-483.

[8] Levy, M.Z., Allsopp, R.C., Futcher A.B., Greider C.W. and Harley C.B. Telomere end-replication problem and cell Aging. *J. Molec. Biol.* **225** (1992), 951-960

[9] Mode, C. *Multitype Branching Processes.* American Elsevier (1971), New York

[10] Olofsson, P. A Branching Process Model of Telomere Shortening. *Communications in Statistics-Stochastic Models* **16(1)** (2000), 167-177

[11] Olofsson, P. and Kimmel M. Stochastic models of telomere shortening. *Math. Biosci.* **1** (1999), 75-92

[12] Rubelj, I. and Vondraček, Z. Stochastic Mechanism of Cellular Aging-Abrupt Telomere Shortening as a Model for Stochastic Nature of Cellular Aging. *J. Theor Biol.* **197** (1999), 425-438

[13] Tan, Z. Intramitotic and Intraclonal Variation in Proliferative Potential of Human Diploid Cells: Explained by Telomere Shortening. *J. Theor Biol.* **198** (1999), 259-268