Stochastic Modeling of *Chlamydia*
Abstract of Ph.D. Thesis

Máté Szalai

Supervisor:

Péter Kevei, associate professor, Ph.D.

Doctoral School in Mathematics
University of Szeged, Bolyai Institute

Szeged

2024

Introduction

Stochastic modeling of biological systems holds significant importance in the natural sciences, as highlighted in [1]. The interest in mathematical modeling of natural systems has significantly increased over the past century. One of the best tools for modeling natural systems is branching processes [3]. In recent times, branching processes have proven to be effective tools for shedding light on various challenges across molecular biology, cell biology, developmental biology, immunology, evolution, ecology, medicine, and related domains. One of the most frequently used types of branching processes is represented by Galton–Watson processes.

The thesis predominantly incorporates real measurement results for modeling, with the tested organism being the bacterial species *Chlamydia trachomatis*. We can observe measurement results obtained from the Department of Medical Microbiology and Immunobiology, University of Szeged. *Chlamydia trachomatis* infections, which are sexually transmitted, pose a significant global public health challenge. These infections affect millions of individuals worldwide, including men, women, and children, often leading to severe medical complications.

The thesis is structured as follows. We provide a Galton–Watson model for the growth of a bacterial population in the presence of antibiotics. We assume that bacterial cells either die or duplicate, and the corresponding probabilities depend on the concentration of the antibiotic. Assuming that the mean offspring number is given by $m(c) = 2/(1 + \alpha c^{\beta})$ for some α, β, where c stands for the antibiotic concentration we obtain weakly consistent, asymptotically normal estimator both for (α, β), and for the minimal inhibitory concentration (MIC), a relevant parameter in pharmacology. For the measurements of *Chlamydia* growth quantitative polymerase chain reaction (qPCR) technique was used. The 2-parameter model fits remarkably well to the biological data.

Assuming other measurement results, we estimated the probability of extinction. The model assumption is entirely similar. We assume that bacterial cells either die or duplicate, with probabilities $p_0(c)$, and $p_2(c)$, where $p_2(c) = 1/(1 + \alpha c^{\beta})$ for some positive real numbers α, β. Using measurements based on colony counting method we obtain weakly consistent, asymptotically normal estimator for the parameters.

We explore the unique life cycle of *Chlamydia*. We model the population growth by a 2-type discrete-time branching process, where the probability of duplication depends on the state. Maximizing the EB production leads to a stochastic optimization problem. Simulation study shows that our novel model is able to reproduce the main features of the development of the population, deterministic models had not been able to achieve until now.

Then we establish a connection with our previous findings. Specifically, at a given antibiotic concentration, we determine the optimal transition of *Chlamydia* from the RB form to the EB form. We assume that the antibiotic solely affects the RB body, but not the EB body. This assumption is biologically plausible since EB bodies have the capability to form inclusions, aiding their survival under adverse conditions. In this scenario, we can numerically determine the optimal strategy. In the alternate case, we assume that the antibiotic affects both the RB and EB bodies. To the best of our knowledge, there is no real-world measurement data available for these models. The dissertation is based on three articles of the author. These publications are the following:

[1] A. Bogdanov, P. Kevei, **M. Szalai**, D. Virok: Stochastic modeling of in vitro bactericidal potency. *Bulletin of Mathematical Biology* 84 (6), 2022.

[2] **M. Szalai**, P. Kevei: Estimation of in vitro bactericidal potency based on colony counting method. *22nd European Young Statisticians Meeting – Proceedings*, pages 133 – 137. Panteion University of Social and Political Sciences, 2021.

[3] P. Kevei, **M. Szalai**: Branching model with state dependent offspring distribution for Chlamydia spread. *Mathematical Modelling of Natural Phenomena*, 2024, 19: 14.

Stochastic Modeling of In Vitro Bactericidal Potency

We assume that the bacterial population is homogeneous, all the cells behave similarly. In particular, there is no resistant type. As mutation is rare under normal conditions and in short time, this is a natural assumption for our data set.

In the experiments growth of *Chlamydia trachomatis* bacterial population was analyzed by quantitative PCR (qPCR) method with 12 different antibiotic concentrations and 2 different antibiotics. Adding a measurement error, the measurements have the form

$$C_i(c, x_0) = a - \log_2 Z^{(i)}_{n;c,x_0} + \varepsilon_{i;c}, \qquad i = 1, \dots, N,$$

where measurement error $\varepsilon_{i;c}$ is assumed to be Gaussian with mean zero, and variance σ^2_ε. This simple linear model is suggested by Yuan et al. [7]. Due to the measurement method, lower values means higher genome concentration.

We consider a simple Galton–Watson branching process where the offspring distribution depends on the antibiotic concentration $c \geq 0$. Each bacteria either dies (leaves no offspring), survives (leaves 1 offspring), or divides (leaves 2 offsprings) with respective probabilities $p_0 = p_0(c)$, $p_1 = p_1(c)$, and $p_2 = p_2(c)$. Let $f(s) = f_c(s)$ denote the offspring generating function and $m = m(c)$ the offspring mean if the antibiotic concentration is c, i.e.

$$\begin{aligned} f(s) = f_c(s) = \mathbf{E}s^{\xi_c} = \sum_{i=0}^{2} p_i(c)s^i, \quad s \in [0,1], \\ m = m(c) = f_c'(1) = \mathbf{E}\xi_c, \end{aligned} \tag{1}$$

where ξ_c is the number of offsprings. The process starts with $X_0 = x_0$ initial individuals, and

$$X_{n+1;c} = \sum_{i=1}^{X_{n;c}} \xi^{(n)}_{i;c},$$

where $\{\xi_c, \xi^{(n)}_{i;c} : i \geq 1, n \geq 1\}$ are iid random variables with generating function f_c.

Using the qPCR method the observed quantity is the genom of all individual bacteria, which is a constant times the *total* number of bacteria, that is alive and

dead cells together. Therefore, we have to keep track the dead bacterias too. In order to do this we consider a two-type Galton–Watson branching process $\mathbf{X}_n = (X_n, Y_n)$, $n \geq 0$, where X_n, Y_n stands for the number of alive, dead bacterias respectively, in generation n. Then the total number of bacteria at generation n is $Z_n = X_n + Y_n$. We also write Z_{n,x_0} to emphasize that $X_0 = x_0$. The process evolves as

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i^{(n)}$$

$$Y_{n+1} = Y_n + \sum_{i=1}^{X_n} \eta_i^{(n)}, \quad n \geq 0,$$

$(X_0, Y_0) = (x_0, 0)$, where $(\xi, \eta), (\xi_i^{(n)}, \eta_i^{(n)})$, $n = 1, 2, \ldots$, $i = 1, 2, \ldots$ are iid random vectors such that $\mathbf{P}((\xi, \eta) = (0, 1)) = p_0$, $\mathbf{P}((\xi, \eta) = (1, 0)) = p_1$, $\mathbf{P}((\xi, \eta) = (2, 0)) = p_2$.

Lemma 1. *If $x_0 = 1$ then for the mean we have $\mathbf{E}X_n = m^n$, and $\mathbf{E}Y_n = p_0(1 + m + \ldots + m^{n-1})$, thus*

$$\mu_n := \mathbf{E}Z_{n,1} = \begin{cases} m^n \left(1 + \frac{p_0}{m-1}\right) - \frac{p_0}{m-1}, & m \neq 1, \\ 1 + p_0 n, & m = 1. \end{cases}$$

The strong law of large numbers and the central limit theorem imply that for each fixed n as $x_0 \to \infty$

$$\frac{Z_{n,x_0}}{x_0} \to \mu_n \quad \text{a.s.}$$

and

$$\frac{Z_{n,x_0} - x_0 \mu_n}{\sqrt{x_0}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_n^2),$$

Estimation of the offspring mean

Put

$$\log_2 \widehat{\mu}_n = a - \log_2 x_0 - \frac{\sum_{i=1}^N C_i(c, x_0)}{N}.$$

Proposition 1. *As first $x_0 \to \infty$ and then $N \to \infty$*

$$\log_2 \widehat{\mu}_n \xrightarrow{\mathbf{P}} \log_2 \mu_n,$$

which implies that $\widehat{\mu}_n$ is weakly consistent estimatior of μ_n. Furthermore, as first $x_0 \to \infty$ and then $N \to \infty$

$$\frac{1}{\sigma_\varepsilon \mu_n \log 2} \sqrt{N} \left(\widehat{\mu}_n - \mu_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Thus we can estimate μ_n. The problem is that μ_n does not determine uniquely m, only gives a possible range for it, see Figure 1.

To overcome this difficulty, we assume that $p_1 \equiv 0$. Then μ_n in Lemma 1 simplifies to

$$\mu_n(m) = \frac{m}{2}\left(m^{n-1} + \ldots + 1\right) + 1 = \begin{cases} \frac{m(m^n - 1)}{2(m-1)} + 1, & m \neq 1, \\ \frac{n}{2} + 1, & m = 1. \end{cases}$$

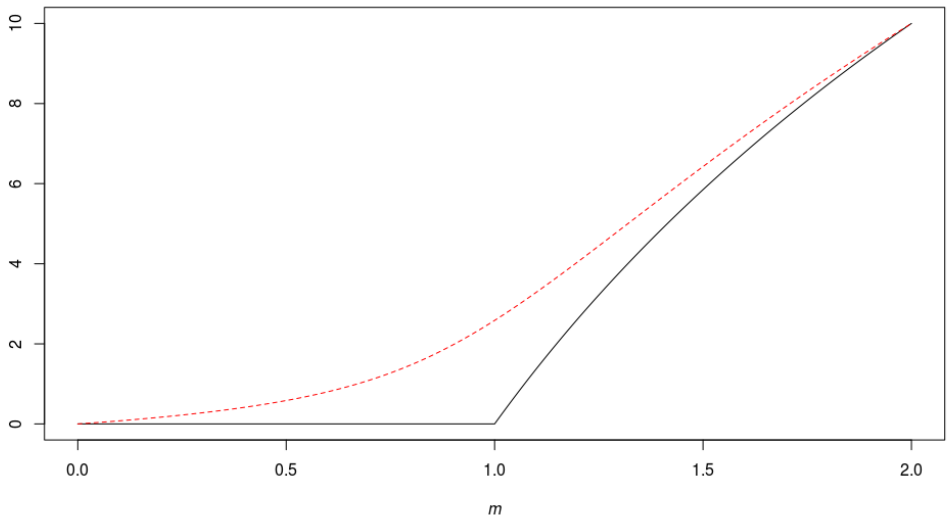

Figure 1: Upper and lower bound for $\log_2 \mu_n$ for $n = 10$.

Proposition 2. *Assume that $p_1 = 0$. As first $x_0 \to \infty$ and then $N \to \infty$, the estimate $\widehat{m}$ is a weakly consistent estimator of m, and*

$$\frac{\mu_n'(m)}{\sigma_\varepsilon \mu_n(m) \log 2} \sqrt{N}(\widehat{m} - m) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

The dependence of m on the antibiotic concentration

Assuming $p_1 \equiv 0$ we can estimate the mean for $c > 0$ fixed as described in Proposition 2. Next we combine our estimator for different concentrations. We assume that the offspring mean as a function of c can be described by the Hill function [6].

$$m(c) = \frac{2}{1 + \alpha c^\beta} \tag{2}$$

for some unknown parameters $\alpha > 0$, $\beta > 0$. This is a quite flexible model, and we show that empirical data fits very well to this model. Rewriting (2)

$$\log \alpha + \beta \log c = \log \left(\frac{2}{m(c)} - 1 \right). \tag{3}$$

Assume that we have measurements for $K \geq 2$ different concentrations $c_1 < c_2 < \ldots < c_K$, and we obtain the estimator for the offspring mean $\widehat{m}(c_i)$, $i = 1, 2, \ldots, K$. Standard least square theory implies that the expression

$$\sum_{i=1}^{K} \left(\log \left(\frac{2}{\widehat{m}(c_i) - 1} \right) - \beta \log c_i - \log \alpha \right)^2$$

attains it minimum at $(\alpha, \beta) = (\widehat{\alpha}, \widehat{\beta})$, with

$$\widehat{\beta} = \frac{K \sum_{i=1}^K h_i \ell_i - \sum_{i=1}^K h_i L_1}{K L_2 - L_1^2}, \qquad \widehat{\alpha} = \exp \left\{ \frac{\sum_{i=1}^K h_i - \widehat{\beta} L_1}{K} \right\}, \tag{4}$$

where to ease notation we write $h_i = \log\left(\frac{2}{\widehat{m}(c_i)} - 1\right)$, and $\ell_i = \log c_i$, furthermore $L_1 = \sum_{i=1}^K \ell_i$, and $L_2 = \sum_{i=1}^K \ell_i^2$.

The *minimal inhibitory concentration* (MIC) is the smallest antibiotic concentration that stops bacteria growth. In mathematical terms $\vartheta := \text{MIC} = \min\{c : m(c) \leq 1\}$, which, under the assumption (2) $\vartheta = \text{MIC} = \alpha^{-1/\beta}$. Define the estimator $\widehat{\vartheta} = \widehat{\alpha}^{-1/\widehat{\beta}}$.

In the following statement we summarize the main properties of these estimators. Introduce the notation

$$k_i = \frac{2}{m(c_i)(2 - m(c_i))} \frac{\sigma_\varepsilon \mu_n(m(c_i)) \log 2}{\mu_n'(m(c_i))}, \quad i = 1, 2, \ldots, K.$$

Proposition 3. *Assume that first $x_0 \to \infty$ and then $N \to \infty$. Then the estimates $\widehat{\alpha}, \widehat{\beta}$, and $\widehat{\vartheta}$ are weakly consistent estimators of the corresponding quantities. Furthermore, as $x_0 \to \infty$ and then $N \to \infty$*

$$\sqrt{N}(\widehat{\alpha} - \alpha, \widehat{\beta} - \beta) \xrightarrow{\mathcal{D}} (U, V),$$

where (U, V) is a two-dimensional normal random vector with mean 0 and covariance matrix

$$\begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix},$$

where

$$\begin{aligned}
\sigma_\alpha^2 &= \frac{\alpha^2}{(KL_2 - L_1^2)^2} \sum_{i=1}^K k_i^2 (L_2 - L_1 \ell_i)^2 \\
\sigma_{\alpha\beta} &= \frac{\alpha}{(KL_2 - L_1^2)^2} \sum_{i=1}^K k_i^2 (K\ell_i - L_1)(L_2 - L_1 \ell_i) \\
\sigma_\beta^2 &= \frac{1}{(KL_2 - L_1^2)^2} \sum_{i=1}^K k_i^2 \left(K\ell_i - L_1\right)^2,
\end{aligned}$$

and

$$\sqrt{N}(\widehat{\vartheta} - \vartheta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_\vartheta^2),$$

with

$$\sigma_\vartheta^2 = \frac{\vartheta^2 (\log \alpha)^2}{\beta^2 (KL_2 - L_1^2)^2} \sum_{i=1}^K k_i^2 \left(\frac{L_2 - L_1 \ell_i}{\log \alpha} - \frac{K\ell_i - L_1}{\beta}\right)^2.$$

In Figures 2 and 3 we see the estimated means and the corresponding fitted curve $m(c)$, where the parameters α, β are estimated as described in (4).

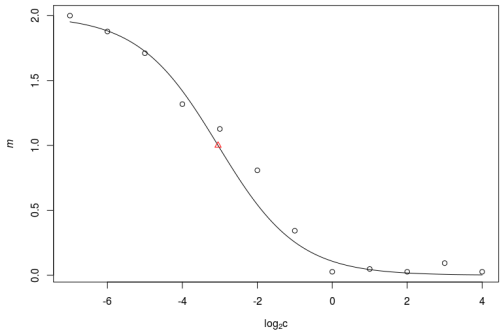

Figure 2: Estimated means and the fitted curve for azithromycin.

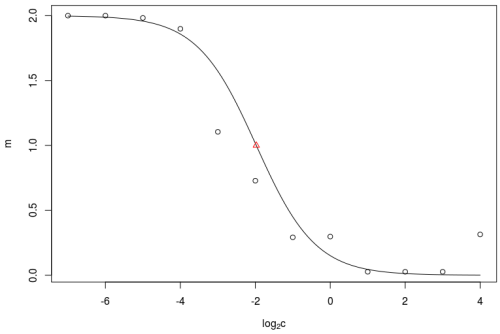

Figure 3: Estimated means and the fitted curve for ciprofloxacin.

Estimation of in vitro bactericidal potency based on colony counting method

The qPCR method measures the total bacterial genom, which is the total number of *dead and alive* bacterial cells multiplied by a constant. On the other hand, colony counting gives an estimator for the extinction probability. The basic experiment is the following. Originally, x_0 bacterial cells (e.g. *Escherichia coli*) are inoculated onto agar plates containing a series of antibiotic concentration, and after the incubation period all the viable colonies are enumerated, see e.g. Liu et al. [5].

Consider a simple Galton–Watson branching process as in (1), where each bacteria either dies (leaves no offspring) or divides (leaves 2 offsprings) with respective concentration dependent probabilities $p_0 = p_0(c)$, and $p_2 = p_2(c) = 1 - p_0(c)$.

Let $f(s) = f_c(s) = p_0 + p_2 s^2$ denote the offspring generating function and $m = m(c) = 2p_2(c)$ the offspring mean if the antibiotic concentration is c. The process starts with a single ancestor $X_{0;c} = 1$, and

$$X_{n+1;c} = \sum_{i=1}^{X_{n;c}} \xi_{i;c}^{(n)},$$

where $\{\xi_c, \xi_{i;c}^{(n)} : i \geq 1, n \geq 1\}$ are iid random variables with generating function f_c. We further assume that the offspring distribution is given by

$$p_2(c) = \frac{1}{1 + \alpha c^\beta}, \tag{5}$$

where $\alpha > 0$, $\beta > 0$ are unknown parameters. Note that as $m = 2p_2$ this is the same assumption as earlier. Under this model the MIC, the smallest antibiotic concentration preventing bacterial growth, is the smallest c for which $m(c) = 1$, that is $\alpha^{-1/\beta}$.

If $m \leq 1$ then the process dies out almost surely, while if the process is supercritical, i.e. $m > 1$ then the probability of extinction is the smaller root of $f_c(q) = q$, which is in our setup

$$q(c) = \begin{cases} \frac{1-p_2(c)}{p_2(c)}, & \text{if } p_2(c) > 1/2, \\ 1, & \text{if } p_2(c) \le 1/2. \end{cases} \tag{6}$$

Estimation of the parameters

Assume that the initial number of bacterial cells is x_0, that is we observe x_0 independent copies of the Galton–Watson process $(X_{n;c})$. Then the number Y_c of living colonies has binomial distribution with parameters x_0 and $1 - q(c)$. Therefore, the natural estimator for $q(c)$ is

$$\widehat{q}(c) = 1 - \frac{Y_c}{x_0}.$$

The law of large numbers and the central limit theorem implies that $\widehat{q}(c)$ is a weakly consistent estimator, and as $x_0 \to \infty$

$$\frac{\sqrt{x_0}}{\sqrt{q(c)(1-q(c))}} \left(\widehat{q}(c) - q(c)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{7}$$

From (6) we see that we can estimate $p_2(c)$ only if $q(c) < 1$, or equivalently $m(c) > 1$, in which case

$$\widehat{p_2}(c) = \frac{1}{1 + \widehat{q}(c)}. \tag{8}$$

We assume that the offspring mean as a function of c satisfies (5) for some unknown parameters $\alpha > 0$, $\beta > 0$. Assume that we have measurements for $K \ge 2$ different concentrations $c_1 < c_2 < \ldots < c_K$, such that $m(c_K) > 1$. As in (8), we obtain the estimator $\widehat{p}_2(c_i)$ at different concentrations, from which, using simple least squares estimator we obtain the estimator

$$\widehat{\beta} = \frac{K \sum_{i=1}^K h_i \ell_i - \sum_{i=1}^K h_i L_1}{K L_2 - L_1^2},$$
$$\widehat{\alpha} = \exp\left\{ \frac{\sum_{i=1}^K h_i - \widehat{\beta} L_1}{K} \right\},$$

where to ease notation we write $h_i = \log\left(\frac{1}{\widehat{p}_2(c_i)} - 1\right)$, $\ell_i = \log c_i$, $L_1 = \sum_{i=1}^K \ell_i$ and $L_2 = \sum_{i=1}^K \ell_i^2$.

Under the assumption (5) the MIC equals $\vartheta = \alpha^{-1/\beta}$, therefore its natural estimator is

$$\widehat{\vartheta} = \widehat{\alpha}^{-1/\widehat{\beta}}.$$

Using (7), as in Proposition 3 we can prove that these estimators are asymptotically normal. Introduce the notation

$$k_i = \frac{p_2(c_i)}{1 - p_2(c_i)} \sqrt{q(c_i)(1 - q(c_i))}, \quad i = 1, 2, \ldots, K.$$

Proposition 4. *Assume that $c_1 < \ldots < c_K$ are given concentrations such that $m(c_K) > 1$. Then as $x_0 \to \infty$, $\widehat{\alpha}, \widehat{\beta}$, and $\widehat{\vartheta}$ are weakly consistent estimators of the corresponding quantities. Furthermore, as $x_0 \to \infty$*

$$\sqrt{x_0}(\widehat{\alpha} - \alpha, \widehat{\beta} - \beta) \xrightarrow{\mathcal{D}} (U, V),$$

where (U, V) is a two-dimensional normal random vector with mean 0 and covariance matrix $\begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix}$, where

$$\sigma_\alpha^2 = \frac{\alpha^2}{(KL_2 - L_1^2)^2} \sum_{i=1}^{K} k_i^2 (L_2 - L_1\ell_i)^2,$$

$$\sigma_{\alpha\beta} = \frac{\alpha}{(KL_2 - L_1^2)^2} \sum_{i=1}^{K} k_i^2 (K\ell_i - L_1)(L_2 - L_1\ell_i),$$

$$\sigma_\beta^2 = \frac{1}{(KL_2 - L_1^2)^2} \sum_{i=1}^{K} k_i^2 \left(K\ell_i - L_1\right)^2,$$

and

$$\sqrt{x_0}(\widehat{\vartheta} - \vartheta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_\vartheta^2),$$

as $x_0 \to \infty$, with

$$\sigma_\vartheta^2 = \frac{\vartheta^2 (\log \alpha)^2}{\beta^2 (KL_2 - L_1^2)^2} \sum_{i=1}^{K} k_i^2 \left(\frac{L_2 - L_1\ell_i}{\log \alpha} - \frac{K\ell_i - L_1}{\beta} \right)^2.$$

Branching model with state dependent offspring distribution for *Chlamydia* spread

Chlamydiae are obligate intracellular bacteria which have a unique two-stage developmental cycle, with two forms, the elementary body (EB) and the reticulate body (RB). The EB is the infectious form but it is incapable of reproducing. Once the EB infects the host cell, it transforms into an RB. The RB multiplies in the host cell by binary fission. After some time RBs redifferentiate to EBs. The EBs are then released from the host cell ready to infect new host cells. It was shown recently by Lee et al. [4] using 3D electron microscopy method and manual counting that this conversion occurs asynchronously, so that some RBs are converting into EBs, while others continue to divide.

Consider a two-type discrete-time Galton–Watson branching process $\mathbf{X}^\pi = (\mathbf{X}_n^\pi)_n = (X_n^\pi, Y_n^\pi)_n$, $n \geq 0$, together with a sequence of probabilities $\pi = (p_n)_n$. We assume that π is adapted to the natural filtration $(\mathcal{F}_n)_n$ generated by $\mathbf{X}$, i.e. $\mathcal{F}_n = \sigma(\mathbf{X}_k^\pi, k \leq n)$. Initially $\mathbf{X}_0^\pi = (1, 0)$, and the process evolves as

$$X_{n+1}^\pi = \sum_{i=1}^{X_n^\pi} \xi_{n,i},$$

$$Y_{n+1}^\pi = Y_n^\pi + \sum_{i=1}^{X_n^\pi} \left(1 - \frac{\xi_{n,i}}{2}\right), \quad n \geq 0,$$

where $(\xi_n, \xi_{n,i}), n = 1, 2, \ldots, i = 1, 2, \ldots$ are conditionally independent random variables given $(p_n)_n$, for fix n the variables $(\xi_n, \xi_{n,i}), i = 1, 2, \ldots$ are identically distributed, such that $\mathbf{P}(\xi_n = 2|p_n) = p_n$, $\mathbf{P}(\xi_n = 0|p_n) = 1 - p_n$. Here X_n^π stands for the number of RBs and Y_n^π for the number of EBs in generation n.

The process ends at a random time $T \in \{1, 2, \ldots\}$ when the infected host cell dies. The aim of the bacterial population is to produce as many EBs as possible, that is to maximize $\mathbf{E}(Y_T^\pi)$ over all possible strategies (p_n). Denoting by $\mathcal{P}$ the set of all strategies, a strategy $\mathbf{q}$ is *optimal*, if

$$\sup_{\pi \in \mathcal{P}} \mathbf{E}(Y_T^\pi) = \mathbf{E}(Y_T^{\mathbf{q}}).$$

Assume that the host cell's death time T is independent of the process $\mathbf{X}^\pi$. Introduce the notation $\pi_\ell = (1, 1, \ldots, 1, 0, 0, \ldots)$, where the first $\ell \geq 0$ components are 1.

Theorem 1. *Assume that $T \geq 1$ is bounded and it is independent of $\mathbf{X}^\pi$. Let ℓ be such that*

$$2^\ell \mathbf{P}(T > \ell) = \sup_{k \geq 0} 2^k \mathbf{P}(T > k).$$

Then π_ℓ is an optimal strategy, with optimal value

$$\sup_{\pi \in \mathcal{P}} \mathbf{E}(Y_T^\pi) = \sup_{k \geq 0} 2^k \mathbf{P}(T > k).$$

Now we assume that T, the death time depends on the process $\mathbf{X}^\pi$. Given that the host cell is alive in generation $n-1$, the probability that it dies in the next step is $d(X_n^\pi, Y_n^\pi)$, that is

$$\mathbf{P}(T = n | T > n - 1, \mathcal{F}_n) = d(X_n^\pi, Y_n^\pi).$$

Assume that

$$\exists\, C > 0 \;\text{ such that }\; d(x, y) = 1 \;\text{ whenever }\; x + y \geq C. \tag{9}$$

Let $\widetilde{X}_n = X_n \mathbb{I}(T > n)$, $\widetilde{Y}_n = Y_n \mathbb{I}(T \geq n)$. Note that $\widetilde{X}_T = 0$, $\widetilde{Y}_T = Y_T$, and $\widetilde{Y}_{T+1} = 0$, which is convenient at the definition of the reward function in (10). Define a Markov chain $(\widetilde{X}_n, \widetilde{Y}_n)_n$ on the state space $\{0, 1, \ldots\}^2$, where the possible controls are given by the duplication probabilities $p_n \in [0, 1]$. The *reward function* (-1 times the cost function in [2]) gives the number of EBs upon cell's death, that is

$$c(x, y) = \begin{cases} y, & x = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Define the *value function*

$$h(x, y) = \begin{cases} \sup_{\pi \in \mathcal{P}} \mathbf{E}\left[\sum_{n=0}^\infty c(\widetilde{X}_n, \widetilde{Y}_n) | (\widetilde{X}_0, \widetilde{Y}_0) = (x, y)\right], & d(x, y) < 1, \\ y, & d(x, y) = 1. \end{cases}$$

Theorem 2. *Assume that* (9) *holds. Then* $h(x,y) = y$ *if* $x+y \geq C$, *and* $h(0,y) = y$ *for any* y. *Assume that* $h(x,y)$ *is determined whenever* $x+y \geq m$ *for some* $m \leq C$, *and let* $x+y = m-1$. *Then*

$$h(x,y) = \max_{p\in[0,1]} \sum_{j=0}^{x} \binom{x}{j} p^j (1-p)^{x-j} \times \left[d(2j, y+x-j)(y+x-j) + (1-d(2j, y+x-j))h(2j, y+x-j) \right], \tag{11}$$

where all the values of h *on the right-hand side are determined. The maximum in* p *of the continuous function on the right-hand side of* (11) *is attained at* $p(x,y)$, *which gives the optimal strategy.*

Branching model for the spread of *Chlamydia* under the influence of antibiotics

Now we investigate the behavior of the combined model, specifically determining the optimal spread of *Chlamydia* in the presence of antibiotic.

Consider a three-type discrete-time Galton–Watson branching process $\mathbf{X}^\pi = (\mathbf{X}_n^\pi)_n = (X_n^\pi, Y_n^\pi, Z_n^\pi), n \geq 0$, together with a sequence of probabilities $\pi = (p_n)_n$. We assume that π is adapted to the natural filtration $(\mathcal{F}_n)_n$ generated by $\mathbf{X}^\pi$, i.e. $\mathcal{F}_n = \sigma(\mathbf{X}_k^\pi, k \leq n)$. Initially $\mathbf{X}_0^\pi = (1,0,0)$, and the process evolves as

$$X_{n+1}^\pi = \sum_{i=1}^{X_n^\pi} \xi_{n,i}, \qquad Y_{n+1}^\pi = Y_n^\pi + \sum_{i=1}^{X_n^\pi} \eta_{n,i}, \qquad Z_{n+1}^\pi = Z_n^\pi + \sum_{i=1}^{X_n^\pi} \zeta_{n,i}, \qquad n \geq 0,$$

where (ξ_n, η_n, ζ_n), $(\xi_{n,i}, \eta_{n,i}, \zeta_{n,i})$ $n = 1, 2, \ldots$, $i = 1, 2, \ldots$ are conditionally independent random variables given $(p_n)_n$, for fix n the variables are identically distributed, such that

$$\begin{aligned} &\mathbf{P}\left((\xi_n, \eta_n, \zeta_n) = (2,0,0) | p_n\right) = p_n, \\ &\mathbf{P}\left((\xi_n, \eta_n, \zeta_n) = (0,1,0) | p_n\right) = 1 - p_n - p_c, \\ &\mathbf{P}\left((\xi_n, \eta_n, \zeta_n) = (0,0,1) | p_n\right) = p_c. \end{aligned}$$

X_n^π and Y_n^π again stands for the number of RBs and number of EBs in generation n, while Z_n^π denotes the number of dead bacteria in generation n. In generation n each RB duplicates with probability p_n, or die, with a predetermined probability p_c, or converts into EB with probability $1 - p_n - p_c$. The probability p_c depends on antibiotic concentration $c \geq 0$.

Theorem 3. *Assume that* $\exists C$ *such that* $d(x,y,z) = 1$ *whenever* $x+y+z \geq C$. *Then* $h(x,y,z) = y$ *if* $x+y+z \geq C$, *and* $h(0,y,z) = y$ *for any* y, *and for any* z. *Assume that* $h(x,y,z)$ *is determined whenever* $x+y+z \geq m$ *for some* $m \leq C$, *and*

let $x+y+z=m-1$. Then

$$h(x,y,z)=\max_{p\in[0,1-p_c]}\sum_{j=0}^{x}\sum_{k=0}^{x-j}\binom{x}{j}p^j\binom{x-j}{k}p_c^k(1-p-p_c)^{x-j-k}$$
$$\times[d(2j,y+x-j-k,z+k)(y+x-j-k)$$
$$+(1-d(2j,y+x-j-k,z+k))h(2j,y+x-j-k,z+k)],$$

where all the values of h on the right-hand side are determined. The maximum of the continuous function is attained at $p(x,y,z)$, which gives an optimal strategy.

The theoretical model, when antibiotic has effect on both types

Now we presume that the antibiotic exerts an influence on the RB body, and on the EB body as well. Consider a three-type discrete-time Galton–Watson branching process $\mathbf{X}^\pi=(\mathbf{X}_n^\pi)_n=(X_n^\pi,Y_n^\pi,Z_n^\pi), n\geq 0$. The process evolves as

$$X_{n+1}^\pi=\sum_{i=1}^{X_n^\pi}\xi_{n,i},\quad Y_{n+1}^\pi=\sum_{i=1}^{X_n^\pi}\eta_{n,i}+\sum_{i=1}^{Y_n^\pi}(1-\zeta'_{n,i}),\quad Z_{n+1}^\pi=Z_n^\pi+\sum_{i=1}^{X_n^\pi}\zeta_{n,i}+\sum_{i=1}^{Y_n^\pi}\zeta'_{n,i},$$

where $(\xi_n,\eta_n,\zeta_n,\zeta'_n)$, $(\xi_{n,i},\eta_{n,i},\zeta_{n,i},\zeta'_{n,i})$ $n=1,2,\ldots$, $i=1,2,\ldots$ are conditionally independent random variables given $(p_n)_n$, for fix n the variables are identically distributed, such that

$$\mathbf{P}\left((\xi_n,\eta_n,\zeta_n)=(2,0,0)|p_n\right)=p_n,$$
$$\mathbf{P}\left((\xi_n,\eta_n,\zeta_n)=(0,1,0)|p_n\right)=1-p_n-p_c,$$
$$\mathbf{P}\left((\xi_n,\eta_n,\zeta_n)=(0,0,1)|p_n\right)=p_c,$$

and $\mathbf{P}(\zeta'_n=0)=1-q_c$, $\mathbf{P}(\zeta'_n=1)=q_c$. In generation n each RB duplicates with probability p_n, or die, with a predetermined probability p_c, or convert into EB with probability $1-p_n-p_c$. In generation n each EB dies with a determined probability q_c.

Theorem 4. *Assume that $\exists C$ such that $d(x,y,z)=1$ whenever $x+y+z\geq C$, and N is given. Then for any $n\leq N$, $h(x,y,z;n)=y(1-q_c)^{N-n}$, if $x+y+z\geq C$, and $h(0,y,z;n)=y(1-q_c)^{N-n}$ for any y, and for any z. Assume that $h(x,y,z;n)$ is determined whenever $x+y+z+n\geq m$ for some $m\leq C+n$, and let $x+y+z+n=m-1$. Then for all $n\leq N-1$*

$$h(x,y,z;n)=\max_{p\in[0,1-p_c]}\sum_{j=0}^{x}\sum_{k=0}^{x-j}\sum_{\ell=0}^{y}\binom{x}{j}\binom{x-j}{k}\binom{y}{\ell}$$
$$\times p^j p_c^k(1-p-p_c)^{x-j-k}q_c^\ell(1-q_c)^{y-\ell}$$
$$\times[d(2j,y-\ell+x-j-k,z+k+\ell)(y-\ell+x-j-k)(1-q_c)^{N-(n+1)}$$
$$+(1-d(2j,y-\ell+x-j-k,z+k+\ell))$$
$$\times(h(2j,y-\ell+x-j-k,z+k+\ell;n+1)(\mathbb{I}(n+1<N)$$
$$+(y-\ell+x-j-k)\mathbb{I}(n+1=N)))],$$

where all the values of h on the right-hand side are determined. The maximum of the continuous function is attained at $p(x, y, z)$, which gives an optimal strategy.

Összefoglaló

A disszertáció a *Chlamydia* baktériumfaj sztochasztikus modellezését mutatja be. Először megadunk egy Galton–Watson modellt, mely egy baktériumpopuláció növekedését írja le antibiotikum jelenlétében. A sztochasztikus modellünk sokkal természetesebb a korábbi determinisztikus modellekhez képest (ld. Liu és munkatársai cikkében [5]). Feltettük, hogy az utódok várható értékét az $m(c) = 2/(1 + \alpha c^{\beta})$ formula adja meg, ahol c az antibiotikum koncentráció, valamint $\alpha > 0, \beta > 0$ ismeretlen paraméterek. A qPCR technikában figyelembe véve a mérési hibát, különböző antibiotikum koncentráció esetén gyengén konzisztens, valamint aszimptotikusan normális becsléseket kaptunk az ismeretlen (α, β) paraméterekre. Szimulációs eredményeink azt mutatják, hogy a modellünk jól írja le a baktérium valódi viselkedését.

Ezután a modellfeltevésünket meghagyva, azonban más mérési eljárást feltételezve a kihalási valószínűséget tudtuk becsülni. A kolóniaszámlálás során, ha x_0 számú egyedet oltanak rá egy sor antibiotikumot tartalmazó agarlemezre, akkor az inkubációs időszak végén az összes életképes telepet megszámolják. Feltettük, hogy az utódok eloszlását a $p_2(c) = m(c)/2$ formula adja meg. Gyengén konzisztens, valamint aszimptotikusan normális becsléseket kaptunk az (α, β) paraméterekre, valamint a MIC-re.

A *Chlamydia* populációk evolúciójának leírására egy új elágazó modellt adunk meg. Ebben a modellben az állapotfüggő utódeloszlás meghatározása sztochasztikus optimalizációs probléma megoldásával történik. A folyamatról az egyetlen bemeneti információnk a d halálozási függvény, mely megadja annak a valószínűségét, hogy a gazdasejt adott állapotban meghal. Természetes halálozási függvényt választva, a szimulációs eredmények azt mutatják, hogy a folyamat képes megfogni a baktériumsejtek aszinkron viselkedését, amit nem régen kísérletekkel is alátámasztottak [4]. A szimulált adataink rendkívül jól illeszkednek a valós mért adatokhoz. Legjobb tudomásunk szerint ez az első olyan matematikai modell, amely reprodukálja ezt a jelenséget.

Végül korábbi eredményeinket kapcsoltuk össze. Megvizsgáltuk a *Chlamydia* optimális terjedését abban az esetben, ha jelen van az antibiotikum. Megjegyzendő, hogy a legjobb ismereteink szerint valós mérési adatok nincsenek. Első esetben a modellünket annyiban egyszerűsítettük, hogy az antibiotikum csupán az RB alakra fejti ki hatását, majd vizsgáltuk azt az esetet is, amikor mindkét típusra hatással van az antibiotikum.

References

[1] N. S. Goel and N. Richter-Dyn. *Stochastic models in biology.* Elsevier, 2013.

[2] O. Hernández-Lerma and J. B. Lasserre. *Discrete-time Markov control processes: basic optimality criteria*, volume 30. Springer Science & Business Media, 2012.

[3] M. Kimmel and D. E. Axelrod. *Branching processes in biology*, volume 19 of *Interdisciplinary Applied Mathematics*. Springer, New York, second edition, 2015.

[4] J.K. Lee, G.A. Enciso, D. Boassa, C.N. Chander, T.H. Lou, Pairawan S.S., M.C. Guo, Wan F.Y.M., M.H. Ellisman, C. Sütterlin, and M. Tan. Replication-dependent size reduction precedes differentiation in chlamydia trachomatis. *Nature Communications*, 45(9):3884–3891, 2018.

[5] Y. Q. Liu, Y. Z. Zhang, and P. J. Gao. Novel concentration-killing curve method for estimation of bactericidal potency of antibiotics in an in vitro dynamic model. *Antimicrobial Agents and Chemotherapy*, 48(10):3884–3891, 2004.

[6] M. Santillán. On the use of the Hill functions in mathematical models of gene regulatory networks. *Mathematical Modelling of Natural Phenomena*, 3(2):85–97, 2008.

[7] J. S. Yuan, A. Reed, F. Chen, and C. N. Stewart. Statistical analysis of real-time PCR data. *BMC Bioinformatics*, 7(85), 2006.

Társszerzői nyilatkozat

Kijelentem, hogy ismerem Szalai Máté Ph.D. fokozatra pályázó Stochastic Modeling of *Chlamydia* című értekezését, amelyet a Szegedi Tudományegyetemre nyújt be. A következő cikk eredményeiben a pályázó hozzájárulása meghatározó volt:

[1] A. Bogdanov, P. Kevei, M. Szalai, D. Virok: Stochastic modeling of in vitro bactericidal potency. *Bulletin of Mathematical Biology* 84 (6), 2022.

Szalai Máté hozzájárulása a fenti cikkhez 25%.
Kijelentem, hogy a fenti eredményeket nem használtam fel, és nem is fogom felhasználni tudományos fokozat megszerzéséhez.

Szeged, 2024. október 16.

Bogdanov Anita
Szegedi Tudományegyetem

Kevei Péter
Szegedi Tudományegyetem

Virok Dezső
Szegedi Tudományegyetem

Társszerzői nyilatkozat

Kijelentem, hogy ismerem Szalai Máté Ph.D. fokozatra pályázó Stochastic Modeling of *Chlamydia* című értekezését, amelyet a Szegedi Tudományegyetemre nyújt be. A következő cikkekből felhasznált eredményekben a pályázó hozzájárulása meghatározó volt:

[2] M. Szalai, P. Kevei: Estimation of in vitro bactericidal potency based on colony counting method. *22nd European Young Statisticians Meeting – Proceedings*, pages 133 – 137. Panteion University of Social and Political Sciences, 2021.

[3] P. Kevei, M. Szalai: Branching model with state dependent offspring distribution for Chlamydia spread. *Mathematical Modelling of Natural Phenomena*, 2024, 19: 14.

Szalai Máté hozzájárulása a fent felsorolt cikkekhez 50-50%.
Kijelentem, hogy a fenti eredményeket nem használtam fel, és nem is fogom felhasználni tudományos fokozat megszerzéséhez.

Szeged, 2024. október 16.

Kevei Péter
Szegedi Tudományegyetem