Modelling and parameter estimation of bacterial growth with distributed lag time

PhD Thesis

by

József Baranyi

Doctoral School of Informatics

University of Szeged, Hungary

2010

Acknowledgement

The study embraces the last two decades I worked for the Institute of Food Research, UK. For the opportunity to become a research scientist, I am indebted to Dr Terry Roberts, the former head of the Microbiology Department there, who decided that further advances in food microbiology was impossible without a dedicated mathematician.

Content

Most frequently used variables and parameters

$\alpha(t)$	A quantity characterising the physiological state of the bacterial culture at the time t ($0 \le \alpha(t) \le 1$) . In simplest form, $\alpha(t)=\mu(t)/\mu_{max}$.
α_0	Value of $\alpha(t)$ (at the moment of inoculating a bacterial culture in a new environment). $\alpha_0=e^{-\mu \cdot \lambda}$
α_N	(Initial) Physiological state parameter for a population consisting of N cells at inoculation.
h_0	$= \mu \cdot \lambda = -\ln \alpha_0$
λ	Generic notation for the lag time (defined in various, empirical or mechanistic ways).
L, L_g	In the exponential phase, the cell population grows, on the log scale, as the delayed linear function $y(t) = \mu \cdot (t-L)$. The $_g$ subscript refers to the fact that this is a rather "geometrical" (opposed to "physiological") definition.
$L^{(i)}$	The L-lag for the cell i . Though strictly speaking τ_i , the physiological lag, is not the same as $L^{(i)}$ we identify them unless their very difference is studied. Numerical simulation can prove that $E(L^{(i)})=E(\tau_i)$ and their variance is close to each other, provided the generation times in the exponential phase do not have too big variance.
L_N	"Geometrical" lag as above, emphasizing that it is generated by N initial cells.
$\mu(t)$	Instantaneous specific growth rate at the time t : $\mu(t)= dx/dt /x(t) = d(\ln x(t)) / dt$
μ_{max} , μ	(Maximum) specific growth rate
m	Curvature parameter for our sigmoid growth model, characterizing the transition from the exponential to the stationary phase.
ν_b	Specific accumulation rate of a bottle-neck intracellular substance needed to overcome the lag time
n_C	Curvature parameter characterizing the transition from the lag to the exponential phase. In our model $n_C=\nu_b/\mu_{max}$
N , N_0	Inoculation level of a batch culture. x_0 is also used.
τ_i	A random variable, the "physiological" lag time of the single cell i. The time to the first division is the sum of τ_i and the first (random) generation time.
t	time
T_{det}	Detection time. $x(T_{det})= X_{det}$
$x(t)$	Cell concentration at the time t
x_0	Initial cell concentration, $x(0)$
x_{max}	Maximum population density; the carrying capacity of the environment.
X_{det}	Detection level (constant): $x(T_{det})= X_{det}$
$y(t)$	Natural logarithm of cell concentration at the time t : $y(t)=\ln x(t)$
y_0	Natural logarithm of initial cell conc.: $y_0 = y(0)$
y_{max}	Natural logarithm of maximum population density: $y_{max} =\ln X_{maxt}$
y_{det}	natural logarithm of the detection level: $y_{det} =\ln X_{det}$

Introduction

A highly interdisciplinary and fast developing area of applied mathematics is mathematical biology, sometimes also termed as biomathematics. The Biotechnology and Biological Sciences Research Council of the UK, sponsoring among others the Institute of Food Research in Norwich, where the author of this thesis works, has always been in the forefront of supporting biological research that applies advanced mathematical modelling techniques. (A good example for this was Ronald A. Fisher (1890-1962), one of the pioneers of mathematical statistics, who worked for the predecessor of BBSRC for many years).

Food security, defined as a sustainable supply of "sufficient, affordable, nutritious and safe food, adapting to a rapidly changing world", was identified as the No. 1 research priority of the BBSRC Strategic Plan 2010 – 2015 (http://www.bbsrc.ac.uk/publications/policy/strategy/index.html). Food safety and, generally, food microbiology problems belong to this strategic area, in which the increasingly extensive application of mathematical modelling techniques is expected to play a primary role (e.g. on modelling the microbial ecology of foods and the gastrointestinal tract).

In the 80's, with powerful desktop computing becoming everyday use, the name 'predictive microbiology' was coined for this area of mathematical microbiology. Its closest relative can be found in biotechnology with its mathematical models well elaborated in the 60's. When the author of this thesis started his work in this field, ideas regarding microbial growth models applied in biotechnology were taken from the classic papers of Tsuchiya et al (1966), Frederickson et al (1967), Turner et al (1976), Roels and Kossen (1978), Srivastava and Volesky (1990), Nilsen and Villadsen (1992).

However, three major points of difference had to be taken into account (Baranyi et al, 1994):

- The bacterial concentration region of concern (usually 1 - 10^9 cells/gram) in food microbiology is much broader than the concentration region that biotechnologists are interested in. It may be inappropriate therefore to extrapolate the mathematical models applied in biotechnology. Another numerical consequence of the broad region of interest is that, for statistical/computational reasons, it is not the cell concentration but its logarithm which is the preferred variable to be modelled.

- The aims are different. Generally, the aim of food microbiologists is to prevent bacterial growth (rather than for example to optimize it). The region for food safety problems are typically below 10^2 cell/gram; food spoilage issues focus on cell concentrations below $10^5 – 10^7$ cell/gram. Even the latter one is lower than the typical cell concentrations in biotechnological studies. Therefore, in food microbiology, the data at the initial stage of the bacterial growth, at low cell concentrations, are deemed to be uncertain but their study is of great importance.
- In food microbiology, the environmental factors influencing growth are generally heterogeneous, their measurements are inaccurate and often inadequate; sometimes, such as in the case of food structure, rather poorly quantified.

These are the main reasons why new approaches were needed when modelling bacterial growth and survival in food. The first attempts to put down a mathematically solid framework for bacterial growth in food were Zwietering et al (1990) and Van Impe et al (1992). Models taking the above differences into consideration, in a dynamic model described by Ordinary Differential Eequations, were published first by the author of present thesis in a series of papers (Baranyi et al, 1993a, 1993b, 1994, 1995a, 1995b).

In the following chapters we show the process of model development, from deterministic models of population growth, with special attention to the bacterial lag, to stochastic models of the kinetics of single cells. We demonstrate how to integrate the two and what numerical algorithms they imply for practical applications. We point out the differences between growth- and (seemingly mirror-image) survival- modelling.

Chapter 1 introduces some classical deterministic growth models, some of them with a history that spans through centuries. One of the main drives behind our modelling approach (a main difference from the models analysed by Zwietering et al, 1990) is the recognition that the classical models should not be directly applied to bacterial growth on the log scale, especially if the lag time has a significant role.

Chapter 2 discusses deterministic models of bacterial batch cultures commonly generated in food microbiology laboratories. As we will see, they are (on the log scale) of sigmoid shape but, from mechanistic point of view, it is incorrect to model them by one of the well-known growth functions described in Chapter 1 (such as the Gompertz or the logistic model). On the log-scale, those classical growth models have monotone decreasing slope (they are sigmoid on

the arithmetical scale only). The reason for the real curve being of sigmoid shape is that the cell population needs to adapt to the new environment after inoculation. This adjustment period, somewhat incorrectly, is traditionally called the lag period. In fact it is more like a transient period connecting two autonomous systems and a lag parameter is for convenience only rather than for mechanistic interpretation. Its connection to the physiological lag concept interpreted at the single cell level is far from straightforward, as we will see in chapters that follow.

Chapter 3 concentrates on the implementation of the new model, from regression point of view. We put special emphasis on optimal reparametersisations, for the sake of numerical stability. We also report on the possible biological interpretations of some of the newly introduced parameters.

Integrating single-cell and cell-population models is the main topic of Chapter 4. Assuming, in the exponential phase, a delayed $\mu(t\text{-}L)$ linear function for the logarithm of a single-cell-generated subpopulation, where the L delay parameter is a stochastic variable, the lag appears at population level as a sort-of (non-arithmetical!) average of the individual lag times. We will give an approximation for the expected value of the population lag, finding its main links with the distribution of the single cell lag times. The theory provides a set of tools to investigate the variability of single-cell kinetics, which will be explored in the following chapters.

Chapter 5 refines the theory introduced above and analyses the “compatibility” between parallel growth and survival models. One of its main messages is that the analogy breaks down in many aspects.

In Chapter 6 and 7 we show how the developed models can be used in practical applications, to estimate statistical and kinetic parameters of bacterial populations. The methods have been extensively used in laboratories since their publication.

Chapter 8 generalizes the short-term (non-genetic) adaptation model introduced earlier, for future research. It emphasizes the importance of databases, the handling of data deluge and studying its implications in the near future.

The thesis follows the time-sequence of the publications of the author, spanning through 18 years. An overall conclusion inferred by these papers is that a healthy modelling framework cannot be completed without observing and modelling intracellular processes during the lag time. Accordingly, in the research group led by the author, laboratory measurements and modelling effort have been initiated and being carried out at the moment at molecular level. Concepts and computational methods borrowed from Network Science is being applied to the dynamics how a cell genetic (more precisely transcriptomic) intracellular networks change during the lag time. The work is designed to lead to a complete analysis of bacterial adjustment, from theory to application. We hope, with this, to contribute to the development of predictive food microbiology and quantitative microbial risk assessment, and as a final goal, to improve the microbiological safety and quality of our foods.

Chapter 1

Modelling population growth

1.1. A deterministic approach

The simplest model of population growth was derived by Malthus (1798), after observing that a homogeneous population in a constant environment grows exponentially (al least until the carrying capacity of the environment inhibiting its growth). That is, the variation of the *x(t)* cell number with time can be described by a linear differential equation with a constant coefficient:

$$\frac{dx}{dt} = \mu\, x \qquad (1.1)$$

The evidence is strong that μ, called the specific growth rate of the population, is characteristic to the species and the environment (Rubinow, 1984). Because of

$$\mu(t) = \frac{\frac{dx(t)}{x(t)}}{dt} = \frac{d \ln x(t)}{dt} \qquad (1.2)$$

μ can be conceived as a "sort-of" average number of divisions for a cell per unit time (the reason, why this is only a "sort-of", is that this average is not necessarily arithmetical; see Chapters 4-6). Indeed, for bacterial populations, the statistical distribution of the number of division per unit time can be measured (Baranyi, 2005) and described by a stochastic birth process. Using an analogy: just like humans can have maximum about one child/female a year, in the same way a bacterium like *Escherichia coli* can divide maximum ca 3 times in an hour, due to its biological make-up (Koch, 1988).

The specific growth rate, as a function of environmental factors such as temperature, pH etc, can be very accurately predicted by observing individual cells (Pin and Baranyi, 2008). The problem is that the environment-dependent specific growth rate is, in practice, an instantaneous parameter; in fact it is very difficult to make sure that the environment is reasonably constant for sufficiently long time, so that μ can be measured really accurately. If yes, the easiest approach is to fit a linear function to the exponential phase of the "ln *x(t) v.* time" curve:

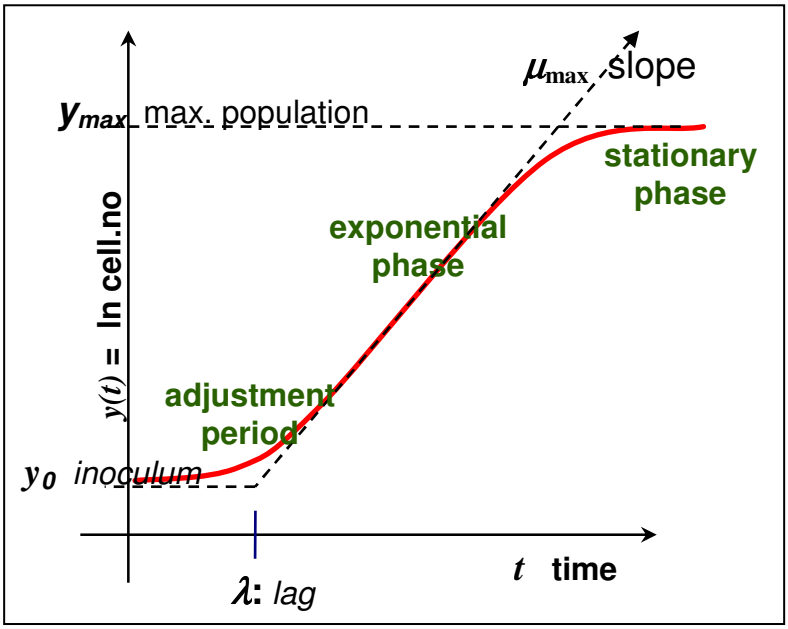


Fig. 1.1. Phases of a typical bacterial growth curve commonly analyzed in food microbiology

In food microbiology research, experimenters typically sample the cell population after applying a series of dilutions, so that they can count ca 50-300 colonies, each produced by a single. Therefore both the measurement errors and the expected growth kinetics justify that the logarithm of cell concentrations are considered as raw data. The accuracy of the measured concentrations can be assessed by the observation that the number of significant digits are typically around two.

These curves commonly show a sigmoid shape (see Fig.1.1). The raw data are distributed around them with more or less normally distributed error.

Following the notations of Fig.1.1

$y(t)$ is the natural log of the population at the time t.

$y_0 = \ln x_0$ - natural log of the initial population;

μ_{max} - maximum specific growth rate ($\mu(t)$ will denote the instantaneous specific growth rate at the time t);

λ - duration of the lag time

$y_{max} = \ln x_{max}$ - natural log of the maximum concentration the culture can reach;

For convenience, a so-called lag-parameter, λ, is also defined for sigmoid curves; it is empirically described as the time when the interception of the tangent drawn to the inflexion crosses the inoculum level (Pirt, 1975). The fact that this definition does not have mechanistic basis and a possible correction will be a main point of the first three chapters.

1.2. Classical sigmoid models of growth

There are several useful sigmoid growth models that could be fitted to the bacterial growth curves, but those were originally conceived for the number of individuals in the growing population, not for its logarithm, and this makes their use an empirical curve fitting exercise only. However, can we modify them in order to obtain a suitable modelling framework?

Let the starting point be the well-known generic autonomous growth model (Vance, 1990):

$$\frac{dx}{dt} = \mu(x)\, x \tag{1.3}$$

$$x(0) = x_0 \qquad (0 < x_0 < x_{max} \;; \qquad x_{max} \text{ is fixed}) \tag{1.4}$$

where

- μ: $(0, x_{max}] \longrightarrow \mathrm{R}$
- $\mu(x)$ is continuously differentiable on $(0, x_{max})$
- $d\mu/dx$ is strictly negative on $(0, x_{max})$
- $\mu(x_0) > 0$ and $\mu(x_{max}) = 0$

As well-known, under these conditions the above differential equation has a unique solution, which is monotone increasing and converges to x_{max} as $t \to \infty$.

Turner et al. (1976) published a formula for $\mu(x)$, called 'generic equation', which is general enough to include most of the well-known sigmoid functions: Gompertz (1825); Pearl-Verhulst (or logistic: Pearl, 1927); Richards (1959), etc. For example, for the Gompertz function:

$$\mu(x) = c \ln\left(\frac{x_{\max}}{x}\right) \tag{1.5}$$

Introduce $\mu(x) = \mu \cdot u(x)$, where μ is the maximum specific growth rate, as defined earlier, and $u(x)$ is the so-called inhibition function. For Richards' model:

$$\mu(x) = \mu_{\max} \cdot u(x) = \mu_{\max} \cdot \left(1 - \left(\frac{x}{x_{\max}}\right)^m\right) \tag{1.6}$$

where m is a curvature parameter, characterising the "abruptness" of the transition to the stationary phase. For m=1, the renowned logistic model is obtained. Fig.1.2. shows the

obvious difference between the inhibition functions depending on whether we consider it as a function of the x population size or its logarithm.

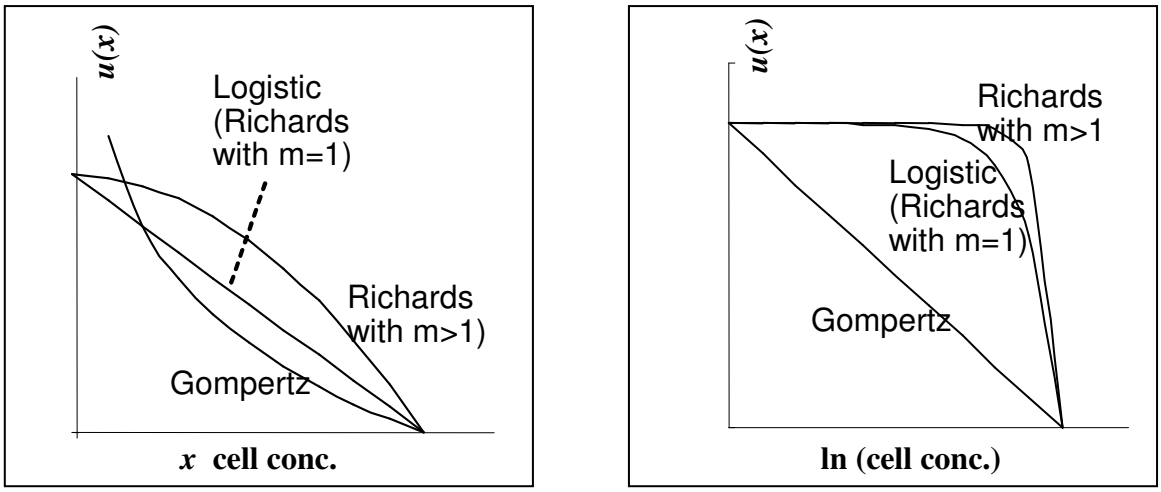


Fig.1.2. Some classical sigmoid models. They differ in their inhibition functions.

1.3. Are the classical growth models suitable to model bacterial growth in food?

The autonomous nature of the model (1.3) – (1.4) means that the $\mu(x)$ specific rate of the population depends only on the actual number of cells and it does not depend, for example on their internal physiological state. That this cannot hold for our problem can be easily seen on two growth curves measured in our laboratory (http://www.ifr.ac.uk/safety/comicro/; thousands of such curves can be freely downloaded from www.combase.cc) :

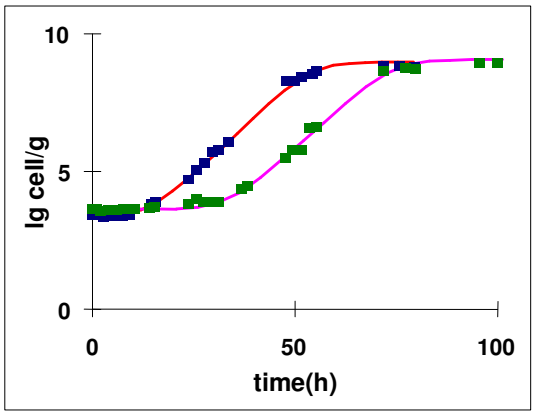


Fig.1.3. Two different growth curves of the same organism, same environment, except that the cells along the red curve (first to grow) had more favourable history than those in the other culture. The maximum specific rates are the same but the lag (adjustment) times are markedly different.

The specific rate of these curves obviously must depend on something else, too, not only on the cell concentration level. So we have to give up the autonomous property, at least if the number of variables is only one (as known, every Ordinary Differential Equation can be made autonomous, at the expense of introducing an extra variable).

Non-autonomous growth models have already been analysed for example by Vance (1990) and Vance and Coddington (1989), but their “non-autonomy” came from the seasonal changes

of the environment. Coleman (1978) also analyzed non-autonomous logistic equations to characterize the adjustment of the population to environmental changes. Our problem is more like the “distributed delay” studied by McDonald (1978).

Before the 90’s, several food microbiology papers used classical sigmoid growth models as above, but for the *logarithm of the cell concentrations*. This was obviously of rather limited use, making mechanistic interpretation impossible. For details, see Zwietering et al. (1990), who compared these models from fitting and robustness point of view.

Chapter 2

Deterministic modelling of adaptation

2.1. A non-autonomous model

The problem outlined above was solved in a series of papers published by the author of this thesis The model described by Baranyi and Roberts (1994) is commonly called the "Baranyi-model" in predictive microbiology (see for example Buchanan et al, 1997; Grijspeerdt and Vanrolleghem, 1999; Xiong et al, 1999a; Vadász and Vadász 2007).

We assumed that the effect of the actual environment is described by the autonomous ODE (1.3). Depending on the choice of $\mu(x)$, different sigmoid functions can be obtained for solutions. If $f(t)$ denotes the particular solution for which $f(0)=x_0$ (see (1.4), then the general solution of (1.3) takes the form $f(t-L)$, where L is a delay parameter. The solutions of (1.3) are monotone increasing and converging to x_{max} as $t\to\infty$, with a sigmoid shape (Vance, 1990).

Note that the condition $d\mu/dx < 0$ means that the larger the population density the lower its specific growth rate. As can be shown, this means that the derivative of the logarithm of the solution is strictly monotone decreasing. However, due to the history-effect (see the background of the lag concept in the Introduction), growth curves in food microbiology show sigmoid shape on the log-scale, too! The (1.3) autonomous growth model will be a suitable description in asymptotic sense only.

Apply a multiplier, a so-called adjustment function, to the $\mu(x)$ function:

$$\frac{d}{dt}x = \alpha(t)\mu(x)\cdot x \qquad (0 \le t < \infty; \quad 0 < x) \tag{2.1}$$

$$x(0) = x_0 \qquad (0 < x_0 < x_{\max}) \tag{2.2}$$

where $\alpha(t)$ is a sufficiently smooth function. It can depend on the pre-inoculation environment, too; besides we require that

$0 \le \alpha(t) \le 1 \quad (0 \le t < \infty)$

$\alpha(t) \to 1$ monotone increasingly as $t \to \infty$

We will say that the model defined by (1.3) - (1.4) describe the *potential growth*, and (2.1) - (2.2) define the *actual growth* in the current environment.

The biological background can be summarized as follows: The environment changes suddenly at the time $t = 0$. This happens when food becomes contaminated, or a pathogenic cell gets in the gastrointestinal track, or the cells are inoculated in a new, fresh medium under laboratory conditions. Let $\mathbf{E_1}$ denote the pre-inoculation environment ($t < 0$) , and $\mathbf{E_2}$ the actual growth environment ($t \geq 0$). $\mathbf{E_1}$ and $\mathbf{E_2}$ are commonly represented by a vector of environmental variables. In predictive microbiology, $\mathbf{E_i}$=(temperature, pH, Aw, Air, preservatives … etc) , where Aw denotes the water availability, Air characterizes the air composition.

The inoculum level, x_0, uniquely determines a potential growth curve, according to which the population would be able to grow if the previous environment had been the same as the present environment ($\mathbf{E_1 = E_2}$: no need to adjust, that is $\alpha(t) \equiv 1$ and $\lambda=0$). The potential growth of the population is described by the autonomous model (1.3). The actual growth, however, is described by (2.1), which means that initially, for small t values, the cells' growth is heavily inhibited because they need to spend their energy for adaptation. Later, however, the effect of the previous environment diminishes, until, some time after the inoculation, it has little or no effect at all and the cells grow essentially at their potential growth rate, $\mu(x)$, defined by the new environment, E_2. Therefore the ratio of the actual and the potential growth rate, *i.e.* the adjustment function, is expected to increase from close to zero (no growth) to 1 (total adjustment).

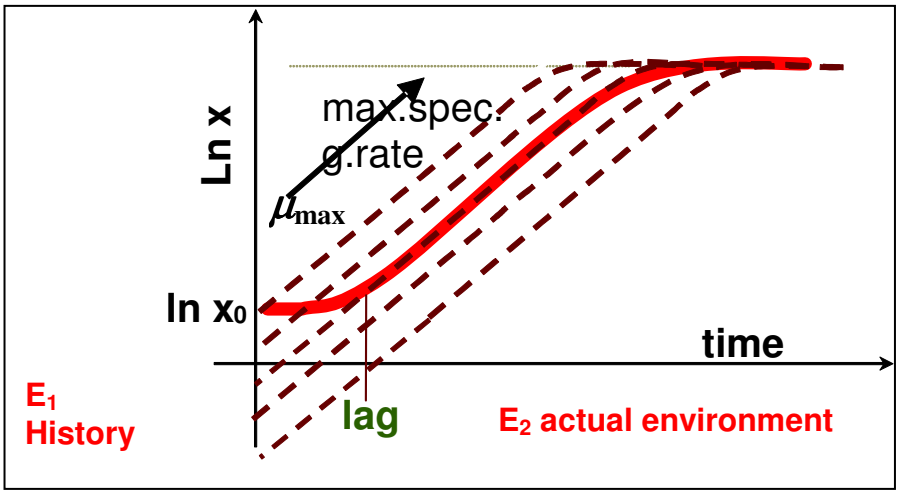


Fig.2.1. The reason for the sigmoid shape is the adjustment to the actual environment, which can be described by a classical growth model, an autonomous ODE.

It can be readily checked that the solution $g_\alpha(t)$ for the (2.1) non-autonomous model can be expressed by means of the solution of the autonomous model, f, and the adjustment function, α. Its form is:

$$g_\alpha(t) = f(A(t)) \qquad (2.3)$$

where

$$A(t)=\int_0^t \alpha(s)ds \qquad (2.4)$$

(see Baranyi et al (1993b)

2.2. Using the Hill-function to model the adaptation period

In Baranyi et al (1993a), the Hill-function (Hill, 1910) was tested for the role of the adjustment function:

$$\alpha(t)=\frac{t^n}{\lambda^n+t^n}=\frac{(t/\lambda)^n}{1+(t/\lambda)^n}=\frac{z^n}{1+z^n} \qquad (z=t/\lambda) \qquad (2.5)$$

where n is a positive integer, λ is a parameter. It is evident that it satisfies the criteria of an adjustment function; besides

$\alpha(0) = 0$; $\alpha(\lambda) = ½$; $\alpha(t) \to 1 \quad (t \to \infty)$; $\left.\frac{d\alpha}{dt}\right|_{t=0} = 0$.

Furthermore, it has an inflexion point at $\Lambda_n = \lambda \cdot \sqrt[n]{\frac{n-1}{n+1}}$. As was pointed out in Baranyi et al (1993b), the λ parameter is a suitable definition of the lag time, since for $t>\lambda$ the $g_\alpha(t)=f(A(t))$ solution of the non-autonomous growth model is close to a solution of its autonomous counterpart of the form $f(t-\Lambda_n)$. It was shown that

$A(t) = t - B_n(t/\lambda)$ where

$$B_n(z)=\int_0^z \frac{1}{1+s^n}ds \qquad (2.6)$$

For n=4, the most commonly used curvature parameter, when $\Lambda_4 = 0.88 \cdot \lambda$, the above integral is

$$B_4(z)=\frac{1}{2\sqrt{2}}\left(\frac{1}{2}\ln\frac{z^2+\sqrt{2}z+1}{z^2-\sqrt{2}z+1}+\gamma(z)\right) \qquad (2.7)$$

where

$$\gamma(z)=\begin{cases}\arctan\dfrac{\sqrt{2}z}{1-z^2} & (t<1)\\ \dfrac{\pi}{2} & (t=1)\\ \arctan\dfrac{\sqrt{2}z}{1-z^2}+\pi & (t>1)\end{cases} \qquad (2.8)$$

Though this specific form of the model, published in Baranyi et al (1993a), was used and cited by many papers in the 90's, it had some disadvantages. The (2.5) adjustment function resulted in recursive formulae (from n to n+1) for the needed integral defined by (2.4). The (2.8) formula was not straightforward to program correctly, e.g. in an ordinary spreadsheet, because of its relative complexity of. Another, more serious problem with it was the lack of mechanistic interpretation. It was dubious how to apply the model, if the environment changed dynamically during lag.

2.3. The most commonly used adjustment function

Baranyi and Roberts (1994) introduced an adjustment function with some mechanistic model assumptions:

Suppose that an intracellular substance is the bottle-neck in the process of growth (such as certain enzymes to metabolize new substrates in the new environment). Denote its time-dependent *per cell* concentration by $P(t)$. Suppose that the specific growth rate of the cells is affected by $P(t)$ according to the well-known Michaelis-Menten kinetics:

$$\alpha(t) = \frac{P(t)}{K_P + P(t)} \qquad (2.9)$$

where K_P is the Michaelis-Menten constant and that, after the inoculation, the accumulation of $P(t)$ follows a first order kinetic process

$$\frac{d}{dt}P = \nu_b P \qquad (2.10)$$

where the rate of the bottle-neck substance, ν_b, is characteristic of the actual environment:

$$\nu_b = \nu_b(\mathbf{E_2}). \qquad (2.11)$$

The adjustment function depends on the ratio $q(t)=P(t)/K_P$. A rescaling of $q(t)$:

$$\alpha(t) = \frac{q(t)}{1 + q(t)} \qquad (2.12)$$

can be used to characterize the gradual adjustment of the cells to the new environment.

Namely:

$$\frac{d}{dt}x = \frac{q(t)}{1 + q(t)}\mu(x) \cdot x \qquad (2.13)$$

$$\frac{d}{dt}q = \nu_b \cdot q \qquad (2.14)$$

with the initial values $x(0)=x_0$; $q(0)=q_0=P(0)/K_p$

If the actual environment, $\mathbf{E_2}$, is constant then $P(t)$ grows exponentially at a constant specific rate, v_b, and our new adjustment function can be obtained in the form:

$$\alpha(t) = \frac{q_0}{q_0 + e^{-v_b \cdot t}} \tag{2.15}$$

therefore

$$A(t) = t + \frac{1}{v}\ln\left(\frac{e^{-v_b t} + q_0}{1 + q_0}\right) \tag{2.16}$$

The parameter

$$\lambda = \frac{\ln(1+1/q_0)}{v_b} \tag{2.17}$$

turned out to be a good definition for the lag time: $f(A(t))$ will be a sigmoid curve on the log scale (Fig. 2.2), converging to $f(t-\lambda)$

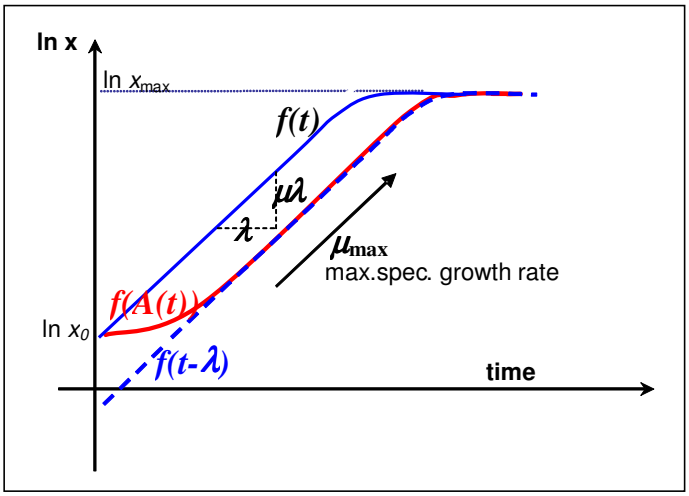


Fig. 2.2. The relationship between $f(t)$, the potential growth curve, its delayed version, $f(t-\lambda)$, and $f(A(t))$, the solution of the non-autonomous model. The sigmoid shape is due to the inhibited growth before and after the linear segment (exponential growth).

Chapter 3

Model implementation and numerical considerations

3.1. A versatile sigmoid function

Consider the model of Richards (1959) for (2.1):

$$\frac{dx(t)}{dt} = \mu_{\max} x(t)\left(1 - \left(\frac{x(t)}{x_{\max}}\right)^m\right) \qquad (m>0) \qquad (3.1)$$

As can be checked, it has algebraic solution. Express it for $y(t)$, the natural logarithm of the cell concentration:

$$y(t) = y_0 + \mu_{\max} t - \frac{1}{m}\ln\left(1 + \frac{e^{m\mu_{\max} t} - 1}{e^{m(y_{\max} - y_0)}}\right) \qquad (3.2)$$

Most frequently, the parameter m is taken to be 1, in which case the well-known logistic growth can be obtained for the bacterial concentration.

Using

$$u(x) = \left(1 - \left(\frac{x(t)}{x_{\max}}\right)^m\right), \qquad (3.3)$$

the model that combines the logistic growth model with our adjustment function, detailed by (2.12) – (2.14) , can be written as

$$\frac{dx(t)}{dt} = \frac{q(t)}{1+q(t)}\mu_{\max} x \cdot u(x) \qquad (3.4)$$

$$\frac{dq(t)}{dt} = \nu_b \qquad (3.5)$$

$$x(0) = x_0 \quad ; \quad q(0)=q_0 \qquad (3.6)$$

The solution for $y(t)=\ln x(t)$ is

$$y(t) = y_0 + \mu_{\max} A(t) - \frac{1}{m}\ln\left(1 + \frac{e^{m\mu_{\max} A(t)} - 1}{e^{m(y_{\max} - y_0)}}\right) \qquad (3.7)$$

Here, $A(t)$ is as defined by (2.16).

3.2. Reparameterisations for numerical stability and for biological interpretation

Another parameterization of (2.16) is:

$$A(t) = t - \lambda + \frac{\ln\left(1 - e^{-\nu_b \cdot t} + e^{-\nu_b (t-\lambda)}\right)}{\nu_b} \qquad (3.8)$$

where

$$\lambda = -(\ln \alpha_0)/\nu_b = h_0/\nu_b \qquad (3.9)$$

$$\alpha_0 = q_0/(1+q_0) \qquad (3.10)$$

It can be easily seen that the parameter λ defined by (3.9) is very close to the traditional definition of the lag time given in the Introduction under Fig.1.1.

The above reparameterization can be useful when selecting an optimal arrangement of the expressions for numerical stability. Namely

$$A(t) = t - \lambda + \frac{\ln\left(1 - e^{-\nu_b \cdot t}(1 - e^{h_0})\right)}{\nu_b} \qquad (3.11)$$

From this rearrangement, due to $\ln(1-\varepsilon) \approx -\varepsilon$ if ε is small, it is obvious, that $A(t)$ is exponentially converging to $t-\lambda$ as t is increasing. What is more, the e^{h_0} term is a stable parameter, since for the majority of experimental curves $h_0 < 5$, regardless of the experimental conditions (see Fig. 3.1), therefore a generic value can be chosen for a threshold t from which $A(t)=t-\lambda$ can be used.

The

$$\alpha_0 = \frac{q_0}{1+q_0} \qquad (3.12)$$

$$h_0 = -\ln \alpha_0 \qquad (3.13)$$

reparameterization of the q_0 initial value also have biological interpretations. One is given by Baranyi and Pin (1999): if only α_0 fraction of the inoculum grows, but *without* lag, they would provide the same growth in the end as the actual total population *with* lag (Fig. 4.1).

Baranyi and Roberts (1994) pointed out that the $\nu_b = \mu$ assumption is a useful one, in which case

$$h_0 = \frac{\lambda}{1/\mu} \qquad (3.14)$$

where $1/\mu$ is in fact a "sort-of" generation time in the exponential phase (the exact statement depends on the distribution of the generation times of single cells – see later). The parameter

h_0 can also be considered as a backlog, by how many generations the culture is behind a potential one, which would NOT have needed to adapt. In other words: the delay due to the need to adapt can be measured by h_0 number of "generation-backlog". As Fig.3.1 shows, this is independent of the temperature that affects only the speed of the process, in accord with intuition.

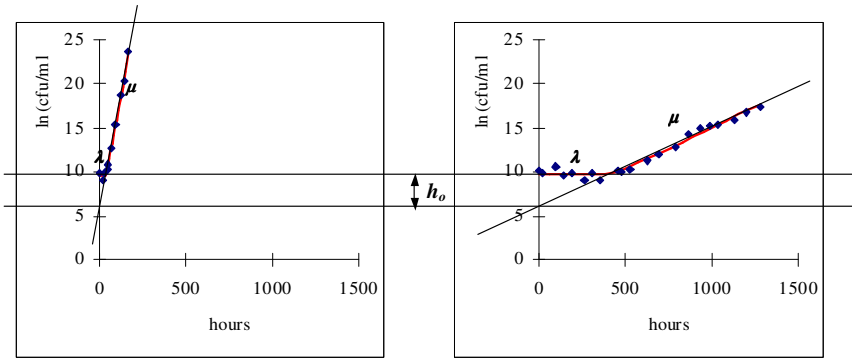


Fig. 3.1 The $h_0=\lambda\cdot\mu$ parameter is the same for the two bacterial curves, though they were generated at very different temperatures (15 and 5 C). In both cases, the actual growth curve is ca 4 generations behind the potential growth curve.

The obtained (3.7)-(3.8) equations describe a sigmoid curve with six parameters. Out of these, four (y_0, y_{max}, μ_{max} and λ) are familiar; their role is similar to other four-parameter sigmoid curves. The fifth, the parameter m affects the curvature of the curve linking the linear segment and the upper asymptote. Similarly, the parameter ν affects the curvature from the lag to the linear phase. In practice, parallel to the definition of m, a dimensionless curvature parameter can be introduced also for the pre-exponential transition phase:

$$n_C = \nu_b / \mu_{\max} \qquad (3.15)$$

3.3. Regression by the algebraic solution – how many parameters?

With the notations above, a versatile sigmoid function is obtained, where the abruptness of the transition phases are controlled by the respective curvature parameters:

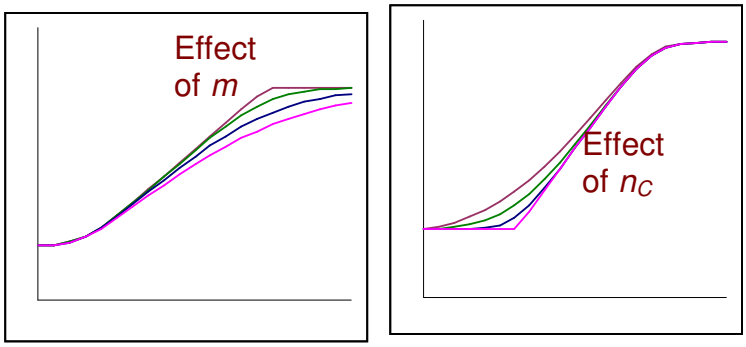


Fig. 3.2. Effect of the curvature parameters. The bigger they are, the more abrupt the transition period.

In practice, the curve fitting has to take into consideration that, with commonly available measurements of bacterial growth, the curvature estimates are not very robust. It is much more efficient to fix them at a certain value and concentrate on the remaining four parameters only. Then one can follow the method well-known in polynomial regression. Statistical tests, such as

the F-test can decide whether the parameters y_{max} or λ can be added to an originally linear fit, as shown below:

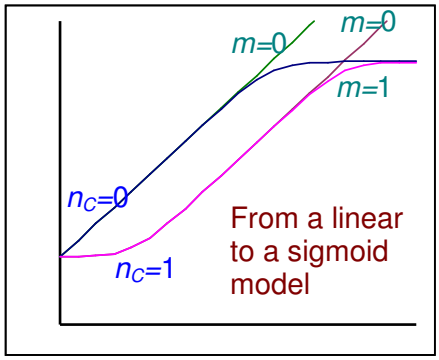


Fig. 3.3. Symbolically, we call the respective curvature parameters zero, if there is no transition. This is equivalent to the omitting a parameter of the sigmoid curve. This technique enables users to fit sigmoids progressively, starting from a linear (two-parameter-) model and adding the other two parameters one by one, provided that a statistical test approves it.

If one wishes to use the model to fit data other than bacterial counts, it is useful to consider the reparameterization of the equations in such a way that the optimal curvature parameters do not depend on the scale on which the raw data were measured. $h_0 = v_b\lambda$ instead of n_C, and $M=m(y_{max}-y_0)$ instead of m, can be such substitutions. The reparameterized form of the equations, with these h_0 and M new curvature parameters, can be written as:

$$y(t) = y_0 + \mu_{max} A(t) - \frac{y_{max} - y_0}{M} \ln\left(1 - e^{-M} + e^{-M\frac{y_{max}-y_0-\mu_{max}A(t)}{y_{max}-y_0}}\right) \qquad (3.16)$$

where

$$A(t) = t - \lambda\left(1 - \frac{1}{h_0}\ln\left(1 - e^{-h_0\frac{t}{\lambda}} + e^{-h_0\left(\frac{t}{\lambda}-1\right)}\right)\right) \qquad (3.17)$$

Following numerical experience, $5 < h_0 < 10$, $5 < M < 10$ are good empirical constants, to describe gentle curvatures.

It is important to see that the parameter μ_{max} is not necessarily the steepest slope of the actual growth curve, but expresses a 'potential' maximum specific growth rate that is characteristic to the population only if the initial concentration level is not very high. The difference between the steepest slope and μ_{max} is noticeable if the curvature parameters are small and/or y_0 is close to y_{max}.

Keeping robust numerical algorithms in mind when using the algebraic expression of the model for regression, it is useful to consider another form of (3.16). There, the sigmoid shape is partly due to a modification of the delayed linear segment, $y(t) = y_0+\mu_{max}A(t)$, with a logarithmic term. A rearranged form

$$y(t) = y_{\max} - \frac{1}{m}\ln\left(1 + \frac{e^{m(y_{\max}-y_0)}-1}{e^{m\mu_{ma}A(t)}}\right) \qquad (3.18)$$

shows the modification of its upper asymptote, $y_{\max}$. As t increases, the denominator in the logarithmic term in the latter arrangement can become very big, for which case a control should be introduced in the fitting program, applying the $\ln(1+\varepsilon)\approx\varepsilon$ approximation for small ε. That is, for large t : $y(t)=y_{\max}$.

3.4. Simulating bacterial growth under dynamic conditions

One of the most important features of the developed model is that the lag is considered as an adjustment process, where the rate of adjustment is ν_b. If the **E** actual environment changes during the lag time, while still remains in a growth-supporting region, then it is a reasonable assumption that the $\nu_b(\mathbf{E})$ and $\mu(\mathbf{E})$ instantaneously take up the values characteristic of the actual environment. The "comfort zone" of this assumption was studied by Baranyi et al (1995).

The existence of algebraic solution was significant for data fitting. However, if the environment changes during the bacterial response, algebraic solution generally does not exist. With the assumption above, we can still use the original ODE form (for the sake of simplicity, with the $\nu_b=\mu$ assumption):

$$\frac{dx(t)}{dt} = \frac{q(t)}{1+q(t)}\mu(\mathbf{E}(t))x\cdot u(x) \qquad (3.19)$$

$$\frac{dq(t)}{dt} = \mu(\mathbf{E}(t)) \qquad (3.20)$$

This model was tested by Baranyi et al (1995). Individual growth curves generated under different but constant **E** environments were fitted by the algebraic form, from which a $\mu=\mu(\mathbf{E})$ so-called secondary model was established. As have been mentioned, $h_0=\mu_{max}\cdot\lambda$ could be taken as a constant, from which the q_0 initial value was calculated by (3.9)-(3.10). The carrying capacity of the environment was also taken as constant (this quantity does not have significance in food microbiology anyway, because by the time bacteria reach the carrying capacity of the environment, the food is inedible anyway).

The result is demonstrated in Fig. 3.4. The environment was characterised by temperature only. The simulation predicted the bacterial growth very well, even during the adjustment (lag) period. This prediction would not have been possible with former predictive models used in predictive microbiology (see a summary of those in Zwietering et al, 1990). The key was the dynamic (ODE) background and the introduction of the physiological state characterised by a critical substance that is gradually improving at the same rate as the specific rate.

Note that the α_0=const assumption proved to be valid only in a "smooth" growth-supporting region of the temperature (see the original publication). In addition, recent studies have shown that the assumption does not hold for other environmental factors, like pH.

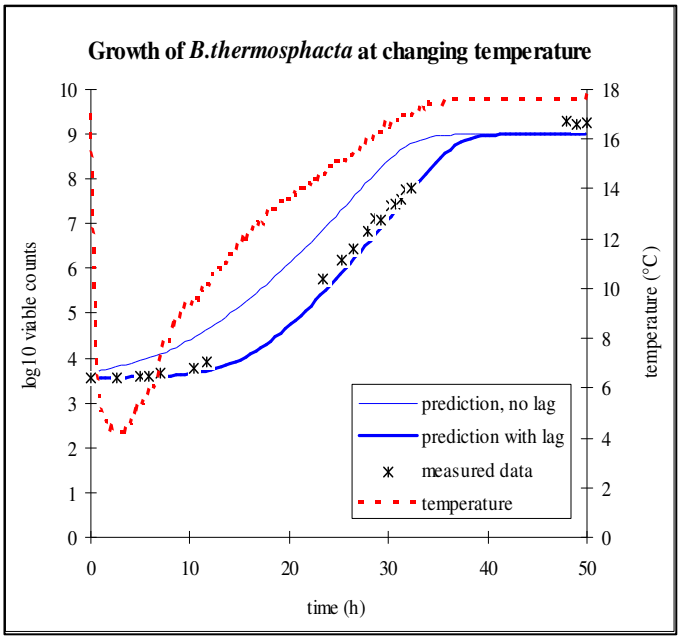


Fig. 3.4. Bacterial growth when the temperature changes (during the lag time, too). The temperature profile (dotted red line) is recorded in a fridge, the door of which was left open after 2 hrs, so the food put inside first cooled down then increased back to ambient. The data (black stars) are not fitted but predicted (thick blue line) from parameters measured at constant temperatures.
For demonstration, the thin blue line is the predicted bacterial growth, had the physiological state been 100% suitable to the new environment (i.e. $\alpha_0 = 1$; $h_0 = 0$: the no-lag situation.)

Chapter 4

A stochastic model of adaptation

4.1. A stochastic model of the lag time of individual cells

Baranyi (1998) introduced a stochastic birth process with random lag time which was distributed over the initial cell population. The paper became the basis of further papers detailed in later chapters.

Let the initial number of a growing cell population be $N_0=N$. Because here we omit the stationary phase, the end of the lag phase will be defined as the breakpoint of the bi-phasic linear function fitted to the $y(t)=\ln x(t)$ growth curve (see Fig.6.1, where more details are discussed about this definition). This *population lag,* L_N , is generated by N initial cells; as can be seen from the notation, we cannot guarantee that the population lag is independent of the initial counts, if the cells divide randomly.

Though the (rather geometrical) definition of the population lag has been used extensively in microbiology, its connection with the lag time of the individual cells has not been analysed sufficiently. The main reason why this should be looked into is that lag studies typically focus on low bacterial counts. As Renshaw (1991) pointed out, stochastic models should be used when studying small populations.

The physiological lag for a single cell is not necessarily the same as L_1. The end of the physiological lag, τ, is not marked by any special event; the only thing we know is that the time to the first division is the sum of the physiological lag and the following first generation time. This is why it is difficult to validate any physiological cell concept. The definition of the above defined "L-lag", on the other hand assumes and requires information on the generated subsequent subpopulation. Still, simulations show that the difference between the two is negligible, unless the generation times in the exponential phase are variable or/and their length is about the same as the lag. This is discussed in more details in Chapter 6.
For the sake of compatibility with the original publications, we denote the L_i -value for the cell i also by τ_i unless we want to emphasize their differences.

Denote the single cell lag time of the cell i of the initial population by τ_i. We assume that τ_i (i=1,2...N) are identically distributed independent random variables, and their expected value $E(\tau_i)$ is τ.

Suppose that a cell population, size of which is $x(t)$ at the time t, grows according to the classical Poisson birth process after the lag, with a constant birth intensity parameter, μ. Let $y(t) = \ln x(t)$ denote the logarithm of that population. Suppose that the subsequent subpopulations of each single cell of the initial culture grow together, but independently of each other. Simulate the population growth as above and measure the individual lag times by means of the fitted bi-phasic linear function where there is no growth in the first phase:

$$y(t) = y_0 + \max(\mu(t-L), 0) \quad (4.1)$$

Here the y_0 initial value is $y_0=\ln(1)=0$. If the exponential phase is long enough then the fitted μ parameter will be an accurate estimation of the birth intensity. As mentioned, the estimated L parameter is a close representation of the individual lag time: $L^{(i)} = L_1^{(i)}$ varies around τ.

Similarly, for N=1,2... , fit the bi-phasic linear function given by

$$y_N(t) = \ln(N) + \max(\mu(t-L), 0) \quad (4.2)$$

to the growth curves of the populations generated by N=1,2... cells. The expected value for the L parameter is L_N, the population lag produced by N cells. This way, the L_N sequence is defined for the expected population lag as a function of the initial cell number, N.

4.2. An analysis of the stochastic lag

Our investigation is focused on the L_N sequence. Because only the population lag can be measured by traditional microbiology methods, it is important to know, whether the measured lag depends on N, the initial cell number (in this model, the cells grow independently, so they do not 'know' if they are part of a small or a large population). Another, perhaps even more important, question raised here is: what can be inferred, from the measured population lag, to the individual lag?

Following the definitions above, the size of the bacterial population, as a function of time, can be approximated by a sum of the exponent of bi-phasic linear functions. These bi-phasic functions estimate the growth curves of the subpopulations generated by the individual cells of the initial culture, consisting of N cells:

$$x_N(t) = \sum_{i=1}^{N} e^{\mu \cdot \max(t-\tau_i, 0)} \tag{4.3}$$

Take the logarithm of the population size: $y_N(t)=\ln x_N(t)$. For sufficiently large time values, the formula can be rearranged to a linear form :

$$y_N(t) = \ln(N) + \mu \left(t - \frac{-\ln \frac{\sum_{i=1}^{N} e^{-\mu \cdot \tau_i}}{N}}{\mu} \right) \qquad (t > \max \tau_i) \tag{4.4}$$

Therefore, for the population lag produced by N initial cells, the following formula can be given:

$$\lambda(N) = -\frac{1}{\mu} \ln \frac{\sum_{i=1}^{N} e^{-\mu \tau_i}}{N} \tag{4.5}$$

Introducing the $\alpha_i=\exp(-\mu\tau_i)$ (i=1,2,...N) notation, the α_i variables are also identically distributed, so their arithmetical mean obtained in the right hand side of the above formula converges to a limit value as N increases. This limit is the common expected value of the α_i variables; let it be denoted by α.

It follows from our formula (4.5) that

$$L_N \xrightarrow{N \to \infty} \frac{-\ln(\alpha)}{\mu} \tag{4.6}$$

Remember that we used $\alpha = \exp(-\mu\lambda)$ to quantify the 'suitability' of the population to the new environment. Similarly,

$$\alpha(N) = e^{-\mu \cdot L_N} \tag{4.7}$$

will quantify the suitability of the initial population consisting of N cells. A rearrangement of our formula (4.5) is

$$e^{-\mu L_N} = \frac{\sum_{i=1}^{N} e^{-\mu \tau_i}}{N} \tag{4.8}$$

i.e.

$$\alpha(N) = \frac{\sum_{i=1}^{N} \alpha_i}{N} \tag{4.9}$$

This says that the suitability of the population is the arithmetical average of the suitability of the individual cells, for any initial cell number (Fig.4.3b).

This result can help to estimate the rate of convergence in (4.6) via the variance of the arithmetical mean:

$$\mathrm{Var}[\alpha(N)] = \frac{\mathrm{Var}(\alpha_i)}{N} \tag{4.10}$$

The population lag can be written as

$$L_N = \frac{-\ln \alpha(N)}{\mu} \tag{4.11}$$

(see formula 4.9). For explicit formulae consider the following line of thoughts:
As it is well-known, the individual generation time of a Poisson birth process follows the exponential distribution. Suppose that the individual lag times also follow the exponential distribution, with the mean value τ. In this case, as can be calculated, the probability distribution function of α_i is :

$$F_\alpha(s) = \Pr(\alpha_i < s) = \begin{cases} 0 & \text{if } s \le 0 \\ s^r & \text{if } 0 < s \le 1 \\ 1 & \text{if } 1 \le s \end{cases} \tag{4.12}$$

where $r=(\mu\tau)^{-1}$.

By means of the $F_\alpha(s)$ distribution function, the expected value and the variance of the α_i variables can be calculated as

$$\alpha = \mathrm{E}(\alpha_i) = r\int_0^1 s^r \,\mathrm{ds} = \frac{1}{1+\mu\cdot\tau} \tag{4.13}$$

$$\mathrm{Var}(\alpha_i) = r\int_0^1 s^{r+1} \,\mathrm{ds} - \left(\frac{1}{1+\mu\cdot\tau}\right)^2 = \frac{1}{1+2\cdot\mu\cdot\tau} - \left(\frac{1}{1+\mu\cdot\tau}\right)^2 \tag{4.14}$$

After substitution:

$$L_N \xrightarrow{N\to\infty} \frac{\ln(1+\mu\cdot\tau)}{\mu} \tag{4.15}$$

The above results solved our basic problem. The population lag converges to the above value, while the formulae (4.10), (4.11) and (4.14) define the rate of that convergence (Fig.4.3a).

4.3. Comparison with a deterministic model

Now, we show a deterministic compartment-model-equivalent of the above problem.
Let the cell population be divided into two compartments: cells which are still in the lag phase, $x_{\mathrm{Lag}}(t)$, and cells which are in the exponential phase, $x_{\mathrm{Exp}}(t)$. Assume that the cells transform from the lag to the exponential phase at a constant rate, κ. (Remember that the duration of the lag is smaller than the time to the first division, which includes also the first generation time).

Assuming that all the initial cells (x_0=N) are in the lag phase at the time zero, our compartment model can be described by a system of two linear differential equations with two initial values:

$$\frac{dx_{\mathrm{Lag}}}{dt} = -\kappa\cdot x_{\mathrm{Lag}} \tag{4.16}$$

$$\frac{dx_{\mathrm{Exp}}}{dt} = \mu\cdot x_{\mathrm{Exp}} + \kappa\cdot x_{\mathrm{Lag}} \tag{4.17}$$

$$x_{\mathrm{Lag}}(0) = x_0 \tag{4.18}$$

$$x_{\mathrm{Exp}}(0) = 0 \tag{4.19}$$

The solution of this initial value problem is

$$x_{\mathrm{Lag}}(t) = x_0 e^{-\kappa\cdot t} \tag{4.20}$$

$$x_{\mathrm{Exp}}(t) = x_0 \frac{\kappa}{\mu+\kappa}(e^{\mu\cdot t} - e^{-\kappa\cdot t}) \tag{4.21}$$

Therefore, the total cell population can be described by

$$x(t) = x_{\mathrm{Lag}}(t) + x_{\mathrm{Exp}}(t) = x_0\left(\frac{\kappa}{\mu+\kappa}e^{\mu\cdot t} + \frac{\mu}{\mu+\kappa}e^{-\kappa\cdot t}\right) \tag{4.22}$$

This result can be obtained also by the following way: Because the daughter cells do not need the adjustment (lag-) time anymore, the probability that, at time *t*, a cell is still in the lag phase, is equal to the proportion of these cells in the present total population *i.e.* to $x_0e^{-\kappa t} / x(t)$.

Incorporate the lag in the Poisson model. The probability that a cell got over the lag phase is $1-x_0e^{-\nu t}/x(t)$, therefore

$$\frac{\Delta x}{x} = \left(1 - \frac{x_0 e^{-\nu t}}{x}\right)\mu\Delta t \qquad (4.23)$$

Transform this equation back to its infinitesimal form:

$$\frac{dx}{dt} = \mu\,(x - x_0 e^{-\nu t}) \qquad (4.24)$$

As can be checked, the solution of this differential equation is really (4.22).

Let $y(t)$ be the natural logarithm of $x(t)$ and let $y_0 = \ln x_0$. Then for the solution (4.22) we can write:

$$y(t) = \ln x(t) = y_0 + \mu\left(t - \frac{-\ln\left(\frac{\kappa}{\mu+\kappa} + \frac{\mu}{\mu+\kappa}e^{-(\mu+\kappa)t}\right)}{\mu}\right) \qquad (4.25)$$

Therefore $y_\infty(t) \to y_0 + \mu(t-\lambda)$ as $t \to \infty$ (4.26)

where

$$\lambda = \frac{\ln\left(1 + \frac{\mu}{\kappa}\right)}{\mu} \qquad (4.27)$$

If the individual lag times follow the exponential distribution with $\tau = 1/\kappa$ as expected value, than we get the formula (4.15) back for the limit value of the population lag.

4.4. Comparing the developed stochastic and deterministic models

When studying growth by both deterministic and stochastic models, the expected growth predicted by the stochastic process is often the same as the deterministic prediction. For certain growth parameters, however, sometimes surprising results can be obtained. As has been mentioned, for example the doubling time for the deterministic exponential growth was $\ln 2/\mu$, which is about 30% less than the mean generation time, $1/\mu$, of the analogous Poisson growth. As Rubinow (1984) writes, the fast growing cells contribute to the total population

more than the slow cells, this is why this unexpected result. From our model, an analogous result can be inferred, since we have obtained:

$$\frac{\lambda}{\tau} = \frac{\ln(1+\mu\tau)}{\mu\tau} \qquad (4.28)$$

As can be seen, the ratio between the population lag and the mean individual lag depends only on the factor $h' = \tau/(1/\mu)$, which is the ratio between the mean lag and generation time of the individual cells. Because the right hand side of the above equation cannot be bigger than 1, the population lag is always less than the average individual lag. Besides, from the above formula, λ/τ converges to zero as μ is increasing. Therefore, at fixed individual lag-distribution, the population lag can be any small, as long as the subsequent exponential growth can be chosen sufficiently high.

The analysed process can be easily simulated on computer, by means of a sequence of independent random numbers, R_i (i=1,2..N), uniformly distributed between 0 and 1. The values τ_i=-τ ln(R_i) will be then independent, exponentially distributed numbers, with the expected value τ, and α_i=$R_i^{\mu\tau}$ are the appropriate random numbers for the simulation of the physiological state parameters of the cells.

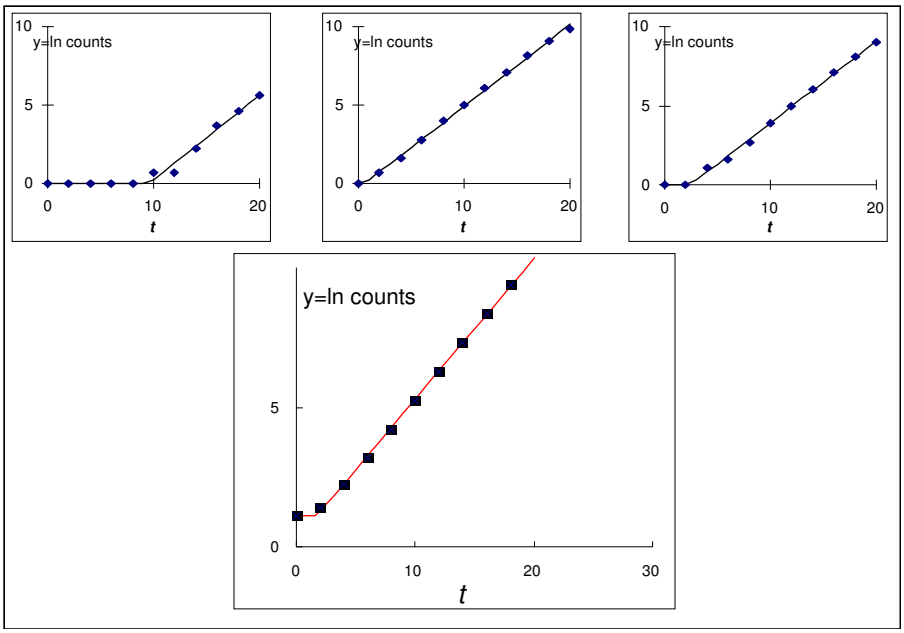


Fig 4.1. Simulated growth curves of the subsequent subpopulations of three single cells, generated independently of each other (small plots), and the joint growth curve of the unified population of the three subpopulations (large plot). The (population) lag of the joint growth curve, with y0=ln N = ln 3 initial cell number, is about 2, while the average of the individual lag times of the three cells is about 4.

The figures 4.1 -4.2 were prepared by means of the above random number generating procedure, with $\tau = 3$ and $\mu=0.5$. Fig.4.1 shows the independent and joint growth of the subpopulations generated by $N=3$ cells. Fig.4.2 demonstrates the inoculum-dependence of the expected population lag, described by the L_N sequence. Here, the subsequent growth of 64 cells were simulated and the generated subpopulations were grouped as 1·64 ($N=1$) , 4·16 ($N=4$) … 32·2 ($N=32$) , and the respective population lag times were fitted for these N values. As can be seen, from lower inoculum, higher population lag can be expected, with higher variance.

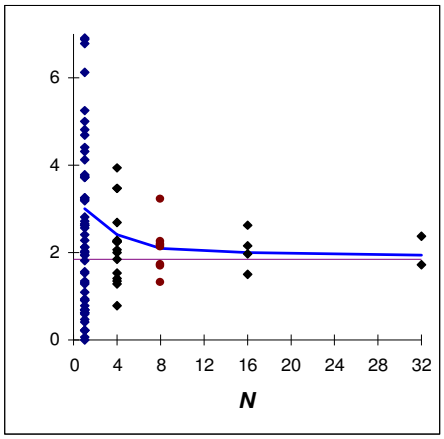


Fig 4.2. a/ The population lag, L_N, and its variance, as a function of the initial cell number, demonstrated by simulated data.

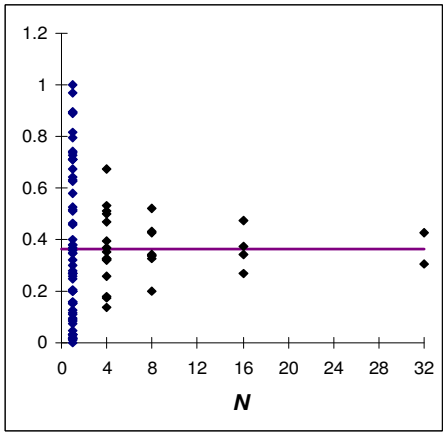


Fig 4.2.b/ The mean physiological state parameter, $\alpha(N)$, and its variance, demonstrated by the same data.

It is important to see that the decrease of the population lag with higher inoculum is caused by purely the randomness of the process and not because the cells have information on the size of the surrounding population (which would be the so called quorum-sensing). The population lag converges to a limit value given by formula (4.15). However, the rate of this convergence depends on the specific growth rate (although the generation time is independent of the individual lag time!) and on the distribution of the individual lag times of the cells.

The results have numerical effects on those studies which are based on the detection time of bacterial populations as defined by the interval elapsed from the inoculation to the time when the population reaches a certain level. Assuming pure exponential growth until the time of detection, preceded by the lag period only, the distribution of the detection times reflect the distribution of the population lag times which is more and more scattered for lower inoculum levels. The detection time, for lower inoculum, has higher expected value and variance, which has obvious consequences, for example, when estimating the time until certain symptoms appear caused by growing bacteria. We return to this question in Chapter 6.

Chapter 5

Generalizations and analogues with survival models

5.1. Bacterial survival curves

We have seen that for food safety modelling, it is vital to study and model the adaptation time when bacteria are inoculated in a new environment. The situation is similar with survival curves where a so-called "shoulder" period precedes the exponential decay. Mirror images of growth functions are also used to model survival curves (Xiong *et al*., 1999b) but a parallel study on growth and survival models outlined in this chapter shows that the symmetry is not that straightforward as it looks at the first sight.

For both growth and survival models, we call the "ln counts *v*. time" curve the *population curve*. When the cells' environment suddenly becomes inactivating (such as thermal treatment of food), the "time to death" for a single cell will be called the *individual survival time*.

The main questions to be answered in this chapter are: What is the relationship between the distribution of individual survival times and the survival curve of the population; furthermore between the distribution of individual lag times and the growth curve of the population.

We are particularly interested in developing formulae between the parameters of the distribution functions of individual survival/lag times and the parameters of the survival/growth curves of the population. For this purpose, we introduce the following definition. If the growth/survival curve (meant as the "*ln counts v. time*" curve, $y(t)$) asymptotically converges to a linear function

$$y_a(t) = \ln N + \mu \cdot (t - \lambda), \qquad (5.1)$$

where N is the initial counts, μ and λ are constant and do not depend on N, (*i.e.* $|y(t) - y_a(t)| \to 0$ as $N \to \infty$, $t \to \infty$), then μ and λ are called the *limit slope* and *limit shoulder / lag* parameter of the survival/growth curve, respectively (see Fig.5.1).

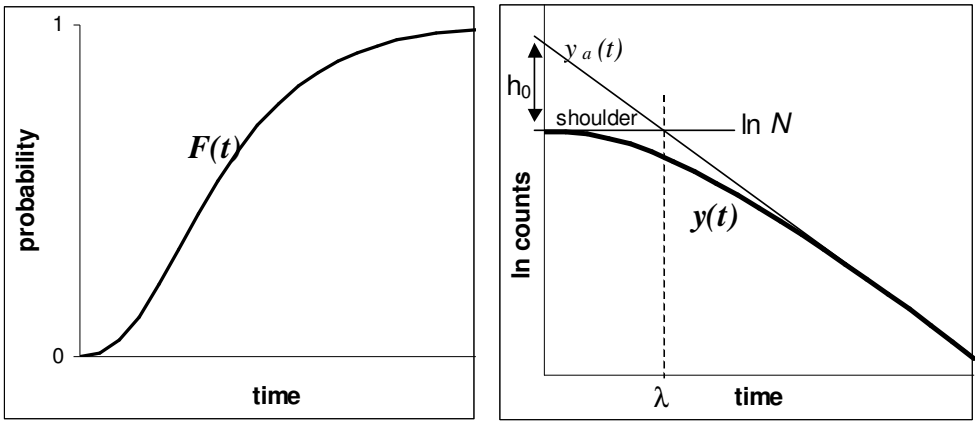


Fig.5.1. There is a one-to-one mapping between $F(t)$, the cumulative distribution function of the survival times of individual cells and $y(t)$, the survival curve of the population. If $y(t)$ asymptotically converges to a linear function, $y_a(t)$, then a shoulder parameter, λ , is interpreted for the population curve. The parameter h_0 can be interpreted as in the growth scenario.

Denote the survival time of the i-th cell, during an inactivation process, by τ_i (i=1…N). As before, for the sake of compatibility with the respective publication (Baranyi and Pin, 2001), we equate τ_i with the "geometrical" definition of lag, defined by a bi-phasic function fitted to the logarithm of the generated subpopulation.

Suppose that the population is homogeneous, that is the τ_i random variables are independent and identically distributed, with τ mean value. Denote their common probability density function (pdf) and cumulative distribution function (cdf) by $f(t)$ and $F(t)$, respectively.

5.2 Inactivation after a shoulder period

For inactivation, the answer to our basic question is easy and well known. Because

$$x(t) = \sum \mathrm{P}(\tau_i >= t) = N \cdot (1 - F(t)) \qquad (5.2)$$

so the population survival curve is (see Fig.5.1):

$$y(t) = \ln N + \ln(1 - F(t)) \qquad (5.3)$$

Consider some specific examples for $F(t)$.

Case A_1.

The most cited situation is when τ_i are exponentially distributed with the mean value τ (pure Poissonian death process). In this case, $F(t)=1-e^{-t/\tau}$, so the survival curve is

$$y(t) = \ln N - t/\tau \qquad (5.4)$$

Therefore, the survival curve is linear; the reciprocal of its slope is the mean value of the exponentially distributed survival times of the individual cells, and there is no positive shoulder parameter in this situation: $\lambda=0$.

Case A_2.

A possible generalisation of *Case A_1* is when τ_i ($i=1\ldots N$) follow the gamma distribution, with the parameters $p\geq 1$ and $\nu>0$, where the expected value of the survival time for a cell is $\tau=p/\nu$. This can be interpreted in the following way: the cell needs p damaging hits (see the multi-hit theory in Casolari, 1988), and the times θ_j ($j=1...p$) between the hits are independent, exponentially distributed variables, with a common mean value $\theta=1/\nu$. Then the survival time of the i-th cell is

$$\tau_i = \sum_{j=1}^{p} \theta_j \qquad (5.5)$$

therefore τ_i ($i=1\ldots N$) are gamma distributed. In this case,

$$y(t) = \ln N + \ln (1 - \Gamma_{p,\nu}(t)) \qquad (5.6)$$

where $\Gamma_{p,\nu}(t)$ is the cdf of the gamma distribution, with parameters p and ν, and mean value p/ν.

We show below that, though the derivative of $y(t)$ converges to a constant (namely ν), a limit shoulder still does not exist in the $p>1$ case. This, however, has only theoretical significance. With practical data, the value of $y(t)$ cannot be arbitrarily small and therefore, though the shoulder λ does not exist in mathematical limit sense (apart from the $p=1$, exponential distribution case), in practice it would be simply determined by ν and the smallest measurable value of $y(t)$.

► Let the individual survival times be gamma-distributed, with $p \geq 1$, $\nu > 0$ parameters:

$$\Pr(\tau < t) = \Gamma_{p,\nu}(t) = \frac{\nu^p}{\Gamma_p} \int_0^t s^{p-1} e^{-\nu s} ds \qquad (5.7)$$

where the constant Γ_p is defined as

$$\Gamma_p = \int_0^\infty t^{p-1} e^{-t} dt \qquad (5.8)$$

Remember that for an integer p, $\Gamma_p = (p-1)!$ and for $p=1$ the exponential distribution is obtained.

The mean value and the variance of τ are

$$\mathrm{E}(\tau) = \frac{p}{\nu} \qquad \mathrm{Var}(\tau) = \frac{p}{\nu^2} \tag{5.9}$$

The survival curve of the population is

$$y(t) = \ln N + \ln\left(1 - \Gamma_{p,\nu}(t)\right) \tag{5.10}$$

For its derivative, one can calculate that

$$\frac{dy}{dt} = \frac{-f(t)}{1-F(t)} = \frac{t^{p-1}e^{-\nu t}}{\int_t^\infty s^{p-1}e^{-\nu s}ds} \tag{5.11}$$

Applying the L'Hospital rule, as $t\to\infty$, the ratio of the counter and denominator converges to the same value as the ratio of their derivative:

$$\lim_{t\to\infty}\frac{dy}{dt} = \lim_{t\to\infty}\frac{(p-1)t^{p-2}e^{-\nu t} - \nu t^{p-1}e^{-\nu t}}{t^{p-1}e^{-\nu t}} = \lim_{t\to\infty}\left(\frac{p-1}{\nu} - \nu\right) = -\nu \tag{5.12}$$

As was expected, the derivative of the survival curve converges to $-\nu$. However, this does not ensure that asymptotic linear function exists. Indeed, suppose that, for a constant b:

$$\ln(1\text{-}F(t)) \;\to\; -\nu t + b \qquad (t\to\infty) \tag{5.13}$$

That would be equivalent to

$$\frac{\int_t^\infty \frac{\nu^p}{\Gamma_p} s^{p-1}e^{-\nu s}ds}{e^{-\nu t}} \to e^b \qquad (t\to\infty) \tag{5.14}$$

Again, apply the L'Hospital rule:

$$\frac{-\frac{\nu^p}{\Gamma_p}t^{p-1}e^{-\nu t}}{-\nu e^{-\nu t}} = \frac{\nu^{p-1}}{\Gamma_p}t^{p-1} \to e^b \qquad (t\to\infty) \tag{5.15}$$

This expression, however, is convergent only for p=1, when b=0 and the survival curve describes a pure exponential death, with $y(t)=y_a(t)=-\nu t$ (straight line through the origin). If $p>1$ then the expression cannot be convergent, so though the slope of the survival curve is convergent, the shoulder is not. ◄

Case A_3

Another possible generalisation of *Case A_1* is the so-called multi-target model (Hermann and Horst, 1970). According to that, a cell has p targets that are being hit synchronously (not consecutively as in the previous case). The times needed to destroy the targets, θ_j (j=1...p), are independent, exponentially distributed variables with the common mean value of $1/\nu$ and the cell is live until all targets are inactivated. Therefore, the survival time of the i-th cell is

$$\tau_i = \max_{1\le j\le p} \theta_j \qquad (5.16)$$

and $F(t)=(1-e^{-\nu t})^p$, where $p \ge 1$. The population survival curve is now

$$y(t) = \ln N + \ln (1 - (1-e^{-\nu t})^p) \qquad (5.17)$$

which will converge to the linear asymptote $y_a(t) = \ln N - \nu \cdot (t - \ln p / \nu)$, as shown below:

► The population growth curve in this case is

$$y(t) = \ln N + \ln (1 - (1-e^{-\nu t})^p) \qquad (5.18)$$

Standard analysis can show that this is a monotone decreasing function of t, starting from $y(0)=\ln N$. Its derivative also decreases monotonically, to the asymptotic final slope, $-\nu$:

$$\ln\left(1-(1-e^{-\nu t})^p\right) = \ln p - \nu t + \ln \frac{1-(1-e^{-\nu t})^p}{p e^{-\nu t}} \qquad (5.19)$$

According to the L'Hospital rule, the last term converges to ln(1)=0:

$$\lim_{t\to\infty} \frac{1-(1-e^{-\nu t})^p}{p e^{-\nu t}} = \lim_{t\to\infty} \frac{-\nu p (1-e^{-\nu t})^{p-1} e^{-\nu t}}{-\nu p e^{-\nu t}} = 1 \qquad (5.20)$$

Therefore $y(t)$ converges monotonically to the linear function

$$y_a(t) = \ln N + \ln p - \nu t \qquad (5.21)$$

In this case, the limit shoulder parameter can be calculated as

$\lambda = \ln p / \nu$ ◄

5.3. Growth following lag

The situation is more complex when studying growth after a lag period. Let τ_i be the lag time of the *i*-th cell, where τ_i (*i*=1…*N*) are independent, identically distributed random variables, with common pdf and cdf (*f(t)* and *F(t),* respectively). Assuming that, after the lag period, the subpopulation generated by the cell grows at a constant specific growth rate, μ, the expected population *x(t)* consists of two subpopulations: those that are over the lag period (or already daughter cells) and those that are the original cells, still in the lag period. Therefore

$$x(t) = \sum_{i=1}^{N} x_i(t) = N\left(\int_0^t e^{\mu\cdot(t-s)} f(s)ds + \int_t^\infty f(s)ds\right) \qquad (5.22)$$

that is:

$$y(t) = \ln N + \ln\left(\int_0^t e^{\mu\cdot(t-s)} f(s)ds + 1 - F(t)\right) \qquad (5.23)$$

Just like in the case of inactivation curves, the above formula provides a one-to-one mapping between the distribution of the individual lag times and the population growth curve. The growth curve converges to a straight line as *t* approaches infinity, so (not surprisingly) it is the curvature from the lag to the exponential phase that characterizes the *F(t)* distribution. This is similar to that situation when, in inactivation models, a Laplace-transformation maps the distribution of resistance of individual cells onto a survival curve (Körmendy *et al.*, 1998).

If one could observe the lag times of individual cells and one or two subsequent divisions in the exponential phase (to estimate μ), then, by the equation (5.23), the growth curve of the whole population could be predicted. It should be also possible to estimate the distribution of the lag times of individual cells from traditional viable count growth curves, if there is sufficient amount of accurate measurements before the exponential phase. However, equation (5.23) is a one-to-one mapping only in theory. In practice, it is not feasible to identify the distribution of the individual lag times from traditional viable count growth curves. We give a numerical example for this below.

Consider the growth curve on Fig.5.2/B, with one of the observed growth curves in the paper of McClure *et al* (1993). Test the exponential distribution for *f(t)*:

$$f(t) = \nu\, e^{-\nu t}$$

where ν is the reciprocal of the average lag time of the individual cells: $\nu = 1/\tau$. Our equation (5.23) reads as follows:

$$y(t) = y_0 + \mu \cdot t + \ln\left(\frac{\nu}{\mu+\nu} + \frac{\mu}{\mu+\nu} e^{-(\mu+\nu)t}\right) \qquad (5.24)$$

After converting the raw data into natural logarithm of the cell concentrations, the three parameters, y_0 , μ, ν, can be fitted by least squares method. The fitted curve, after transforming it back to $\log_{10}$-scale, can be seen in Fig.5.2/B. Though the fit is obviously good, one must note that this is the best fit if *f(t)* is chosen from the family of exponential distributions for the individual lag times. Indistinguishably similar fits can be obtained if *f(t)* describes gamma (of which exponential is only a special case) or lognormal distribution (Fig.5.3/A).

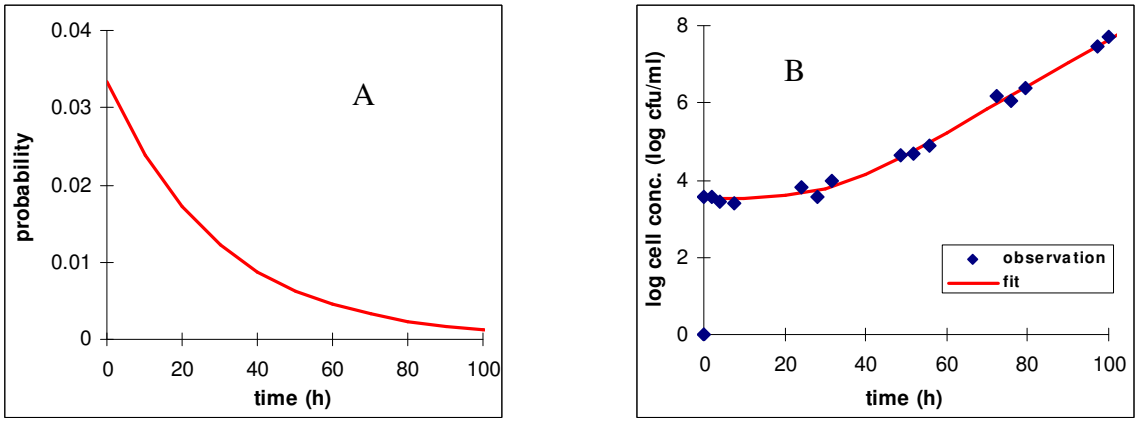


Fig.5.2. One-to-one mapping between (A) the probability density function of the exponential distribution assumed for the individual lag times and (B) the growth curve characterizing the transition of the population from the lag to the exponential phase.

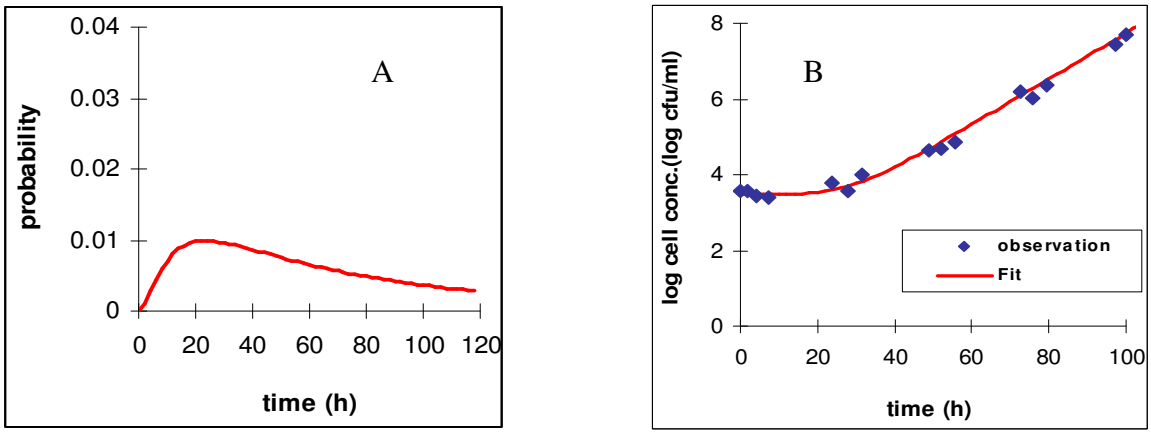


Fig.5.3. As in Fig.5.2, but log-normal assumption for the distribution of the indiviual lag times. The two distributions (Figs.5.2/A and 5.3A) are markedly different, still the respective two population growth curves (Figs.5.2/B and 5.3/B) are indistinguishable from each other.

Therefore, though equation (5.23) provides a theoretical background to make equivalence between growth curves and lag distributions, the standard curve fitting procedures, numerically, are not robust enough on commonly measured growth data. The traditional viable count curves are not suitable to identify what type of distribution is followed by the individual cells. New approaches are necessary in the measurement techniques, too (image analysis, flow cytometry, *etc*), in order to study individual cell kinetics.

By rearranging (5.23), we obtain:

$$y(t) = \ln N + \mu \cdot t + \ln\left(\int_0^t e^{-\mu \cdot s} f(s) ds + e^{-\mu \cdot t} \left(1 - F(t)\right) \right) \qquad (5.25)$$

As can be seen, for $t \to \infty$ the population curve $y(t)$ asymptotically converges to the linear function

$$y_a(t) = \ln N + \mu \cdot \left(t + (\ln \alpha) / \mu \right) \qquad (5.26)$$

where

$$\alpha = \int_0^\infty e^{-\mu \cdot s} f(s) ds \qquad (5.27)$$

As the formula shows, the parameter α is the (common) mean value of the $\alpha_i = e^{-\mu \cdot \tau_i}$ variables, which is the mean physiological state.

Note that this is a new proof for the result of the previous chapter that, for sufficiently large t, the $y(t)$ population curve converges to the linear function

$$y_a(t) = \ln N + \mu \cdot t + \ln \frac{\sum_{i=1}^{N} e^{-\mu \tau_i}}{N} \qquad (5.28)$$

Because the argument of the "ln" function is the arithmetical mean of the $\alpha_i = e^{-\mu \cdot \tau_i}$ (independent, identically distributed) variables, it converges to their common mean value, α. Therefore the limit lag parameter of the population is related to the individual lag times indirectly, through the mean of the individual physiological states. The wanted formula is

$$\lambda = -\frac{\ln \alpha}{\mu} \qquad (5.29)$$

so we obtained the lag introduced in Chapters 2-3 as a special case.

This relationship was deduced without making any use of some particular form of the individual lag distribution, exploiting only that the mean value of the α_i variables exists. This was not the case with survival curves because there the linear asymptote was not guaranteed.

Because the growth situation is much more complex than the parallel equations for survival, it is worth investigating possibilities for simplification.

Case B_1

Suppose that the individual lag times are exponentially distributed with the mean τ. Then, with the notation $\kappa=1/\tau$,

$$\int_0^t e^{\mu\cdot(t-s)} f(s)ds + 1 - F(t) = \int_0^t e^{\mu\cdot(t-s)} \nu e^{-\kappa\cdot s} ds + e^{-\kappa\cdot t} = \frac{\nu e^{\mu\cdot t} + \mu e^{-\kappa\cdot t}}{\mu + \kappa} \qquad (5.30)$$

Therefore

$$y(t) = \ln N + \ln \frac{\nu e^{\mu\cdot t} + \mu e^{-\kappa\cdot t}}{\mu + \kappa} \qquad (5.31)$$

This confirms another formula of the previous chapter when we derived this by using two compartments (lag and exponential phase cells) to model the population growth curve, assuming that the specific transition rates between the compartments were constant. However, this result holds only if the individual lag times are exponentially distributed. After calculating α from τ and μ, the wanted relation between the limit population lag and the mean individual lag time is:

$$\lambda = \frac{\ln(1+\mu\tau)}{\mu} \qquad (5.32)$$

Case B2

Suppose that the τ_i (i=1…N) individual lag times follow the gamma distribution, with the parameters p and κ; so the mean individual lag is $\tau=p/\kappa$. The biological interpretation of this scenario is analogous to *Case A_2* : the cells have to carry out p consecutive tasks, and the

times θ_j (j=1...p) required by the individual tasks are independent, exponentially distributed variables, with the mean time $\theta = \tau/p = 1/\kappa$. Therefore, the τ_i individual lag time is the sum of the times required by the p subtasks.

In this case, the common mean of the α_i=exp(-$\mu\tau_i$) variables is α=(1+μ/κ)$^{-p}$ as shown below:

► Let $\kappa/\mu = r$, where μ is the subsequent constant specific rate in the exponential phase. Then, as can be checked, the respective pdf of the individual physiological states, $\alpha_i = \exp(-\mu\tau_i)$ is:

$$f(t) = \frac{r^p}{\Gamma_p} t^{r-1}(-\ln t)^{p-1} \qquad (0 < t \le 1) \qquad (5.33)$$

The mean and variance of the above pdf is calculated by means of two substitutions: first $s = -\ln t$ is applied to both; then the $u = (r+1)s$ transformation is applied to the expected value, and $u = (r+2)s$ to the variance:

$$\alpha = \mathrm{E}[\alpha_i] = \int_0^1 t\, f(t)dt = \int_0^1 t \frac{r^p}{\Gamma_p} t^{r-1}(-\ln t)^{p-1} dt =$$

$$= \frac{r^p}{\Gamma_p}\int_0^\infty e^{-sr} s^{p-1} e^{-s} ds = \frac{r^p}{\Gamma_p}\int_0^\infty e^{-u}\left(\frac{u}{r+1}\right)^{p-1}\frac{1}{r+1}du = \left(\frac{r}{r+1}\right)^p \qquad (5.34)$$

Similarly,

$$\mathrm{Var}[\alpha_i] = \int_0^1 t^2 f(t)dt - \mathrm{E}^2(\alpha_i) = \int_0^1 t^2 \frac{r^p}{\Gamma_p} t^{r-1}(-\ln t)^{p-1} dt - \left(\frac{r}{r+1}\right)^{2p} =$$

$$= \frac{r^p}{\Gamma_p}\int_0^\infty e^{-s(r+1)} s^{p-1} e^{-s} ds - \left(\frac{r}{r+1}\right)^{2p} =$$

$$= \frac{r^p}{\Gamma_p}\int_0^\infty e^{-u}\left(\frac{u}{r+2}\right)^{p-1}\frac{1}{r+2}du - \left(\frac{r}{r+1}\right)^{2p} = \left(\frac{r}{r+2}\right)^p - \left(\frac{r}{r+1}\right)^{2p} \qquad (5.35)$$

We obtain the relation for the limit population lag and the mean individual lag time:

$$\lambda = \frac{\ln\left(1+\frac{\mu}{p}\tau\right)}{\mu/p} \qquad (5.36)$$

This equation is, indeed, a generalization of the formula used so far with p=1, when the gamma distribution is reduced to exponential distribution (the cell has one task to be carried out during the lag). ◄

Case B₃

Suppose that the τ_i (i=1…N) individual lag times are obtained analogously to *Case A₃*: the cells have to carry out p subtasks simultaneously during the lag, and the times θ_j (j=1...p) required by the individual subtasks are independent, exponentially distributed variables, with the mean time $\theta = 1/\kappa$. Then τ_i is the maximum of the times required by the p subtasks.

As we will show, now the expected value of the population mean of α_i will be

$$\alpha = p\frac{\Gamma_p \Gamma_{\mu\tau+1}}{\Gamma_{p+\mu\tau+1}} \qquad (5.37)$$

where

$$\Gamma_p = \int_0^\infty t^{p-1}e^{-t}dt \qquad (5.39)$$

► The cdf of the individual lag times is $F(t)=(1-e^{-\kappa t})^p$. With the notation $\kappa/\mu = r$, the cdf of $\alpha_i = \exp(-\mu\tau_i)$ is:

$$F(s) = 1-\left(1-s^r\right)^p \qquad (\ s\in(0,1]\ ) \qquad (5.39)$$

Its expected value is therefore

$$\alpha = \mathrm{E}[\alpha_i] = pr\int_0^1 s(1-s^r)^{p-1}s^{r-1}ds \qquad (5.40)$$

This definite integral can be calculated in an explicit form, after $u=s^r$ substitution and considering the relationship between the beta and gamma functions:

$$\int_0^1 u^{1/r}(1-u)^{p-1}du = \mathrm{B}_{p,\frac{1}{r}+1} = \frac{\Gamma_p \Gamma_{\frac{1}{r}+1}}{\Gamma_{p+\frac{1}{r}+1}} \tag{5.41}$$

hence

$$\alpha = p\frac{\Gamma_p \Gamma_{\frac{1}{r}+1}}{\Gamma_{p+\frac{1}{r}+1}} \tag{5.42}$$

Similarly, the variance can be calculated from the connection between the gamma and beta functions:

$$Var[\alpha_i] = p\left(\frac{\Gamma_p \Gamma_{\frac{2}{r}+1}}{\Gamma_{p+\frac{2}{r}+1}} - \left(\frac{\Gamma_p \Gamma_{\frac{1}{r}+1}}{\Gamma_{p+\frac{1}{r}+1}}\right)^2\right) \tag{5.43}$$

We obtain the relation for the population lag and the mean individual lag time:

$$\lambda = \frac{1}{\mu}\ln\left(p\frac{\Gamma_{p+\mu\tau+1}}{\Gamma_p \Gamma_{\mu\tau+1}}\right) \tag{5.44}$$

Again this is a generalisation of earlier results, when the cell had only one task to be carried out during the lag (p=1). ◄

5.4. An important by-product

In Chapters 2-3, we introduced the concept of 'physiological state' or "suitability parameter" quantified by α_0=exp(-$\mu\lambda$) , where λ is the traditionally defined lag time of a growing bacterial culture. In Chapter 4, we saw a good use of it when studying the kinetics of dividing individual cells. Here we extended the approach to survival kinetics and pointed out the analogy as well as the difference.

We defined the shoulder and lag parameters assuming that the initial population, N , is big. This is frequently a pre-condition for the applicability of population parameters. A commonly used formula, for example, that the expected time required by an exponentially growing population to double is $t_d = \ln 2 / \mu$ (Rubinow, 1984). However, this is obviously not true, for example, for a single cell, if the division time is exponentially distributed; then the

expected time to division is $1/\mu$. In fact, we show that the expected time for N cells to double is

$$t_d = \frac{1}{\mu}\sum_{i=0}^{N-1}\frac{1}{N+1} \qquad (5.45)$$

converging to $\ln 2 /\mu$ as $N\to\infty$. Therefore, the expected doubling time is a function of the number of cells and it decreases from $1/\mu$ (for N=1) to $\ln 2 /\mu$ (for $N\to\infty$).
The proof is partly (up to the formula 5.50) is the work of my co-author Carmen Pin (see Baranyi and Pin, 2001).

► Suppose that all the N cells are in the exponential growth phase and the individual division times, d_i (i=1...N) are independent, exponentially distributed variables, with the common parameter μ (Poisson birth process). We wish to estimate the time required to obtain $2N$ cells; *i.e.* the sum of the first N time intervals, each of which is defined by two consecutive divisions in the population:

$$t_d(N) = \sum_{i=1}^{N} t_i \qquad (5.46)$$

For the time to the first division, t_1, we obtain

$$t_1=\min\{d_1 \ldots d_N) \qquad (5.47)$$

It can be shown that t_1 follows the exponential distribution with the parameter $N\mu$:

$$F(t) = \Pr(t_1 < t) = 1-\Pr(d_1 > t\ ;\ \ldots\ ;\ d_N > t) = 1-(e^{-\mu t})^N = 1-e^{-N\mu t} \qquad (5.48)$$

therefore

$$E(t_1)=(N\mu)^{-1} \qquad (5.49)$$

The time between the first and second division (t_2) is the minimum division time of the resulted N+1 cells. Because of the so-called "memory-less" feature of the exponential distribution, we can apply the previous result for $N+1$ cells, so the cdf of t_2 is

$F(t) = 1-e^{-(N+1)\mu t}$ and $E(t_2)=((N+1)\mu)^{-1}$.

By induction, it follows that the expected value of $t_d(N)$ is

$$E[t_d(N)] = E(t_1 + t_2 + .. + t_i + .. + t_N) = \frac{1}{\mu}\sum_{i=0}^{N-1}\frac{1}{N+i} \qquad (5.50)$$

As a special case, if N=1, then the doubling time is $1/\mu$ (as well-known about exponentially distributed times).

Let $H_N = \sum_{i=1}^{N} \frac{1}{i}$ so $E[t_d(N)] = \frac{H_{2N-1} - H_N}{\mu}$

We show that the right hand side converges to ln2/μ as $N \to \infty$.

Using a simple stepwise rectangular approximation for the integral of the $1/x$ function, one can obtain:

$$H_{2N} - H_N < \int_N^{2N} \frac{1}{x} dx < H_{2N-1} - H_{N-1} \tag{5.51}$$

That is, ln2 separates the two series:

$$H_{2N} - H_N < \left(\ln 2N - \ln N\right) = \ln 2 \quad < H_{2N-1} - H_{N-1} \tag{5.52}$$

We just need to prove that the difference between the two series converges to zero:

$$(H_{2N-1} - H_{N-1}) - (H_{2N} - H_N) = (H_N - H_{N-1}) - (H_{2N} - H_{2N-1}) = \frac{1}{N} - \frac{1}{2N}$$

The right hand side converges to zero, as $N \to \infty$, so the proof is complete. ◄

Chapter 6

Estimating the distribution parameters of individual lag times in a cell-population

6.1 Can we measure the distribution of single cell lag times?

Automated measures are commonly used to estimate bacterial growth parameters. Unfortunately, little information is obtained on the lag phase (λ) because the change in the physical properties of a culture (turbidity, conductance, *etc*.) is detectable only at high cell concentrations. This problem is serious, for example, in food microbiology, where predicting the end of the lag phase is of great importance.

The parameter α=exp(-$\mu\lambda$) was introduced in Chapters 2-3 to quantify the *physiological state* of the initial population. As has been shown in Chapter 4, the lag parameter of a bacterial growth curve, also termed population lag, is not a simple arithmetical average of the lag times of the individual cells, τ_i (i=1...N , where N denotes the initial cell number). The physiological state of the inoculum, however, is equal to the arithmetical mean of the physiological states of the individual cells, the α_i=exp(-$\mu\tau_i$) quantities. We refer to this as *physiological state theorem.* Note that the theorem is generally valid for "smooth" distributions of the individual lag times.

Consider the biological interpretation of the physiological state of the inoculum as below. Let $y_0 = \ln N$ is the natural logarithm of the inoculum level, and let the lag period be λ. Find another, hypothetical growth curve which will be identical to the previous 'real' growth curve in its exponential phase but has no lag (Fig. 6.1). It can be readily seen that α=exp(-$\mu\lambda$) is the factor by which if N decreased, the hypothetical growth curve crosses the vertical axis. In other words, α expresses the potential fraction of the initial counts which, without lag, could "catch up" with the real growth curve which does have lag. The extreme values of this fraction are 0 and 1, corresponding to the situations that the real growth curve has 'infinitely long lag' and 'no lag', respectively. The α physiological state is a dimensionless parameter quantifying the 'suitability' of the culture to the actual environment. It is, in fact, an initial value, just like the inoculum level, from which the lag parameter can be derived by $\lambda = -\ln\alpha/\mu$, expressing that the lag is inversely proportional to the maximum specific growth rate and depends on the physiological state of the inoculum, too, not only on the actual environment.

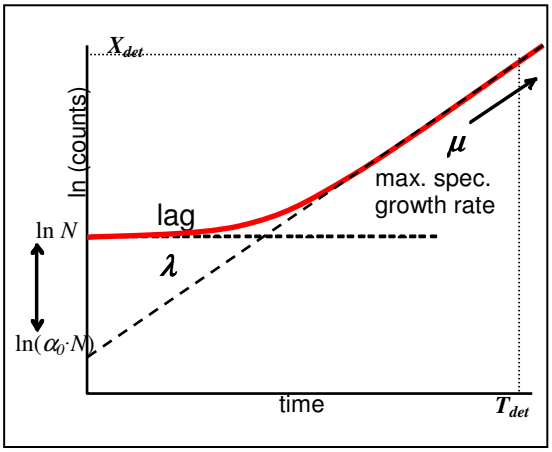


Fig.6.1. The intercept of the inoculum level with the tangent drawn to the exponential phase of the growth curve marks the end of the lag phase. The detection time, T_{det}, depends linearly on the lag, λ. The parameter α_0, can be used to define a hypothetical inoculum level from which a growth curve, without lag, is able to catch up with the real curve with lag.

In this Chapter, we highlight a useful feature of the physiological state parameter. We develop a new method, based on the physiological state theorem and an ANOVA procedure, to estimate the maximum specific growth rate and the lag time of a homogeneous bacterial population. The advantage of the method is that it uses detection times, which are the first data available when recording bacterial growth, and it allows for the estimation of the within-population variance of lag times.

6.2 A consequence of the physiological state theorem.

Let $\alpha_i = \exp(-\mu\tau_i)$ $(i=1...x_0)$ denote the individual physiological states. If L_N denotes the population lag generated by N cells, then $\alpha(x_0) = \exp(-\mu L_N)$ is the physiological state of the inoculum consisting of x_0 cells. According to the physiological state theorem:

$$e^{-\mu L_N} = \alpha(N) = \frac{\sum_{i=1}^{N} \alpha_i}{N} = \frac{\sum_{i=1}^{N} e^{-\mu\tau_i}}{N} \qquad (6.1)$$

Therefore:

(i) the expected value of the physiological state of the initial population is the same as the (common) expected value, $\hat{\alpha}$, of the individual physiological states: $E(\alpha(x_0))=E(\alpha_i)= \hat{\alpha}$;

(ii) with higher initial counts, the physiological state of the initial population closer to $\hat{\alpha}$;

(iii) denoting the (common) variance of the individual physiological states by v, the rate of the above convergence can be estimated by the relation that the variance of the physiological state of the initial population is x_0-times smaller than the variance of the individual physiological states: $\mathrm{Var}(\alpha(x_0)) = \mathrm{Var}(\alpha_i)/x_0 = v/x_0$

6.3. An ANOVA protocol.

We use the population state theorem to develop an ANOVA procedure for our method. We use the indices *i, j, k* to differentiate between *N* cells of an inoculum (*i*=1...*N*); between *n* detection times generated by $N^{(j)}$ initial cells (*j*=1...*n*); and between m groups of identical inoculum levels (*k*=1...*m*).

Suppose that a culture, growing from x_0 initial counts, reaches a certain detection level, X_{det}, at the time T_{det}, while still in the exponential phase. As can be followed from Fig.5.1:

$$T_{\mathrm{det}}(N) = L_N + \frac{\ln \frac{X_{\mathrm{det}}}{N}}{\mu} \qquad (6.2)$$

From this equation, it follows that the detection time does not depend on both X_{det} and *N* independently, but only on the ratio

$$r = \frac{N}{X_{\mathrm{det}}}$$

that we call the dilution ratio. If the variance of the dilution ratio is negligible, the distribution of the lag times, apart from a constant additive term, is identical to the distribution of the detection times.

Suppose that we measure the $T^{(j)}$ detection times for some subcultures generated by $N=x_0^{(j)}$ initial counts (*j*=1,2...*n*) . Denote $\alpha^{(j)}=\alpha(x_0^{(j)})$. Then, from equation (6.2),

$$\alpha^{(j)} = \frac{e^{-\mu T^{(j)}}}{r^{(j)}} \qquad (6.3)$$

where

$$r^{(j)} = x_0^{(j)} / X_{\mathrm{det}} \qquad (6.4)$$

Besides, as we have seen,

$$Var(\alpha^{(j)}) = \frac{v}{X_{\mathrm{det}} r^{(j)}} \qquad (6.5)$$

(recall that *v* is the common variance of the $\exp(-\mu\tau_i)$ individual physiological states).

For the expected value of the physiological state, an efficient estimation is the weighted average of the $\alpha^{(j)}$ values, where the weights are proportional to the reciprocals of the respective variances. After simplification, we obtain:

$$\overline{\alpha} = \frac{\sum_{j=1}^{n} r^{(j)} \alpha^{(j)}}{\sum_{j=1}^{n} r^{(j)}} = \frac{\sum_{j=1}^{n} e^{-\mu T^{(j)}}}{r_{sum}} \tag{6.6}$$

where

$$r_{sum} = \sum_{j=1}^{n} r^{(j)} \tag{6.7}$$

The variance of the $\overline{\alpha}$ estimator is:

$$\mathrm{Var}(\overline{\alpha}) = \frac{\sum_{j=1}^{n} \left(r^{(j)}\right)^2 \mathrm{Var}\left(\alpha^{(j)}\right)}{\left(\sum_{j=1}^{n} r^{(j)}\right)^2} = \frac{\sum_{j=1}^{n} \left(r^{(j)}\right)^2 \frac{v}{r^{(j)} X_{det}}}{\left(\sum_{j=1}^{n} r^{(j)}\right)^2} = \frac{v}{X_{det} r_{sum}} \tag{6.8}$$

Suppose that we dilute a culture from the detection level, X_{det} , and we obtain $n=n_1+\ldots+n_m$ subcultures, where

- n_1 subcultures belong to Group 1 characterized by the $r^{(1)} = r_{1,1} = \ldots = r_{1,n_1}$ dilution ratio.

  Denote their detection times by $T_{1,1}, \ldots, T_{1,n_1}$;

.....

- n_m subcultures belong to Group m characterized by the $r^{(m)} = r_{m,1} = \ldots = r_{m,n_m}$ dilution ratio.

  Denote their detection times by $T_{m,1}, \ldots, T_{m,n_m}$.

Let $\overline{\alpha}^{(k)}$ be the mean of the k-th group ($1 \le k \le m$) of the physiological state observations and let

$$\alpha_{k,j} = \frac{e^{-\mu T_{k,j}}}{r^{(k)}} \tag{6.9}$$

Then, from equation (3):

$$\overline{\alpha}^{(k)} = \frac{\sum_{j=1}^{n_k} \alpha_{k,j}}{n_k} = \frac{\sum_{j=1}^{n_k} e^{-\mu T_{k,j}}}{n_k r^{(k)}} \tag{6.10}$$

As we know, the variance of $\overline{\alpha}^{(k)}$ is

$$\mathrm{Var}(\overline{\alpha}^{(k)}) = \frac{v}{X_{det} n_k r^{(k)}} \tag{6.11}$$

Define the *V*-value as the variance ratio

$$V = \frac{\sum_{k=1}^{m} \frac{\left(\overline{\alpha}^{(k)} - \overline{\alpha}\right)^2}{\mathrm{Var}\left(\overline{\alpha}^{(k)} - \overline{\alpha}\right)}}{\sum_{k=1}^{m}\sum_{j=1}^{n_k} \frac{\left(\alpha_{k,j} - \overline{\alpha}^{(k)}\right)^2}{\mathrm{Var}\left(\alpha_{k,j} - \overline{\alpha}^{(k)}\right)}} \tag{6.12}$$

Following the standard ANOVA technique,

$$\mathrm{Var}\left(\overline{\alpha}^{(k)} - \overline{\alpha}\right) = \mathrm{Var}\left(\overline{\alpha}^{(k)}\right) + \mathrm{Var}\left(\overline{\alpha}\right) - 2\mathrm{E}\left[\left(\overline{\alpha}^{(k)} - \hat{\alpha}\right)\left(\overline{\alpha} - \hat{\alpha}\right)\right] \tag{6.13}$$

As can be checked,

$$\mathrm{E}\left[\left(\overline{\alpha}^{(k)} - \hat{\alpha}\right)\left(\overline{\alpha} - \hat{\alpha}\right)\right] = \frac{v}{nX_{\mathrm{det}}r^{(k)}} \tag{6.14}$$

therefore

$$Var\left(\overline{\alpha}^{(k)} - \overline{\alpha}\right) = \frac{v}{X_{\mathrm{det}}}\left(\frac{1}{n_k r^{(k)}} + \frac{1}{r_{sum}} - \frac{2}{nr^{(k)}}\right) \tag{6.15}$$

Calculate the variances in the denominator of (6.12):

$$\mathrm{Var}\left(\alpha_{k,j} - \overline{\alpha}^{(k)}\right) = \frac{v}{X_{\mathrm{det}}}\left(\frac{1}{r^{(k)}} - \frac{1}{n_k r^{(k)}}\right) \tag{6.16}$$

After substitution, we obtain:

$$V = \frac{\sum_{k=1}^{m} \frac{r^{(k)}\left(\overline{\alpha}^{(k)} - \overline{\alpha}\right)^2}{\frac{1}{n_k} + \frac{r^{(k)}}{r_{sum}} - \frac{2}{n}}}{\sum_{k=1}^{m} \frac{r^{(k)}}{1 - 1/n_k} \sum_{j=1}^{n_k} \left(\alpha_{k,j} - \overline{\alpha}^{(k)}\right)^2} \tag{6.17}$$

The distribution of the V variance ratio, if completed with the respective degrees of freedom, is very close to that of the *F*-distribution, except that the differences in the summations are not normally distributed. Even so, due to the robustness of *F*-statistics, the maximum specific growth rate, μ, can be estimated by minimizing the *V* variance ratio. The advantage of this is that

V is dimensionless, independent of the X_{det} and $x_0^{(j)}$ values and depends only on the dilution ratios and the detection times.

6.4. Practical implementation

Baranyi and Pin (1999) used *Pseudomonas* strains to record the turbidity of the cultures inoculated in Bioscreen. Microtiter plates, with 300 µl/well, were incubated and the detection times were measured needed to reach $OD_{det} = 0.15$, equivalent to approximately $X_{det} \approx 10^7$ cells/well. This estimate was checked by making a series of dilutions from a culture grown in nutrient-rich broth at 25 °C for 24 h. The ODs of the dilutions were monitored by the Bioscreen while bacterial counts were estimated by ordinary plating technique.

A culture, whose turbidity was $OD_{det}=0.15$ was used to produce altogether $m=7$ groups of subcultures with different inoculum levels. The groups k=1...7 were characterized by $r_1 ... r_7$ dilution ratios where $r_1 = 10^{-3} \cdot 2^{-6}$ and $r_k = r_{k-1}/2$ (k=2...7), because of consecutive binary dilutions. Note that the exact value of X_{det} belonging to OD_{det} is not necessary for our method; it is enough to know that X_{det} is reached in the exponential phase.

The data were collected in a Microsoft Excel spreadsheet and the 'Solver' add-in of the software was used to minimize the calculated V variance ratio with respect to the maximum specific growth rate, μ. A sample data and calculations are given in Tables 6.1 and 6.2.

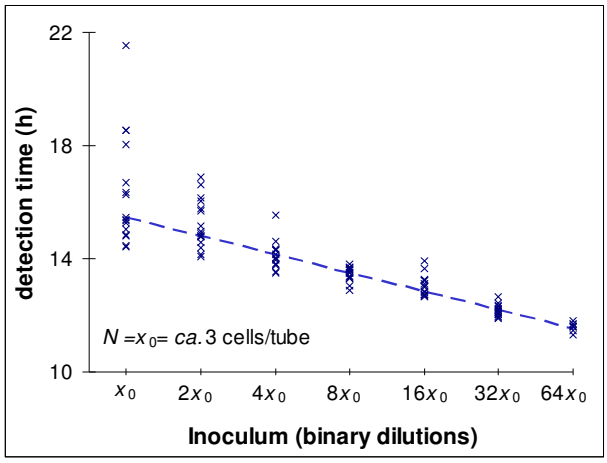


Fig.6.2. Detection times from different inoculum levels obtained by a series of binary dilutions.
The lower the inoculum level the larger is the scatter of the detection times.

Legend:
Cross (×): measured detection times.
Broken line (---): detection time predictions

The observed detection times belonging to seven groups of initial counts are shown in Fig. 6.2. The groups are characterized by the $r_1=10^{-3} \cdot 2^{-6}$ ... $r_7=10^{-3} \cdot 2^{-12}$ dilution ratios. Calculating with $X_{det}=10^7$ cells/well detection level corresponding to the $OD_{det} = 0.15$ turbidity, the initial counts in the wells of the lowest inoculum level was around $N = x_0 = r_7 X_{det} = 10^{-3} \cdot 2^{-12} X_{det} = 2.44$

cells/well. As has been said, however, the actual values of x_0 or X_{det} were not used to estimate the maximum specific growth rate and the population lag.

The ANOVA procedure described above estimated μ=1.07 h^{-1} for the maximum specific growth rate of *pseudomonad* at 25 °C. From μ, the mean physiological state of the inoculum was estimated as $\hat{\alpha} = 0.27$ (see the grand mean in Fig.6.3). The population lag was calculated as $\lambda = -\ln\hat{\alpha}/\mu = 1.21$ (h).

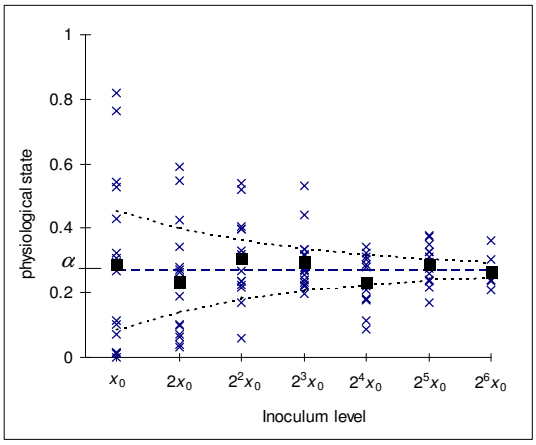


Fig.6.3. Physiological state values, obtained by transforming the detection times by means of the respective dilution ratios and the estimated specific growth rate. Legend:
Cross (×) : Physiological state data.
Filled squares connected by continuous line (■): Group mean of the physiological state data generated by the same inoculum level.
Broken line (---) : Grand mean, ; the estimate for the mean physiological state of the initial population.
Dotted curve (....) : the expected theoretical deviation from the grand mean calculating with $X_{det}=10^7$ detection level and assuming exponential distribution for the individual lag times.

To demonstrate, how robust the technique is, Fig. 6.4 shows the scatter and trend of the physiological state values at two specific growth rates which were obtained by perturbing the calculated μ value. If the specific rate is chosen about 10% lower or higher, then the (group means of the) physiological states show a downwards or upwards tendency.

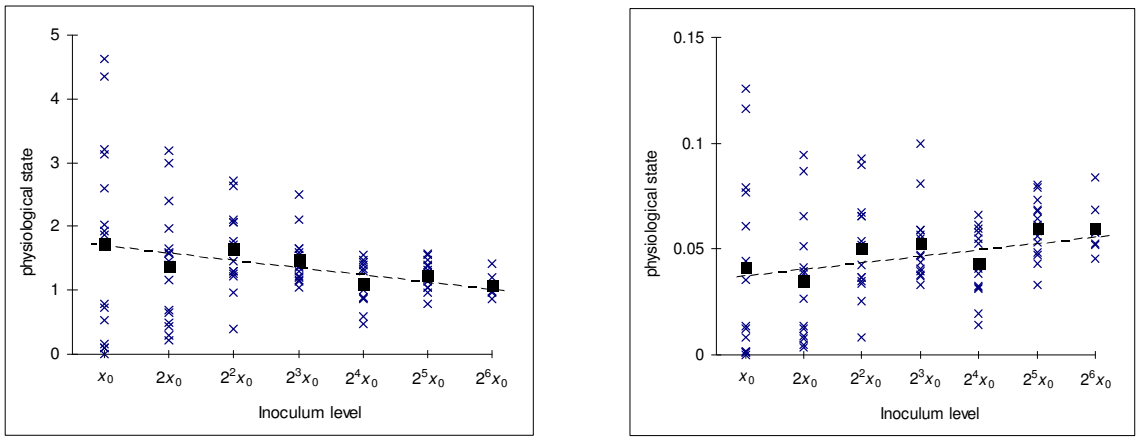


Fig.6.4. Demonstrating the robustness of the method. If the maximum specific growth rates are slightly perturbed (μ(A)=0.95 h^{-1} , μ(B)=1.2 h^{-1} ; instead of the correct μ=1.07 h^{-1}), then the physiological state values show strong downwards and upwards tendency, respectively. Legend as in Fig.6.3.

6.5. Discussing the use of detection time measurements for single cell studies

Detection times from different $x_0^{(j)}$ initial levels, have been used by other authors to estimate bacterial growth parameters. Unfortunately, the variance of the observed detection times

increases as the inoculum size decreases. We overcome this problem by applying the physiological state theorem. One of its consequences is that the variance of the $\alpha = \exp(-\mu T_{det})/r$ value is inversely proportional to the $r=x_0/X_{det}$ dilution ratio. This relationship was used to develop an ANOVA procedure.

To apply our method, the detection level should be in the exponential phase. If, for example, X_{det} is close to the stationary phase then the method underestimates the real specific rate. Another source for uncertainty is the possible error in the r dilution ratio.

The physiological state theorem is valid irrespective of the distribution of the lag times, τ_i, of the individual cells. An important case, however, when those are exponentially distributed, deserves special attention. In that case, the mean individual lag time can be estimated by

$$\tau = \frac{\frac{1}{\alpha_0} - 1}{\mu}$$

with the same τ variance, while the variance of the individual physiological states is

$$\mathrm{Var}(\alpha_i) = v = \frac{1}{(1+2\mu\tau)\left(1+\frac{1}{\mu\tau}\right)^2}$$

(Note that the mean individual lag time is longer than the population lag).

Applying the above formulae to our numerical results, the average of the lag times of the individual cells was 2.5 hours, with v=0.084 variance (*ca.* 0.29 hours standard deviation). By means of v and the estimated $X_{det}=10^7$ detection level, the standard deviations of the $\bar{\alpha}^{(k)}$ group means can be calculated (assuming exponential distribution for the individual lag times) by equation (5.11) . These estimated standard deviations are represented by the differences between the dotted lines and the grand mean of the α values in Fig. 6.3. The fact that they are close to the standard deviations of the groups (which can be calculated simply from the raw data, irrespective of the exponential assumption) suggests that the distribution of the lag times of the individual cells are, indeed, close to exponential.

An important point in the applicability of the method is that, as follows from the assumptions of the physiological state theorem, the total number of cells in a homogeneous living space that should be considered for the inoculum, as well as for the detection level (cells/well), and not just the density of the inoculum. Therefore, a population of, say, 1 cell/ml concentration in a liter volume (1000 cells altogether) should produce the same lag as 10^3 cell/ml

concentration in a ml volume. This relationship does not hold in practice because the cells do not grow independently but exchange chemical signals whose effectiveness does depend on the actual size of the living space.

The distribution of the detection times of cultures with low initial numbers has not been previously examined in detail and has the potential to be used in the development of stochastic approaches. For practical applications, see a simple spreadsheet model described by Tables 6.1 and 6.2, using the syntax of Microsoft Excel, being aware that food microbiology labs are commonly collecting their data using this application.

A	B	C	D	E	F	G	H	I	J	K
k	**j**	**N_k**	**r_kj**	**T_kj**	**beta_kj**	**alpha_kj**	**alpha_k**		**Between**	**Within**
1	1	4	2.44E-07	21	1.05E-11	4.29E-05	3.06E-03	9.1E-06	1.438E-12	6.9E-12
1	2		2.44E-07	18	3.88E-10	1.59E-03	3.06E-03	2.2E-06		
1	3		2.44E-07	17	1.29E-09	5.30E-03	3.06E-03	5.0E-06		
1	4		2.44E-07	17	1.29E-09	5.30E-03	3.06E-03	5.0E-06		
2	1	3	4.88E-07	18	3.88E-10	7.95E-04	2.03E-03	1.5E-06	3.319E-13	1.7E-12
2	2		4.88E-07	17	1.29E-09	2.65E-03	2.03E-03	3.8E-07		
2	3		4.88E-07	17	1.29E-09	2.65E-03	2.03E-03	3.8E-07		
					n	**r_sum**	**alpha**		**mu**	**V**
					7	2.44E-06	2.44E-03		1.2E+00	2.1E-01

Table 6.1. Example data and their organization given in an Excel sheet to follow the ANOVA procedure. The raw data are given in the D3-E9 region. The calculations can be followed in the F3-K12 region. The numerical results shown here are generated by the respective Excel statements given in Table 6.2. The bold letters indicate Excel variable names referring to the arrays below them, which were used in the formulae in Table 6.2.

F	G	H	I	J	K
beta_kj	**alpha_kj**	**alpha_k**		**Between**	**Within**
				=r_k*(alpha_k-	=r_k/
		=AVERAGE	=(alpha_kj-	alpha)^2/(1/n_k+	(1-1/n_k)*
=EXP(-mu*T_kj)	=beta_kj/r_kj	(G3:G6)	alpha_k)^2	r_k/r_sum-2/n)	SUM(I3:I6)
			=(alpha_kj-		
=EXP(-mu*T_kj)	=beta_kj/r_kj	=$H$3	alpha_k)^2		
			=(alpha_kj-		
=EXP(-mu*T_kj)	=beta_kj/r_kj	=$H$3	alpha_k)^2		
			=(alpha_kj-		
=EXP(-mu*T_kj)	=beta_kj/r_kj	=$H$3	alpha_k)^2		
				=r_k*(alpha_k-	=r_k/
		=AVERAGE	=(alpha_kj-	alpha)^2/(1/n_k+	(1-1/n_k)*
=EXP(-mu*T_kj)	=beta_kj/r_kj	(G7:G9)	alpha_k)^2	r_k/r_sum-2/n)	SUM(I7:I9)
			=(alpha_kj-		
=EXP(-mu*T_kj)	=beta_kj/r_kj	=$H$7	alpha_k)^2		
			=(alpha_kj-		
=EXP(-mu*T_kj)	=beta_kj/r_kj	=$H$7	alpha_k)^2		
N	**r_sum**	**Alpha**		**mu**	**V**
		=SUM(beta_kj)/			=SUM(Between)/
=SUM(n_k)	=SUM(r_kj)	r_sum		1.204	SUM(Within)

Table 6.2. Respective Excel statements generating the numerical results given in Table 6.1. The Solver add-in minimized the value of the V variance ratio in the K12 cell by varying the value of μ in the J12 cell.

Chapter 7

Numerical estimations for the statistical distributions of individual lag times

7.1. Single cell lag time – a review on its definition

In this chapter, we will denote the lag parameter defined at population level, by L_g (see Fig.7.1), referring to its rather "geometric" definition. The $L_g(N_0)$ notation indicates that its value depends on the initial cell number, N_0.

In Fig. 7.1, the horizontal "steps" represent the intervals when no division occurred. The "physiological" lag for the initial cell was a random number generated to follow the Gamma distribution with mean=5.4 h, σ=1.8 h. After the lag, the subsequent times for division of the daughter cells were also generated by the Gamma distribution, with mean=1 h and σ=0.26 h. The time to the first division (equal to the "detection time" when the detection level is y_{det}=ln 2) is the sum of τ and the first generation time. The parameters of the simulation were taken from flow-chamber experiments presented in Elfwing (2004).

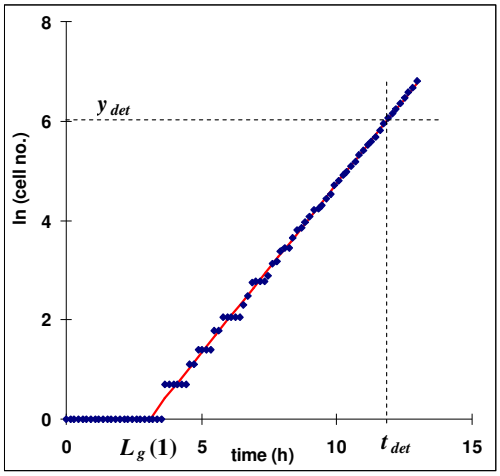


Fig.7.1. Simulated growth curve of a one-cell-generated population growing as a result of subsequent divisions.

As have been said, no observable event happens at the end of the τ_i period. The time to the first division is in fact the sum of τ_i and the first generation time. When interpreting the single cell lag time as above, it is important to see that $L_g(1)$ is not identical to τ_i. In fact, $L_g(1)$ is not a parameter of the original single cell, but that *of the sub-population generated by the single cell*. Namely, due to its definition, the estimation is affected by the variability of the successive generation times of the

daughter cells. The distribution of the detection times, t_{det}, reflects the distribution of the L_g "geometrical" lag and not necessarily the physiological lag.

To analyse this difference, consider the simulated, single-cell-generated population growth curves of an exponentially growing culture in Fig.7.2. It was assumed that the initial cell had no lag , however, the resulting geometrical lag times, L_g, are different from zero because of the variability of the successive generation times of the daughter cells. The estimated L_g values are dispersed around zero, with a variance that depends on the generation time distribution. The L_g parameter has a considerable spread (Fig. 7.3), due to the variance of the first few generation times.

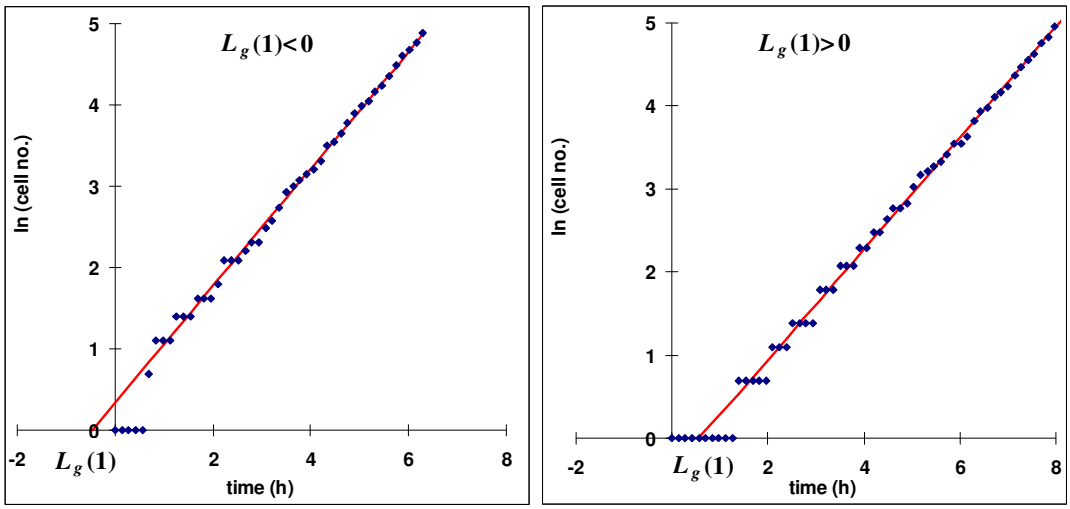


Figure 7.2. Two simulated examples for a one-cell-generated population with no lag. The distribution of the generation times is as in Fig. 7.1. The distribution of the L_g parameter comes exclusively from the distribution of the subsequent generation times, since the physiological lag time is zero. The two panels demonstrate that both (i) $L_g(1)<0$ and (ii) $L_g(1)>0$ are possible for the random population growth process

The fact that L_g is not a physiologically interpretable single-cell lag time is not a disadvantage. With the L_g parameter, a simple bi-phasic model can be used to predict the time and its distribution by which the population would reach a given (e.g. harmful) level of bacteria; risk assessors need exactly this.

7.2. Detection time

In practice, it is difficult to follow the division of individual cells with automated measurements. Suppose that $y_{det}= \ln(x_{det})$ is the level at which a homogeneous bacterial culture can be detected, for example, by measuring its turbidity in a liquid

medium (see the previous Chapters). Suppose that y_{det} is in the exponential phase and let T_{det} denote the detection time at which the single-cell generated subpopulation reaches this y_{det} threshold value. Then the detection time is a shifted version of the L_g parameter. However, this value depends on the number of initial cells and if this is unknown, it is not easy to infer to the single cell lag times.

Our main objective is to use detection time data to give an accurate estimation procedure for the distribution parameters of the single cell lag times, when the initial cell number is random. We will also discuss some practical implications regarding Quantitative Microbial Risk Assessment.

7.3. Practical considerations and measurements

Simulation studies can prove that, under reasonable assumptions, $L_g(1)$ can be well approximated by a shifted Gamma distribution, if the single cell generation times are also Gamma-distributed (not necessarily with the same parameters as the distribution of $L_g(1)$). This is demonstrated in Fig.7.3.

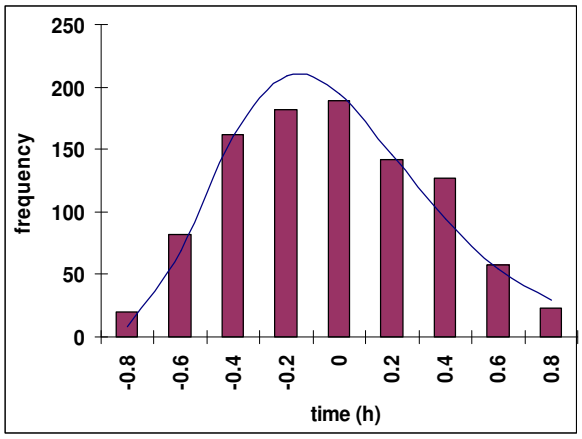


Fig. 7.3. Distribution of the $Lg(1)$ single cell lag times with the same simulation parameters as in Fig. 6.2, based on 1000 replicates of a Monte-Carlo simulation. The age of the initial cell was also picked randomly. The cells had no physiological lag time, so this distribution of the geometrical lag time is a sole consequence of that of the subsequent generation times. It is compared with a zero-centered, shifted Gamma distribution (continuous line), where the shift is equal to the mean of the generation times (1 h) but the deviation is higher (0.4 h vs. 0.26) than that of the generation times.

Another source of variability could be the specific growth rate of the single-cell generated population. As the simulation studies showed, the variance of the specific growth rate is negligible in comparison to the variance of L_g and the first few generation times (Métris et al 2003). This can also be seen intuitively, since the specific rate is determined by many cells, therefore its variance is much less than that of the division time of a single cell. Hence the estimation of the specific growth rate

is hardly affected by the variability of the single cell generation times in the exponential phase.

First (Case A), the detection times of a *Listeria innocua* culture (Métris et al, 2006) were used to test the new method. The detection times refer to the times when the turbidity as measured by Bioscreen C plate reader (an automated optical density detector) reached a fixed level (optical density, OD=0.11) which was shown to be equivalent to ca $10^{7.7}$ cell/ml concentration.

In the second experiment (Case B), *L. innocua* was inoculated in the same way, but after receiving a mild heat shock, so the lag was significantly longer.

The maximum specific growth rate of the organism was measured in an independent experiment, by means of traditional plate count technique. The parameter, at 22°C, was estimated as $\mu = 0.45$ h^{-1} from the plot of colony forming units against time, using the curve-fitting program DMFit, available at www.ifr.ac.uk/safety/DMFit.

If the initial number of cells in a well is exactly one and the detection level is in the exponential phase, then the difference between the detection time and the $L_g(1)$ parameter of the initial single cell depends only on the specific growth rate of the organism, which is constant as the population is in a homogeneous environment (see above). In this case the detection time differs from the $L_g(1)$ single cell lag time only by an additive constant.

There is no easy way to sort single cells into the wells of microtitre plates. Methods available include sorting by flow cytometer, which is expensive and can introduce bias during the process, e.g. if the cells are sorted according to size. Instead cultures are diluted to a level such that a sample in a well should contain only a few cells. With a sufficiently high dilution factor, most wells will receive zero or one cell. This technique was applied by Augustin (1999) and Francois (2005) to evaluate the variation of the lag times of single cells The disadvantage is that many wells will be empty and, for a statistically robust distribution estimation, it is desirable to have as many positive wells as possible, a minimum of about one hundred (Baranyi, 2005).

7.4 Estimation procedures

Cell cultures were obtained after successive dilutions, and placed into W number of wells (W = 200 for the microtitre plates of the Bioscreen). Hence, the initial number of cells (N_0) in a well follows the Poisson distribution, i.e.

$$P_k = P(N_0 = k) = \exp(-\rho)\frac{\rho^k}{k!} \qquad (k=0, 1, \ldots) \qquad (7.1)$$

where the expected value of N_0 is $E(N_0)=\rho$. The fraction of empty wells can be used to estimate ρ by $\hat{\rho} = -\ln(W_0 / W)$, where W_0 is the number of negative (empty) wells.

A linear method relating detection time and single cell lag time. Consider the stochastic delayed birth process model proposed in Chapter 4. For a cell in the initial population, the first division happens after a time interval, which is the sum of its physiological lag time and the first generation time. At time point t in the exponential phase, the natural logarithm of the expected population size in a well can be calculated from

$$y(t;N_0) = \ln(N_0) + \mu\,(t - L_g(N_0)) \qquad \text{for } t > L_g \qquad (7.2)$$

where $L_g(N_0)$ is the "geometrical" lag in the well, and μ is the specific growth rate of the population. The notation indicates that the L_g variable depends on the N_0 initial number of cells in a well. We assume that, for single cells, the distribution of $L_g=L_g(1)$ can be well described by a shifted Gamma distribution, i.e. for a T_{shift} parameter, $L_g(1) - T_{\text{shift}}$ is Gamma distributed (Fig.7.3). In order to create a simple bi-phasic model for the growth curve the T_{shift} parameter compensates for the randomness of the first few division times. Alternatively, if T_{shift} is significantly different from zero, then it can be interpreted as a 'compulsory repair time' for a period after injury during which the probability of division is zero for any cell. An example will be shown for this later (Case B), with heat-shocked cells.

The linear method transforms the $T_{det}(N_0)$ detection times into individual lag times using the estimation for ρ :

$$L_g^{(LIN)} = T_{\det}(N_0) - \frac{y_{\det} - \ln(\hat{\rho})}{\mu} \quad (7.3)$$

where $L_g^{(LIN)}$ is the linear estimation for the $L_g(1)$ single cell lag time, y_{det} is the natural logarithm of the cell concentration at the detection level. This linear method substitutes the random variable, N_0, by an estimation of its mean, therefore the variability of $L_g(1)$ comes only from that of the observed detection times. If N_0 was fixed at 1, then the difference between $T_{det}(1)$ and $L_g(1)$ would be only the additive term y_{det}/μ .

A 'moments-based' method. Define

$$\alpha(N_0) = \exp(-\mu L_g(N_0)) \quad (7.4)$$

as the *physiological state parameter* for the population generated by N_0 initial cells in a well (Baranyi, 1998). Let $S_\alpha(N_0)$ denote the sum of the physiological state parameters of the one-cell generated subpopulation:

$$S_\alpha(N_0) = \sum_{j=1}^{N_0} \alpha_j(1) \quad (7.5)$$

As we have seen in Chapter 4, the physiological state parameter of a population is the arithmetical average of the physiological state parameters of the constituent subpopulations,

$$\frac{S_\alpha(N_0)}{N_0} = \alpha(N_0) \quad (7.6)$$

and using the relationship

$$y_{det}(N_0) = \ln(N_0) + \mu\,[T_{det}(N_0) - L_g(N_0)] \quad (7.7)$$

we obtain

$$S_\alpha(N_0) = e^{y_{\det}(N_0) - \mu \cdot T_{\det}(N_0)} \quad (7.8)$$

This idea was used by Métris et al (2006). In Baranyi et al (2009) we developed the method further. We utilised the fact that N_0 is Poission-distributed and we can get explicit formulae for the $\mathbf{M_1}$, $\mathbf{M_2}$, $\mathbf{M_3}$ moments of $S_\alpha(N_0)$:

► The first three moments of the random variable $e^{-\mu\tau}$, where τ is Gamma distributed with scale parameter θ and shape parameter β:

$$m_1 = (1+\mu\theta)^{-\beta} \tag{7.9}$$

$$m_2 = (1+2\mu\theta)^{-\beta} \tag{7.10}$$

$$m_3 = (1+3\mu\theta)^{-\beta} \tag{7.11}$$

Suppose that the $L_g = L_g(1)$ geometrical lag has a (constant) time shift parameter, i.e. L_g-T_{shift} follows the original Gamma distribution. Therefore, the respective moments of the physiological state $\alpha = e^{-\mu L_g}$ are

$$M_1 = \mathbf{E}(e^{-\mu L_g}) = e^{-\mu T_{\text{shift}}}(1+\mu\theta)^{-\beta} \tag{7.12}$$

$$M_2 = \mathbf{E}(e^{-2\mu L_g}) = e^{-2\mu T_{\text{shift}}}(1+2\mu\theta)^{-\beta} \tag{7.13}$$

$$M_3 = \mathbf{E}(e^{-3\mu L_g}) = e^{-3\mu T_{\text{shift}}}(1+3\mu\theta)^{-\beta} \tag{7.14}$$

The moments of the compound random variable $S_\alpha(N_0)$ can be calculated by means of the moments of N_0 and α_j:

$$\mathbf{M}_1(S_\alpha) = \mathbf{E}\left(\sum_{j=1}^{N_0}\alpha_j\right) = \mathbf{E}(N_0)\cdot M_1 = \rho\, M_1 \tag{7.15}$$

$$\mathbf{M}_2(S_\alpha) = \mathbf{E}\left(\sum_{j=1}^{N_0}\alpha_j{}^2\right) = \mathbf{E}(N_0)\cdot M_2 + \left(\mathbf{E}(N_0)\, M_1\right)^2 = \rho\, M_2 + \rho^2\, M_1{}^2 \tag{7.16}$$

$$\mathbf{M}_3(S_\alpha) = \mathbf{E}\left(\sum_{j=1}^{N_0}\alpha_j{}^3\right) = \rho\, M_3 + 3\rho^2 M_1 M_2 + \rho^3 M_1{}^3 \tag{7.17}$$

Combining the two sets of equations:

$$\mathbf{M}_1(S_\alpha) = e^{-\mu T_{\text{shift}}}\rho(1+\mu\theta)^{-\beta} \tag{7.18}$$

$$\mathbf{M}_2(S_\alpha) = e^{-2\mu T_{\text{shift}}}\left(\rho(1+2\mu\theta)^{-\beta} + \rho^2(1+\mu\theta)^{-2\beta}\right) \tag{7.19}$$

$$\mathbf{M}_3(S_\alpha) = e^{-3\mu T_{\text{shift}}}\left(\rho(1+3\mu\theta)^{-\beta} + 3\rho^2(1+\mu\theta)^{-\beta}(1+2\mu\theta)^{-\beta} + \rho^3(1+\mu\theta)^{-3\beta}\right) \tag{7.20}$$

Let the $L_g(1)$- T_{shift} Gamma distributed variable have the scale and shape parameters θ and β, respectively. Make the moments of S_α equal to the empirical moments of S_α :

$$e^{-\mu T_{\text{shift}}}\rho(1+\mu\theta)^{-\beta} = \frac{\sum_{j=1}^{n} e^{y_{\text{det}}{}^{(j)} - \mu \cdot T_{\text{det}}{}^{(j)}}}{n} \tag{7.21}$$

$$e^{-2\mu T_{\text{shift}}}\left(\rho(1+2\mu\theta)^{-\beta} + \rho^2(1+\mu\theta)^{-2\beta}\right) = \frac{\sum_{j=1}^{n} e^{2\left(y_{\text{det}}{}^{(j)} - \mu \cdot T_{\text{det}}{}^{(j)}\right)}}{n} \tag{7.22}$$

$$e^{-3\mu T_{\text{shift}}}\left(\rho(1+3\mu\theta)^{-\beta} + 3\rho^2(1+\mu\theta)^{-\beta}(1+2\mu\theta)^{-\beta} + \rho^3(1+\mu\theta)^{-3\beta}\right) = \frac{\sum_{j=1}^{n} e^{3\left(y_{\text{det}}{}^{(j)} - \mu \cdot T_{\text{det}}{}^{(j)}\right)}}{n} \tag{7.23}$$

where $j=1\ldots n$ numbers the observations,

As mentioned, ρ can be estimated by the number of empty wells, W_0, out of the total wells, W. The specific rate is calculated from different data, so three equations remain with 3 parameters, T_{shift}, β and θ. The estimate for $L_g(1)$ will be

$$L_g{}^{(M)} = T_{\text{shift}} + \beta\theta \tag{7.24}$$

with the fitted parameters.

7.4 Experimental validation

The author of this thesis coordinated a large EU-funded project, called Bacanova, in which seven institutes of different countries analysed bacterial lag times of single cells in order to provide a solid basis for quantitative microbial risk assessment (see Baranyi, 2005). One of the major outputs of the project was the paper of Elfwing et al (2004), reporting on a so-called flow-cell-based method. The output of the system was a series of images that made it possible to observe the statistical distribution of the

division times of single cells (Fig. 7.4 – experimental results from microbiologist colleagues):

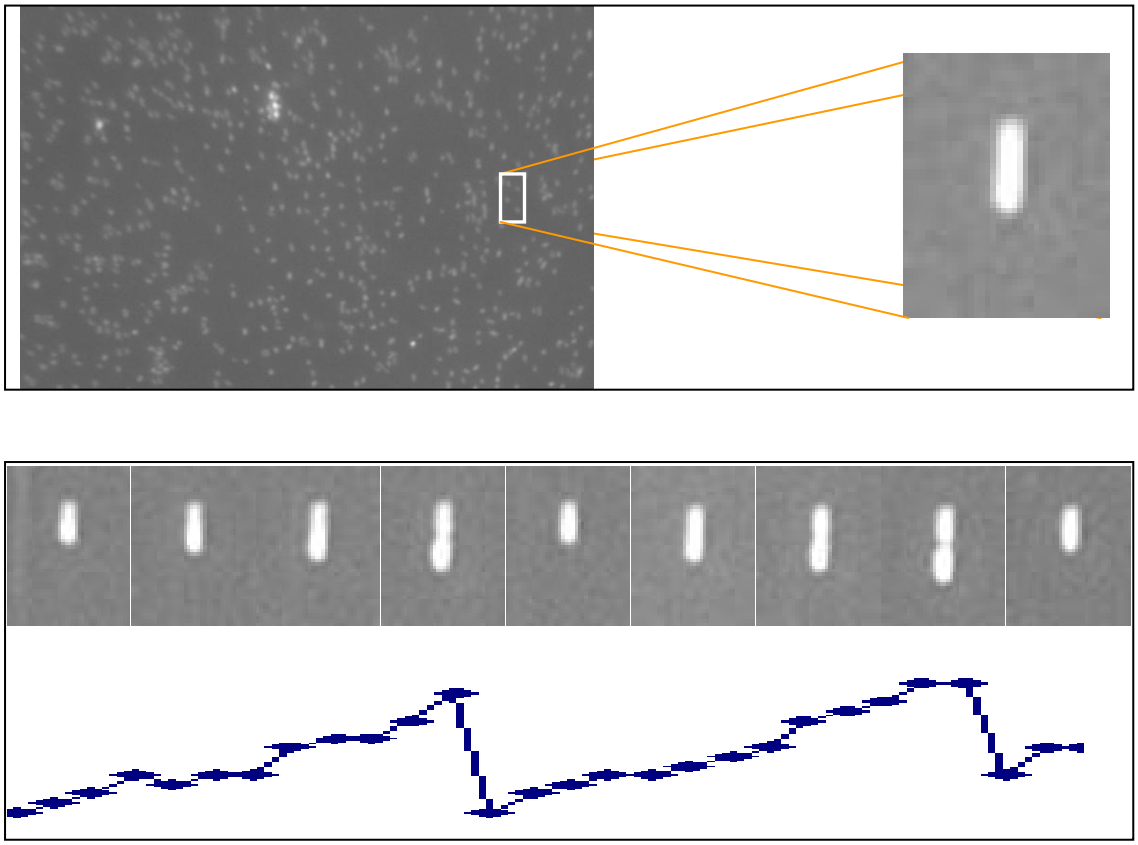


Fig.7.4. The flow chamber technique of Elfwing et al (2004) projects the cell division into a series of images from which the distribution parameters of single cell division times can be estimated.

The novelty of the method was that the cells were immobilised in the chamber, and the flow of nutrients was exactly so strong that during division one daughter cell attached to the surface remained immobile while the other was flushed away by the shear-power. Above them, a digital camera took photos at regular intervals and the images were analysed by analysis software. The sizes of hundreds of cells were translated into pixel sizes and the series of photos revealed the stochastic dynamics of cell division.

The data acquired this way was the basis of a series of papers (Métris et al, 2003, 2006; Kutalik et al 2004a, 2004b, 2005; Pin and Baranyi 2006, 2008). One of the important conclusions of the data was that the division time distribution of the cells gradually converged to the one they have in the exponential phase. Fig. 7.5 demonstrates this process.

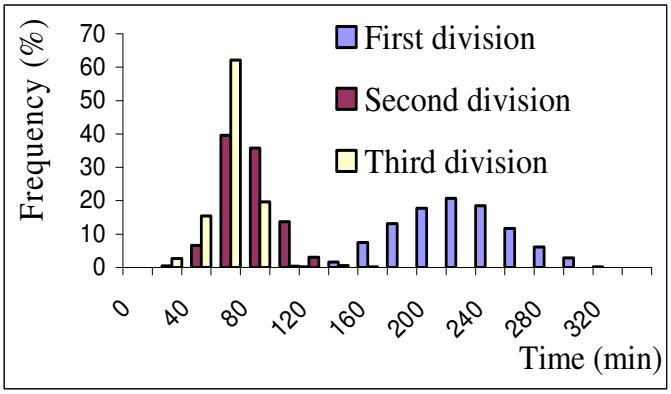


Fig.7.5 Cells need three-four divisions to arrive at the exponential phase (Métris et al, 2006), where their average division time is ca 3-4 times shorter, depending on the environment, than the time to the first division.

This result confirms that the physiological lag time is not a well-modellable variable, and highlights the usefulness of the L_g concept for the lag time.

7.6. Application

The linear method is reliable only if the average number of initial cells is very low, so the majority of the wells contain one cell only. We demonstrate below that there can already be a significant difference between the true $L_g(1)$ values and the linear estimate produced by Eq. (7.3) when the average initial number of cells in a well is between 2 and 3.

The simulation programs and the analysis below were prepared by my co-author, Z. Kutalik, of the Baranyi et al (2009) paper. Two cases for the Gamma parameters of the physiological single cell lag times ere simulated:

Simulation_1: $\tau \sim \Gamma(2,1)$ [$E(\tau) = 2h$, $SD(\tau) = 1.4h$]

Simulation_2: $\tau \sim \Gamma(2,10)$ [$E(\tau) = 20h$, $SD(\tau) = 14$ h]

The "coefficient of variation" values of these Gamma distributed variables are the same and, from observed data, realistic (Kutalik et al, 2005a). The difference between them is the ratio of their scale parameters to their means and to the subsequent specific growth rate, which is fixed at $\mu = 0.45$ (h^{-1}).

Assume that the cells are inoculated into 200 wells, with $E(N_0) = \rho = 2.5$ cells per well. Simulate the theoretical outcome considering only wells with less than 5 cells (88% of the non-empty wells). Table 7.1 shows the mean and SD values of the respective

distributions based on one million simulations. With Simulation_1, the SD of the linearly transformed T_{det} (last row) is smaller, while with Simulation_2 it is much bigger than that of $L_g(1)$ (which is represented in the first row, by the one-cell-generated subpopulations). There is no doubt that the distribution of the detection time, $T_{det}(N_0)$, is not a very good approximation of the distribution of the single cell lag time.

Number of initial cells in a well.	Proportion of wells with k cells among the positive wells	Simulation_1 $E(\tau) = 2$ h $\sigma(\tau) = 1.4$ h	Simulation_2 $E(\tau) = 20$ h $\sigma(\tau) = 14$ h
$N_0 \sim Poisson(2.5)$	$P_k = P(N_0=k) =$	(lag ≈ gen.time)	(lag >> gen.time)
$N_0 = k$ $(0<k<5)$	$(\rho^k e^{-\rho} / k!) / (1-e^{-\rho})$	$\sigma(L_g(k))$	$\sigma(L_g(k))$
$K = 1$	**$P_1 = 0.223564$**	**1.41**	**14.15**
$K = 2$	$P_2 = 0.279455$	0.86	8.25
$K = 3$	$P_3 = 0.232879$	0.67	6.12
$K = 4$	$P_4 = 0.145549$	0.56	5
Convolution $(k < 5)$	**0.881447**	$\sigma(L_g(N_0)) \approx$ **1.54**	$\sigma(L_g(N_0)) \approx$ **10.36**

Table 7.1. The detection time variability (i.e. the variability of the $L_g(N_0)$ lag time, where N_0 is random) can significantly differ from the variability of single cell lag times. If the single cell lag time is comparable to the generation times in the exponential phase (Simulation_1), then the difference is around 10%; however if the lag time is much longer than the subsequent generation times (Simulation_2), then the relative difference between the two can be as high as 40-50%.

It is evident that the higher the Poisson parameter, ρ, the less it is true that the linear method estimates the distribution of the single cell lag time. However, we demonstrate below that the new method is also suitable for $\rho>1$ cases. We applied it to two experimental data sets obtained with *L. innocua*.

Case A. The first example is the same as used by Métris et al (2006). Out of the W=200, 146 were positive, so the average initial number of cells per well was $\hat{\rho} = -\ln(54/200) = 1.31$. The specific rate was fixed at the value μ=0.15 h^{-1} as reported in

the original paper. The physiological state of the cell population could be calculated by using the estimate for the mean of N_0 :

$$\alpha = \frac{\sum_{j=1}^{n} e^{y_{\det}{}^{(j)} - \mu \cdot T_{\det}{}^{(j)}}}{\rho \cdot n} = 0.019$$

The standard deviation of the detection times was SD(T_{det})=12.5 h while our method resulted in SD(L_g(1))=32.23 h for the single cell lag times. T_{shift} was insignificant, and its value could be fixed at 0. After fitting the β, θ parameters, the average single cell lag time was calculated as $T_{\text{shift}}+\beta\theta$ = 55.5 h. Note that the lag of the population was $-\ln(\alpha)/\mu$ = 26.5, about half of the mean single cell lag time. These results coincide with those of Métris et al (2006), because the T_{shift} parameter was zero.

Case B. In the second example, after a mild heat shock, cells were inoculated into wells as described earlier. In this case, 189 wells out of W=198 were positive, so the average initial counts per well was $\hat{\rho}$ = -ln(9/198)=3.09. Equipped with the linear method only, experiments with such a high average initial count would normally be discarded as unsuitable to measure the variability of single cell lag times. Using our method, we were able to estimate that the standard deviation of the detection times was SD(T_{det})=15.52 h and that of the individual lag times was SD(L_g(1))=23.6 h. Here, the T_{shift} parameter was significant, its estimate was 6.6 hours. The physiological state parameter and the lag of the population was ca. 10^{-4} and 20.6 h, respectively, the latter being less than half of the mean of the single cell lag times that was estimated as 49h. These result could not have been obtained without introducing the T_{shift} parameter.

7.6. Comparing the two methods

Numerical stability. The estimation procedure requires the solving of the three equations (7.21)-(7.23) for T_{shift} , θ and β. The system of equations is highly non-linear and a common solver algorithm tends to find extreme values of the parameters, when all terms disappear. To avoid this, we minimise the sum of squares of the differences of the logarithm of the equality of the respective sides of those equations. If that minimum is sufficiently small, then we accept the obtained parameters as solutions.

Even in this case, a global minimisation algorithm can still cause problems because of the extreme parameter values. A local minimum is needed which requires good initial estimations of the parameters. This is quite demanding, and probably there is no better way than examining the sum of squares on a feasible grid of the parameters. If no local minimum was found then the data are not compatible with the basic assumptions of the method (Poisson distribution for N_0 and shifted Gamma distribution for $L_g(1)$).

As shown in Fig.7.4, the distribution of the single cell lag times can be either wider or narrower than the distribution of the $T_{det}(N_0)$ detection times, depending mainly on N_0 and the relation between the lag and the growth rate. With N_0 increasing, the estimation for ρ is less accurate; as are the estimates for the moments. However, when cells grow together in a well, then their population lag has a smaller variability than that of a single cell lag time. This increases the accuracy of the estimation procedure.

In our two examples, $SD(T_{det}(N_0)) < SD(L_g(1))$. The relatively wide distribution of the single cell lag time means that if cells grow together then the population lag of $N_0>2$ cells is dominantly determined by the subpopulation of the first dividing initial cells (by the time the slowest cells divide, the others are in the exponential phase). In this case, it is crucial to know the specific growth rate *a-priori*. Here we used viable count growth curves for its accurate estimation, but Baranyi and Pin (1999) showed a method to also measure the specific rate as well as determining the lag time distribution using the Bioscreen.

An obvious alternative to fitting a distribution to $T_{det}(N_0)$, instead of the moment-based method, is the Maximum Likelihood (ML) method. We compared our results with those generated by an ML program. The difference in the parameter estimates was insignificant but the computational cost (time, resources) of the ML method was much higher.

7.7. Application of single cell lag variation in Microbial Risk Assessment

The results can be used in simulations for Quantitative Microbial Risk Assessment. A framework for this is demonstrated in Fig. 7.6: Suppose that the initial number of cells on a portion of food is a random variable following a distribution that we call

"vertical" for obvious reasons. This distribution is the result of the history of the cells. Let y_0=lnN_0 denote the natural logarithm of the initial counts. Each cell goes through a lag period, L_g, which follows a distribution we call "horizontal", then, during the exponential phase, the log number increases linearly. The question a risk assessor is most interested in is the probability that the contamination level at the time of consumption, t_c, is above a certain threshold. The “vertical” distribution of this contamination level, $y(t_c)$ is affected by the “vertical” distribution of y_0 , the (horizontal) distribution of L_g and the specific growth rate that we consider constant here. The distribution of the bacterial population level can seldom be calculated explicitly by algebraic operations but Monte-Carlo simulations provide a straightforward solution to study them. For those simulations, the underlying distributions should be measured as accurately as possible. Unfortunately, the region of interest is at relatively low cell concentrations, where automated measurements, which are necessary for sufficient data to obtain robust distribution estimates, are not available. At high concentrations, however, assuming that the detection level y_{det} is still in the exponential phase, one can use the "horizontal" distribution of the T_{det} detection times instead of the L_g lag times, as we did here.

Generally, only vertical distributions are measurable in food, while broth cultures are suitable to determine both horizontal and vertical distributions. The transformation between them can be made by means of the specific growth rate, since the two types of distributions are rotated by 90^{o}.

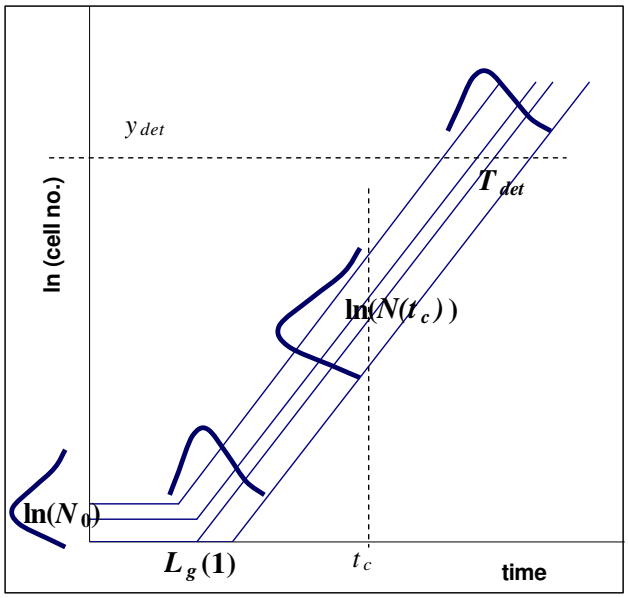


Fig. 7.4. Schematic representation of the horizontal and vertical distributions for a Monte-Carlo simulation for Quantitative Microbial Risk Assessment. The vertical distribution of the Ln $N(t_c)$ log cell number, where *tc* is the time of consumption, depends both on the vertically distributed initial load, y_0, and the horizontally distributed L_g "geometrical" lag time. The horizontal distribution of the detection time is measurable at higher cell concentration. The vertical and horizontal distributions can be transformed into each other by means of the specific growth rate, μ, which is assumed to be constant.

The lag time distribution has significance in predicting contamination levels only if the estimation of the specific growth rate is accurate. When this is the case, the ideas shown in this chapter can be built into Monte-Carlo simulations as the "computational engine" behind Microbial Risk Assessment programs.

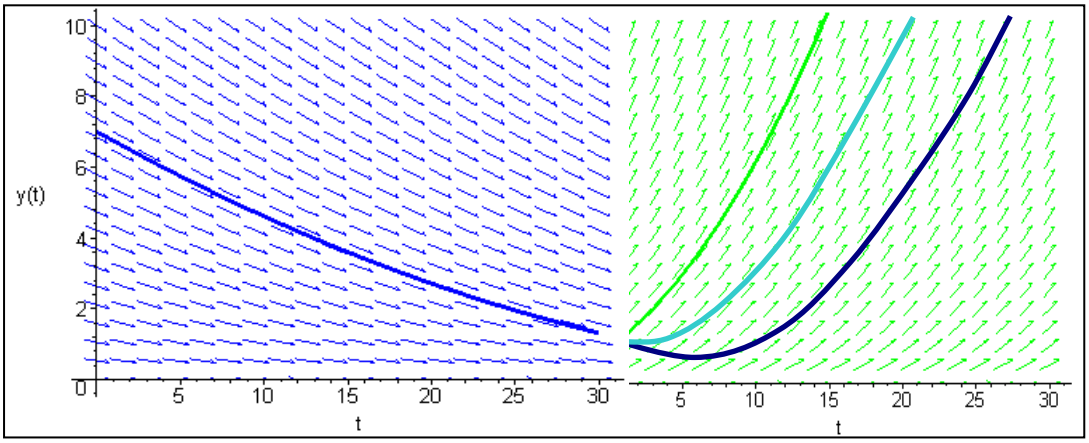
Chapter 8

Future research and conclusions

8.1. A predictable development

Physiological state of bacteria and their role in adaptation have not been modelled before. In fact, strictly speaking, the results above do not show a mechanistic model of the physiological state either, rather a summary of the desired and needed analytical properties of a mechanistic model. The computational analysis of their fitting properties showed that the deterministic model is useful from empirical curve fittings, too; while the stochastic model is suitable to assess the kinetics and the distribution of the kinetic parameters of individual cells (Baranyi et al 2009).

The deterministic model introduced in Chapter 2 can be embedded in the following problem:



$\frac{d}{dt}x(t) = f(x(t))$ $\frac{d}{dt}x(t) = g(x(t))$

In the Environment $\mathbf{E_1}$ (history) the system is described by the autonomous system

$$\frac{d}{dt}x(t) = f(x(t))$$

In the environment $\mathbf{E_2}$ (actual environment) it is

$$\frac{d}{dt}x(t) = g(x(t))$$

At time t=0, when the environment suddenly changes from $\mathbf{E_1}$ to $\mathbf{E_2}$, the adjustment is controlled by an internal process, which is a "work to be done" during adaptation. Modelling this internal process is the key to understand adaptation.

A typical example for the above framework is when the cells are under sub-lethal or lethal stress (starving, heat) and the survival fraction goes through a recovery period, then start to growth. Many laboratory growth curves show these characteristics and a suitable growth model to describe growth curves of this shape (starting with decline) can be derived similarly as shown in Chapter 5.

Over the last 10 years, a wide range of sophisticated methods has been developed to observe the growth and division of single cells. (Elfwing *et al*, 2004; Brehm-Stecher and Johnson, 2004). Consequently, more and more studies have concentrated on modelling the lag times of single cells (Prats *et al* 2008; Pin and Baranyi, 2008). Stochastic models play a major role in the new approach. Taking into account the variability of single cells is a must for risk assessors focussing on the presence or absence of a few pathogenic cells in food.

However, a healthy mechanistic approach to modelling adaptation must acquire information also on the intracellular processes during lag. Such molecular processes are being studied at the moment at the research group of the author of this thesis (http://www.ifr.ac.uk/safety/comicro/). We use network science methods to characterize the cells' physiological state by the topological properties of the cells' transcriptomic network; similarly the adaptation process would be described by the dynamic change of that network. The first attempt to characterize the physiological state by the topological properties of the cells' transcriptomic network was published in Pin *et al* (2009).

8.2 Conclusions

Predicting microbial responses to the food environment has long been on the wish-list of researchers, industrialists and regulatory officers interested in the microbial quality and safety of food (McMeekin et al, 1993). Traditionally, when a new technology or a new formulation for a food product is introduced, it is subjected to a challenge test for shelf-life and safety, a process that is expensive, time-consuming and often does not reflect the numbers of potentially pathogenic microbes in the food or the conditions of use of the product. Hence challenge tests provide only modest assurance that a

product formulation will be safe in the food chain. There has always been a question hanging over such experiments: couldn't the results be predicted/estimated, from appropriate laboratory data, as occurs in scientific literature, or from the outcome of other microbiological tests?

The key to this is mathematical modelling (McKellar and Lu, 2004). It is important to appreciate that modelling is not simply compiling a set of equations. Modelling, in some way, is the art of omitting the unnecessary. It will never be an automated process, because the extent and nature of those omissions depend on the purpose of the modelling.

The first mathematical models in food microbiology described the inactivation of pathogens at constant heating temperatures by log-linear models. Analogously to this, log-linear models were applied also to growth temperatures, when the logarithm of the microbial counts was expected to increase linearly with time. This simple law was applicable only to idealistic situations, when the environment was constant around the microbial population and the cells could grow according to their maximum ability characteristic of that environment. In practice, however, the growth phase is preceded by a certain "preparation" time. This “lag” (or "adjustment") time, which is a transition from the previous to the new environment, has become the centre of research for understandable reasons: in applications, the lag time is the very parameter the user would like to see extended as much as possible. This complication was also added to simple linear kinetics and today the non-linear growth / inactivation models are the norm rather than the exception.

It is important to see, however, that only those new non-linear approaches have received long-lasting acknowledgment, in which the classical linear kinetics can be embedded as a special case. In other words: when more complex models are introduced, it is desirable to ensure that they can accommodate the scenarios for which previously simpler models were "good enough". This evolution process is typical in science. To draw a parallel with a distant but well-known field: quantum mechanics and relativity theory both involve the classical Newton mechanics as a special case.

An interesting by-product of this evolution is that food microbiologists, in order to communicate with modellers and with each other, had to learn "to omit the unnecessary". Namely, not only did mathematical models need to be formulated, but also systematic computerisation of microbiological data (which is a pre-requisite for creating mathematical models and, ultimately, predictive software from them), database structures and "templates" had to be created in order to pool experimental results from different laboratories. The quantity of data necessitates the use of relational databases, with rigorous structure that would not just swallow any information indiscriminately. The database builder must select what pieces of information are relevant and has to omit the unnecessary here as well; otherwise the creation would be a datadump rather than a database (McMeekin et al, 2005). An example where this selection process has taken place is the ComBase database (see www.combase.cc), which is a joint effort between several international laboratories.

Interdisciplinary attitude has played a key role in the development of predictive microbiology. It is not difficult to predict that this will be true in the future, too. For this, just as food microbiologists need to learn to omit the unnecessary, we (statisticians, mathematicians, as this author) need to exercise empathy: to see not only data but also the process and the circumstances behind the data, in order to suggest what is unnecessary.

The mathematical models developed by the author, and their numerical implementation, have been overviewed in this thesis. The application area is food microbiology which was an almost exclusively empirical science until some decades ago. Partly the reason for this is that the food environment can be very complex and it may be difficult to quantify or even to categorise some of its features and/or their potential effects on microbial population dynamics or the ability to recover a target organism from a food sample. An example is food structure, which may affect environmental limits for growth, but little progress has been made to model this effect.

The process, through which modelling microbial adaptation has been going through is typical for life sciences:

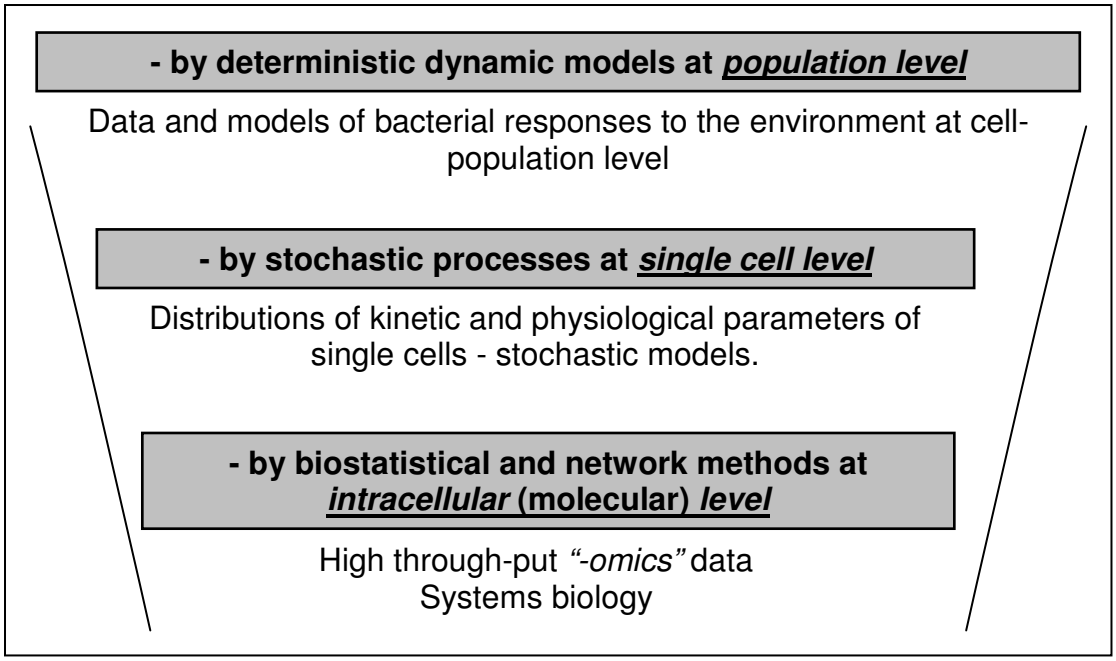


Fig. 8.1. Reductionist, followed by a systems approach.

An additional difficulty is that, with the background information on the environment and with currently available techniques to measure microbial responses, both variability and uncertainty may be large. Uncertainty also arises when information is missing or conflicting, events that regularly cause consternation in the conduct of quantitative risk assessments. In such situations, it is essential to accumulate as much information as possible.

A database pooling such information is ComBase (Combined, or Common i.e. joint, dataBase of microbial responses to food environments), which was primarily developed by the author of this thesis. Its technical details can be read in Baranyi and Tamplin (2004) and on the website, www.combase.cc .

The implementation of the model of Baranyi and Roberts (1994) in the ComBase system contributed to its popularity. Users can freely access >50,000 growth and survival curves and can fit them interactively, via the web. The stability of those fits have greatly improved by the rearrangement technique described in Chapter 3. Similarly, publicly available Risk Assessment Model is being commonly used that is based on the results of Chapter 6-7.

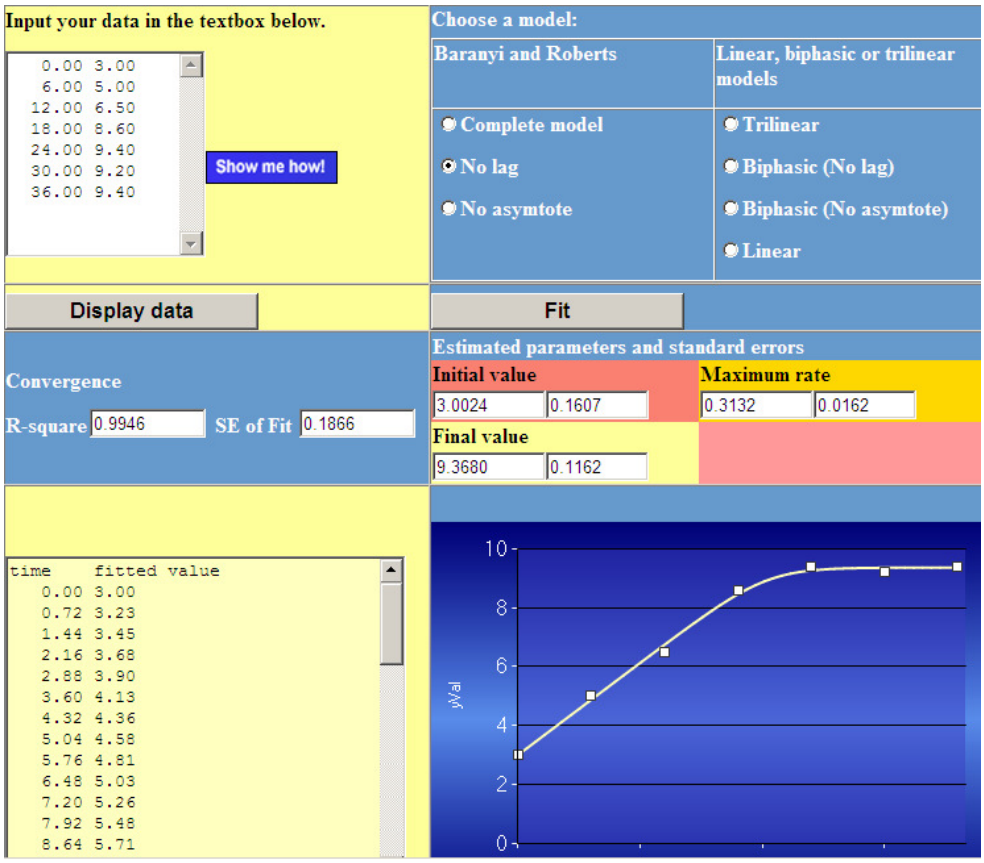


Fig. 8.2. Fitting a growth curve on www.combase.cc using the significant parameters only (in this case no lag). The non-significant parameters are omitted by means of an F-test.

Mathematical modelling techniques inevitably permeate all sciences, such as food microbiology, that was a purely descriptive and empirical science just some decades ago. The present thesis gave a snapshot of the results with which the author hoped to contribute to this process.

REFERENCES

Augustin J-C., Rosso L. and Carlier V. (1999). Estimation of temperature dependent growth rate and lag time of *Listeria monocytogenes* by optical density measurements. J. Microbiol. Methods, 38: 137-146.

Baranyi J., Roberts T.A. and McClure P.J. (1993a). A non-autonomous differential equation to model bacterial growth. *Food Microbiol. 10*, 43-59.

Baranyi J., Roberts TA. and McClure PJ. (1993b). Some properties of a non-autonomous deterministic growth model describing the adjustment of the bacterial population to a new environment. *IMA J. of Mathematics Applied in Medicine and Biology 10*, 293-299.

Baranyi J. and Roberts TA. (1994). A dynamic approach to predicting bacterial growth in food. *Int. J. Food Microbiol. 23*, 277-294.

Baranyi J. and Roberts TA. (1995). Mathematics of predictive food microbiology. *Int. J. Food Microbiol.* ***26***. 199-218.

Baranyi J., Robinson TP., Kaloti A. and Mackey BM. (1995). Predicting growth of *Brochothrix thermosphacta* at changing temperature. *Int. J. Food Microbiol. 27*. 61-75.

Baranyi J. (1998). Comparison of stochastic and deterministic concepts of bacterial lag. *J.Theor.Biol. 192*. 403-408.

Baranyi J. and Pin C. (1999). Estimating bacterial growth parameters by means of detection times. *Appl. Env. Microbiol. 65*. 732-736.

Baranyi J. and Pin C. (2001). A parallel study on modelling bacterial growth and survival curves. *J.Theor.Biol.* 210. 327-336.

Baranyi J. and Tamplin M. (2004). ComBase: A Common Database on Microbial Responses to Food Environments. J. Food Prot. 67, 1967–1971.

Baranyi J. (ed. 2005). Final report on the EU-QLRT-2000-01145: Optimisation of safe food processing methods based on accurate characterisation of bacterial lag time using analysis of variance techniques: BACANOVA.

Baranyi J., George S. and Kutalik Z. (2009). Parameter estimation for the distribution of single cell lag times. *J.Theor.Biol*. 258.

Brehm-Stecher BF and Johnson EA., 2004. Single-cell microbiology: tools, technologies, and applications. Microbiol. and Mol. Biol. Reviews, 68: 528–559.

Buchanan RL., Whiting RC. and Damert, WC. (1997). When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves. *Food Microbiol*. 14. 313-26.

Casolari A. (1988). Microbial death. Chapter 7 in: Physiological models in microbiology. Eds: Bazin, M.J. and Prosser, J.I. Vol. II. CRC Peress, Boca Raton, Fla.

Coleman BD. (1978). Nonautonomous logistic equations as models of the adjustment of populations to environmental change. *Math. Biosci.* 45. 159-173.

Elfwing A., Le Marc Y., Baranyi J., and Ballagi A., 2004. Observing the growth and division of large number of individual bacteria using image analysis. Appl. Environ. Microbiol., 70: 675–678.

Francois K., Devlieghere F., Smet K., Standaert AR., Geeraerd AH., Van Impe JF., and Debevere J., 2005. Modelling the individual cell lag phase: effect of temperature and pH on the individual cell lag distribution of *Listeria monocytogenes* Int. J. Food Microbiol.

Frederickson AG., Ramkrishna D. and Tsuhiya HM. (1967). Statistics and Dynamics of Procaryotic Cell Populations. *Math. Biosci.* 1. 327-374.

Gompertz BJ. (1825). On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies. *Phil. trans. Roy. Soc.*115. 513-85.

Grijspeerdt K. and Vanrolleghem P. A. (1999). Estimating the parameters of the Baranyi model for bacterial growth. *Food Microbiol*. 16. 593-605.

Hill AV. (1910). The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *J. Physiol.* 40.

Hermann D. and Horst J. (1970). Molecular Radiation Biology. Springer, Berlin.

Koch AL (1988). Why can't a cell grow infinitely fast? *Can.J.Microbiol.* 34. 421-426.

Körmendy I., Körmendy L. and Ferenczy A. (1998). Thermal inactivation kinetics of mixed microbial populations. A hypothesis paper. J.Food Eng. 38. 439-453.

Kutalik Z., Razaz M. and Baranyi J. (2005a). Connection between stochastic and deterministic modelling of bacterial growth. J. Theor. Biol. 232 (2), 283-297.

Kutalik Z., Razaz M., Elfwing A., Ballagi A. and Baranyi J. (2005b). Stochastic modelling of individual cell growth using flow chamber microscopy images. Int. J. Food Microbiol., 105: 177-190.

Malthus TR. (1798) An Essay on the principle of population, 1st ed., printed for J. Johnson in St Paul's Churchyard, London.

MacDonald N. (1978). Time lags in biological models. Springer. Berlin.

McClure PJ., Baranyi J., Boogard E. Kelly TM. and Roberts, TA. (1993) A predictive model for the combined effect of pH, sodium chloride and storage temperature on the growth of *Brochothrix thermosphacta*. Int. J. Food Microbiol. 19, 161-178.

McKellar, RC. and Lu, X. (eds). (2004). Modelling Microbial Responses in Foods. CRC Press, Boca Raton, Fla.

McMeekin TA., Olley J.N., Ross, T. and Ratkowsky, D.A. (1993). Predictive Microbiology. John Wiley & Sons Ltd. Chichester, UK.

McMeekin TA., Baranyi J., Zwietering M., Ross T., Dalgard P., Bowman J. and Kirk M. (2005). Information systems in food safety management. *Int. J. Food Microbiol.* 112, 181–194

Métris A., George S.A. Peck M.W. and Baranyi J. (2003). Distribution of turbidity detection times produced by single cell-generated bacterial populations. Journal of Microbiological Methods, 55, 821-827.

Métris, A., George, S.M., and Baranyi, J. (2006). Using optical density detection times to assess the effect of acetic acid on single cell kinetics Appl. Env. Microbiol. 72, 10.

Nielsen, J. and Villadsen, J. (1992). Modelling of microbial kinetics. Chem. Eng. Sci. 47. (4225-4270).

Pearl R. (1927). The growth of populations. *The Quarterly Review of Biology* II 4, 532-548.

Pin C. and Baranyi J. (2006). Kinetics of single cells: Observation and modeling of a stochastic process. Appl.Env.Microbiol. 72, 2163-2169.

Pin, C. and Baranyi J. (2008). Single-cell and population lag times as a function of cell age. *Appl. Env. Microbiol*. 74, 2534-2536.

Pin C, Rolfe MD, Muñoz-Cuevas M, Hinton JCD, Peck MW, Walton NJ and Baranyi J. (2009). Network analysis of the transcriptional pattern of young and old cells of Escherichia coli during lag phase. BMC Systems Biology, 3. 108-125.

Pirt SJ., (1975). Principles of microbe and cell cultivation. Blackwell. London.

Prats, C., Giró, A., Ferrer, J., López, D. and Vives-Rego, J., 2008. Analysis and IbM simulation of the stages in bacterial lag phase: basis for an updated definition. Journal of Theoretical Biology 252, 56-68.

Renshaw E. (1991). Modelling biological populations in space and time. Cambridge University Press. Cambridge.

Richards FJ. (1959). A flexible growth function for empirical use. *J. Experimental Botany* 10. 290-300.

Roels, JA. and Kossen, N.W.F. (1978) On the modelling of microbial metabolism. In: Progress in industrial microbiology, 14. Ed: Bull, M.J. Elsevier. Amsterdam.

Rubinow SI. (1984) Cell Kinetics. In Mathematical Models in Molecular and Cell Biology (ed L.A. Segel). Cambridge University Press, Cambridge. Chapter 6.6.

Srivastava, AK. and Volesky, B. (1990) Characterization of transient cultures of Clostridium acetobutylicum. Biotechnol. Prog. 6, 408-420.

Tsuchiya, HM., Frederickson, AG. and Aris, R., (1966). Dynamics of microbial cell populations. In: Advances in Chemical Engineering 6, 125-206.

Turner ME., Bradley E.L., Kirk KA., & Pruitt KM., (1976). A theory of growth. *Math. Biosci.* 29. 367-373

Van Impe, JF., Nicolai, BM., Martens, T., Baerdemaeker, J. and Vandewalle, J. (1992) Dynamic mathematical model to predict microbial growth and in-activation during food processing. Appl. Environ. Microbiol. 58, 2901-2909.

Vadasz P. and Vadasz AS. (2007). Biological Implications from an Autonomous Version of Baranyi and Roberts Growth Model. *Int. J. Food Microbiol*. 114. 357-365.

Vance RR., (1990). Population growth in a time-varying environment. *J. Theor. Biol.* 37. 438-454.

Vance RR. and Coddington EA., (1989). A nonautonomous model of population growth. *J. Math. Biol.* 27. 491-506.

Xiong R., Xie G., Edmondson A. S., Linton R. H. and Sheard MA. (1999a). Comparison of the Baranyi model with the modified Gompertz equation for modelling

thermal inactivation of *Listeria monocytogenes* Scott A. *Food Microbiology*, 16. 269-279.

Xiong R., Xie G., Edmondson A.E. and Sheard MA. (1999b). A mathematical model for bacterial inactivation. Int.J.Food Microbiol. 46. 45-55.

Zwietering MH., Jongenburger I., Rombouts FM., and van't Riet K. (1990). Modelling of the bacterial growth curve. *Appl. Environ. Microbiol.* 56. 1875-1881.

Declaration

The work in this thesis is based on research carried out at the Institute of Food Research UK. No part of this thesis has been submitted elsewhere for any other degree or qualification. The results are my own work unless referenced otherwise in the text.

Summary

It has long been recognized that mathematical modelling is key to predicting microbial responses to the food environment. In the 80's, with powerful desktop computing becoming everyday use, the name 'predictive microbiology' was coined for this area of mathematical food microbiology. Perhaps the Quantitative Microbial Ecology of Food would have been a more appropriate name but by now the original term has set strong roots in the literature. The closest relative to the field is probably biotechnology, but because of the differences detailed in the Introduction, it was necessary to develop new models specifically for applications in food microbiology.

In this thesis we showed a process of model development, from deterministic models of population growth of bacteria, with special attention to the bacterial lag, to their stochastic models. We analysed the differences between growth- and survival modelling. We demonstrated how to link the population (deterministic) and single cell (stochastic) models. We developed several numerical algorithms to implement the models for practical applications.

Deterministic modelling of lag

In Chapters 1-2 we pointed out that the classical autonomous growth models (Logistic, Gompertz, etc) are not suitable for lag time modelling. We introduced a non-autonomous model, where an $\alpha(t)$ factor, a so-called adjustment function describes the transition from the lag to the exponential phase. We showed the connection between the non-autonomous model and its autonomous counterpart. Two classes of adjustment functions: that of Hill's and of Michaelis-Menten was analysed in detail.

Numerical considerations

The Michaelis-Menten adjustment function turned out to be more user-friendly and stable for numerical applications as shown in Chapter 3. A by-product was a versatile sigmoid function with six parameters, quantifying the initial value (y_0); the maximum specific growth rate (μ_{max}); the length of the lag period (λ); the maximum carrying capacity (y_{max}) of the environment; and the abruptness of the two transition phases:

from the lag to the exponential phase (n_C) and from the transition to the stationary phase (m).
The model proved to be useful for general parameter-by-parameter curve fitting, too, because the parameters λ and y_{max} can be omitted if data do not justify their use, thus reduce the complexity of the model. For such regression, an F-test can be used to control the elimination, similarly to the significance analysis of the coefficients when fitting polynomials of higher degrees.

Biological interpretation and experimental validation

An important point discussed was that the lag depends on an initial physiological state (α_0) of the culture via $\lambda=-\ln(\alpha_0)/\mu$. The α_0 value is an initial value, just like the inoculum size. It depends on the difference between the history and the current environment. Instead of the lag as a delay in time, we suggested that the delay measured by delays in generations (h_0) is a more convenient concept to characterise adaptation.

After studying large datasets, we established that the h_0 parameter is relatively constant for various environmental conditions. We interpreted this as the manifestation that h_0 is in fact an initial value at the time zero, a link from the history to the current growth environment. By means of this thought and assuming that the maximum specific rate as well as the rate of adjustment instantaneously responds to changing temperatures, we successfully predicted the growth of a spoilage organism even when the temperature continuously changed during the lag time.

Stochastic modelling of lag

In Chapter 4, we showed that the commonly used definition of the lag is less and less compatible with its physiological concept as the initial population decreases. We focussed on bi-phasic growth models where the duration of the first phase (lag, denoted by L) is random; so is the $\alpha_0=e^{-\mu L}$ physiological state parameter. Our most important theorem was that, due to the subsequent exponential growth, the physiological state parameter of the population converges to the mean of the individual cells' physiological state parameter, as the initial population size increases.

This is a basic theorem in this thesis, with easily interpretable consequences and applications:

- The variance of the population's physiological state can be explicitly calculated;
- The lag of the population is a non-arithmetical mean of the lag of individual cells, in which relationship the subsequent specific rate plays a crucial role.
- The higher the specific growth rate, the closer the population lag to the minimum of the individual cells' lag times.

Comparison between growth and survival modelling

The main questions answered in Chapter 5 were: What is the relationship between the distribution of individual survival times and the survival curve of the population; the distribution of individual lag times and the growth curve of the population. We were particularly interested in developing formulae between the parameters of the distribution functions of individual survival/lag times and the parameters of the survival/growth curves of the population. How extensive is the analogy between the two seemingly perfect inverse problems?

We derived explicit, useful formulae for the problems above. We pointed out that, though there is a one-to-one map between the distribution of the individual lag times and the adaptation period (and similar mapping is valid in the survival case), still in practice, it is not feasible to identify the distribution of the individual lag times from traditional viable count growth curves. Again, we obtained in a different way that the physiological state parameter of the population is a mean value of the individual α_0 values.

An important special case was analysed when the individual lag times followed the gamma distribution. Explicit formulae were derived for the mean and the variance of the physiological state of the population, which later played a significant role in the numerical applications. As an important by-product of our analysis, we gave an elementary proof that for a Poisson birth process, the doubling time of the population is smaller, by an ln2 factor, than the average generation time.

Estimating the distribution parameters of individual lag times from measurable information

Automated measures cannot provide direct information on the lag phase of individual cells because measurable properties of a culture (turbidity, conductance, *etc*.) are detectable only at high cell concentrations.

In Chapter 6 we made good use of the relationships, we obtained for the physiological state parameter, to solve the above problem. We developed a new method, based on the physiological state theorem and an ANOVA procedure, to measure the maximum specific growth rate and the lag time distribution of single cells. The advantage of the method is that it uses detection times, which are the first data available when recording bacterial growth. An important point was made that, though the physiological state theorem is valid irrespective of the distribution of the lag times of the individual cells, in the case, when those follow the gamma distribution, the explicit formulae for the statistical parameters made it possible to do the calculations in a simple spreadsheet model. We noted that, by means of a so-called flow-chamber technique, developed specifically for this purpose in the research group I lead, it was possible to validate the Gamma-assumption.

Further numerical developments

The disadvantage of the methods discussed so far was that we could reliably estimate only the variance of the distribution of single cell lag times, not its shape. In Chapter 7 we developed new formulae for the situation when the initial cell number at the highest dilution is small and random. A great advantage of the procedure that it assumes Poisson distribution for the inoculula that can be achieved by serial dilutions.

Our method was based on the fact that moments of the sum of random variables can be calculated explicitly in those cases, when the number of terms in the sum is also random, namely Poisson-distributed. The moments-based method was validated by the maximum-likelihood method. The results were similar but our method was superior in terms of speed and computational resources. This made it possible to

build it into user-friendly spreadsheet applications, which are the most commonly used in food microbiology laboratories.

Future research

We discussed some practical implications of our findings regarding Quantitative Microbial Risk Assessment. We pointed out that a healthy mechanistic approach to modelling adaptation must gain information on the intracellular processes during lag. For this Pin et al (2009) is an example that reports on the results of a project completed recently in the research group of the author of resent thesis. There, the physiological state was characterised by the topology of the actual genetic (more exactly transcriptomic) network, among others, by methods used in network science. Another field, on which big emphasis has been put recently by the very development of Network science is the sistematicus development of databases, for which a good example is the ComBase system (www.combase.cc). Its predictive numerical tools are based on the results reviewed in this thesis and their future development is planned to follow the directions shown here.

Összefoglaló

Régóta elismert hogy az élelmiszer-mikrobiológia fejlődéséhez is elengedhetetlen új matematikai modellek alkotása. A 80-as években, ahogy egyre hatékonyabbá és hétköznapibbá váltak a személyi számítógépek, a 'prediktív mikrobiológia' nevet kapta a matematikai élelmiszer-mikrobiológia ezen ága. Talán az "élelmiszer kvantitatív mikrobiológiai ökológiája" név helyesebb lett volna, ám mára az eredeti elnevezés gyökeret vert az irodalomban. A legközelebbi rokon tudományág valószínűleg biotechnológia, de a különbségek amelyeket a Bevezetésben részletezünk szükségessé tették hogy matematikai modelleket fejlesszünk külön csak élelmiszer-mikrobiológiai alkalmazásokra.

A jelen disszertációban egy modellalkotási folyamatról számoltunk be, a baktériumpopuláció determinisztikus modelljeitől - különleges figyelemmel a lappangási időre - a sztochasztikus modellekig. Elemeztük a növekedés és pusztulás modellezésének hasonlóságait és különbségeit, demonstráltuk hogyan kapcsolódnak egymáshoz a populációs determinisztikus és az egyéni sejtek sztochasztikus modelljei. Számos numerikus algoritmust dolgoztunk ki a modellek megbízható gyakorlati alkalmazására.

A lappangási idő determinisztikus modellezése

Az első két fejezetben rámutattunk hogy a klasszikus autonóm növekedési modellek (Logisztikus, Gompertz, etc) nem alkalmasak a lappangási idő leírására. Bevezettünk egy nem-autonóm modellt, ahol egy $\alpha(t)$ faktor, ún. adaptációs függvény írja le az átmenetet a lappangási fázisból az exponenciális fázisba. Kielemeztük a nem-autonóm modell és autonóm párja között a kapcsolatot. Az adaptációs függvények két osztályát elemeztük részletesebben: a Hill és a Michaelis-Menten függvényeket.

Numerikus meggondolások

Numerikus szempontból a Michaelis-Menten adaptációs függvény bizonyult a legkezelhetőbbnek, mint az a 3-dik fejezetből kiderül. Mellékeredményként egy gazdag felhasználhatóságú szigmoid függvényt is kaptunk, hat paraméterrel, amely jellemezte a kezdeti szintet (y_0); a maximális érintőt (μ_{max}); a lappangási időt (λ); a

környezet kapacitását (y_{max}); valamint a két fázisátmenet gyorsaságát: a lappangási fázisból az exponenciális fázisba (n_C); s az exponenciális fázisból a stacionárius fázisba (m). A modell hasznosnak bizonyult általános, fokozatos görbeillesztésre is, mert a λ és y_{max} paraméterek elhagyhatók ha az illesztendő adatok nem igazolják használatukat, így a modell komplexitása igénynek megfelelően csökkenthető. Az ilyen regresszióra F-teszt használható hogy a paraméterek elhagyhatóságát eldöntse, hasonlóan ahhoz amikor magasabbrendű polinomok illesztésénél az együtthatók szignifikanciáját vizsgáljuk.

Biológiai értelmezés és kísérleti igazolás

Lényeges megállapítás volt, hogy a lappangási idő egy kezdeti értéktől függ (α_0) ami a (kezdeti) fiziológai állapotot kvantifikálja, s így $\lambda=-\ln(\alpha_0)/\mu$. Az α_0 épp úgy kezdeti érték mint az inokulum nagysága. Valójában az inokuláció előtti és a jelenlegi környezet különbségét jellemzi. Rámutattunk, hogy a lappangás miatti késleltetést célszerűbb azzal mérni hogy hány (h_0) generációval van lemaradva a növekedés ahhoz képest mintha a múlt és a jelen környezet megegyezne.

Megfelelő, nagyszámú adathalmazok tanulmányozása után arra jutottunk, hogy a h_0 paraméter a különféle környezetektől viszonylag független konstans. Ezt úgy értelmeztük hogy h_0 valóban egy kezdeti értéket fejez ki, ami előírt laboratóriumi méréseknél kb ugyanaz. Ezzel, valamint a feltevéssel hogy a maximális specifikus növekedési sebesség és az igazodási ráta késlekedés nélkül felveszi az éppen aktuális környezetre jellemző értéket, sikeresen megjósoltuk egy romlást okozó baktérium növekedését időben változó hőmérséklet mellett, holott a hőmérséklet a lappangási idő alatt is változott.

A lappangási idő sztochasztikus modellezése

A 4-dik fejezetben megmutattuk hogy a lappangási idő tradicionális értelmezése egyre kevésbé összeegyeztethető a fiziológiai értelmezéssel ahogy a kezdeti populációs nagyság csökken. A lappangási idő definíciójára kétfázisú lineáris függvény első fázisát használtuk (L) mint véletlen változót. Ezzel a $\alpha_0=e^{-\mu L}$ fiziológiai állapot is véletlen változó. Legfontosabb tételünk az volt, hogy a rákövetkező exponenciális növekedés miatt, a populáció fiziológiai állapotát jellemző paraméter az egyéni sejtek

hasonló paraméterének átlagához tart amint a kezdeti populációt nagyobbnak választjuk. Ez az egyik alaptétele a disszertációnak, könnyen értelmezhető következményekkel és alkalmazásokkal:

- A populáció fiziológiai állapotának varianciája zárt formában kiszámítható az egyéni sejtek lappangási idejének eloszlásából.
- A populáció lappangási ideje az egyéni sejtek hasonló paraméterének egy nem-aritmetikai átlaga, s az összefüggés képletében a μ specifikus ráta játsza a fő szerepet
- Minél magasabb a specifikus növekedési sebesség, annál közelebb lesz a populációs lag az egyéni sejtek lappangási idejének minimumához.

Pusztulási és növekedési modellek összehasonlítása

Az 5-dik fejezet megválaszolandó kérdése az volt, hogy vajon milyen kapcsolat van az egyéni sejtek tulélési idejének eloszlása és a populáció pusztulási görbéje valamint az egyéni sejtek lappangási idejének eloszlása és a populáció növekedési görbéje között. Különösen az érdekelt bennünket, milyen formulák vezethetők le az egyéni túlélési/lappangási idők eloszlásainak paraméterei és a populáció túlélési/növekedési görbéinek paraméterei között. Mennyire húzható párhuzam a két látszólag teljesen inverz probléma között?

Explicit, könnyen kezelhető formulákat vezettünk le a fenti problémára. Rámutattunk, hogy egy-egy értelmű leképezés létezik az egyéni lappangási idők eloszlása és az adaptáció lefolyása között (ugyanígy a túlélési esetben), mégis a gyakorlatban az egyéni sejtekre vonatkozó eloszlás nem vezethető le a populáció adaptációs szakaszából. Ismét eljutottunk, most egy másik úton, ahhoz a tételhez, hogy a populáció fiziológiai állapota az egyéni sejtek fiziológiai állapotának várható értéke.

Fontos speciális esetként tárgyaltuk azt amikor az egyéni lappangási idők gamma eloszlást követtek. Explicit formulákat vezettünk le a populáció fiziológiai állapotának várható értékére és varianciájára, amely később nagy szerepet játszott numerikus alkalmazásokban. Mint mellékeredmény, elemi bizonyítást adtunk arra

vonatkozólag, hogy a Poisson-féle születési folyamat esetén a populáció duplázási ideje ln2 faktorral kisebb mint az átlagos generációs idő.

Egyéni lappangási idők eloszlás-paramétereinek becslése hozzáférhető mérésekből
Automatizált mérések nem alkalmasak arra hogy egyéni sejtek lappangási idejére vonatkozó információt közvetlen úton nyerjünk, mert a baktériumtenyészet mérhető tulajdonságai (turbiditás, vezetőképesség, *etc*.) csak magasabb koncentráció esetén jelentkeznek.

A 6-dik fejezetben ennek a problémának a megoldására használtuk fel a fiziológiai állapotra kidolgozott összefüggéseket. Kidolgoztunk egy, a fiziológiai állapotra vonatkozó tételen és ANOVA eljáráson alapuló numerikus módszert, amivel a maximális növekedési sebesség és lappangási idők eloszlása megállapítható. A módszer előnye, hogy detektációs időket használ, amely az első mérhető adat a baktériumtenyészet vizsgálata közben. Rámutattunk hogy bár a fiziológiai állapotra vonatkozó tétel az egyéni sejtek lappangási idejének eloszlásától függetlenül érvényes, amikor azok gamma eloszlást követnek az explicit formulák lehetővé teszik a mérések feldolgozását egyszerű táblázatkezelő programokkal. Beszámoltunk arról hogy egy technológia (ún "flow-cell"), amit munkatársaim speciálisan erre a problémára fejlesztettek ki, lehetővé tette a Gamma-hipotézis ellenőrzését.

További numerikus fejlesztések
Az egyéni sejtekre eddig ismertetett numerikus módszerek hátránya volt hogy feltételezte, hogy a kémcsőben vizsgált baktériumtenyészet egy sejtből indult. A 7-dik fejezetben formulákat dolgoztunk ki arra az esetre amikor ez a szám kicsi és véletlen. Nagy előnye az eljárásnak, hogy Poisson-eloszlást tételez fel a kezdeti sejtszámra, ami egyszerű hígítási sorozat végén kapható.

A módszer azon alapult hogy véletlen változók véletlen tagszámú összegének momentumai explicit módon kiszámíthatók, ha a tagszám Poisson eloszlást követ. A momentumon alapuló módszert a maximum-likelihood módszerrel összehasonlítva igazoltuk gyakorlati számításokra. A mi módszerünk előnye a sebesség és az alkalmazhatóság egyszerű, felhasználó-barát, táblázatkezelő programokban, amelyek élelmiszer-mikrobiológiai laboratóriumok leggyakrabban használt számítási eszközei.

Jövőbeni kutatás

Eredményeink gyakorlati következményeire mutattunk rá ebben a fejezetben, különös tekintettel az élelmiszerek mikrobiológiai kockázatának kvantitatív módszereit illetően. Rámutattunk, hogy továbbfejlődés lehetetlen anélkül, hogy ne elemeznénk intracelluláris kvantitatív modellekkel a sejtek fiziológiai állapotát. Erre példa Pin et al (2009), amely egy, a jelen disszertáció szerzőjenek kutatócsoportjában befejezett projekt eredményéről számol be. Ott a fiziológiai állapotot az aktuális genetikai (még pontosabban, transzkripciós) hálózat topológiajával jellemeztük, többek között hálózattudományban használatos módszerekkel.

Másik nagy, éppen a hálózattudomány által egyre nagyobb hangsúlyt kapott terület a szisztematikus adatbázis építés. Erre jó példa a ComBase rendszer (www.combase.cc), amelynek prediktív numerikus eszközei az ebben a dolgozatban ismertetett eredményeken alapulnak, s fejlesztésük is az itt bemutatott iráanyvonalat követik.