

## A CRUMP–MODE–JAGERS BRANCHING PROCESS MODEL OF PRION LOSS IN YEAST

PETER OLOFSSON,\* *Trinity University*

SUZANNE S. SINDI,\*\* *University of California, Merced*

### Abstract

The yeast *Saccharomyces cerevisiae* has emerged as an ideal model system to study the dynamics of prion proteins which are responsible for a number of fatal neurodegenerative diseases in humans. Within an infected cell, prion proteins aggregate in complexes which may increase in size or be fragmented and are transmitted upon cell division. Recent work in yeast suggests that only aggregates below a critical size are transmitted efficiently. We formulate a continuous-time branching process model of a yeast colony under conditions of prion curing. We generalize previous approaches by providing an explicit formula approximating prion loss as influenced by both aggregate growth and size-dependent transmission.

*Keywords:* Branching process; prion; yeast

2010 Mathematics Subject Classification: Primary 60G99

Secondary 60K99; 62P10; 92D25

### 1. Introduction

Prion proteins are responsible for a variety of neurodegenerative diseases in mammals such as Creutzfeldt–Jakob disease in humans and ‘mad-cow disease’ (Bovine Spongiform Encephalopathy or BSE) in cattle [5], [9], [16]. Although the precise mechanisms remain unclear, the infectious agent itself is a protein produced by the cells of the host. While these diseases are fatal to mammals, a host of harmless phenotypes have been associated with prion proteins in the yeast *S. cerevisiae* [7], [20], [22], making yeast an ideal model to study the prion process in isolation.

According to the prion hypothesis, new phenotypes arise when misfolded versions of a protein appear (Figure 1(a)) and are joined together in aggregates. These aggregates are complexes of multiple misfolded proteins and can range in size from tens to hundreds of proteins. The misfolded (prion) state is infectious and can spread to normal proteins within the cell by *converting* their conformation to the misfolded state. During this process, the newly misfolded protein joins the complex and the aggregate increases in size by one unit (see Figure 1(b)).

Since cell division results in an exponentially increasing number of cells, in order for prions to be present throughout a colony, the total number of aggregates must increase. In total, there are four biochemical processes essential to maintaining the prion phenotype in a growing yeast culture (Figure 1(c)): (i) *synthesis* of new normal protein; (ii) *conversion* of normal

---

Received 21 November 2012; revision received 26 June 2013.

\* Postal address: Department of Mathematics, Trinity University, One Trinity Place, San Antonio, TX 78212, USA.

Email address: polofsso@trinity.edu

\*\* Postal address: School of Natural Sciences, University of California, Merced, USA.

Email address: ssindi@ucmerced.edu

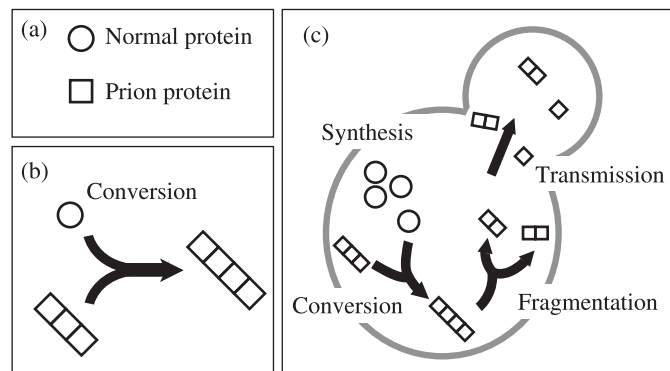


FIGURE 1: The prion pathway in yeast. (a) A prion protein is capable of distinct conformations. (b) The prion form of the protein aggregates into complexes which then may convert normal proteins. (c) Several steps are required for the prion state to persist: normal protein must be continuously synthesized; normal protein must be converted by existing aggregates; aggregates must increase in number by fragmentation, and a subset of the aggregates must be passed on during cell division. Perturbation of any of these parts of the prion pathway results in dilution of existing aggregates due to cell division and loss of the prion state at the population level.

protein to the misfolded state by existing aggregates; (iii) creation of new prion aggregates through *fragmentation* of existing aggregates into smaller complexes, and (iv) *transmission* of aggregates through cell division.

A single protein may correspond to multiple prion strains (distinct misfolded conformations) each of which are associated with biochemical rates of conversion, fragmentation and transmission [6], [21]. In yeast, a common tool used to characterize prion strains is the number of transmissible aggregates within a typical cell. These transmissible aggregates, sometimes called propagons or prion seeds, are the infectious agent of the prion disease [20]. The dominant experimental method used to infer the number of transmissible aggregates in a cell is treatment with guanidine hydrochloride (GdnHCl). Sufficiently high concentrations of GdnHCl have been shown to severely disrupt the fragmentation process [2], thus the total number of aggregates will not change. Exposure to GdnHCl does not alter other biochemical processes, so aggregates may continue to increase in size, through conversion, and will be diluted in the population due to cell division. Over time the population will be *cured* of the prion disease as the fraction of cells with aggregates approaches zero.

Mathematical approaches have been developed to estimate the number of transmissible aggregates in a founder cell by examining loss of aggregates over time due to population growth [3], [4], [12]. These models assume that all aggregates are equally transmissible; however, recent studies have revealed that only aggregates below a threshold size, about 20 monomers, can be efficiently transmitted during cell division [6]. Thus, without size-dependent transmission of aggregates, these previous models underestimate the number of transmissible aggregates [19].

Prion loss with size-dependent transmission has been considered by two recent studies [15], [19]. Palmer *et al.* [15] studied curing through computer simulations of yeast populations, but did not attempt to explicitly represent prion loss as a function of time. Furthermore, their simulations focused on only a single well-studied prion strain, the  $[PSI^+]$  strong strain [21]. As such, the rate of conversion was assumed to be known. Thus, while their simulations yielded a detailed analysis of curing for a specific strain, their simulation-based approach cannot be

used to determine the number of transmissible aggregates for other strains or, more generally, an unknown strain [1]. In [19] Sindi and Olofsson provided analytical solutions for curing, but only under the assumption of discrete-time cell division. Since yeast cells have an asymmetric cell-cycle, this approach greatly simplifies the age-structure of a population [4].

In this study we follow [19] by modeling aggregate conversion as a Poisson process, but now consider this process embedded within a yeast population growing under a continuous-time branching process. We derive a closed-form approximation for the expected rate of curing as a function of time. By using the rate of conversion as an explicit parameter, our model is suitable for analyzing loss for arbitrary strains without resorting to time-intensive simulations. Finally, our explicit formula may provide a way to assess *in vivo* conversion rates; most current estimates of conversion rates are based on *in vitro* analyses [21].

## 2. The branching process model

We model the yeast population in continuous time with a general (Crump–Mode–Jagers) branching process. In such a process, an individual may give birth at different points in time during its life. It applies naturally to budding-yeast populations where there is a clear distinction between a mother and her daughter. The central mathematical object is the *reproduction process* which describes the (stochastic) process according to which offspring are born. For yeast, a newborn cell first needs time to reach maturity after which it enters the cell cycle and eventually produces a daughter cell. Thereafter, the cell enters the cell cycle again and produces daughter cells indefinitely. In reality, a mother cell can only have a finite number of daughter cells, the *replicative lifespan*. For example, Sinclair *et al.* [18] estimated this number to be on average 25, but during the time of a typical curing experiment we can safely neglect this finite replicative lifespan and for practical purposes assume indefinite reproduction; see also [13] and [14]. If needed, the model can easily be adjusted to incorporate the finite lifespan.

For simplicity, we refer to a cell that has not yet reproduced as a ‘daughter’ and a cell that has reproduced at least once as a ‘mother’. Furthermore, we shall refer to the time between any two reproduction events as a ‘division time’ and consider the two cells resulting from a reproduction event (the mother and the newborn daughter) as belonging to the same generation. Assume that a mother cell has a division time  $M$  that has cumulative distribution function (CDF)  $F_M$  and that a daughter cell has a division time  $D$  that has CDF  $F_D$ . The division time of a mother cell is equal to the amount of time to go through the cell cycle. Newly born daughter cells require additional time to reach maturity before dividing, and thus have a distinct CDF for division times. Thus, the reproduction process has points at times  $D, D + M_1, D + M_1 + M_2, \dots$ , where the  $M_j$  are independent with common CDF  $F_M$ . For details, see Appendix A.1.

We assume that (prion) aggregates grow one unit at a time according to a Poisson process with rate  $\beta$ , and once an aggregate has grown to a critical size  $c$ , it can no longer be transmitted to the daughter but must stay in the mother. The population starts from one ancestor cell that contains a number of transmissible aggregates of different sizes. Let  $N_k$  denote the number of aggregates that are  $k$  units (conversion events) away from reaching the critical size,  $k = 1, \dots, c$ , let  $N = (N_1, \dots, N_c)$ , let  $Z_t$  be the number of cells with aggregates at time  $t$ , and let  $Y_t$  be the total number of cells present at time  $t$ . We approximate the expected proportion of cells with aggregates at time  $t$ , conditioned on  $N$  as

$$P_t^N \approx \frac{\mathbb{E}[Z_t | N]}{\mathbb{E}[Y_t]} \quad (2.1)$$

and, unconditionally,

$$P_t \approx \frac{\mathbb{E}[Z_t]}{\mathbb{E}[Y_t]}. \tag{2.2}$$

Following previous approaches [3], [4], [12], [19], (2.1) and (2.2) represent a truncated Taylor series approximation to  $P_t$ . We next explicitly derive  $\mathbb{E}[Z_t]$  and  $\mathbb{E}[Y_t]$ , which requires deriving the division time of the ancestor (founding cell).

**2.1. Expected number of cells in the population**

Since cells from any generation may be present at time  $t$ , we first decompose the population by generation. Note that a given cell in the  $n$ th generation may be described by a sequence consisting of the ancestor,  $d$  daughters, and  $n - d$  mothers, and that there are  $\binom{n}{d}$  such sequences for  $d = 0, 1, \dots, n$ . Of those  $\binom{n}{d}$  sequences,  $\binom{n-1}{d-1}$  have a daughter in the  $n$ th position and  $\binom{n-1}{d}$  have a mother in the  $n$ th position. For the first type of sequence, let  $p_{nd}^D(t)$  be the probability that the  $n$ th generation cell is present at time  $t$ ; for the second type, denote the corresponding probability by  $p_{nd}^M(t)$ . The case  $n = 0$  is special as only the ancestor is present. The ancestor needs to be treated separately as this cell may be sampled from a population and thus will have already started its life at our time  $t = 0$ . Denote its remaining division time at time  $t = 0$  by  $A$  and the CDF of  $A$  by  $F_A$ . The probability that the ancestor is still present at time  $t$  is then  $1 - F_A(t)$  and we obtain

$$\mathbb{E}[Y_t] = 1 - F_A(t) + \sum_{n=1}^{\infty} \sum_{d=0}^n \left[ \binom{n-1}{d-1} p_{nd}^D(t) + \binom{n-1}{d} p_{nd}^M(t) \right],$$

where  $\binom{n-1}{-1} = \binom{n-1}{n} = 0$ . We need to find expressions for  $p_{nd}^M(t)$  and  $p_{nd}^D(t)$ . Let  $F_{n,d}(t)$  denote the CDF of the sum of  $n$  division times,  $d$  of which belong to daughters, that is,

$$F_{n,d}(t) = F_D^{*d} * F_M^{*(n-d)}(t), \tag{2.3}$$

where ‘\*’ denotes convolution. The following expression for  $\mathbb{E}[Y_t]$  is proved in Appendix A.2.

**Proposition 2.1.** *It holds that*

$$\mathbb{E}[Y_t] = 1 - F_A(t) + \sum_{n=1}^{\infty} \sum_{d=0}^n \left[ \binom{n-1}{d-1} (F_A * F_{n-1,d-1}(t) - F_A * F_{n,d}(t)) + \binom{n-1}{d} (F_A * F_{n-1,d}(t) - F_A * F_{n,d}(t)) \right]. \tag{2.4}$$

**2.2. Expected number of cells with aggregates**

Turning next to  $\mathbb{E}[Z_t | N]$ , first note that once an aggregate reaches critical size, it cannot be transmitted to a daughter cell but must remain in the mother. (This limitation on transmission, while strict, is consistent with previously reported findings [6].) For the  $n$ th generation, we divide the set of the  $2^n$  possible sequences of mothers and daughters into subsets  $S_{nld}$  of sequences where the last daughter is in position  $l$  and there is a total of  $d$  daughters. For example, the sequence DDMMM is in  $S_{522}$ , the sequences DMDMM and MDDMM are both in  $S_{532}$ , and so on. Let  $n_{ld}$  be the number of sequences in  $S_{nld}$ ; then  $n_{ld} = \binom{l-1}{d-1}$  for  $l = 0, \dots, n$  and  $d = 0, \dots, l$  (where  $\binom{-1}{-1} = \binom{-1}{0} = 1$  and  $\binom{l-1}{-1} = 0$  for  $l \geq 1$ ). Note that the  $n_{ld}$  do not depend on  $n$  since after the last daughter, there are no further choices. Also note that for fixed  $n$ , the  $n_{ld}$  sum to  $2^n$  as they should.

Let  $C(t)$  denote the event that a cell in  $S_{nld}$  is present in the population at time  $t$  and let  $B$  denote the event that at least one of the initial aggregates is present in the cell. The expected

number of cells with aggregates at time  $t$ , conditioned on  $N$ , is then

$$\mathbb{E}[Z_t | N] = \sum_{n=0}^{\infty} \sum_{l=0}^n \sum_{d=0}^l \binom{l-1}{d-1} \mathbb{P}(B \cap C(t) | N). \tag{2.5}$$

Now note that

$$B = \bigcup_{k=1}^c \bigcup_{j=1}^{N_k} B_{jk}, \tag{2.6}$$

where  $B_{jk}$  is the event that the  $j$ th aggregate among those that are  $N_k$  conversion events away from critical size is present in the cell. Denote the successive division times by  $T_0 \equiv 0, T_1, T_2, \dots$ , and let  $A(\mathbf{u})$  be the event that  $T_j = u_j$  for  $j = l, \dots, n$ , i.e.

$$A(\mathbf{u}) = \{T_l = u_l, \dots, T_n = u_n\}.$$

We start at  $T_l$  and not at  $T_1$  because if an aggregate grows beyond the critical size before the  $l$ th division, it cannot be present in a cell whose lineage has the last daughter at position  $l$ . Furthermore, let  $\mathbf{u} = (u_l, \dots, u_n)$ ,  $S(t) = \{\mathbf{u} : 0 \leq u_l \leq \dots \leq u_n \leq t\}$ , and let  $f$  denote the joint probability distribution function (PDF) of  $(T_l, \dots, T_n)$  to obtain

$$\mathbb{P}(B \cap C(t) | N) = \int_{S(t)} \mathbb{P}(B | C(t) \cap A(\mathbf{u}), N) \mathbb{P}(C(t) | A(\mathbf{u})) f(\mathbf{u}) d\mathbf{u}. \tag{2.7}$$

To deal with the three factors of the integrand, note that  $f(\mathbf{u})$  is easily obtained from the cell cycle distributions as

$$f(\mathbf{u}) = \begin{cases} f_A * f_D^{*(d-1)} * f_M^{*(l-d)}(u_l) f_D(u_{l+1} - u_l) \prod_{k=l+2}^n f_M(u_k - u_{k-1}) & \text{for } l \geq 1, l < n, \\ f_A * f_D^{*(d-1)} * f_M^{*(l-d)}(u_l) & \text{for } l \geq 1, l = n, \\ f_A(u_1) \prod_{k=2}^n f_M(u_k - u_{k-1}) & \text{for } l = 0, \end{cases}$$

where  $f_D^{*(d-1)} * f_M^{*(l-d)}$  is the PDF corresponding to the CDF  $F_{l-1,d}$  in (2.3) and  $f_A$  is the PDF of the remaining cell cycle time of the ancestor (and as usual, an empty product equals 1 and a 0th-order convolution power is a unit point mass at 0).

Given  $A(\mathbf{u})$ , the cell is present at time  $t$  if and only if it has not yet divided at  $t$  and, hence,

$$\mathbb{P}(C(t) | A(\mathbf{u})) = \begin{cases} \mathbb{P}(M > t - u_n) & \text{if } l < n, \\ \mathbb{P}(D > t - u_n) & \text{if } l = n. \end{cases}$$

To obtain an expression for  $\mathbb{P}(B | C(t) \cap A(\mathbf{u}))$ , recall (2.6) and note that the  $B_{jk}$  are independent given  $C(t) \cap A(\mathbf{u})$ . Also, note that the  $B_{jk}$  have the same (conditional) probabilities for  $j = 1, \dots, N_k$ . We obtain

$$\begin{aligned} \mathbb{P}(B | C(t) \cap A(\mathbf{u}), N) &= 1 - \prod_{k=1}^c \prod_{j=1}^{N_k} (1 - \mathbb{P}(B_{jk} | C(t) \cap A(\mathbf{u}))) \\ &= 1 - \prod_{k=1}^c (1 - \mathbb{P}(B_{1k} | C(t) \cap A(\mathbf{u})))^{N_k}. \end{aligned}$$

To deal with  $\mathbb{P}(B_{1k} | C(t) \cap A(\mathbf{u}))$ , condition further on the critical time  $\tau_k$  when the initial aggregate reaches critical size. Since aggregate growth is independent of cell reproduction,

$\tau_k$  is independent of  $C(t) \cap A(\mathbf{u})$ . Furthermore,  $\tau_k$  and  $A(\mathbf{u})$  together identify in which generation critical size is reached, and given this information  $B_{1k}$  is conditionally independent of  $C(t) \cap A(\mathbf{u})$ . If the aggregate reaches critical size before generation  $l$  where the last daughter is, it cannot be passed on to this daughter and will not be present in the given  $n$ th generation cell which gives

$$\mathbb{P}(B_{1k} \mid \{\tau_k < u_l\} \cap A(\mathbf{u})) = 0.$$

If critical size is reached in generation  $l \leq j \leq n-1$ , the aggregate must first reach generation  $j$  by chance after which it will stay in the remaining sequence of mothers. As it must be passed on to a daughter cell  $d$  times and to a mother cell  $j-d$  times, we obtain

$$\mathbb{P}(B_{1k} \mid \{u_j \leq \tau_k \leq u_{j+1}\} \cap A(\mathbf{u})) = p^d(1-p)^{j-d}.$$

Note, as in previous formulations, we assume that aggregates small enough to transmit during cell division are independently transmitted to the daughter cell with probability  $p$ , or retained in the mother cell with probability  $(1-p)$ ; see [3] and [19] for other formulations. Finally, if critical size is reached in generation  $n$  or later, the aggregate must make it to generation  $n$  by chance which gives

$$\mathbb{P}(B_{1k} \mid \{\tau_k \geq u_n\} \cap A(\mathbf{u})) = p^d(1-p)^{n-d},$$

and putting it all together we obtain

$$\mathbb{P}(B_{1k} \mid C(t) \cap A(\mathbf{u})) = \sum_{j=l}^{n-1} p^d(1-p)^{j-d} \mathbb{P}(u_j \leq \tau_k \leq u_{j+1}) + \mathbb{P}(\tau_k \geq u_n) p^d(1-p)^{n-d}.$$

Since conversion occurs according to a Poisson process with rate  $\beta$ , we have  $\tau_k \sim \Gamma(k, \beta)$  since the subscript  $k$  indicates that the aggregate is  $k$  conversion events away from reaching critical size. Thus, with  $F_k$  denoting the CDF of the said gamma distribution,

$$\mathbb{P}(u_j \leq \tau_k \leq u_{j+1}) = F_k(u_{j+1}) - F_k(u_j).$$

This is summarized in the following proposition.

**Proposition 2.2.** *It holds that*

$$\begin{aligned} \mathbb{E}[Z_t \mid \mathbf{N}] &= \sum_{n,l,d} \int_{S(t)} \binom{l-1}{d-1} \left( 1 - \prod_{k=1}^c (1 - \mathbb{P}(B_{1k} \mid C(t) \cap A(\mathbf{u})))^{N_k} \right) \\ &\quad \times \mathbb{P}(C(t) \mid A(\mathbf{u})) f(\mathbf{u}) \, d\mathbf{u}, \end{aligned} \tag{2.8}$$

where  $\mathbb{P}(B_{1k} \mid C(t) \cap A(\mathbf{u}))$ ,  $\mathbb{P}(C(t) \mid A(\mathbf{u}))$ , and  $f(\mathbf{u})$  are given above.

To arrive at the unconditional expectation  $\mathbb{E}[Z_t]$ , let  $\varphi$  denote the joint probability generating function of  $\mathbf{N}$  and let  $b_k = 1 - \mathbb{P}(B_{1k} \mid C(t) \cap A(\mathbf{u}))$  for  $k = 1, \dots, c$  to obtain

$$\begin{aligned} \mathbb{E}[Z_t] &= \mathbb{E}[\mathbb{E}[Z_t \mid \mathbf{N}]] \\ &= \sum_{n,l,d} \int_{S(t)} \binom{l-1}{d-1} \left( 1 - \mathbb{E} \left[ \prod_{k=1}^c (1 - \mathbb{P}(B_{1k} \mid C(t) \cap A(\mathbf{u})))^{N_k} \right] \right) \\ &\quad \times \mathbb{P}(C(t) \mid A(\mathbf{u})) f(\mathbf{u}) \, d\mathbf{u} \\ &= \sum_{n,l,d} \int_{S(t)} \binom{l-1}{d-1} (1 - \varphi(b_1, \dots, b_c)) \mathbb{P}(C(t) \mid A(\mathbf{u})) f(\mathbf{u}) \, d\mathbf{u}. \end{aligned}$$

### 2.3. Division time of the ancestor

Finally, we consider the distribution of the division time of the ancestor with CDF  $F_A$  and PDF  $f_A$ . Realistically, the ancestor is not newborn at time 0 but rather sampled from a cell population and has thus already lived for a period of time. Hence,  $F_A$  is the CDF of the remaining division time of such a randomly sampled cell which can be either a mother or a daughter. To obtain an expression for the PDF  $f_A$ , we apply stable population results from the theory of general branching processes. The following proposition is proved in Appendix A.3.

**Proposition 2.3.** *The PDF of the division time  $A$  of the ancestor equals*

$$f_A(t) = \alpha \int_0^\infty e^{-\alpha s} (f_D(s+t) + f_M(s+t)) ds, \quad t \geq 0,$$

where  $f_D$  and  $f_M$  are the PDFs of division times of daughters and mothers, respectively.

The constant  $\alpha$  is the *Malthusian parameter* which describes the asymptotic growth rate of the population in the sense that as  $t \rightarrow \infty$ ,  $\mathbb{E}[Z_t]$  grows proportionally to  $e^{\alpha t}$ . For the definition of  $\alpha$  and its relation to cell cycle parameters, see Appendix A.1.

Previous studies assume the ancestor has the distribution of the excess life in a renewal process [4]. However, this assumption is questionable as it is valid only if the cell is sampled from a lineage and not from a population. As the latter is arguably more realistic, exponential growth must also be factored in. Although previous approaches such as those used by Cole *et al.* [4] will underestimate  $A$ , as we later investigate, the practical effect of this underestimation may be small.

## 3. Simulation comparison

We next compare (2.5) and (2.8) with the population averages from simulations of the prion curing process. In accordance with our assumptions above and previous simulation studies [19, 6], aggregates smaller than the transmission threshold  $c$  are transmitted to daughter cells with probability  $p = 0.40$ . (This is consistent with the differences in volume between mother and daughter cells [6].) We use Monte Carlo integration (see [17]) to evaluate (2.7). Both our code for simulating yeast populations and evaluating (2.5) and (2.8) are available upon request.

### 3.1. Comparing theory to empirical averages

In Figure 2 we compare our theoretical expressions (solid line) to averages (dots) over 1000 simulated yeast populations and gamma distributed division times, with  $f_M(t; a_M, b)$  and  $f_D(t; a_D, b)$ , where

$$f(t; a, b) = \frac{t^{(a-1)} e^{-t/b}}{\Gamma(a) b^a}. \quad (3.1)$$

We varied the numbers of aggregates, the sizes of aggregates, and the rate of conversion  $\beta$ . In all cases, we found strong agreement between our theory and simulation averages. In Figure 2 we show representative results under the simplifying assumptions that the founder is a newly born cell and all initial aggregates are the same size (i.e. they are the same units away from the transmission threshold  $c$ ). Figures 2(a) and (c) show a box plot of  $Z_t$ , the number of cells with at least 1 aggregate, in simulated yeast populations (box plot) as well as the empirical (dots) and theoretical (solid) expressions for  $\mathbb{E}[Z_t]$ . As expected,  $Z_t$  eventually stops changing as all aggregates exceed the transmission threshold. Figures 2(b) and (d) show curves corresponding to the empirical (dots) and theoretical (solid) expressions for  $\mathbb{E}[Z_t]/\mathbb{E}[Y_t]$ .

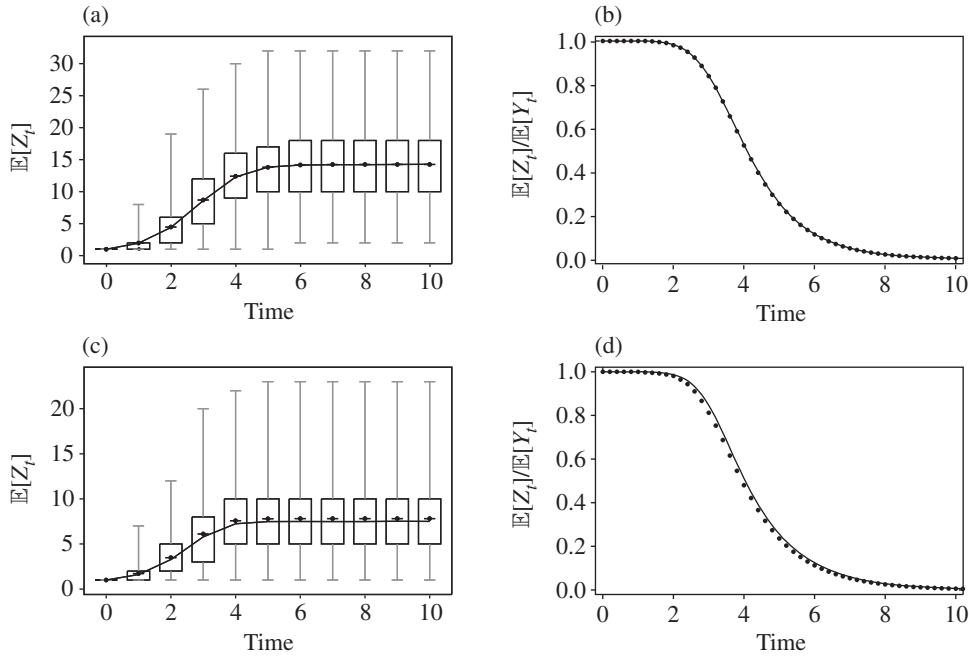


FIGURE 2: Theoretical distributions are compared to empirical averages for different aggregate configurations. (a)–(b): Identical gamma division times,  $\beta = 0.50$ , # Aggregates = 100, units until  $c = 5$ ; (c)–(d) Asymmetric gamma division times,  $\beta = 0.25$ , # Aggregates = 50, units until  $c = 10$ . In all cases our theoretical distributions closely match simulation averages.

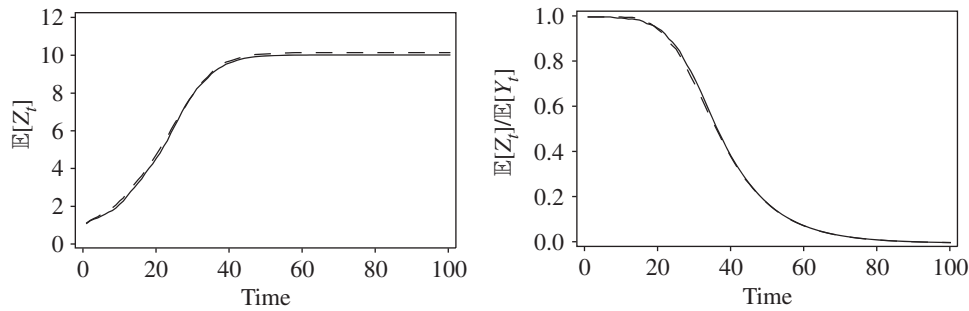


FIGURE 3: Theoretical curing curves under two choices for ancestor division time closely agree. For each case, the rate of conversion was  $\beta = 0.25$  and the population began with 50 aggregates each 10 units away from  $c$ . (See text for more information.)

For both symmetric (Figure 2(a) and (b)) and asymmetric (Figure 2(c) and (d)) cell division, our theoretical expressions for  $\mathbb{E}[Z_t]$  and  $\mathbb{E}[Z_t]/\mathbb{E}[Y_t]$  agree well with the empirical averages.

### 3.2. Incorporating the ancestor division time

We next examine the role of the ancestor distribution. Rather than drawing from a *random lineage*, as in previous approaches [3], [4], [12] we consider the ancestor as a *random cell* from the population. The PDF of the division time of the ancestor cell is given by Proposition 2.3 above and proven in Appendix A. In Figure 3, we compare  $\mathbb{E}[Z_t]$  and  $\mathbb{E}[Z_t]/\mathbb{E}[Y_t]$  under two



choices for the division time of the ancestor. First, as described in [4], the ancestor is chosen from a random lineage (dashed lines). Alternately, the ancestor is chosen randomly from a population of yeast cells as defined in Proposition 2.3 (solid lines). In these results  $f_M(t)$  and  $f_D(t)$  were chosen as identical gamma distributions,  $a = 2$  and  $b = \frac{1}{2}$  in (3.1), and  $f_A(t)$  is defined as in Proposition 2.3. We note that the  $\mathbb{E}[Z_t]$  and  $\mathbb{E}[Z_t]/\mathbb{E}[Y_t]$  closely agree under both scenarios for division time of the ancestor. Further investigation revealed that, unless aggregates are likely to exceed  $c$  in size within a single generation, the impact of the ancestor division time on the curing curve was minimal.

#### 4. Discussion

Prion proteins are responsible for a host of neurological disorders in mammals and a host of phenotypes in yeast. Of particular concern for yeast prion strains, is the number of transmissible aggregates in a typical cell. Previous approaches to computing this number, see [3], [4], [12], neglected continued growth of aggregates and size-limited transmission. Recently, studies have begun exploring these conditions, but have relied on computationally intensive simulations of yeast populations and, as such, required biochemical rates to be specified [6], [15].

Our work represents the first theoretical treatment of prion curing under realistic models of aggregate transmission and cell division. We demonstrated the accuracy of our derivations by comparing to simulations of prion curing. Our formulation enables novel lines of analysis which we plan to explore in future studies. Because we have parameterized our model in terms of the aggregate conversion rate, our work is applicable to arbitrary prion strains. Furthermore, our work offers the potential to estimate *in vivo* aggregate conversion rates, in addition to the number of transmissible aggregates.

### Appendix A

#### A.1. General branching processes

We give a brief introduction to general branching process theory; for a comprehensive account, see [11]. The fundamental object in a general branching process is the reproduction process which describes how offspring are being born. Denoting the reproduction process by  $\xi$ ,  $\xi(t)$  denotes the number of offspring a mother has had before age  $t$ , noting that  $\xi(t)$  is a random variable. The corresponding random measure is  $\xi(dt)$  so that

$$\xi(t) = \int_0^t \xi(du).$$

Individuals reproduce according to independent identically distributed copies of  $\xi$ . As pointed out in Section 2, the reproduction process for yeast cells has points at times  $D, D + M_1, D + M_1 + M_2, \dots$  where  $D$  has CDF  $F_D$ , and the  $M_j$  are independent (and independent of  $D$ ) with common CDF  $F_M$ .

The asymptotic exponential growth rate of the population is described by the Malthusian parameter  $\alpha$ . Let  $\xi$  denote the reproduction process and define its expected Laplace transform evaluated at a point  $r$  as

$$\mathbb{E}[\hat{\xi}(r)] = \int_0^\infty e^{-rt} \mathbb{E}[\xi(dt)].$$

The Malthusian parameter  $\alpha$  is the number satisfying  $\mathbb{E}[\hat{\xi}(\alpha)] = 1$ . To obtain an expression for  $\xi$ , note that offspring are born at times  $D, D + M_1, D + M_1 + M_2, \dots$  so if we let  $S_k = \sum_{j=1}^k M_j$ ,

we obtain the reproduction process

$$\xi(dt) = \delta_D(dt) + \sum_{k=1}^{\infty} \delta_{D+S_k}(dt),$$

where  $\delta_X$  denotes unit point mass at the random variable  $X$  so that

$$\int_0^{\infty} e^{-\alpha t} \delta_X(dt) = e^{-\alpha X}.$$

For a given random variable  $X$  with CDF  $F$ , let

$$\hat{F}(\alpha) = \mathbb{E}[e^{-\alpha X}] = \int_0^{\infty} e^{-\alpha t} F(dt),$$

the Laplace transform of the measure induced by the CDF of the random variable  $X$ . The expected Laplace transform of  $\xi$  is now given by

$$\mathbb{E}[\hat{\xi}(\alpha)] = \hat{F}_D(\alpha) + \hat{F}_D(\alpha) \sum_{k=1}^{\infty} \hat{F}_M(\alpha)^k = \frac{\hat{F}_D(\alpha)}{1 - \hat{F}_M(\alpha)},$$

which gives the following equation for  $\alpha$ :

$$\hat{F}_D(\alpha) + \hat{F}_M(\alpha) = 1. \tag{A.1}$$

(This is the same relation as Equation (2) in [8].) Note that the equation for  $\alpha$  can be written

$$\mathbb{E}[e^{-\alpha D}] + \mathbb{E}[e^{-\alpha M}] = 1,$$

so if  $D$  and  $M$  are constant, we obtain  $e^{-\alpha D} + e^{-\alpha M} = 1$  as in [10]. Also, note that any given cell is a mother at age  $a$  if and only if it has  $D \leq a$ . Hence, the asymptotic proportion of mothers is

$$\int_0^{\infty} e^{-\alpha a} F_D(da) = \hat{F}_D(\alpha),$$

and since  $\hat{F}_D(\alpha) + \hat{F}_M(\alpha) = 1$ , the asymptotic proportion of daughters is  $\hat{F}_M(\alpha)$ . This fact is also pointed out in [4] and is obtainable from Equation (5) of [8] by computing his  $q_0^U + q_0^B$ .

**A.2. Proof of Proposition 2.1**

Recall that a given cell in the  $n$ th generation can be described by a sequence consisting of the ancestor,  $d$  daughters, and  $n - d$  mothers. Of those sequences,  $\binom{n-1}{d-1}$  have a daughter in the  $n$ th position and  $\binom{n-1}{d}$  have a mother in the  $n$ th position. For the former, note that the last cell in the sequence is present if and only if the sum of the division times of (a) the ancestor, (b) the  $d - 1$  daughters, and (c) the  $n - d$  mothers is less than  $t$ , and the sum when the division time of the last cell (which is a daughter) has been added exceeds  $t$ . For the latter, the reasoning is similar, noting that the last cell is now instead a mother. Recall  $F_{n,d}(t)$  from (2.3) and add the division time  $A$  of the ancestor to obtain the CDF  $F_A * F_{n-1,d}$  of the time of birth of an  $n$ th generation cell that has  $d$  daughters in its ancestry. We obtain the probabilities of a cell being present at time  $t$  as

$$p_{nd}^M(t) = F_A * F_{n-1,d-1}(t) - F_A * F_{n,d}(t)$$

and

$$p_{nd}^D(t) = F_A * F_{n-1,d}(t) - F_A * F_{n,d}(t)$$

which proves the proposition.

**A.3. Proof of Proposition 2.3**

There is a general procedure to find asymptotic probabilities of which we present the special case needed for our application. To that end, consider some individual property of interest, call it  $B$ . We want to find the asymptotic probability  $\mathbb{P}(B)$  in an exponentially growing population. Let  $\mathbb{P}(B, s)$  denote the probability that an individual of age  $s$  has property  $B$ ; then

$$\mathbb{P}(B) = \alpha \int_0^\infty e^{-\alpha s} \mathbb{P}(B, s) ds. \tag{A.2}$$

In our case, we denote the remaining division time of an individual by  $Y$ , fix  $t$ , and let  $B$  be the property that  $Y > t$  (easier to work with than  $Y \leq t$ ). Thus, we need to determine the probability that an individual of age  $s$  has  $Y > t$ . For that purpose, denote the time of birth of the  $k$ th daughter by  $T_k$ , where

$$T_k = D + \sum_{j=1}^{k-1} M_j \tag{A.3}$$

for  $k \geq 1$ , and we let  $T_0 \equiv 0$ . For an individual of age  $s$ , the event  $Y > t$  occurs if and only if there is a  $k$  such that  $T_k \leq s$  and  $T_{k+1} > s + t$  which gives

$$\mathbb{P}(Y > t, s) = \sum_{k=0}^\infty \mathbb{P}(T_k \leq s, T_{k+1} > s + t)$$

(there can be at most one such  $k$ , that is, the events in the sum are mutually exclusive). Thus, we obtain

$$\mathbb{P}(Y > t) = \sum_{k=0}^\infty \int_0^\infty \alpha e^{-\alpha s} \mathbb{P}(T_k \leq s, T_{k+1} > s + t) ds.$$

Let us first consider the case  $k = 0$ . We obtain

$$\mathbb{P}(T_0 \leq s, T_1 > s + t) = 1 - F_D(s + t)$$

since  $T_0 \equiv 0$  and  $T_1 = D$ . For  $k \geq 1$ , let  $f_k$  denote the PDF of  $T_k$  and note that  $T_{k+1} = T_k + M_k$ . Condition on  $T_k$  to obtain

$$\begin{aligned} \mathbb{P}(T_k \leq s, T_{k+1} > s + t) &= \int_0^s \mathbb{P}(T_k \leq s, T_{k+1} > s + t \mid T_k = v) f_k(v) dv \\ &= \int_0^s \mathbb{P}(T_k + M_k > s + t \mid T_k = v) f_k(v) dv \\ &= \int_0^s (1 - F_M(s + t - v)) f_k(v) dv. \end{aligned}$$

We now obtain

$$\begin{aligned} &\int_0^\infty \int_0^s e^{-\alpha s} (1 - F_M(s + t - v)) f_k(v) dv ds \\ &= \int_0^\infty e^{-\alpha v} f_k(v) dv \int_0^\infty e^{-\alpha s} (1 - F_M(s + t)) ds \\ &= \hat{F}_D(\alpha) \hat{F}_M(\alpha)^{k-1} \int_0^\infty e^{-\alpha s} (1 - F_M(s + t)) ds \end{aligned}$$

by (A.3) and the product property of Laplace transforms of independent random variables. Now sum over  $k$  to obtain

$$\begin{aligned} & \sum_{k=1}^{\infty} \hat{F}_D(\alpha) \hat{F}_M(\alpha)^{k-1} \int_0^{\infty} e^{-\alpha s} (1 - F_M(s+t)) \, ds \\ &= \hat{F}_D(\alpha) \frac{1}{1 - \hat{F}_M(\alpha)} \int_0^{\infty} e^{-\alpha s} (1 - F_M(s+t)) \, ds \\ &= \int_0^{\infty} e^{-\alpha s} (1 - F_M(s+t)) \, ds \end{aligned}$$

by (A.1). Finally, add the term for  $k = 0$  and we obtain, by (A.2),

$$\mathbb{P}(Y > t) = \alpha \int_0^{\infty} e^{-\alpha s} (2 - F_D(s+t) - F_M(s+t)) \, ds.$$

The final result follows since the PDF of  $Y$  is the derivative of  $-\mathbb{P}(Y > t)$ . Technical conditions needed to differentiate under the integral sign (such as the Leibniz integral rule or dominated convergence) are satisfied for the gamma distribution, as well as for a range of other commonly used distributions.

### Acknowledgement

This work was supported by the NIH grant 1R15GM093957-01.

### References

- [1] BATEMAN, D. A. AND WICKNER, R. B. (2013). The  $[PSI^+]$  prion exists as a dynamic cloud of variants. *PLoS Genetics* **9**, e1003257.
- [2] BYRNE, L. J. *et al.* (2007). Cell division is essential for elimination of the yeast  $[PSI^+]$  prion by guanidine hydrochloride. *Proc. Nat. Acad. Sci. USA* **104**, 11688–11693.
- [3] BYRNE, L. J. *et al.* (2009). The number and transmission of  $[PSI^+]$  prion seeds (propagons) in the yeast *Saccharomyces cerevisiae*. *PLoS ONE* **4**, e4670.
- [4] COLE, D. J. *et al.* (2004). Estimating the number of prions in yeast cells. *Math. Med. Biol.* **21**, 369–395.
- [5] COLLINGE, J. (1999). Variant Creutzfeldt–Jakob disease. *The Lancet* **354**, 317–323.
- [6] DERDOWSKI, A. *et al.* (2010). A size threshold limits prion transmission and establishes phenotypic diversity. *Science* **330**, 680–683.
- [7] FOWLER, D. M. AND KELLY, J. W. (2009). Aggregating knowledge about prions and amyloid. *Cell* **137**, 20–22.
- [8] GREEN, P. J. (1981). Modelling yeast cell growth using stochastic branching processes. *J. Appl. Prob.* **18**, 799–808.
- [9] GRIFFITH, J. S. (1967). Self-replication and scrapie. *Nature* **215**, 1043–1044.
- [10] HARTWELL, L. H. AND UNGER, M. W. (1977). Unequal division in *Saccharomyces cerevisiae* and its implications for the control of cell division. *J. Cell Biology* **75**, 422–435.
- [11] JAGERS, P. AND NERMAN, O. (1984). The growth and composition of branching populations. *Adv. Appl. Prob.* **16**, 221–259.
- [12] MORGAN, B. J., RIDOUT, M. S. AND RUDDOCK, L. W. (2003). Models for yeast prions. *Biometrics* **59**, 562–569.
- [13] OLOFSSON, P. AND BERTUCH, A. A. (2010). Modeling growth and telomere dynamics in *Saccharomyces cerevisiae*. *J. Theoret. Biol.* **263**, 353–359.
- [14] OLOFSSON, P. AND DAILED, R. C. (2011). Budding yeast, branching processes, and generalized Fibonacci numbers. *Math. Magazine* **84**, 163–172.
- [15] PALMER, K. J., RIDOUT, M. S. AND MORGAN, B. J. T. (2011). Kinetic models of guanidine hydrochloride-induced curing of the yeast  $[PSI^+]$  prion. *J. Theoret. Biol.* **274**, 1–11.
- [16] PRUSINER, S. (1982). Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–144.
- [17] ROBERT, C. P. AND CASELLA, G. (2009). *Introducing Monte Carlo Methods with R*. Springer, New York.
- [18] SINCLAIR, D., MILLS, K. AND GUARENTE, L. (1998). Aging in *Saccharomyces cerevisiae*. *Annual Rev. Microbiol.* **52**, 533–560.

- [19] SINDI, S. S. AND OLOFSSON, P. (2013). A discrete-time branching process model of yeast prion curing curves. *Math. Pop. Stud.* **20**, 1–13.
- [20] SINDI, S. S. AND SERIO, T. R. (2009). Prion dynamics and the quest for the genetic determinant in protein-only inheritance. *Curr. Opinion Microbiol.* **12**, 623–630.
- [21] TANAKA, M., COLLINS, S. R., TOYAMA, B. H. AND WEISSMAN, J. S. (2006). The physical basis of how prion conformations determine strain phenotypes. *Nature* **442**, 585–589.
- [22] TUIITE, M. F. AND COX, B. S. (2009). Prions remodel gene expression in yeast. *Nature Cell Biol.* **11**, 241–243.

Copyright of Journal of Applied Probability is the property of Applied Probability Trust School of Mathematics & Statistics and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.