

# EFFECTS OF INITIAL TELOMERE LENGTH DISTRIBUTION ON SENESENCE ONSET AND HETEROGENEITY.

SARAH EUGÈNE, THIBAUT BOURGERON, AND ZHOU XU

**ABSTRACT.** Replicative senescence, induced by telomere shortening, exhibits considerable asynchrony and heterogeneity, the origins of which remain unclear. Here, we formally study how telomere shortening mechanisms impact on senescence kinetics and define two regimes of senescence, depending on the initial telomere length variance. We provide analytical solutions to the model, highlighting a non-linear relationship between senescence onset and initial telomere length distribution. This study reveals the complexity of the collective behavior of telomeres as they shorten, leading to senescence heterogeneity.

Telomeres, the ends of eukaryote chromosomes, are poised in a dynamic equilibrium controlled by two processes: limited telomere shortening at each cell division and elongation by telomerase, a dedicated holoenzyme able to generate *de novo* telomere sequence. When telomerase is not expressed, as in human somatic cells, or is experimentally mutated in model organisms such as *Saccharomyces cerevisiae*, [18], telomeres only shorten and after many divisions the cell enters replicative senescence, a permanent cell cycle arrest induced by short telomeres that elicits a DNA damage response. Replicative senescence is implicated in organismal ageing and is a potent barrier to cancer emergence, but its remarkable asynchrony and heterogeneity remain a challenge for investigating the exact relationship between initial telomere length distribution and senescence onset.

Telomere shortening is the unavoidable consequence of the end replication problem, [22, 33, 31]. In most examined species, telomeres end with a 5' to 3' single-stranded DNA overhang (Fig. 1), [11, 13, 14, 19, 20, 26, 27, 34]. When the replication fork reaches the end of the chromosome, the processing of the last Okazaki fragment leaves a gap at the lagging strand, which recreates the single-stranded overhang of the parental telomere (Fig. 1). On the leading strand, after replication, complex maturation steps involving resection and fill-in also regenerate the overhang structure, [16, 8, 6, 36, 31]. Regardless of these maturation steps, the leading strand template for replication is shorter than the lagging strand one, thus generating after replication two new telomeres of different lengths, one unchanged compared to the parental telomere and the other shorter by exactly the length of the overhang, as illustrated in Fig. 1. Previous mathematical models of telomere shortening also based on the end-replication problem [17, 2, 21, 3] did not consider the maturation of the leading strand telomere that generates a 3'-end overhang identical to the one on the lagging strand. This maturation step is widely conserved throughout species with the notable exception of angiosperm plants that display a blunt end

---

*Date:* February 10, 2020.

at the leading telomere [27]. We also note that other mathematical models examined higher level structures such as t-loops [9, 28], or additional telomere states or breaking mechanisms [15, 29, 24, 25].

On average, telomeres shorten at a constant rate of exactly half of the overhang length per division. However, while studying the average is informative of the global regulation and homeostasis of telomere length, it misses important contributions of the asymmetry of telomere replication mechanism to the overall telomere length distribution and to the heterogeneity of the onset of senescence. Taking this asymmetry into account, the shortening of a telomere in a cell lineage, defined as a random succession of mitotically related cells, [38], is probabilistic and follows a Bernoulli process. Additionally, if the two ends of a given chromosome are considered together, the 3'-end at one telomere belongs to the same DNA strand as the 5'-end on the other telomere of the same chromosome, implying that the asymmetry at one telomere is inverted compared to the other (Fig. 1). We define this relationship between the two ends of the same chromosome as a coupling mechanism, which adds another layer of constrain and will also be modeled here.

In this Letter, we study the consequences of the asymmetry and the coupling on the distribution and the dynamics of telomere length in two distinct phases: at steady state in the presence of telomerase and in a strictly shortening phase without telomerase. We show that the robustness of telomerase recruitment impacts on the variance of the steady-state distribution of telomere length. In turn, this variance defines different regimes of senescence. In a regime of low initial variance, senescence onset cannot be linearly inferred from the average telomere length or even the length of the shortest telomere and we provide an asymptotic expansion to account for this phenomenon. In contrast, a high variance implies a linear correlation between the initial shortest telomere and senescence onset. We provide analytical solutions to the different models we describe and suggest applications for the inference of the initial telomere length distribution from experimental measurements of senescence onset.

We first describe the most general model, corresponding to a lineage of haploid yeast cells dividing in the presence of active telomerase. The two telomeres of a given chromosome at generation  $n$ , called  $(L_n^1, L_n^2)$ , are coupled as defined above and the 32 telomeres of the cell shorten according to a Bernoulli random variable  $B_n$  of parameter  $1/2$ : if  $B_n = 1$ , then  $L_n^1$  is shortened by  $a$  nucleotides, whereas  $L_n^2$  is preserved, and conversely if  $B_n = 0$ . Telomerase adds new telomere sequences preferentially to shorter telomeres, [32, 5], behavior that we capture by introducing  $C_n^i$ ,  $i \in \{1, 2\}$ , Bernoulli random variables of parameter  $f(L_n^i)$ , as done in [37], where  $f$  has the shape shown in Fig. 2 (a) and  $L_n^i$  is the length of the telomere at the extremity  $i$  before replication. The shape of  $f$  is such that below a length threshold  $L_s$ , the Bernoulli random variable  $C_n^i$  equals 1, that is telomerase is always active. For a telomere longer than  $L_s$ , the probability of  $\{C_n^i = 1\}$  decreases to zero, meaning that the longer the telomere, the less likely it is to be elongated by telomerase. Since the number of nucleotides added by telomerase is independent of the length of the telomere, [32, 37], we introduce  $\mathcal{G}_1^n$  and  $\mathcal{G}_2^n$  two independent geometric random variables of parameter  $p$ , independent of all the other quantities (including  $L_n^1, L_n^2$ ), which correspond to the number of nucleotides added by

telomerase. As a result, for any given chromosome the  $\mathbb{N}^2$ -valued process  $(L_n^1, L_n^2)$  follows:

$$(1) \quad \begin{pmatrix} L_{n+1}^1 \\ L_{n+1}^2 \end{pmatrix} = \begin{pmatrix} (L_n^1 - a \cdot B_n)^+ + C_n^1 \cdot \mathcal{G}_1^n \\ (L_n^2 - a \cdot (1 - B_n))^+ + C_n^2 \cdot \mathcal{G}_2^n \end{pmatrix},$$

where  $x^+ = \max(0, x)$  is the positive part of  $x$ . Using the same Bernoulli random variable  $B_n$  for  $L_n^1$  and  $L_n^2$  mathematically defines the coupling between the two telomeres.

To characterize the steady state of telomere length distribution, we focus on one telomere—because as an approximation, telomeres of different chromosomes are assumed to be independent, [30]—, and consider the projection of the first coordinate of a chromosome in order to compute its equilibrium. We will analyze the coupling effect in more depth in the second regime without telomerase. Our model thus becomes:

$$(2) \quad L_{n+1} = (L_n - a \cdot B_n)^+ + C_n \cdot \mathcal{G}_n$$

where  $L_n$  is the length of a given telomere,  $L_{n+1}$  the length of one of the two daughter telomeres,  $\mathcal{G}_n$  a geometric random variable of parameter  $p \in (0, 1)$ . An averaged version of this model has been studied in [37, 7], where instead of being stochastic, telomere shortening was chosen to be deterministic with a constant value of  $a/2$ . To make our computations fully explicit without betraying the principles of the biological mechanism, instead of  $f$ , we consider a sharp threshold at a value  $i_s$  (Fig. 2b). Our model becomes:

$$(3) \quad L_{n+1} = (L_n - a \cdot B_n)^+ + \mathcal{G}_n \cdot \mathbb{1}_{\{L_n \leq i_s\}}$$

Independently of the value of  $i_s$ , the Markov chain  $(L_n)_{n \geq 0}$  defined by (3) has a unique equilibrium distribution  $L_\infty$  and the generating function of  $L_\infty$  is characterized by the equality:

$$(4) \quad \begin{aligned} \forall u \in \mathbb{C}, |u| \leq 1, \quad & (u^a - 1)(1 - uq) \mathbb{E}(u^{L_\infty}) \\ & = -q(1 - u)(u^a + 1)S_{i_s}(u) + pS_{a-1}(1)u^a - pS_{a-1}(u), \end{aligned}$$

where  $\pi_j = \mathbb{P}(L_\infty = j)$ ,  $S_k(u) = \mathbb{E}(u^{L_\infty} \mathbb{1}_{\{L_\infty \leq k\}}) = \sum_{j=0}^k \pi_j u^j$  and  $q = 1 - p$ .

Appendix A gives a proof of this result and explains how to get a fully explicit expression for the distribution  $L_\infty$  from equation (4). This calculation reveals how the parameters  $a$  and  $p$  affect the steady state distribution.

The choice of the value of  $i_s$  according to biological experiments is explained in Appendix B. We find that the variance obtained using the simplified model (3) is significantly smaller than the one with the complete model (2) (37 bp as compared to 101 bp), demonstrating that the residual recruitment of telomerase to rather long telomeres strongly contributes to the spread of the steady state distribution of telomere length. In turn, the variance of this distribution is critical for determining the onset of senescence and its heterogeneity, as we show below. Thus, the mode of recruitment and activation of telomerase, dependent on the biochemical properties of the holoenzyme and on its interactions with telomeric proteins [35], controls key features of senescence once telomerase is removed.

We then analyze the consequences of the steady state distribution on the onset of senescence, meaning the number of generations undergone by a given cell lineage

until it enters senescence. We simply call it time of senescence, denoted by  $T$ . One practical goal of this section is to derive the parameters of the initial distribution from the time of senescence, which is useful for experimentalists. In senescing cells, telomerase is inactive and when the shortest telomere reaches a threshold  $S$ , the cell enters replicative senescence and stops dividing [1, 18, 12, 39, 4]. A haploid yeast cell has 16 chromosomes and thus 32 telomeres. Mathematically, we consider the vector  $(L_n^1, L_n^2, \dots, L_n^{32})$  of these 32 telomere lengths at generation  $n$ . Because each chromosome behaves independently [30], we can start by studying one chromosome and the behavior of the 16 will easily follow. More precisely, the vector  $(L_n^1, L_n^2, \dots, L_n^{32})$  can be seen as a family  $(X_n^i, Y_n^i)_{1 \leq i \leq 16}$  of 16 independent identically distributed couples each representing the two telomeres of a chromosome, with  $(X_0^i, Y_0^i) \stackrel{dist.}{\sim} \Pi$ . The time of senescence is mathematically expressed as:

$$T = \inf \left\{ n \geq 0, \min_{1 \leq i \leq 16} [\min(X_n^i, Y_n^i)] < S \right\}.$$

As telomeres can only shorten, we consider the shortening length to be  $a = 1$ . For numerical estimations of the time of senescence, we will divide our results by  $a = 7$  to obtain biologically relevant values. Moreover, we can choose  $S = 0$  by simply translating the initial state by  $S$ . These assumptions are made in all following calculations unless stated otherwise.

Under these normalization conditions, we find that the distribution of the random variable  $T$  is fully explicit:

$$(5) \quad \mathbb{P}(T > n) = \left[ \sum_{k+l \geq n} \Pi(X_0 = k, Y_0 = l) 2^{-n} \sum_{t=n-l}^k \binom{n}{t} \right]^{16}.$$

In particular, its expectation can be written as a function of  $\pi$  as follows:

$$(6) \quad \mathbb{E}(T) = \sum_{n=0}^{\infty} \left[ \sum_{k+l \geq n} \Pi(X_0 = k, Y_0 = l) 2^{-n} \sum_{t=n-l}^k \binom{n}{t} \right]^{16}.$$

See Appendix C for the proof. Because of the difficulty to invert this formula, we choose to study separately the influence of the mean and the variance of the initial state on the time of senescence. Thus, we first consider a deterministic and constant initial state  $X_0 = x_0$ :

$$k = 1, \dots, 32, \quad L_0^k = \mathbb{E}(L_\infty) =: x_0$$

We define  $T_{x_0}^1$  as the first time one of two coupled telomeres reaches zero both starting from  $x_0$ , and  $T_{x_0}$  as the time of senescence of the whole cell when the initial state is constant and equals  $x_0$ .

Almost surely,  $x_0 \leq T_{x_0}^1 \leq 2x_0$ , this implies that  $\mathbb{P}(T_{x_0}^1 > n) = 0$  for  $n \geq 2x_0$ , and  $\mathbb{P}(T_{x_0}^1 \geq n) = 1$  for  $n < x_0$ . For  $x_0 \leq n \leq 2x_0 - 1$ , the law of  $T_{x_0}^1$  is given by:

$$\mathbb{P}(T_{x_0}^1 > n) = 2^{-n} \sum_{k=n-x_0}^{x_0} \binom{n}{k}.$$

The expected time of senescence is then:

$$(7) \quad \mathbb{E}(T_{x_0}) = x_0 + \sum_{n=x_0}^{2x_0-1} [\mathbb{P}(T_{x_0}^1 > n)]^{16} = x_0 + \sum_{n=x_0}^{2x_0-1} \left[ 2^{-n} \sum_{k=n-x_0}^{x_0} \binom{n}{k} \right]^{16}.$$

We then perform an asymptotic expansion of  $\mathbb{E}(T_{x_0})$  for large values of  $x_0$ , which is numerically justified (Appendix B and Fig. 2c). At the first order the mean behavior prevails:

$$(8) \quad \mathbb{E}(T_{x_0}^1) \sim 2x_0.$$

Concerning the second order, we obtain the following convergence in distribution:

$$(9) \quad \frac{2x_0 - T_{x_0}^1}{\sqrt{x_0}} \xrightarrow[x_0 \rightarrow +\infty]{dist.} |N|, \quad \text{with } N \stackrel{dist.}{\sim} \mathcal{N}(0, 2).$$

See Appendix D for proofs of these results.

The asymptotic development of the time of senescence for one chromosome  $T_{x_0}^1$  allows us to derive an approximation of the expected time of senescence (7) by replacing the law of  $T_{x_0}^1$  by its asymptotic (9):

$$(10) \quad \mathbb{E}(T_{x_0}) \approx x_0 + \sum_{n=x_0}^{2x_0-1} [\mathbb{P}(2x_0 - \sqrt{x_0}|N| > n)]^{16} = x_0 + \sum_{k=0}^{x_0-1} \left[ \text{erf} \left( \frac{k}{2\sqrt{x_0}} \right) \right]^{16},$$

where erf is the error function defined as:

$$\text{erf } x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

We find that the expansion (10) is hardly distinguishable from the theoretical process (6), as shown in simulations (compare grey and dashed black lines in Fig. 3 (a)) and can thus be directly used to estimate the mean of the initial state in experimental studies.

Now, to study only the influence of the initial variance, we consider that each initial telomere is uniformly distributed in the interval  $[\mathbb{E}(L_\infty) - \sigma, \mathbb{E}(L_\infty) + \sigma]$  and simulate the expected time of senescence as a function of  $\sigma$  (Fig. 3b). When  $\sigma$  has large values, there is a higher probability that the initial shortest telomere of  $(L_0^1, \dots, L_0^{32})$  is far from the mean  $\mathbb{E}(L_\infty)$  and, thus, that it remains the shortest one until senescence. We therefore expect, for large enough value of  $\sigma$  (Fig. 3b), that the time of senescence is asymptotically equivalent to the time when the initial shortest telomere reaches zero. As, on average, the number of steps for a simple random walk starting from  $M = \mathbb{E}[\min_{1 \leq k \leq 32} L_k^0]$  to reach zero is  $M/(1/2) = 2M$ , we expect the following result:

$$(11) \quad \lim_{\sigma \rightarrow \mathbb{E}(L_\infty)} \mathbb{E}(T) = 2 \mathbb{E} \left[ \min_{1 \leq k \leq 32} L_k^0 \right].$$

We indeed find this asymptotic behavior by simulations (Fig. 3b), highlighting two regimes that depend on the initial variance. If  $\sigma$  has a small value, which is close to the deterministic initial state studied above (equation (10)), the time of senescence is much smaller than expected by just considering the shortest telomere because of the coupling effect. If  $\sigma$  has a large value, the time of senescence is mainly determined by the shortening of the initial shortest telomere.

We next ask which of these two regimes can be observed in simulated times of senescence from the telomere distributions described in Fig. 2, which are biologically more relevant than the previous distribution models. To do so, we simulate 1000 individual lineages of senescing cells and record their time of senescence, starting by randomly drawing their 32 telomeres from the biologically relevant distribution (large variance, in grey in Fig. 2c), from the simplified distribution (intermediate variance, in black in Fig. 2c) or from a constant distribution (no variance) (Fig. 4). We then compare these simulated times of senescence with those predicted either from the mean behavior of the shortest telomere (equation (11), dashed black lines in Fig. 4) or the asymptotic expansion on the mean of the initial distribution (equation (10), black lines in Fig. 4). The biologically relevant distribution gives simulated times of senescence that are fully predicted by computing the mean behavior of the average initial shortest telomere (Fig. 4a, compare grey and dashed black lines which are superimposed). In contrast, the constant distribution leads to a senescence onset dictated by the asymptotic expansion (Fig. 4c, compare grey and black lines), consistent with the results in Fig. 3a. The simplified distribution produces an intermediate result where the mean behavior of the shortest telomere and the asymptotic expansion lead to similar predictions, which lay close to the simulated times of senescence (Fig. 4b, compare grey, black and dashed black lines). These results show that defining two senescence regimes depending on the initial variance of telomere length distribution is critical for understanding the relevant dynamics of telomere shortening leading to senescence.

In summary, in this Letter, we isolated all the sources of fluctuations of the time of senescence that are dependent on telomere length. To do so, we modeled several molecular mechanisms that contribute at various levels to telomere length distribution and dynamics in *S. cerevisiae*, where they are the most exhaustively and quantitatively described. Among these mechanisms, we found that the asymmetry of telomere replication and the coupling between the two telomeres belonging to the same chromosome significantly contribute to senescence heterogeneity and we formally established their links. We also showed that the mode and robustness of telomerase recruitment control the variance of the steady-state telomere length distribution, which in turn defines two senescence regimes. With a low initial variance, the time of senescence is non-linearly related to the initial mean telomere length. In contrast, a high initial variance leads to a major role of the initial shortest telomere in controlling senescence. Because natural telomere length distributions can vary considerably, even within a species, we suggest that depending on the initial variance, the two regimes we describe may operate at the same time during senescence. This work uncovers a new layer of complexity in the relationship between senescence onset and telomere shortening explained by the asymmetry and coupling mechanisms, and proposes methods for assessing the time of senescence or conversely inferring parameters of the initial telomere length distribution.

**Acknowledgments.** The authors thank Marie Doumic, Lucas Gerin, Philippe Robert, Maria Teresa Teixeira for their critical reading of the manuscript. Z. X.'s research is supported by the Initiative d'Excellence program (Grant DYNAMO, ANR-11-LABX-0011-01), the Edmond de Rothschild foundation and Fondation pour la recherche médicale (FRM équipe labellisée DEQ20160334914 to the Teixeira lab). S. E.'s research is supported by ERC Starting Grant SKIPPER<sup>AD</sup> No. 306321. T. B.'s research is supported by ERC Starting Grant MESOPROBIO.

## 1. FIGURES

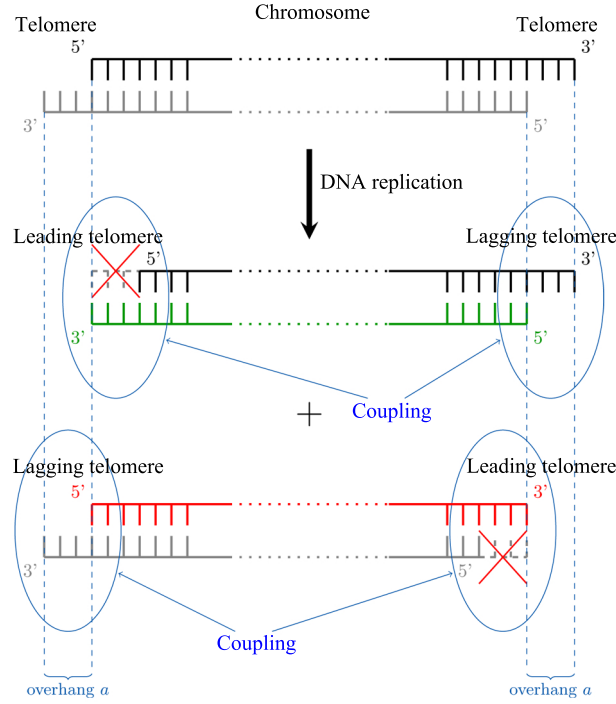


Figure 1: Scheme of a chromosome bearing two telomeres and undergoing replication. Telomeres end with a 3' overhang of length  $a$  (measured to be 5 – 10 nucleotides in yeast [31], chosen here as  $a = 1$  or 7 for theoretical or numerical purposes, respectively). After DNA replication, each telomere generates, through either the leading or the lagging strand replication machineries, two new telomeres of different lengths. The coupling effect between the two ends of the same chromosome imposes that only one of the two is shortened while the other retains the parental length.

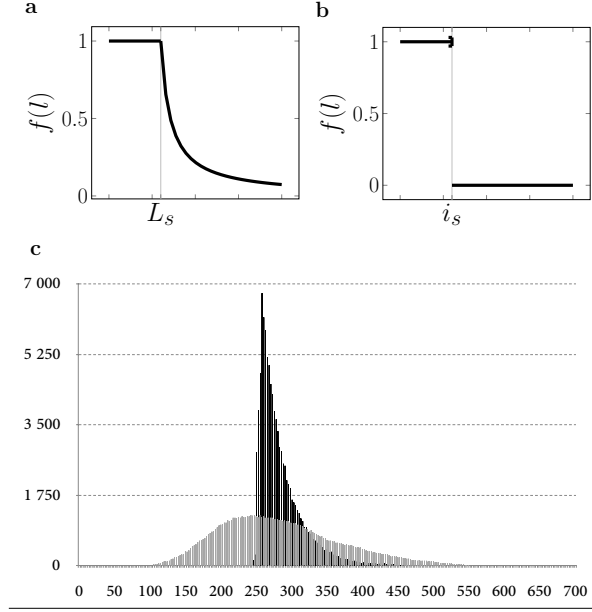


Figure 2: Steady-state telomere length distribution in the presence of telomerase. (a) and (b) Probability of recruitment and action of telomerase, as modeled from [32] with a length threshold  $L_s$  or simplified with a sharp switch occurring at  $i_s$ . (c) Simulation of the steady-state distribution of telomere length using either (a) (black) or (b) (grey) to describe telomerase recruitment.  $i_s$  was set so as to reach the same mean in the steady state distribution (Appendix B).

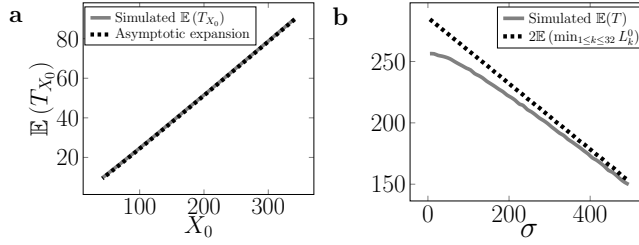


Figure 3: Distinct effects of the mean and variance of the initial distribution on theoretical expressions and numerical simulations of the time of senescence. (a) Starting from a constant distribution  $\mathbb{E}(L_\infty) := x_0$ , the asymptotic expansion in equation (10) is computed and compared to numerical simulations (1000 independent simulations). (b) Starting from a uniform distribution of variance  $\sigma$  and mean  $\mathbb{E}(L_\infty)$ , the time of senescence is computed using equation (11), which takes only the mean behavior of the initial shortest telomere into account, and compared to numerical simulations (1000 independent simulations).



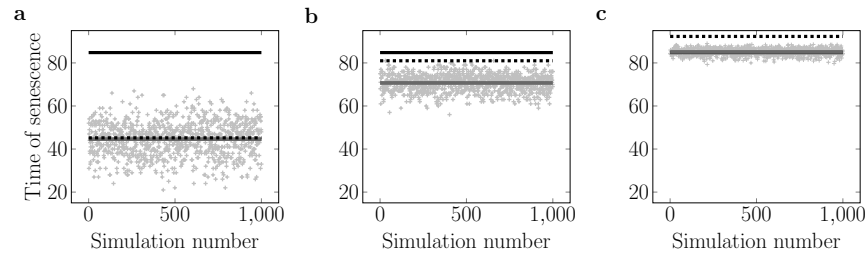


Figure 4: Comparison between simulated times of senescence (grey dots) and predictions from equations (11) (dashed black lines) and (10) (black lines). (a) 32 telomere lengths are randomly drawn from a biologically relevant distribution with a high variance (Fig. 2c, grey distribution) and the time of senescence is simulated to give one data point (grey dot). This process is repeated 1000 times and compared to the two predictions. The grey line represents the average simulated time of senescence. (b) and (c): as in (a), but starting with an intermediate level of variance for the initial telomere length distribution or no variance at all, respectively.

## REFERENCES

- [1] Pauline Abdallah, Pierre Luciano, Kurt W. Runge, Michael Lisby, Vincent Géli, Eric Gilson, and M. Teresa Teixeira. A two-step model for senescence triggered by a single critically short telomere. *Nature Cell Biology*, 11(8):988–993, 2009.
- [2] Ovide Arino, Marek Kimmel, and Glenn F. Webb. Mathematical modeling of the loss of telomere sequences. *Journal of theoretical biology*, 177(1):45–57, 1995.
- [3] Natalie Arkus. A mathematical model of cellular apoptosis and senescence through the dynamics of telomere loss. *Journal of theoretical biology*, 235(1):13–32, 2005.
- [4] Mary Armanios, Jonathan K. Alder, Erin M. Parry, Baktiar Karim, Margaret A. Strong, and Carol W. Greider. Short telomeres are sufficient to cause the degenerative defects associated with aging. *The American Journal of Human Genetics*, 85(6):823–832, 2009.
- [5] Bethan Britt-Compton, Rebecca Capper, Jan Rowson, and Duncan M. Baird. Short telomeres are preferentially elongated by telomerase in human cells. *FEBS Letter*, 583(18):3076–3080, 2009.
- [6] Weihang Chai, Agnel J. Sfeir, Hirotoshi Hoshiyama, Jerry W. Shay, and Woodring E. Wright. The involvement of the Mre11/Rad50/Nbs1 complex in the generation of G-overhangs at human telomeres. *EMBO reports*, 7(2):225–230, 2006.
- [7] Khanh Dao Duc and David Holcman. Computing the Length of the Shortest Telomere in the Nucleus. *Physical Review Letters*, 111(22):228104, 2013.
- [8] Virginie Faure, Stéphane Coulon, Julien Hardy, and Vincent Géli. Cdc13 and telomerase bind through different mechanisms at the lagging-and leading-strand telomeres. *Molecular cell*, 38(6):842–852, 2010.
- [9] Jack D. Griffith, Laurey Comeau, Soraya Rosenfield, Rachel M. Stansel, Alessandro Bianchi, Heidi Moss, and Titia De Lange. Mammalian telomeres end in a large duplex loop. *Cell*, 97(4):503–514, 1999.
- [10] Martin Hairer. Convergence of Markov Processes. Accessed: 2016-06-19, 2016.
- [11] Michael T. Hemann and Carol W. Greider. G-strand overhangs on telomeres in telomerase-deficient mouse cells. *Nucleic acids research*, 27(20):3964–3969, 1999.
- [12] Michael T. Hemann, Margaret A. Strong, Ling-Yang Hao, and Carol W. Greider. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell*, 107(1):67–77, 2001.
- [13] E. R. Henderson and E. H. Blackburn. An overhanging 3’ terminus is a conserved feature of telomeres. *Molecular and Cellular Biology*, 9(1):345–348, 1989.

- [14] Lawrence A. Klobutcher, Marshal T. Swanton, Pierluigi Donini, and David M. Prescott. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proceedings of the National Academy of Sciences*, 78(5):3015–3019, 1981.
- [15] A. Kowald. Possible mechanisms for the regulation of telomere length. *Journal of molecular biology*, 273(4):814–825, 1997.
- [16] Michel Larrivée, Catherine LeBel, and Raymund J. Wellinger. The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. *Genes & Development*, 18(12):1391–1396, 2004.
- [17] Michael Z. Levy, Richard C. Allsopp, A. Bruce Futcher, Carol W. Greider, and Calvin B. Harley. Telomere end-replication problem and cell aging. *Journal of molecular biology*, 225(4):951–960, 1992.
- [18] Victoria Lundblad and Jack W. Szostak. A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell*, 57(4):633–643, 1989.
- [19] Vladimir L. Makarov, Yoko Hirose, and John P. Langmore. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell*, 88(5):657–666, 1997.
- [20] Richard McElligott and Raymund J. Wellinger. The terminal DNA structure of mammalian chromosomes. *The EMBO journal*, 16(12):3705–3714, 1997.
- [21] Peter Olofsson and Marek Kimmel. Stochastic models of telomere shortening. *Mathematical biosciences*, 158(1):75–92, 1999.
- [22] Alexei M. Olovnikov. A theory of marginotomy: the incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *Journal of theoretical biology*, 41(1):181–190, 1973.
- [23] Philippe Robert. *Stochastic Networks and Queues*. Stochastic Modelling and Applied Probability Series. Springer-Verlag, New York, 2003.
- [24] Carole J. Proctor and Thomas B. L. Kirkwood. Modelling telomere shortening and the role of oxidative stress. *Mechanisms of ageing and development*, 123(4):351–363, 2002.
- [25] Carole J. Proctor and Thomas B. L. Kirkwood. Modelling cellular senescence as a result of telomere state. *Aging cell*, 2(3):151–157, 2003.
- [26] Marcela Raices, Ramiro E. Verdun, Sarah A. Compton, Candy I. Haggblom, Jack D. Griffith, Andrew Dillin, and Jan Karlseder. C. elegans telomeres contain G-strand and C-strand overhangs that are bound by distinct proteins. *Cell*, 132(5):745–757, 2008.
- [27] Karel Riha, Thomas D. McKnight, Jiri Fajkus, Boris Vyskot, and Dorothy E. Shippen. Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. *The Plant Journal*, 23(5):633–641, 2000.
- [28] Ignacio A. Rodriguez-Brenes and Charles S. Peskin. Quantitative theory of telomere length regulation and cellular senescence. *Proceedings of the National Academy of Sciences*, 107(12):5387–5392, 2010.
- [29] Ivica Rubelj and Zoran Vondracek. Stochastic mechanism of cellular aging; abrupt telomere shortening as a model for stochastic nature of cellular aging. *Journal of theoretical biology*, 197(4):425–438, 1999.
- [30] Janis Shampay and Elizabeth H. Blackburn. Generation of telomere-length heterogeneity in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 85(2):534–538, 1988.
- [31] Julien Soudet, Pascale Jolivet, and M. Teresa Teixeira. Elucidation of the DNA end-replication problem in *Saccharomyces cerevisiae*. *Molecular cell*, 53(6):954–964, 2014.
- [32] M. Teresa Teixeira, Milica Arneric, Peter Sperisen, and Joachim Lingner. Telomere length homeostasis is achieved via a switch between telomerase -extendible and -nonextendible states. *Cell*, 117(3):323–35, 2004.
- [33] Watson. Origin of concatemeric T7 DNA. *Nat. New Biol.*, 239(94):197–201, 1972.
- [34] Raymund J. Wellinger, Alexander J. Wolf, and Virginia A. Zakian. *Saccharomyces* telomeres acquire single-strand TG 1–3 tails late in S phase. *Cell*, 72(1):51–60, 1993.
- [35] Raymund J. Wellinger and Virginia A. Zakian. Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. *Genetics*, 191(4):1073a–1105, 2012.
- [36] Peng Wu, Hiroyuki Takai, and Titia de Lange. Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell*, 150(1):39–52, 2012.

- [37] Zhou Xu, Khanh Dao Duc, David Holcman, and M. Teresa Teixeira. The Length of the Shortest Telomere as the Major Determinant of the Onset of Replicative Senescence. *Genetics*, 194(4):847–857, 2013.
- [38] Zhou Xu, Emilie Fallet, Camille Paoletti, Steffen Fehrmann, Gilles Charvin, and M. Teresa Teixeira. Two routes to senescence revealed by real-time analysis of telomerase-negative single lineages. *Nature communications*, 6, 2015.
- [39] Ying Zou, Agnel Sfeir, Sergei M. Gryaznov, Jerry W. Shay, and Woodring E. Wright. Does a sentinel or a subset of short telomeres determine replicative senescence? *Molecular biology of the cell*, 15(8):3709–3718, 2004.

*E-mail address:* Sarah.Eugene@inria.fr

*E-mail address:* thibault.bourgeron@ens-lyon.fr

*E-mail address:* zhou.xu@ibpc.fr

(S. Eugène) SORBONNE UNIVERSITÉS, UPMC UNIVERSITÉ PIERRE ET MARIE CURIE, UMR 7598, LABORATOIRE JACQUES-LOUIS LIONS, F-75005, PARIS, FRANCE AND INRIA PARIS, 2 RUE SIMONE IFF, F-75012 PARIS, FRANCE

(T. Bourgeron) ÉCOLE NORMALE SUPÉRIEURE DE LYON, UMR 5569, UNITÉ DE MATHÉMATIQUES PURES ET APPLIQUÉES, 69007 LYON, FRANCE AND INRIA NUMED, 46 ALLÉE D’ITALIE, 69007 LYON, FRANCE

(Z. Xu) SORBONNE UNIVERSITÉS, UPMC UNIV PARIS 06, CNRS, UMR8226, LABORATOIRE DE BIOLOGIE MOLÉCULAIRE ET CELLULAIRE DES EUCARYOTES, INSTITUT DE BIOLOGIE PHYSICO-CHIMIQUE, 75005 PARIS, FRANCE

# EFFECTS OF INITIAL TELOMERE LENGTH DISTRIBUTION ON SENESCENCE ONSET AND HETEROGENEITY.

SARAH EUGÈNE, THIBAUT BOURGERON, AND ZHOU XU

## Appendices

### A. STEADY STATE OF TELOMERES EVOLVING WITH TELOMERASE

First, we prove that both Markov chains defined by (2), (3) are ergodic and, second, we derive formula (4). The ergodicity is a direct consequence of Foster-Lyapunov criteria (Corollary 8.7 p. 214 in [?] or Proposition 1.3 in [?], for instance). The chains are time-homogeneous and, for  $L_0 > a$ :

$$\mathbb{E}(L_1 - L_0) = \mathbb{E}(-a \cdot B_0 + C_0 \cdot \mathcal{G}_0) = -a/2 + f(L_0) \mathbb{E}(\mathcal{G}_0),$$

and this last quantity is negative for large enough values of  $L_0$  because  $f$  tends to zero, either  $f$  is  $f(l) = (1 + \beta(l - L_s))^{-1}$  or  $f(l) = \mathbb{1}_{l \leq i_s}$ .

Moreover, the Markov chains  $(L_n)$  defined by (2), (3) are irreducible and aperiodic. Therefore, in both cases, there exists a unique equilibrium distribution, denoted  $L_\infty$ . For the model (3), it is characterized by the fact that  $L_\infty$  and  $(L_\infty - a \cdot B_0)^+ + \mathcal{G}_0 \mathbb{1}_{\{L_\infty \leq i_s\}}$  have the same probability generating function.

To establish formula (4) we distinguish the three regimes:  $L_\infty < a \cdot B$ ,  $a \cdot B \leq L_\infty \leq i_s$ ,  $L_\infty > i_s$ , and the cases  $B = 0$  or  $B = 1$ . For  $u$  such that  $|u| \leq 1$ , we obtain:

(12)

$$\begin{aligned} 2\mathbb{E}(u^{L_\infty}) &= 2\mathbb{E}\left(u^{(L_\infty - a \cdot B_0)^+ + \mathcal{G}_0 \cdot \mathbb{1}_{L_\infty \leq i_s}}\right) \\ &= 2\mathbb{E}\left(u^{\mathcal{G}_0} \mathbb{1}_{\{L_\infty < a \cdot B_0\}}\right) + 2\mathbb{E}\left(u^{L_\infty - a \cdot B_0 + \mathcal{G}_0} \mathbb{1}_{\{a \cdot B_0 \leq L_\infty \leq i_s\}}\right) + 2\mathbb{E}\left(u^{L_\infty - a \cdot B_0} \mathbb{1}_{\{L_\infty > i_s\}}\right) \\ &= \mathbb{E}\left(u^{\mathcal{G}_0}\right) \left[\mathbb{P}(L_\infty < a) + \mathbb{P}(L_\infty < 0)\right] + \mathbb{E}\left(u^{\mathcal{G}_0}\right) \left[\mathbb{E}\left(u^{L_\infty - a} \mathbb{1}_{\{a \leq L_\infty \leq i_s\}}\right)\right. \\ &\quad \left.+ \mathbb{E}\left(u^{L_\infty} \mathbb{1}_{\{0 \leq L_\infty \leq i_s\}}\right)\right] + \left[\mathbb{E}\left(u^{L_\infty - a} \mathbb{1}_{\{L_\infty > i_s\}}\right) + \mathbb{E}\left(u^{L_\infty} \mathbb{1}_{\{L_\infty > i_s\}}\right)\right] \\ &= \mathbb{E}\left(u^{\mathcal{G}_0}\right) S_{a-1}(1) + \mathbb{E}\left(u^{\mathcal{G}_0}\right) \left[u^{-a} (S_{i_s}(u) - S_{a-1}(u)) + S_{i_s}(u)\right] \\ &\quad + \left[u^{-a} (\mathbb{E}(u^{L_\infty}) - S_{i_s}(u)) + (\mathbb{E}(u^{L_\infty}) - S_{i_s}(u))\right]. \end{aligned}$$

From (12), we get:

$$\mathbb{E}(u^{L_\infty}) (1 - u^{-a}) = S_{i_s}(u) (\mathbb{E}(u^{\mathcal{G}_0}) - 1) (1 + u^{-a}) + \mathbb{E}(u^{\mathcal{G}_0}) (S_{a-1}(1) - u^{-a} S_{a-1}(u)).$$

---

*Date:* February 10, 2020.

As the probability generating function of a geometric distribution is explicitly given by  $\mathbb{E}(u^{\mathcal{G}_0}) = p(1 - uq)^{-1}$ , with  $q = 1 - p$ , we obtain (4) after multiplication by  $(1 - uq)u^a$ .

In order to compute the  $\pi_k$  for all  $k$ , we identify the coefficients of the power series of each side of (4) distinguishing cases for the values of  $k$ . For simplicity we set  $\pi_k = 0$  for  $k < 0$ . Using the identities:  $(u^a - 1)(1 - uq) = -qu^{a+1} + u^a + qu - 1$ ,  $(u - 1)(u^a + 1) = u^{a+1} - u^a + u - 1$ , we get:

$$\begin{aligned}
(13) \quad & \sum_{k=0}^{\infty} (-q\pi_{k-a-1} + \pi_{k-a} + q\pi_{k-1} - \pi_k) u^k = \sum_{k=0}^{\infty} \pi_k (-qu^{k+a+1} + u^{k+a} + qu^{k+1} - u^k) \\
& = (-qu^{a+1} + u^a + qu - 1) \mathbb{E}(u^{L_{\infty}}) \\
& = q(u^{a+1} - u^a + u - 1) S_{i_s}(u) + pS_{a-1}(1)u^a - pS_{a-1}(u) \\
& = q \left[ \sum_{k=a+1}^{i_s+a+1} \pi_{k-a-1} u^k - \sum_{k=a}^{i_s+a} \pi_{k-a} u^k + \sum_{k=1}^{i_s+1} \pi_{k-1} u^k - \sum_{k=0}^{i_s} \pi_k u^k \right] \\
& \quad + pS_{a-1}(1) u^a - p \sum_{k=0}^{a-1} \pi_k u^k.
\end{aligned}$$

The following table gives the recurrence relations obtained after identification in the coefficients of (13).

| $k$                                  | relation  |
|--------------------------------------|---|
| $\llbracket 0, a-1 \rrbracket$       | $\emptyset$   |
| $a$                                  | $\pi_a = 2\frac{1-p}{p}\pi_0 - \sum_{k=1}^{a-1} \pi_k$        |
| $\llbracket a+1, i_s \rrbracket$     | $\pi_k = -2\frac{1-p}{p}\pi_{k-a-1} + \frac{2-p}{p}\pi_{k-a}$ |
| $i_s + 1$                            | $\pi_{i_s+1} = -2q\pi_{i_s-a} + (1+q)\pi_{i_s+1-a}$           |
| $\llbracket i_s+2, i_s+a \rrbracket$ | $\pi_k = -2q\pi_{k-a-1} + (1+q)\pi_{k-a} + q\pi_{k-1}$        |
| $i_s + a + 1$                        | $\pi_{i_s+a+1} = -2q\pi_{i_s} + \pi_{i_s+1} + q\pi_{i_s+a}$   |
| $> i_s + a + 1$                      | $\pi_k = -q\pi_{k-a-1} + \pi_{k-a} + q\pi_{k-1}$              |

Note that for  $k$  such that  $0 \leq k \leq a-1$ , the identification gives no information on the values of the  $\pi_k$ . These formulas show that all the  $\pi_k$  depend linearly on the  $a$  first states,  $\pi_k$ ,  $k = 0, \dots, a-1$ . Now, we indicate how to compute these  $a$  first values.

Hence, using (4), the generating function of  $L_{\infty}$  is only a function of the  $a$  first states. Dividing (4) by  $(1-u)(1-uq)$ , we can find  $\psi : \mathbb{R}^a \times [0, 1] \rightarrow \mathbb{R}$  linear in the first  $a$  coordinates such that:

$$(14) \quad (1 + \dots + u^{a-1}) \mathbb{E}(u^{L_{\infty}}) = \psi(\pi_0, \dots, \pi_{a-1}, u).$$

The  $a-1$  roots of  $R(u) = 1 + \dots + u^{a-1}$  are the  $u_k = e^{2i\pi k/a}$  for  $1 \leq k \leq a-1$ , which are in the unit disk. Therefore, the vector  $(\pi_0, \dots, \pi_{a-1})$  is solution of the system:

$$(15) \quad \psi(\pi_0, \dots, \pi_{a-1}, u_k) = 0 \text{ for } 1 \leq k \leq a-1$$

where  $\pi_0$  is, as usual determined, by the normalization condition. If the  $a \times a$  system (15) to which we added the normalization condition is invertible, then

there exists a unique solution  $(\pi_0, \dots, \pi_{a-1})$ . Having the vector  $(\pi_0, \dots, \pi_{a-1})$ , the generating function  $\mathbb{E}(L_\infty)$  follows from (14).

Finally, we want to explicitly determine the  $\pi_k$  for  $k \geq a$ . The previous table gives homogeneous linear recurrence relations with constant coefficients. For instance  $\pi_k$ , for  $k$  in  $\llbracket a+1, i_s \rrbracket$ , is a linear combination (independent of  $k$ ) of the  $k$ -th powers of the roots of the (conjugate of the) characteristic polynomial  $cX^{a+1} - (1+c)X^a + 1$ , where  $c = 2q/p \in (0, \infty)$ . This polynomial is of degree  $a+1$  and has 1 as a root. This last property is also true for the characteristic polynomials of the relations for the  $\pi_k$  for  $k$  in  $\llbracket i_s+2, i_s+a \rrbracket$  or  $> i_s+a+1$ . As the quartic equation has explicit solutions, the expressions of the  $\pi_k$ ,  $k \geq a$  are fully explicit if  $a+1 \leq 4+1$ , that is  $a \leq 4$ . In particular for  $a = 1$ , the  $(\pi_k)_{k \in \llbracket 1, i_s \rrbracket}$  and  $(\pi_k)_{k > i_s}$  are two geometric progressions:

$$(16) \quad \forall k \in \llbracket 1, i_s \rrbracket, \pi_k = c^k \pi_0, \quad \forall k > i_s, \pi_k = p(1-p)^k \left(\frac{2}{p}\right)^{i_s+1} \pi_0.$$

For  $a > 4$  the roots of these polynomials can be numerically found.

#### B. CHOICE OF $i_s$

To rigorously compare the variance of the simplified model (3) with the one of (2), we choose  $i_s$  so that the ceiling function of the mean of the equilibrium of (3)  $\lceil \mathbb{E}(L_\infty) \rceil$  is the same as the one for (2), *i.e.* 342 bp in [?]. We take the biological parameters obtained in [?, ?] and used in [?]:

$$a = 7, p = 0.026, L_s = 90, \beta = 0.045.$$

In order to compute the corresponding mean of the equilibrium, for each choice of  $i_s$  we run  $10^6$  numerical simulations of (3). This mean is then computed and plotted as a function of  $i_s$ , *cf.* Fig. (1) Finally, we chose the value of  $i_s$  that gives  $\lceil \mathbb{E}(L_\infty) \rceil = 342$  bp. This procedure leads to  $i_s = 308$  bp.

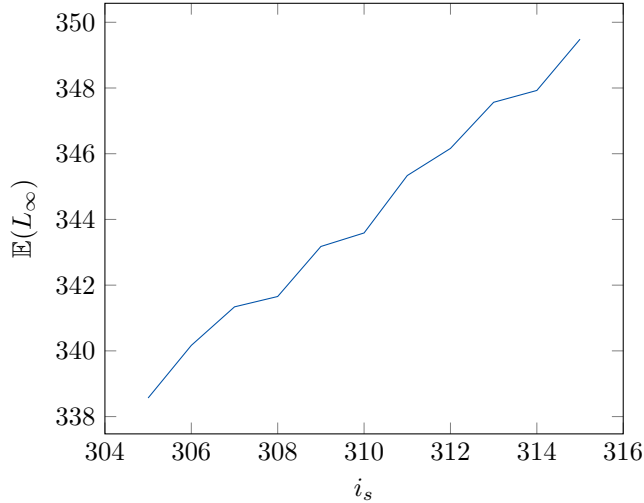


Figure 1: The expected length  $\mathbb{E}(L_\infty)$  as function of  $i_s$ .

### C. EXPECTED TIME OF SENESENCE

The aim of this section is to establish formula (5). Before entering the details just note that taking  $a = 1$  is not strictly equivalent to make a change in the length unit as the lengths of the telomeres are integer-valued. But taking  $a > 1$  only make the results more difficult to state without exhibiting a new behaviour. For the sake of simplicity, we drop the superscript  $i$  in this section and start by studying a typical couple  $(X_n, Y_n)$ . The shortening of these two telomeres can be mathematically translated into the following model:

$$(17) \quad \begin{pmatrix} X_{n+1} \\ Y_{n+1} \end{pmatrix} = \begin{pmatrix} (X_n - a \cdot B_n)^+ \\ (Y_n - a \cdot (1 - B_n))^+ \end{pmatrix},$$

where  $B_n$  is a Bernoulli random variable of parameter  $1/2$ , and  $(X_0, Y_0) \stackrel{dist.}{\sim} \Pi$ .

This process is an oriented simple random walk on  $\mathbb{Z}^2$  until one of the coordinates reaches zero, and can be written explicitly:

$$(18) \quad \begin{aligned} X_n &= X_0 - a \sum_{k=1}^n B_k = X_0 - a B(n, 1/2), \\ Y_n &= Y_0 - a \sum_{k=1}^n (1 - B_k) = Y_0 - a (n - B(n, 1/2)), \end{aligned}$$

where  $B(n, 1/2)$  is a binomial distribution of parameters  $n$  and  $1/2$ . In this case, let us define the first time one of the coordinates reaches zero,  $T^1$ , as:

$$T^1 = \inf \{n \geq 0, \min(X_n, Y_n) < 0\}$$

Then, from (18) and for  $a = 1$ , because  $(X_n)$  and  $(Y_n)$  are non-increasing, we get:

$$(19) \quad \begin{aligned} \mathbb{P}(T^1 > n) &= \mathbb{P}(X_n \geq 0, Y_n \geq 0) = \mathbb{P}(n - Y_0 \leq B(n, 1/2) \leq X_0) \\ &= \sum_{\substack{k+l \geq n \\ k, l \geq 0}} \Pi(X_0 = k, Y_0 = l) 2^{-n} \sum_{t=n-l}^k \binom{n}{t}. \end{aligned}$$

From here, we easily derive the distribution of the time of senescence by considering all 16 independent pairs of telomeres:

$$(20) \quad \begin{aligned} \mathbb{P}(T > n) &= \mathbb{P}(\neg \{\text{senescence at the } n^{\text{th}} \text{ generation}\}) \\ &= \mathbb{P}\left(\min_{1 \leq k \leq 32} L_k^n \geq 0\right) = \mathbb{P}(\forall i \in \llbracket 1, 16 \rrbracket, \min(X_n^i, Y_n^i) \geq 0) \\ &= \mathbb{P}(\min(X_n, Y_n) \geq 0)^{16} = \mathbb{P}(T^1 > n)^{16}. \end{aligned}$$

Formulas (19) and (20) lead to (5), which gives the expected time of senescence (6) using  $\mathbb{E}(T) = \sum_{n=0}^{\infty} \mathbb{P}(T > n)$ .

### D. ASYMPTOTICS OF THE MEAN TIME OF SENESENCE

The aim of this section is to prove formulas (9) and (10). To find the second order of the expansion, we consider a coupling  $(\bar{X}_n, \bar{Y}_n)$  such that for  $n \leq T_{X_0}^1$ ,

$$\mathbb{P}[(X_n, Y_n) = (x, y)] = \mathbb{P}[(\bar{X}_n, \bar{Y}_n) = (x, y)]$$

and  $(\bar{X}_n, \bar{Y}_n)$  is the unbounded random walk on  $\mathbb{Z}^2$ .

For convenience, the subscript  $n$  of the random walks  $(X_n, Y_n)$  and  $(\bar{X}_n, \bar{Y}_n)$  will

from now on be put as an argument, *i.e* we will rather write  $(X(n), Y(n))$ .

We take an affine interpolation of the discrete random walks  $(\bar{X}_n, \bar{Y}_n)$  to make time continuous and we use classical Donsker's theorem, which shows for  $(B_i)_i$  a sequence of independent Bernoulli random variables of parameter  $1/2$ , for the convergence in distribution of processes,

$$\left( \frac{\sum_{i=1}^{\lfloor tX_0 \rfloor} (B_i - 1/2)}{\sqrt{X_0}} \right)_{X_0 \rightarrow +\infty} \xrightarrow{\mathcal{L}} \frac{1}{2}(\mathcal{W}_t)$$

where  $(\mathcal{W}_t)$  is a standard Brownian Motion.

Therefore, for the convergence in distribution, of processes again,

$$\lim_{X_0 \rightarrow +\infty} \left( \frac{X_0 - \bar{X}(\lfloor tX_0 \rfloor) - \lfloor tX_0 \rfloor/2}{\sqrt{X_0}}, \frac{X_0 - \bar{Y}(\lfloor tX_0 \rfloor) - \lfloor tX_0 \rfloor/2}{\sqrt{X_0}} \right) = \left( \frac{1}{2}\mathcal{W}_t, -\frac{1}{2}\mathcal{W}_t \right)$$

We can now compute the asymptotic distribution of  $T_{X_0}^1$ . By definition, since  $(X_n)$  and  $(Y_n)$  are decreasing, and by the previous coupling,

$$\begin{aligned} \mathbb{P} \left( \frac{2X_0 - T_{X_0}^1}{\sqrt{X_0}} < w \right) &= \mathbb{P} \left( T_{X_0}^1 > 2X_0 - w\sqrt{X_0} \right) = \mathbb{P} \left( T_{X_0}^1 > \left\lfloor 2X_0 - w\sqrt{X_0} \right\rfloor \right) \\ &= \mathbb{P} \left( X \left( \left\lfloor X_0 \left( 2 - \frac{w}{\sqrt{X_0}} \right) \right\rfloor \right) \geq 0, Y \left( \left\lfloor X_0 \left( 2 - \frac{w}{\sqrt{X_0}} \right) \right\rfloor \right) \geq 0 \right) \\ &= \mathbb{P} \left( \bar{X} \left( \left\lfloor X_0 \left( 2 - \frac{w}{\sqrt{X_0}} \right) \right\rfloor \right) \geq 0, \bar{Y} \left( \left\lfloor X_0 \left( 2 - \frac{w}{\sqrt{X_0}} \right) \right\rfloor \right) \geq 0 \right). \end{aligned}$$

Therefore,

$$\begin{aligned} \lim_{X_0 \rightarrow \infty} \mathbb{P} \left( \frac{2X_0 - T_{X_0}^1}{\sqrt{X_0}} < w \right) &= \lim_{X_0 \rightarrow \infty} \mathbb{P} \left( X_0 - \frac{\lfloor (2 - w/\sqrt{X_0})X_0 \rfloor}{2} - \sqrt{X_0} \frac{1}{2} \mathcal{W}_2 \geq 0, \right. \\ &\quad \left. X_0 - \frac{\lfloor (2 - w/\sqrt{X_0})X_0 \rfloor}{2} + \sqrt{X_0} \frac{1}{2} \mathcal{W}_2 \geq 0 \right) \\ &= \mathbb{P} (w + \mathcal{W}_2 \geq 0, w - \mathcal{W}_2 \geq 0) \\ &= \mathbb{P} (|\mathcal{W}_2| \leq w) \end{aligned}$$

so that for the convergence in distribution, when  $X_0$  tends to infinity:

$$\frac{2X_0 - T_{X_0}^1}{\sqrt{X_0}} \longrightarrow |\mathcal{W}_2|$$

and:

$$(21) \quad T_{X_0}^1 \underset{X_0 \rightarrow \infty}{\overset{\mathcal{L}}{\sim}} 2X_0 - \sqrt{X_0} |\mathcal{W}_2|,$$

where  $\mathcal{W}_2$  is distributed according to  $\mathcal{N}(0, 2)$ . The speed of convergence is given by the Berry-Essen theorem in terms of the cumulative distribution function of  $(2x_0 - T_{x_0}^1)/\sqrt{x_0}$ , say  $F_{x_0}$ :

$$\forall x \in \mathbb{R}, \forall x_0 \in \mathbb{N} \quad |F_{x_0}(x) - \text{erf } x| \leq \frac{C}{\sqrt{x_0}},$$



where  $C$  is an absolute constant. This inequality should provide a bound to the difference in equation (10)

$$\left| \mathbb{E}(T_{x_0}) - \left( x_0 + \sum_{k=0}^{x_0-1} \left[ \operatorname{erf} \left( \frac{k}{2\sqrt{x_0}} \right) \right]^{16} \right) \right|.$$

*E-mail address:* `Sarah.Eugene@inria.fr`

*E-mail address:* `thibault.bourgeron@ens-lyon.fr`

*E-mail address:* `zhou.xu@ibcp.fr`

(S. Eugène) SORBONNE UNIVERSITÉS, UPMC UNIVERSITÉ PIERRE ET MARIE CURIE, UMR 7598, LABORATOIRE JACQUES-LOUIS LIONS, F-75005, PARIS, FRANCE AND INRIA PARIS, 2 RUE SIMONE IFF, F-75012 PARIS, FRANCE

(T. Bourgeron) ÉCOLE NORMALE SUPÉRIEURE DE LYON, UMR 5569, UNITÉ DE MATHÉMATIQUES PURES ET APPLIQUÉES, 69007 LYON, FRANCE AND INRIA NUMED, 46 ALLÉE D'ITALIE, 69007 LYON, FRANCE

(Z. Xu) SORBONNE UNIVERSITÉS, UPMC UNIV PARIS 06, CNRS, UMR8226, LABORATOIRE DE BIOLOGIE MOLÉCULAIRE ET CELLULAIRE DES EUCARYOTES, INSTITUT DE BIOLOGIE PHYSICO-CHIMIQUE, 75005 PARIS, FRANCE