CWI Tracts

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O).

CWI Tract 58

A random model for plant cell population growth

M.C.M. de Gunst

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: Primary: 60G55, 60F17. Secondary: 62P10, 62J10.
ISBN 90 6196 365 6
NUGI-code: 811

Printed in the Netherlands

ACKNOWLEDGEMENTS

I wish to thank Prof.dr. W.R. van Zwet and Prof.dr. K.R. Libbenga for their contributions to the preparation of this monograph. I am indebted to Dr. P.A.A. Harkes, E. de Graaff and the people of the Plant Cell Group of the Department of Biochemical Engineering of the Technical University of Delft for their assistence with the experiments.

The simulations and numerical computations in Section 3.5 were performed by J. Val; K. in 't Hout and E. Pieters took care of the numerical and computational aspects of the least squares estimation and the corresponding figures in Chapter 6. The figures were drawn by G.P.G. Hock.

Finally, I thank the Centre of Mathematics and Computer Science for the opportunity to publish this monograph in their series CWI-tracts and all those who have contributed to its realization.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1. RANDOM CELL CYCLE MODELS FOR PLANT CELLS; A BRIEF REVIEW

The use of mathematical models for the description of biological events has often turned out to be fruitful for both biology and mathematics. It has frequently led to a better understanding of the underlying biological mechanisms on the one hand and to the development of mathematical theories on the other. For a long time now, the growth of a population by division of its individuals has been considered as a biological random process and mathematical random processes involving random generation times are used to describe the phenomenon of division and growth. For instance KELLY & RAHN (1932) introduced the concept of chance in relation to the generation time of bacteria and KENDALL (1948; 1952) used the data of these authors to illustrate the mathematical properties of a stochastic birth process with variable generation times; POWELL (1955) continued this discussion from the biological point of view and HARRIS (1959) referred to both KELLY & RAHN (1932) and POWELL (1955) when he proposed his mathematical model for binary fission.

After the concept of the cell cycle was introduced in HOWARD & PELC (1953), many papers on cycle times of animal, yeast and plant cells have appeared. In studies of animal and yeast cells the generation or cycle time is generally considered as a random variable having a probability density function which appears to be skewed to the right (see KILLANDER & ZETTERBERG (1965), SISKEN & MORASCA (1965), KUBITSCHEK (1971) among others). However, most studies of plant cell cycles merely determine sample means and sometimes also sample standard deviations of the duration of the cell cycle and its constituent phases in order to estimate the real durations (see for example WIMBER (1960), VAN'T HOF (1965), ERIKSSON (1967), FRIEDBERG & DAVIDSON (1970), EVANS & REES (1971)). It is not clear whether these durations were considered as (unknown) constants or as random variables assuming different values with certain probabilities.

CLOWES (1961) and THOMPSON & CLOWES (1968) are sometimes mentioned as the first papers in which the variability of plant cell cycles is established. However, the variability these papers deal with is certainly of a different kind than that considered in for instance KILLANDER & ZETTERBERG (1965) and here. Indeed, CLOWES (1961) and THOMPSON & CLOWES (1968) show that cells in different regions of the root meristem have different mean cycle times, whereas the variability we are concerned with is that of cycle times of plant cells within one region or, more generally, of plant cells under one set of equal external conditions.

Exceptions to this non-random approach in plant cell studies are BARLOW & MACDONALD (1973) and BAYLISS (1976). In the first paper curves of the fraction of labelled mitoses of pulse labelling experiments on the root meristem of *Zea mays* are analysed using a theoretical analysis developed in MACDONALD (1970). This method also provides estimates for population means and standard deviations, but the estimation is based on the assumption that the cycle time is a random variable. BAYLISS (1976) showed by means of a probit transformation of frequencies of cells of the initially predominant ploidy level at different times during colchicine treatment, that a normal distribution could quite accurately describe the distribution of cycle times of *Daucus carota* cells in suspension cultures. However, since the observed cells were growing asynchronously at the time the colchicine was added, the frequencies used for the calculations reflect not only variability in cycle times, but also in 'starting points'.

In 1973, when SMITH & MARTIN proposed their transition probability model (t.p.m.) for the cell cycle, attention was focussed on the variability concept again. The t.p.m. postulates that the cell cycle consists of two phases, A and B . Some time after mitosis a cell enters state A , which is thus positioned in the G_1-phase of the cell cycle. In the A-state a cell is not progressing towards division. The probability per unit time of leaving the A-state, the so-called transition probability P , is constant, which implies that a cell may remain in state A for any length of time. After leaving the A-state a cell goes into state B , where the normal proliferative events occur; a cell's sojourn in B-state is of (nearly) constant duration. This model originated from the fact that the duration of the G_1-phase was always observed to be more variable than that of the other phases of the cell cycle (S, G_2, M). No biological interpretation of the random transition from A to B was given by SMITH & MARTIN (1973). Still the t.p.m. quickly became quite popular. It could be tested easily by means of the so-called α-curve, a plot of the logarithm of the proportion of not yet divided cells against time. Provided that all cells start at time $t = 0$ at the beginning of their cycle, the graph must be constant on a time interval starting at time $t = 0$, and decline linearly afterwards with slope $-\lambda = \log(1 - P)$.

Although many publications on the t.p.m. and analyses of α-curves have appeared for animal and yeast cells, (e.g. BROOKS (1976), ROBINSON et al. (1976), SHIELDS & SMITH (1977), JIMENEZ DE ASUA et al. (1977), SHIELDS et al. (1978)), only few botanists have related the transition probability concept to the plant cell cycle (GOULD, 1977; EVANS & VAN'T HOF, 1975; WEBSTER, 1979). The only analysis of α-curves for plant cells can be found in KING (1980). Here the data from BAYLISS (1976) are used and found not to contradict the t.p.m.

In the meantime there has been much discussion on the validity and use of the t.p.m. The main objection seems to be the lack of a biological interpretation of the random transition; also the use of α-curves has been criticized (see NELSON & GREEN (1981)). Several alternative models have been proposed, for example a two-transition model (BROOKS et al., 1980), a G_1-rate model (CASTOR, 1980) and a continuum model (COOPER, 1982). Detailed reviews of other approaches can be found in EISEN (1979), BERTUZZI et al. (1981), BROOKS (1981) and BERTUZZI & GANDOLFI (1983). Again botanical literature is strikingly absent.

1.2. POPULATION SIZE MODELS

The small number of investigations of the plant cell cycle from a stochastic point of view is probably due to the fact that there is no way to observe individual plant cells over sufficiently long periods. Therefore inferences must be based on samples from the population as a whole.

Indeed, in many studies on the regulation of cell division, cell growth, etc. of plant cells, cell suspension cultures are used and measurements are taken from the total population. Growth curves of all kinds of population characteristics are the result of this. In the present study the characteristic of interest is the population size. The models most frequently used to describe the increase of cell number are deterministic population size models of the following form.

$$\frac{dN}{dt}(t) = \mu N(t) \quad ,$$

where $N(t)$ denotes the number of cells at time t and μ is the growth rate, which is generally thought to depend on some substrate concentration S :

$$\mu = \mu_{max} \frac{S}{S + k_s} \quad .$$

Here μ_{max} is the maximum growth rate and k_s equals the substrate concentration for which the growth is half of its maximum. This type of model is due to Monod and is related to the Michaelis-Menten kinetics (see e.g. ROELS (1983)). A survey of population size models (deterministic and stochastic) can be found in EISEN (1979).

1.3. CHOICE OF MODEL TYPE

Why then this insistence on random cell cycle models, which are formulated on the level of the individual cell, when we can only expect observations on the population level?

Since the intention of our study is not only to provide an adequate description of plant cell population growth, but also to understand the population growth in terms of the individual cell behaviour, the model we choose should arise naturally from hypotheses concerning the underlying biological mechanisms. In particular, every model which attempts to relate the population behaviour to underlying mechanisms should start with the description of the behaviour of the population members. This individual behaviour should be related as much as possible, given present knowledge, to processes which occur within the individual. Only then will the model enable us to make predictions concerning the properties of the biological processes of interest and to suggest where experiments can help to distinguish between different hypotheses. This is why we formulate a cell cycle model for the individual cell behaviour in which the duration of the cycle depends on internal events.

Moreover, plant cells under equal external conditions do not all act in the same way. Especially the processes we consider, namely growth processes of plant cell populations in batch suspension, which are the result of the division of the individual cells, show great variability. In such circumstances stochastic models to describe the situation are, in contrast to deterministic ones, known to be very effective. To quote JAGERS

(1983): "...only a probabilistic description can catch the regularities yet allow the actual variation". Thus, the choice of a random model instead of a deterministic one is obvious in our case.

Having formulated the random cell cycle model, we can derive the growth of the population. Indeed, a model for the cell cycle informs us not only about the mechanisms which regulate the individual cell behaviour, but also about those which control the population behaviour. For, obviously, the population behaviour is the result of what its members do, and not the other way round.

Often it turns out that, if either the number of individuals or time tends to infinity (practically speaking: is very large), the random population behaviour, as predicted by the stochastic cell cycle model, converges to a deterministic limit. Only when this is the case, will the use of a deterministic model to describe population growth be justified. Otherwise, the stochasticity in the population behaviour is essential and a random model to describe it should be used. However, even in the case where the limiting population behaviour is deterministic, we know much more when we start with the random individual cell cycle model and derive the correct limiting population size model from it, than when we start with this limiting model right away: we know *why* it is adequate.

1.4. OUTLINE OF THIS STUDY

The research which led to the writing of this study started because some specific problems, which seemed to be of general interest for the knowledge of the regulation of cell division, cell differentiation and cell growth, could not be solved by means of the available techniques. First, (see Section 2.1 for an explanation of the terminology) in studies of the growth of a batch culture of plant cells, a lag phase was observed of which the nature was obscure. Next, the character and explicit role of the rate limiting factors for this growth, for which the medium components sucrose and hormone were thought to be the main candidates, appeared to need clarification.

These two problems are certainly not restricted to plant cell population growth, but also extend to the growth of cultures of microorganisms which are frequently used in biotechnology. Plant cells in culture, however, have the peculiar ability to differentiate, presumably depending on the hormone level in the medium. It seemed desirable to go into the subject of the presence of two types of cells and their interrelationship.

To gain some insight into these problems, we shall start with the formulation of a random cell cycle model, which combines the ideas of the transition probability model of Smith & Martin (see Section 1.1) and the Monod-population size model (see Section 1.2), and which is based on the biological facts already known for our specific experimental system (see Section 2.1). Next, we shall consider the consequences of this model for the behaviour of the total population. At the same time we shall derive some mathematical results which may be of independent interest. Finally, we shall discuss some further experiments, which were suggested by the comparison of the model population behaviour with already available experimental data.

The order in which these topics are treated in this study is as follows.

In Chapter 2 a detailed description of the model is given. Section 2.1 describes the biological situation, gives the main idea of the model and explains the biological grounds for choosing just this model. In Section 2.2 the model is formulated in mathematical terms.

The consequences of the model for the total population behaviour is considered in Chapter 3 . The first section of this chapter is devoted to the derivation of several probability inequalities, which enable us in Section 3.2 to establish the convergence of the stochastic population behaviour to a deterministic limit behaviour when the initial cell number of the population tends to infinity. The rate of this convergence is also determined. In Section 3.3 a central limit theorem is derived: it is proved that the difference between the random population number and its deterministic limit, multiplied by a suitable power of the initial cell number, has, again for the initial cell number tending to infinity, a limiting distribution which is the same as that of a certain Gaussian process. In Section 3.4 we determine the asymptotic behaviour of the time at which the population stops growing, in terms of the initial cell number. In the final section of the chapter the stochastic population behaviour and the deterministic limit behaviour are compared graphically.

Some of the asymptotic results derived in Chapter 3 can also be obtained by the methods developed in KURTZ (1981; 1983) and ETHIER & KURTZ (1986) for general population processes. This we do in Chapter 4 .

Ways to obtain processes with a known limiting distribution different from the one in Section 3.3 are described in Chapter 5 . In Section 5.1 this is done under the assumption that for each division it is known to which of the two types of cells mentioned above, each of the two originating cells belongs. In the remainder of the chapter we drop this assumption and provide an approximation for the stochastic process which determines the rate of the stochastic growth of the process, the so-called intensity process. This yields yet another limiting distribution result.

In Chapter 6 experiments which were performed in order to compare our model with actual data, are described. The results are discussed and statistically analysed. We conclude with some general considerations on the consequences of this study.

CHAPTER 2

THE MODEL

This chapter is divided into two parts. In the first part the model is explained from a biological point of view; in the second the model is formulated in mathematical terms.

2.1. DESCRIPTION OF BIOLOGICAL SITUATION AND MODEL

The phenomenon which is studied in this monograph is the growth of a batch culture of plant cells. In particular we are interested in the role substrate (sucrose) and hormone play during this growth.

What is 'a batch culture of plant cells' ? By this we mean a culture of isolated plant cells or very small cell aggregates of for example *Nicotiana* (tobacco), *Cinchona*, *Catharanthus*, etc., which remain dispersed as they grow in a liquid medium. Usually the cells grow in an erlenmeyer flask which is continuously shaken, or in a fermentor in which the medium is continuously stirred. In contrast to a continuous culture, a batch culture neither has an inflow of fresh medium, nor an outflow of culture (STREET, 1973).

At some time, which we shall call time $t = 0$, the cells, say n in number,' are transferred to a fresh medium of known composition. After a certain time period, the so-called lag phase, the first cell divisions are observed. The number of cells, in the sequel denoted by N_n , is seen to increase; slowly in the beginning, then more quickly, then slowly again, until the population seems to have stopped growing: the cell population is in its stationary phase. When the cells are transferred to a fresh medium again, this growth pattern or growth cycle as it is frequently called, repeats itself. In Fig. 2.1 a typical growth curve of a plant cell population is shown. Usually a free-hand curve is drawn through the, generally few, data points, ignoring the hump at the beginning which is mostly ascribed to measurement error (see Fig. 2.1).

As explained in Chapter 1 , we formulate a model for the behaviour of the individual cell and consider the consequences of this model for the population behaviour. The main features of the model are as follows.

The idea is that a plant cell population in suspension culture consists of two types of cells: actively cycling cells, called type-A cells, and 'resting' or differentiating cells, called type-B cells. At time $t = 0$, i.e. just after the transfer to a fresh medium, the population is assumed to consist of type-A cells only. A random time after its birth a type-A cell receives an endogenous stimulus. A fixed time c later it will divide. Thus, the A-cells can be divided into two classes, to wit A^0-cells which are waiting for a stimulus, and A^1-cells which have already received a stimulus but have not yet divided.

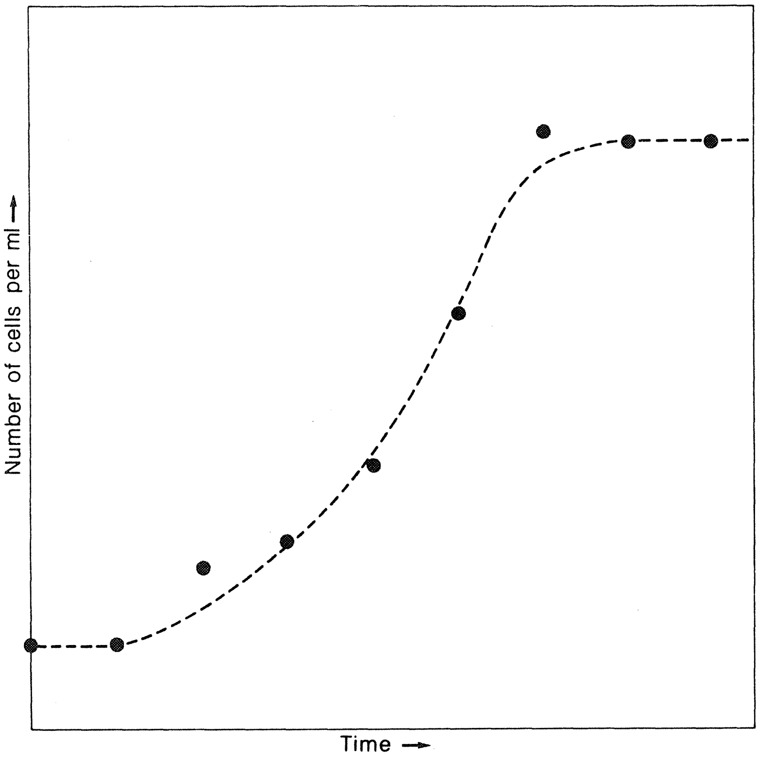


Fig.2.1. Growth curve of a batch culture of plant cells.
● ● ● data; — — — smoothed version.

Since a waiting A^0-cell receives a stimulus at a random time after its birth, the time between birth and division differs from cell to cell. The length of the time period between birth and receipt of stimulus is thought to depend on the substrate concentration in the medium: the higher the concentration, the shorter the mean length. It is assumed that a fixed amount (y_s^{-1}) of substrate is consumed when a cell receives a stimulus, and as a result this mean length becomes longer when time progresses. When all substrate is used up, no further stimuli will be received, so that there will be no more divisions.

The amount of hormone is not thought to influence the length of the cell cycle, but it determines whether a cell, just being born, will be a cycling A-cell or not: depending on the hormone concentration at the time a cell divides, an originating cell will be a cycling cell or will differentiate. The higher the hormone concentration, the more A-cells will be born. Also the amount of hormone is supposed to decrease with fixed amounts (y_h^{-1}), but now at each division. Hence when time goes on, relatively fewer A-cells are produced. When the hormone has run out, only B-cells are born.

When the increase of the total cell number N_n as predicted by this model is simulated (see Section 3.5 , Fig. 3.1), the growth pattern looks like the experimentally observed one in Fig. 2.1 indeed. First, we have a lag phase, because cells, even if they receive a stimulus immediately at $t = 0$, need at least the fixed time c before division. Next, we see a continuously changing growth rate caused mainly by the increase in cell number in the beginning and by the decrease of substrate and hormone concentration later on. In the end the growth process of the model stops, either because the substrate has run out or because the A-cells have become extinct. The simulated curves only differ from the smooth one in Fig. 2.1 in that the former show a somewhat stepwise increase at the start, which fades away gradually when time goes on. This indicates that the hump in the first data after the lag-phase in Fig. 2.1 is probably not due to measurement error, but represents a very real phenomenon. This will be discussed further in Chapter 6 .

Is this model realistic? Observations of small samples from the cultures under a microscope do indeed indicate the presence of two types of cells: some cells are larger and more stretched out than others (see Fig. 2.2), which points to differentiation; the number of these blown up cells is seen to increase with time. Unfortunately, we do not have an adequate differentiation marker yet, but secondary product formation, which also increases with time (KING & STREET, 1973), seems to be coupled to differentiation. On the other hand, it is experimentally established that cells which are kept in stationary phase for some time, have to be transferred to a fresh medium in order to induce dedifferentiation prior to

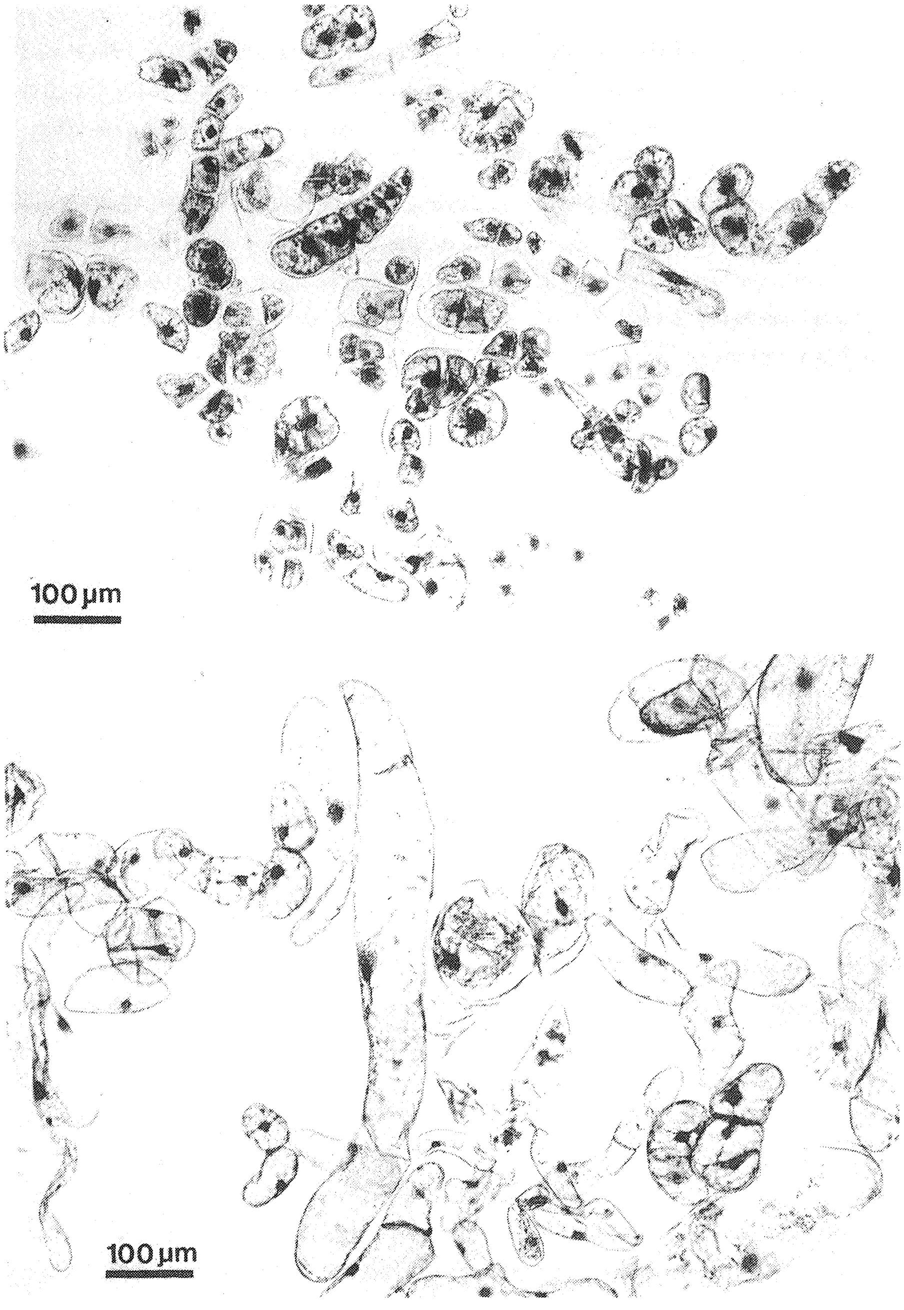


Fig.2.2. Different types of plant cells in batch culture (tobacco). Top: sample of cells drawn from a culture 68 h after inoculation showing active growth by cell division (A-cells). Bottom: sample drawn at 188 h after inoculation showing growth by cell enlargement during stationary phase (B-cells).

cell division resumption (KING & STREET, 1973). Transfer to a fresh medium perhaps brings about the same alterations in differentiated cells as injury in intact plants (ALBERTS et al., 1983). This is why we assume that at $t = 0$ all cells are of type-A.

The choice for a random cycle duration is prompted by the observed fact that 'synchronized' cell populations always loose their synchrony within a few cycles (KING & STREET, 1973). The endogenous stimulus we think of in this context, might be the production of an unstable trigger protein with a high turnover rate (cf. ALBERTS et al., 1983). We assume that it is provided at a variable time after a cell's birth, and that a fixed time (c) later the cell divides. This is in accordance with the observation that the variability of the last part of the cell cycle (S, G_2, M) is less than that of the first part (G_1) (see e.g. GOULD et al. (1974) and BAYLISS (1975)). Another reason is that many cell cycle studies (also for animal cells) mention a major 'point of no return' or 'control point' in the cell cycle somewhere in G_1 (PARDEE, 1974; WADA et al., 1984; BAYLISS, 1985; VAN'T HOF, 1985).

Nutrient depletion is a well known phenomenon too (STREET, 1973). For convenience we take the simplest way of nutrient consumption, namely, for both substrate and hormone we telescope the process of nutrient intake into one crucial event and we suppose that a fixed quantity is used up at a certain point in the division cycle. If, instead, it would be assumed that the consumption is spread out over the cycle, then the level of nutrient concentration would be somewhat different throughout the process. Though this would affect the population growth somewhat, the picture would qualitatively still be the same. Moreover, the price to be paid for such an improvement of the model seems excessive in terms of mathematical complexity.

There are strong indications that both sucrose and hormone belong to the main factors which regulate the plant cell cycle. Although the picture is not altogether clear yet, it is usually argued that the sucrose concentration influences the cell cycle duration by means of regulation of the rate of some biochemical processes taking place in G_1 , whereas hormones would act as permissive factors, altering the physiological state of the cells (bringing about dedifferentiation), so that cell division can occur (FOSKET, 1977; ALBERTS et al., 1983; BAYLISS, 1985; TREWAVAS, 1985). The way we let the two substances act, is derived from the well known Michaelis-Menten/Monod kinetics (see e.g. ROELS (1983)).

Of course the model yields a process which is much simpler than the actual situation it attempts to describe. For instance, we have assumed that the type-B cells do not consume substrate. This is probably not realistic, since differentiating cells grow and accumulate starch and/or sucrose. However, since the number of B-cells is only very large at the end of the growth process, correction on this point most likely will result in reaching the stationary phase earlier, but will not change the growth pattern drastically. Another aspect we have neglected, is the substrate consumption for maintenance energy requirement. We supposed that the quantity of substrate used for maintenance energy requirement is negligible with respect to that needed for the creation of new cells. However, if we do want to build this phenomenon into the model, the effect is not immediately clear, because in this case all cells are involved. Nevertheless, for the moment the model supplies our needs reasonably well (see Chapter 6).

2.2. MATHEMATICAL FORMULATION OF THE MODEL

For $n = 1,2,\ldots$ we consider a model starting with $N_n(0) = N_{A,n}(0) = n$ cells in the A-state. The total number of cells at time t and the number of type-A cells at time t are denoted by $N_n(t)$ and $N_{A,n}(t)$, respectively. In practice the cell numbers are usually very large (10^5 - 10^6 ml^{-1}). Moreover, as we already mentioned in Section 1.2 , deterministic models are often used to describe plant cell population growth. So one of our aims is to investigate whether the population behaviour, as predicted by our stochastic model, converges to a deterministic limit behaviour as the number of individuals tends to infinity. The obvious quantity to which the number of individuals is related is the initial cell number n , which is also usually very large in practice. Hence we shall commit ourselves to asymptotics in n . This is why all random variables in the model are given the index n . Note that in this context a quantity being fixed or constant means that it does not depend on n ; it may of course depend on the type of plant cells used for the experiments and/or be unknown in practice.

2.2.1. THE CELL CYCLE MODEL

The growth of the whole cell population in the model is determined by the behaviour of the individual cells, which is assumed to be ruled by the following regime.

Fix c, independent of n, and define the time of the i-th division by

$$\tau_{i,n} = \inf\{\, t : N_n(t) = n + i \,\} \;, \qquad i = 1,2,\dots \;. \tag{2.2.1}$$

A cell dividing at time $\tau_{i,n}$ has started its mitotic cycle at time $(\tau_{i,n} - c)$, because it received a stimulus at this time. A type-A cell receives this stimulus at the time of the first event in a counting process starting at the birth of the cell and having a variable, random rate $Q_n(t)$. For different cells the processes are only coupled through this rate, but otherwise independent.

In providing a stimulus a fixed quantity y_s^{-1} of substrate is used up. The amount of substrate $S_n(t)$ at time t is given by

$$S_n(t) = \begin{cases} S_n(0) = nb_s y_s^{-1} \;, & 0 \le t < \tau_{1,n} - c \;, \\ S_n(0) - iy_s^{-1} = (nb_s - i)y_s^{-1} \;, & \tau_{i,n} - c \le t < \tau_{i+1,n} - c \\ & i = 1,2,\dots,n_{s,n} - 1 \;, \\ S_n(0) - n_{s,n}y_s^{-1} = (nb_s - n_{s,n})y_s^{-1} \;, & t \ge \tau_{n_{s,n},n} - c \;, \end{cases} \tag{2.2.2}$$

where b_s is fixed and $n_{s,n} = [nb_s]$. Here $[a]$ denotes the integral part of a. It is assumed that the rate $Q_n(t)$ depends on $S_n(t)$ in the following way.

$$Q_n(t) = \frac{S_n(t)}{d(S_n(t) + nk_s)} 1_{[y_s^{-1},\infty)}(S_n(t))$$

$$(2.2.3) \quad = \begin{cases} \dfrac{nb_s}{d(nb_s + na_s b_s)} \equiv Q_{0,n} \equiv Q_0 , & 0 \le t < \tau_{1,n} - c , \\ \dfrac{nb_s - i}{d(nb_s + na_s b_s - i)} \equiv Q_{i,n} , & \tau_{i,n} - c \le t < \tau_{i+1,n} - c \\ & i = 1,2,\dots,n_{s,n} - 1 , \\ 0 \equiv Q_{i,n} , & t \ge \tau_{n_{s,n},n} - c \\ & i = n_{s,n}, n_{s,n} + 1, \dots , \end{cases}$$

where d and $a_s = k_s y_s b_s^{-1}$ are fixed, k_s is a constant, and $1_{\mathfrak{U}}$ denotes the indicator function of the set $\mathfrak{U}$. This means that the total number of stimuli is never larger than $n_{s,n}$, since $Q_n(t) = 0$ for $t \ge \tau_{n_{s,n},n} - c$.

In the model the amount of hormone is given by

$$(2.2.4) \quad H_n(\tau_{i,n}) = \begin{cases} H_n(c) = nb_h y_h^{-1} , & i = 1 , \\ H_n(c) - (i - 1)y_h^{-1} = (nb_h - i + 1)y_h^{-1} , & \\ & i = 2,3,\dots,n_{h,n} , \\ H_n(c) - n_{h,n} y_h^{-1} = (nb_h - n_{h,n})y_h^{-1} , & \\ & i = n_{h,n} + 1, n_{h,n} + 2, \dots , \end{cases}$$

where y_h and b_h are independent of n and $n_{h,n} = [nb_h]$. The two cells originating at time $\tau_{i,n}$ independently become type-A cells with probability $P_{i,n}$ and type-B cells with probability $(1 - P_{i,n})$, where $P_{i,n}$ depends on the amount of hormone present:

$$P_{i,n} = \frac{H_n(\tau_{i,n})}{H_n(\tau_{i,n}) + nk_h} 1_{[1,n_{h,n}]}(i)$$

$$(2.2.5) \qquad = \begin{cases} \dfrac{nb_h - i + 1}{nb_h + na_h b_h - i + 1} \; , & i = 1,2,\dots,n_{h,n} \; , \\ 0 \; , & i = n_{h,n}+1, n_{h,n}+2, \dots \; , \end{cases}$$

where k_h and $a_h = k_h y_h b_h^{-1}$ are fixed. Note that the processes may continue after $n_{h,n}$ divisions, even though $P_{n_{h,n}+k,n} = 0$, $k = 1,2,\dots$. Then only type-B cells are produced.

2.2.2. IMPLICATIONS FOR THE TOTAL POPULATION

Now that the model is formulated on the individual level, we next consider what this means for the total population behaviour.

Let $t \geq c$. Before or at time $(t - c)$, $(N_n(t-c) - n)$ type-A cells have divided. Therefore the number of type-A cells at time $(t - c)$ is given by

$$(2.2.6) \qquad N_{A,n}(t-c) = 2n - N_n(t-c) + \sum_{j=1}^{N_n(t-c)-n} Z_{j,n} \; ,$$

where $Z_{j,n}$ is the number of newly born A-cells at the j-th division. Thus $Z_{1,n}$, $Z_{2,n}$, ... are independent random variables and $Z_{j,n}$ has a binomial distribution with parameters 2 and $P_{j,n}$, $j = 1,2,\dots$. Since $P_{j,n} = 0$ for $j > n_{h,n}$, $Z_{j,n} = 0$ a.s. for $j > n_{h,n}$. We shall find it convenient to write

$$\hat{Z}_n = (Z_{1,n} \, , Z_{2,n} \, , \dots) \; ,$$

$$Z_n(t) = \sum_{j=1}^{N_n(t)-n} Z_{j,n} \; .$$

Let the number of type-A cells waiting for a stimulus at time t , i.e. the type-A cells which have not received a stimulus before t , be denoted by $N^0_{A,n}(t)$. Then, since at time $(t - c)$ there are $(N_n(t) - N_n(t-c))$ A-cells which have received a stimulus and have not yet divided,

$$(2.2.7) \qquad N^0_{A,n}(t-c) = 2n - N_n(t) + Z_n(t-c) \quad .$$

Since at time $(t - c)$ all of these $N^0_{A,n}(t-c)$ cells are waiting for stimuli arriving independently with rate $Q_n(t-c)$, and a cell divides time c after receiving a stimulus, the process $N_n(t)$ grows at a variable random rate $N^0_{A,n}(t-c)Q_n(t-c)$. The two factors entering into this rate are given by (2.2.3) and (2.2.7) . Note that the observed process $N_n(t)$ together with the unobserved random vector $\hat{Z}_n$ determine all other random quantities discussed so far.

The process $N_n(t)$ stops at the first time when either $N_{A,n}(t) = 0$ or $N_n(t) = n + n_{s,n}$. Note that $N^0_{A,n}(t-c) = 0$ is not sufficient for N_n to stop growing at time t , since new A-cells may be born between time $(t - c)$ and t . Thus from (2.2.6) we see that the process stops at the random time T_n defined by

$$(2.2.8) \qquad T_n = \inf\{ t : \sum_{j=1}^{N_n(t)-n} (Z_{j,n} - 1) = -n \} \wedge \inf\{ t : N_n(t) = n + n_{s,n} \},$$

where $(a \wedge b)$ means the smaller of a and b . The random level which $N_n(t)$ has reached by then is

$$(2.2.9) \quad N_n(T_n) = n + \inf\{ m : \sum_{j=1}^{m} (Z_{j,n} - 1) = -n \} \wedge n_{s,n} \quad ,$$

and we see that given $\hat{Z}_n$, this final level is a non-random quantity.

Let us translate this informal description into a more formal mathematical model. For $n = 1,2,\ldots$, let Ω_n be the outcome space of the process $\{N_n(t)\}_{t\geq 0}$ and the sequence $\hat{Z}_n = (Z_{1,n} , Z_{2,n} , \ldots)$, i.e. Ω_n consists of non-decreasing right-continuous step functions with step heights 1 coupled with sequences with elements 0 , 1 or 2 . For $t \geq 0$, let $\mathcal{G}_{t,n}$ be the σ-algebra generated by $\{N_n(s)\}_{s\leq t}$ and $\hat{Z}_n$, and $\mathcal{G}_n$ the smallest σ-algebra containing $\bigcup_{t\geq 0} \mathcal{G}_{t,n}$. Define a probability measure $\mathcal{P}_n$ on $(\Omega_n,\mathcal{G}_n)$ such that $Z_{1,n} , Z_{2,n} , \ldots$ are independent and $Z_{j,n}$ has a binomial distribution with parameters 2 and $P_{j,n}$ given by

(2.2.5) ; conditionally on $\hat{Z}_n$, $\{N_n(t) - n\}_{t\geq 0}$ is distributed under P_n as a stopped counting process with intensity process $\Lambda_{z,n}(t)$ given by the left-continuous version of

$$\Lambda_{z,n}(t) = \lim_{h\downarrow 0} h^{-1} P_n (N_n(t+h) - N_n(t) \geq 1 \mid G_{t,n})$$

$$(2.2.10) \quad = N^0_{A,n}(t-c)Q_{N_n(t)-n,n} 1_{[c,\infty)}(t)$$

$$= (2n - N_n(t) + Z_n(t-c))\, Q_{N_n(t)-n,n}\, 1_{[c,\infty)}(t) \quad .$$

The process stops at time T_n given by (2.2.8) and (2.2.9) or equivalently by

$$T_n = \inf\{ t : N_n(t) = n + \inf\{ m : \sum_{j=1}^{m} (Z_{j,n} - 1) = -n \} \wedge n_{s,n} \}$$

which is clearly a stopping time with respect to $\{G_{t,n}\}_{t\geq 0}$.

To avoid measurability problems we replace (Ω_n, G_n, P_n) by its completion and extend all σ-algebras $G_{t,n}$, $t \geq 0$, to contain all P_n-null sets, while retaining the original notation.

Finally, let let (Ω,G) be the product measurable space of (Ω_n,G_n) , $n = 1,2,\ldots$, and P any probability measure on (Ω,G) with marginals P_n on (Ω_n,G_n) , $n = 1,2,\ldots$. Thus we have defined $\{N_n(t)\}_{t\geq 0}$ and $\hat{Z}_n$ for $n = 1,2,\ldots$ on a single probability space (Ω,G,P) . Again we replace (Ω,G,P) by its completion.

In the sequel we shall occasionally also need the similarly completed versions of the σ-algebras $F_{t,n}$ generated by $\{N_n(s)\}_{s\leq t}$ and the smallest σ-algebra F_n containing $\bigcup_{t\geq 0} F_{t,n}$.

Let $A_{z,n}(t)$ be the integrated intensity process of $(N_n(t) - n)$ given $\hat{Z}_n$, i.e.

$$(2.2.11) \quad A_{z,n}(t) = \int_0^t \Lambda_{z,n}(s)ds \quad .$$

Define the inverse process $A^{-1}_{z,n}$ by

$$(2.2.12)\quad A^{-1}_{z,n}(u) = \begin{cases} \inf\{\ t \geq c : A_{z,n}(t) \geq u\ \} , & 0 \leq u \leq A_{z,n}(T_n) , \\ \infty , & u > A_{z,n}(T_n) , \end{cases}$$

and the process N^*_n by

$$(2.2.13)\quad N^*_n(u) = \begin{cases} N_n(A^{-1}_{z,n}(u)) - n , & 0 \leq u \leq A_{z,n}(T_n) , \\ \Pi_n(u - A_{z,n}(T_n)) + N_n(T_n) - n , & u > A_{z,n}(T_n) , \end{cases}$$

where Π_n is a Poisson process with intensity 1 defined on $[0,\infty)$, independent of N_n and $\Lambda_{z,n}$. Then given $\hat{Z}_n$, N^*_n is a Poisson process with intensity 1 on $[0,\infty)$, and hence independent of $\hat{Z}_n$. It follows that unconditionally N^*_n is also a unit Poisson process. Note that the stopping time T_n , however, is a function of $\hat{Z}_n$. For a detailed discussion on counting processes and other concepts mentioned above see for instance JACOBSEN (1982) or AALEN & HOEM (1978).

2.2.3. RESULTS FOR $n \to \infty$

We now give a summary of the results for the asymptotic situation where the initial cell number n tends to infinity. In the following chapters this is considered in detail.

First of all, the population behaviour does converge to a deterministic limit behaviour. In fact it is proved in Section 3.1 (cf. THEOREM 3.1.2) that there exists a non-random differentiable and non-decreasing function X such that for $n = 1,2,\ldots$ and $x \geq 0$,

$$P\,(\sup_{t\geq 0} |n^{-1}(N_n(t) - n) - X(t)| \geq x) \leq A\exp\{-\alpha x^2 n\}$$

for positive numbers A and α . But this means (cf. THEOREM 3.2.1) that we have the following rates of convergence of $n^{-1}(N_n - n)$ to X :

$$\sup_{t\geq 0} |n^{-1}(N_n(t) - n) - X(t)| = \mathcal{O}_P(n^{-1/2}) ,$$

$$\sup_{t\geq 0} |n^{-1}(N_n(t) - n) - X(t)| = \mathcal{O}((n^{-1}\log n)^{1/2}) \quad \text{a.s.}$$

These results suggest that there exists a process V, which does not depend on n, such that

$$n^{1/2}(n^{-1}(N_n - n) - X) \overset{\mathcal{D}}{\to} V , \quad (n \to \infty) \quad .$$

We prove this in Section 3.3 ; the convergence is in the sense of the Skorohod metric, and V is a Gaussian process with mean zero (see THEOREM 3.3.1).

In Section 3.4 we study the order of magnitude of the time T_n of the final division or the duration of the process N_n. This is closely related to the rate at which N_n or, equivalently, X reaches its final level. When either the number of A-cells becomes extinct or the substrate is used up, but not both, we find (cf. LEMMA 3.4.2 and THEOREM 3.4.1) that there exist positive constants C_0, C_1, ε_0, ε_1, a and A such that

$$C_0\exp\{-\varepsilon_0 t\} \le X(\infty) - X(t) \le C_1\exp\{-\varepsilon_1 t\} \quad , \qquad t \ge 0 \quad ,$$

$$0 < a \le \liminf T_n(\log n)^{-1} \le \limsup T_n(\log n)^{-1} \le A < \infty \quad \text{a.s.}$$

When both the number of A-cells becomes extinct and the substrate is used up, it takes longer for the process to stop and more variation in the duration is possible. In fact we can prove in this case (see LEMMA 3.4.4 and THEOREM 3.4.1) that there exist positive constants C_0, C_1, ε_0, ε_1, b and B such that

$$C_0(1 + \varepsilon_0 t)^{-1} \le X(\infty) - X(t) \le C_1(1 + \varepsilon_1 t)^{-1} \quad , \qquad t \ge 0 \quad ,$$

$$\liminf T_n(n^{-1}\log n)^{1/2} \ge b > 0 \quad \text{a.s.} \quad ,$$

$$\limsup T_n(n\log n)^{-1} \le B < \infty \quad \text{a.s.}$$

Finally, note that in (2.2.10) we defined the intensity process $\Lambda_{z,n}$ which is conditional on $\hat{Z}_n$. Analogously, the unconditional intensity process Λ_n is defined by

$$(2.2.14) \quad \Lambda_n(t) = \lim_{h\downarrow 0} h^{-1} \mathcal{P}\,(N_n(t+h) - N_n(t) \ge 1 \mid \mathcal{F}_{t,n}) \quad .$$

In Chapter 5 we find a process $\hat{\Lambda}_n$, which does not depend on $\hat{Z}_n$, such that for each $M \geq 0$,

$$\sup_{0 \leq t \leq M} |\Lambda_n(t) - \hat{\Lambda}_n(t)| = o_P(n^{1/2}) \quad .$$

This result yields that for $c \leq t \leq M$,

$$n^{-1/2}(N_n(t) - n - \int_c^t \hat{\Lambda}_n(s)ds) \overset{\mathcal{D}}{\to} W_0(X(t)) \;, \quad (n \to \infty) \quad ,$$

where W_0 is a standard Wiener process, X is the limit function as earlier, and the convergence is in the sense of the Skorohod metric (cf. (5.3.21) , (5.4.8) and THEOREM 5.4.1).

CHAPTER 3

THE BEHAVIOUR OF N_n FOR $n \to \infty$

This chapter deals with the asymptotic behaviour of N_n for $n \to \infty$. Convergence of $n^{-1}(N_n - n)$ to some non-random function X is proved and at the same time an exponential bound is provided for $P(\sup_{t\geq c} |n^{-1}(N_n(t) - n) - X(t)| \geq x)$. This allows us to compute rates of convergence for almost sure convergence as well as for convergence in probability. The limit distribution of $n^{1/2}(n^{-1}(N_n - n) - X)$ is studied and the order of the duration T_n of N_n is determined. Finally, simulations of N_n are graphically compared with $n(X + 1)$.

3.1. PRELIMINARIES

In this section we prove several probability inequalities which enable us to derive an exponential bound for the probability that $\sup_{t\geq 0} |n^{-1}(N_n(t) - n) - X(t)|$ exceeds some value x. Here X is a non-random differentiable function of t, which is to be defined later (LEMMA 3.1.4). From now on we shall write

$$X_n(t) = n^{-1}(N_n(t) - n) .$$

LEMMA 3.1.1 For $n = 1,2,\ldots$ and $0 \leq x \leq 2b_s$, we have

$$P(\sup_{t\geq c} |X_n(t) - n^{-1}A_{z,n}(t)| \geq x) \leq \tfrac{11}{3}\exp\{-x^2 n/(32b_s)\} . \tag{3.1.1}$$

PROOF. We have $X_n(t) \leq n^{-1}n_{s,n} = n^{-1}[nb_s] \leq b_s$ and hence

$$\sup_{t\geq c} |X_n(t) - n^{-1}A_{z,n}(t)| \leq \sup_{\{u:N_n^*(u)\leq nb_s\}} n^{-1}|N_n^*(u) - u| ,$$

where N_n^* is the unit Poisson process defined in (2.2.13). The lemma now follows from LEMMA A1.1 in Appendix A1.

□

Define

$$(3.1.2)\qquad \Lambda_{p,n}(t) = (2n - N_n(t) + 2 \sum_{j=1}^{nX_n(t-c)} P_{j,n})Q_{nX_n(t),n}1_{[c,\infty)}(t) \quad ,$$

and note that $\Lambda_{p,n}$ is obtained from $\Lambda_{z,n}$ in (2.2.10) by replacing the $Z_{j,n}$ by their expected values $2P_{j,n}$. We have

<u>LEMMA</u> <u>3.1.2</u> For $n = 1,2,\dots$ and $x \ge 0$,

$$(3.1.3)\qquad P\,(\max_m n^{-1}|\sum_{j=1}^{m}(Z_{j,n} - 2P_{j,n})| \ge x) \le 4\exp\{-x^2 n/(18b_h)\} \quad ,$$

$$(3.1.4)\qquad \begin{aligned} &P\,(\sup_{t\ge c} n^{-1}|\Lambda_{z,n}(t) - \Lambda_{p,n}(t)| \ge x) \\ &\qquad\qquad \le 4\exp\{-d^2(1 + a_s)^2x^2n/(18b_h)\} \quad . \end{aligned}$$

<u>PROOF.</u> $Z_{1,n}$, $Z_{2,n}$, ... are independent and $Z_{j,n}$ has a binomial distribution with parameters 2 and $P_{j,n}$, $j = 1,2,\dots$; for $j > n_{h,n}$, $P_{j,n} = 0$ and $Z_{j,n} = 0$ a.s. . Because $n_{h,n} = [nb_h] \le nb_h$, we may invoke LEMMA A1.2 (see Appendix A1) to obtain (3.1.3) .

Furthermore, we see from definitions (2.2.10) , (3.1.2) and (2.2.3) that for $t \ge c$,

$$|\Lambda_{z,n}(t) - \Lambda_{p,n}(t)| = |\sum_{j=1}^{nX_n(t-c)}(Z_{j,n} - 2P_{j,n})Q_{nX_n(t),n}|$$

$$\le (d(1 + a_s))^{-1} \max_{1\le m\le n_{h,n}} |\sum_{j=1}^{m}(Z_{j,n} - 2P_{j,n})| \quad .$$

Hence (3.1.4) follows from (3.1.3) .

□

Define $F : [0,\infty) \times D[0,\infty) \to \mathbb{R}$ by

$$\begin{aligned} F(t,x) &= \{1 - x(t) + 2(x(t-c) \wedge b_h) \\ &\quad - 2a_h b_h \log\left[\frac{b_h + a_h b_h}{b_h + a_h b_h - (x(t-c) \wedge b_h)}\right]\} \\ &\quad \cdot \frac{b_s - (x(t) \wedge b_s)}{d(b_s + a_s b_s - (x(t) \wedge b_s))} 1_{[c,\infty)}(t) \qquad (3.1.5) \\ &= \{1 - x(t) + 2\int_0^{x(t-c)} P(u)du\} Q(x(t)) 1_{[c,\infty)}(t) \\ &= \{1 - x(t) + 2\bar{P}(x(t-c))\} Q(x(t)) 1_{[c,\infty)}(t) \quad , \end{aligned}$$

where $D[0,\infty)$ is the space of right-continuous, $\mathbb{R}$-valued functions on $[0,\infty)$ with left-hand limits everywhere, and Q, P and $\bar{P} : [0,\infty) \to [0,\infty)$ are defined by

$$Q(u) = \frac{b_s - (u \wedge b_s)}{d(b_s + a_s b_s - (u \wedge b_s))} \quad ,$$

$$(3.1.6) \qquad P(u) = \frac{b_h - (u \wedge b_h)}{b_h + a_h b_h - (u \wedge b_h)} \quad ,$$

$$\bar{P}(u) = \int_0^u P(s)ds = (u \wedge b_h) - a_h b_h \log\left[\frac{b_h + a_h b_h}{b_h + a_h b_h - (u \wedge b_h)}\right] \quad .$$

Note that

$$(3.1.7) \qquad Q_{N_n(t)-n,n} = Q_{nX_n(t),n} = Q(X_n(t)) 1_{[0,n^{-1}n_{s,n})}(X_n(t)) \quad ,$$

$$(3.1.8) \qquad P_{N_n(t)-n,n} = P_{nX_n(t),n} = P(X_n(t)) + \mathcal{O}(\tfrac{1}{n}) \quad ,$$

so that $F(t,X_n)$ may be expected to approximate $n^{-1}\Lambda_{p,n}(t)$. In fact, we have

<u>LEMMA</u> <u>3.1.3</u> For $n = 1,2,\ldots$,

$$(3.1.9)\qquad \sup_{t\geq c} |n^{-1}\Lambda_{p,n}(t) - F(t,X_n)| \leq 2(1 + b_s)(da_s b_s n)^{-1} \quad .$$

<u>PROOF.</u> By (3.1.6)

$$\sum_{j=1}^{m} P_{j,n} - n\bar{P}(m/n) = \sum_{j=1}^{m} (P_{j,n} - n\int_{(j-1)/n}^{j/n} P(u)du) \quad .$$

Because P is a non-negative and non-increasing function with $P((j-1)/n) = P_{j,n}$ for $j \leq n_{h,n}$, it follows that for $m \leq n_{h,n}$,

$$(3.1.10)\qquad 0 \leq \sum_{j=1}^{m} P_{j,n} - n\bar{P}(m/n) \leq \sum_{j=1}^{m} (P((j-1)/n) - P(j/n)) \leq P(0) \quad .$$

Since $P_{j,n} = 0$ for $j > n_{h,n}$ and $P(u) = 0$ for $u \geq b_h$, we see that for $m > n_{h,n}$,

$$(3.1.11)\qquad \sum_{j=1}^{m} P_{j,n} - n\bar{P}(m/n) = \sum_{j=1}^{n_{h,n}} P_{j,n} - n\bar{P}(n_{h,n}/n) - n(\bar{P}(b_h) - \bar{P}(n_{h,n}/n)) \, .$$

Obviously,

$$0 \leq n(\bar{P}(b_h) - \bar{P}(n_{h,n}/n)) \leq (nb_h - n_{h,n})P(n_{h,n}/n) \leq P(0)$$

and (3.1.10) - (3.1.11) imply that

$$(3.1.12)\qquad \max_{m} | \sum_{j=1}^{m} P_{j,n} - n\bar{P}(m/n)| \leq P(0) = (1 + a_h)^{-1} \quad .$$

Next, we recall that (cf. (2.2.3))

$$\sup_{t\geq c} Q(X_n(t)) \leq (d(1 + a_s))^{-1}$$

and in view of (3.1.2) , (3.1.5) and (3.1.12) we conclude that for $t \geq c$ and $X_n(t) < n_{n,s}/n$,

$$|n^{-1}\Lambda_{p,n}(t) - F(t,X_n)| = 2|n^{-1}\sum_{j=1}^{nX_n(t-c)} P_{j,n} - \bar{P}(X_n(t-c))|Q(X_n(t))$$

$$\le 2(d(1 + a_h)(1 + a_s)n)^{-1} \le 2(1 + b_s)(da_sb_sn)^{-1} \quad .$$

Finally, if $X_n(t) = n_{s,n}/n$, then $\Lambda_{p,n}(t) = 0$ by (2.2.3) and $Q(X_n(t)) \le (da_sb_sn)^{-1}$, so that

$$|F(t,X_n)| \le (X_n(t) + 1)Q(X_n(t)) \le (1 + b_s)(da_sb_sn)^{-1} \quad .$$

□

Consider the system

$$(3.1.13) \quad \begin{cases} x(t) = \displaystyle\int_c^t F(s,x)ds \quad , & t \ge c \quad , \\ x(t) = 0 \quad , & t < c \quad , \end{cases}$$

with $x \in D[0,\infty)$ and F as defined in (3.1.5) . Then the following holds.

LEMMA 3.1.4 There exists a unique solution X of (3.1.13) in $D[0,\infty)$. The function X is continuous, non-negative, non-decreasing and bounded on $(0,\infty)$, and differentiable on (c,∞) with a continuous, positive and bounded derivative. Hence X is strictly increasing on (c,∞) and

$$(3.1.14) \quad \lim_{t\to\infty} X(t) = \inf\{ z \in \mathbb{R} : 2\bar{P}(z) - z \le -1 \} \wedge b_s \quad .$$

There exist positive constants C_0 and ε_0 such that for $0 < t < \infty$,

$$(3.1.15) \quad \lim_{\tau\to\infty} X(\tau) - X(t) \ge C_0\exp\{-\varepsilon_0 t\} \quad .$$

PROOF. For $t \ge 0$ and continuous functions x and y on $[0,\infty)$, (3.1.5) implies that

$$\sup_{s\le t} |F(s,x)| \le \frac{1}{d}(1 + \sup_{s\le t} |x(s)|)$$

and since the partial derivatives of F with respect to $x(t)$ and $x(t-c)$ are bounded

$$
\begin{aligned}
|F(t,x) - F(t,y)| &\le C|x(t) - y(t)| + C'|x(t-c) - y(t-c)| \\
&\le C^* \sup_{s\le t} |x(s) - y(s)|
\end{aligned} \tag{3.1.16}
$$

for positive constants C, C' and C^*. It follows from LEMMA A2.2 in Appendix A2 that a unique solution $X \in D[0,\infty)$ exists, and that X is continuous on $(0,\infty)$. Since $X'(t) = F(t,X)$ on (c,∞) and $F(t,X) = 0$ if $X(t) \ge b_s$, we see that X is bounded on $(0,\infty)$ and differentiable on (c,∞) with a continuous and bounded derivative.

The remaining assertions of the lemma will follow if we show that $X'(t) > 0$ for $c < t < \infty$ and establish (3.1.14) and (3.1.15). We begin by defining

$$\phi(t) = 1 - X(t) + 2\bar{P}(X(t-c)) \qquad \text{for } t \ge c$$

and noting that ϕ is continuous with $\phi(c) = 1$, and

(i) $\phi(t) > 0$ for $c \le t \le 2c$;

(ii) $\phi(t) \ge 0$ for $t > 2c$;

(iii) If $\phi(t_0) = 0$ for some $t_0 > 2c$, then $P(X(t_0-c))Q(X(t_0-c)) = 0$.

To see that (i) holds, notice that $\phi(t) = 1 - X(t)$ and $F(t,X) = (1 - X(t))Q(X(t))$ for $c \le t \le 2c$, so that $X'(t)(1 - X(t))^{-1}$ is bounded on $[c,2c]$ whenever $X(t) \ne 1$. But this means that $X(t) \ne 1$ on $[c,2c]$ since otherwise the integral of the bounded function $X'(t)(1 - X(t))^{-1}$ over an interval of length $\le c$ would diverge. As ϕ is continuous with $\phi(c) = 1$ this implies (i).

To prove (ii), suppose to the contrary that $\phi(t) \ge 0$ for $c \le t \le t_0$, $\phi(t_0) = 0$ and $\phi(t) < 0$ for $t_0 < t < t_0 + \varepsilon$ for some $t_0 > 2c$ and $0 < \varepsilon < c$. Then for $t_0 < t < t_0 + \varepsilon$, we have $\phi(t) < 0$, $\phi(t-c) \ge 0$, and

$$
\begin{aligned}
\phi'(t) &= -X'(t) + 2P(X(t-c))X'(t-c) \\
&= -\phi(t)Q(X(t)) + 2P(X(t-c))\phi(t-c)Q(X(t-c)) \ge 0 ,
\end{aligned}
$$

which is an obvious contradiction. Similarly, (iii) is proved by assuming that $\phi(t) > 0$ for $c \le t < t_0$ and $\phi(t_0) = 0$ for some $t_0 > 2c$, and noting the contradiction that $\phi'(t_0) > 0$ unless $P(X(t_0-c))Q(X(t_0-c)) = 0$. Since P and Q are non-increasing, this also holds for every subsequent zero of ϕ.

Next we consider the function g on $(0,\infty)$ defined by

$$g(z) = (1 - z + 2\overline{P}(z))Q(z)$$

and note that $g(X(t)) \ge X'(t) = \phi(t)Q(X(t)) \ge 0$ for $t > c$ because of (3.1.5). As g has a bounded derivative, (3.1.5) also ensures that there exists $\alpha > 0$ such that

$$|(g(X(t)))^{-1}X'(t)g'(X(t))| \le \alpha \qquad \text{for } t > c ,$$

whenever $g(X(t)) \ne 0$. Integrating over (c,t) we see that $|\log[g(X(t))/g(0)]| \le \alpha(t - c)$ and as $g(0) > 0$ there exists $A > 0$ such that

$$g(X(t)) \ge A\exp\{-\alpha t\} \qquad \text{for } t > c .$$

It follows that $Q(X(t)) > 0$ and $1 - X(t) + 2\overline{P}(X(t)) > 0$ for $t > c$. Suppose that $\phi(t_0) = 0$ for some $t_0 > 2c$. As $Q(X(t_0-c)) > 0$, (iii) implies that $P(X(t_0-c)) = 0$, so that $\overline{P}(X(t_0)) = \overline{P}(X(t_0-c))$ and $\phi(t_0) = 1 - X(t_0) + 2\overline{P}(X(t_0)) > 0$. This contradiction shows that $\phi(t) > 0$ for $t \ge c$, and hence that $X'(t) = \phi(t)Q(X(t)) > 0$ for $t > c$.

As X is increasing and bounded and $X''(t) = \frac{d}{dt}F(t,X)$ is easily seen to be bounded too for $t > 2c$, it follows that $X'(t)$ tends to zero as $t \to \infty$. Writing $X(\infty)$ for $\lim_{t\to\infty} X(t)$, we find that

$$\lim_{t\to\infty} X'(t) = (1 - X(\infty) + 2\overline{P}(X(\infty)))Q(X(\infty)) = 0 ,$$

and since for $t > c$,

$$(1 - X(t) + 2\overline{P}(X(t)))Q(X(t)) \ge X'(t) > 0 ,$$

we have proved (3.1.14).

Finally, we note that for $t > c$,

$$(X(\infty) - X(t)) \sup_z |g'(z)| \geq g(X(t)) - g(X(\infty))$$

$$= g(X(t)) \geq A\exp\{-\alpha t\}$$

and as g' is bounded, this establishes (3.1.15) .

□

Having made these preparations, we can now quickly prove one of the main results of this section.

THEOREM 3.1.1 Let X be the unique solution of (3.1.13) and let $M \geq 0$. Then there exist positive numbers A and α depending on M such that for $n = 1,2,\ldots$ and $x \geq 0$,

$$(3.1.17) \quad P\left(\sup_{t\leq M} |X_n(t) - X(t)| \geq x\right) \leq A\exp\{-\alpha x^2 n\} \quad .$$

PROOF. Fix $M \geq c$, let C^* be as in (3.1.16) , $C = 6b_s\exp\{C^*(M - c)\}$ and $6(1 + b_s)(M - c)\exp\{C^*(M - c)\}(da_s b_s n)^{-1} \leq x \leq C$. Note that $C \geq 6b_s$. Also, for every $n = 1,2,\ldots$, it follows from LEMMAS 3.1.1 , 3.1.2 , 3.1.3 and inequality (3.1.16) that, except on a set of probability not larger than

$$\frac{11}{3}\exp\{-x^2 n/\left[288b_s\exp\{2C^*(M - c)\}\right]\}$$
$$+ 4\exp\{-d^2(1 + a_s)^2 x^2 n/\left[162b_h(M - c)^2\exp\{2C^*(M - c)\}\right]\} \quad ,$$

the following holds for all $c \leq t \leq M$:

$$|X_n(t) - X(t)|$$

$$\leq |X_n(t) - n^{-1}\Lambda_{z,n}(t)| + n^{-1}\int_c^t |\Lambda_{z,n}(s) - \Lambda_{p,n}(s)|ds$$

$$+ \int_c^t |n^{-1}\Lambda_{p,n}(s) - F(s,X_n)|ds + \int_c^t |F(s,X_n) - F(s,X)|ds$$

$$\leq 3\,\frac{x}{3}\,\exp\{-C^*(M - c)\} + C^*\int_c^t \sup_{\tau\leq s} |X_n(\tau) - X(\tau)|\,ds$$

$$\leq x\,\exp\{-C^*(M - c)\}\exp\{C^*(t - c)\} \leq x \quad ,$$

where the third inequality follows from LEMMA A2.1 in Appendix A2 .

Obviously, A can be chosen sufficiently large for (3.1.17) to hold for $0 \leq x < 6(1 + b_s)(M - c)\exp\{C^*(M - c)\}(da_s b_s n)^{-1}$ and since $X_n(t) \leq b_s$ and $X(t) \leq b_s$ for all $t \geq c$, (3.1.17) is trivially satisfied for $x > C \geq b_s$. By definition $X_n(t) = X(t) = 0$ for $0 \leq t < c$ and all n , and the theorem follows.

□

In order to obtain a probability inequality like (3.1.17) for the supremum over all t , we first investigate what happens when $t \to \infty$. We start by proving the following lemma.

<u>LEMMA 3.1.5</u> There exist positive constants α and B such that for $n = 1,2,\ldots$ and $x \geq 0$,

$$\text{(i)}\quad P\,\Big(n^{-1}\Big|\Big(n + \inf\{\, m : \sum_{j=1}^{m} (Z_{j,n} - 1) = -n \,\} \wedge n_{s,n}\Big)$$

$$- \Big(n + \inf\{\, m : \sum_{j=1}^{m} (2P_{j,n} - 1) \leq -n \,\} \wedge n_{s,n}\Big)\Big| \geq x\Big)$$

$$\leq 4\exp\{-\alpha x^2 n\} \quad , \tag{3.1.18}$$

$$\text{(ii)}\quad \Big|n^{-1}\Big(n + \inf\{\, m : \sum_{j=1}^{m} (2P_{j,n} - 1) \leq -n \,\} \wedge n_{s,n}\Big)$$

$$- \Big(1 + \inf\{\, z \in \mathbb{R} : 2\bar{P}(z) - z \leq -1 \,\} \wedge b_s\Big)\Big| \leq Bn^{-1} \quad .$$

<u>PROOF.</u> (i) First note that $|a \wedge c - b \wedge c| \leq |a - b|$. Moreover, since (i) is trivially true for $0 \leq nx \leq 1$ for an appropriate choice of α , we may assume without loss of generality that nx is a positive integer. Similarly, we allow ourselves in the proof of (i) to bound the probability of strict inequality. Hence we have to prove that for some $\alpha > 0$, $n = 1,2,\ldots$, $x = \frac{2}{n}, \frac{3}{n}, \ldots$,

$$\begin{aligned}\mathcal{P}\,\Big(n^{-1}\Big|\inf\{\, m : \sum_{j=1}^{m} (Z_{j,n} - 1) = -n \,\}\\ - \inf\{\, m : \sum_{j=1}^{m} (2P_{j,n} - 1) \le -n \,\}\Big| > x\Big) \le 4\exp\{-\alpha x^2 n\}\end{aligned} \tag{3.1.19}$$

Let $m_0 = \inf\{\, m : \sum_{j=1}^{m} (2P_{j,n} - 1) \le -n \,\}$. Then

$$n^{-1} \sum_{j=1}^{m_0} (2P_{j,n} - 1) \le -1 \quad , \qquad n^{-1} \sum_{j=1}^{m_0 - 1} (2P_{j,n} - 1) > -1 \quad . \tag{3.1.20}$$

Since $P_{j,n}$ decreases in j and $m_0 \le 2n_{h,n} + n = \mathcal{O}(n)$,

$$2P_{m_0,n} - 1 \le -\frac{n}{m_0} \le -\gamma < 0 \tag{3.1.21}$$

and for $x = \frac{2}{n} , \frac{3}{n} , \ldots ,$

$$n^{-1} \sum_{j=1}^{m_0 + nx} (2P_{j,n} - 1) \le -1 - \gamma x \quad . \tag{3.1.22}$$

Using this and (3.1.3) , we get

$$\mathcal{P}\,\Big(n^{-1}\inf\{\, m : \sum_{j=1}^{m} (Z_{j,n} - 1) = -n \,\} > n^{-1} m_0 + x\Big)$$

$$= \mathcal{P}\,\Big(n^{-1} \min_{k \le m_0 + nx} \sum_{j=1}^{k} (Z_{j,n} - 1) > -1\Big)$$

$$\begin{aligned}&\le \mathcal{P}\,\Big(n^{-1} \sum_{j=1}^{m_0 + nx} (Z_{j,n} - 1) > -1\Big)\\ &= \mathcal{P}\,\Big(n^{-1} \sum_{j=1}^{m_0 + nx} (Z_{j,n} - 2P_{j,n}) > -1 + n^{-1} \sum_{j=1}^{m_0 + nx} (1 - 2P_{j,n})\Big)\\ &\le \mathcal{P}\,\Big(n^{-1} \sum_{j=1}^{m_0 + nx} (Z_{j,n} - 2P_{j,n}) \ge \gamma x\Big)\end{aligned} \tag{3.1.23}$$

$$\le 4\exp\{-\gamma^2 x^2 n/(18 b_h)\} \quad .$$

On the other hand, because $\sum_{j=1}^{m}(2P_{j,n} - 1)$ is a concave function of m which vanishes for $m = 0$, $m^{-1}\sum_{j=1}^{m}(2P_{j,n} - 1)$ is a decreasing one. Since

$$-n < \sum_{j=1}^{m_0-1}(2P_{j,n} - 1) \le -n + 1 \le 0 \quad ,$$

it follows that for some positive δ and $1 \le k \le m_0 - 1 - nx$, $x = \frac{2}{n}, \frac{3}{n}, \ldots$,

$$\begin{aligned} n^{-1}\sum_{j=1}^{k}(2P_{j,n} - 1) &\ge k(m_0 - 1)^{-1}n^{-1}\sum_{j=1}^{m_0-1}(2P_{j,n} - 1) \\ &\ge (m_0 - 1 - nx)(m_0 - 1)^{-1}n^{-1}\sum_{j=1}^{m_0-1}(2P_{j,n} - 1) \qquad (3.1.24) \\ &\ge -(1 - nx(2n_{h,n} + n)^{-1}) \ge -1 + \delta x \quad , \end{aligned}$$

provided $m_0 - 1 - nx \ge 0$. But this and (3.1.3) imply

$$\begin{aligned} &\mathcal{P}\left(n^{-1}\inf\{\, m : \sum_{j=1}^{m}(Z_{j,n} - 1) = -n \,\} < n^{-1}m_0 - x\right) \\ &= \mathcal{P}\left(n^{-1}\min_{k \le m_0 - nx - 1}\ \sum_{j=1}^{k}(Z_{j,n} - 1) \le -1\right) \\ &\le \mathcal{P}\left(n^{-1}\min_{k \le m_0 - nx - 1}\ \sum_{j=1}^{k}(Z_{j,n} - 2P_{j,n}) \le -\delta x\right) \qquad (3.1.25) \\ &\le 4\exp\{-\delta^2x^2n/(18b_h)\} \quad . \end{aligned}$$

For $m_0 - 1 - nx \le 0$ the probability on the left equals zero and (3.1.25) is trivially correct. Hence (3.1.23) and (3.1.25) together prove (3.1.19).

(ii) Since $2\bar{P}(z) - z$ is a concave function of z which vanishes in $z = 0$, and its derivative with respect to z is bounded, we may as well consider $n^{-1}\inf\{ m \in \mathbb{N} : 2\overline{P}(\frac{m}{n}) - \frac{m}{n} \le -1 \}$, instead of $\inf\{ z \in \mathbb{R} : 2\overline{P}(z) - z \le -1 \}$. Also, because $b_s - n^{-1}n_{s,n} < n^{-1}$, it suffices (cf. (i)) to prove that for $n = 1,2,\ldots$ and some $B > 0$,

$$\begin{aligned} &n^{-1}\Big|\inf\{ m : \sum_{j=1}^{m} (2P_{j,n} - 1) \le -n \} \\ (3.1.26)\qquad &\quad - \inf\{ m \in \mathbb{N} : 2\overline{P}(\tfrac{m}{n}) - \tfrac{m}{n} \le -1 \}\Big| < Bn^{-1} . \end{aligned}$$

Let $x = \frac{2}{n}, \frac{3}{n}, \ldots$. Then in view of (3.1.12) and (3.1.22)

$$\begin{aligned} &2\overline{P}(n^{-1}(m_0 + nx)) - n^{-1}(m_0 + nx) \\ &= \Big(2\overline{P}(n^{-1}(m_0 + nx)) - 2n^{-1}\sum_{j=1}^{m_0+nx} P_{j,n}\Big) + n^{-1}\sum_{j=1}^{m_0+nx} (2P_{j,n} - 1) \\ &\le 2(1 + a_h)^{-1}n^{-1} - 1 - \gamma x \le -1 , \end{aligned}$$

provided $x \ge 2(1 + a_h)^{-1}\gamma^{-1}n^{-1}$.

Similarly, invoking (3.1.12) and (3.1.24) we find for $k \le m_0 - 1 - nx$,

$$2\overline{P}(\tfrac{k}{n}) - \tfrac{k}{n} \ge - 2(1 + a_h)^{-1} n^{-1} - 1 + \delta x > -1 ,$$

provided $x > 2(1 + a_h)^{-1}\delta^{-1}n^{-1}$. Thus (3.1.26) holds.

□

<u>COROLLARY</u> <u>3.1.1</u> There exist positive constants A and α such that for $n = 1,2,\ldots$ and $x \ge 0$,

$$\begin{aligned} &P\Big(\Big|n^{-1}(n + \inf\{ m : \sum_{j=1}^{m} (Z_{j,n} - 1) = -n \} \wedge n_{s,n}) \\ (3.1.27)\qquad &\quad - (1 + \inf\{ z \in \mathbb{R} : 2\overline{P}(z) - z \le -1 \} \wedge b_s)\Big| \ge x\Big) \\ &\le A\exp\{-\alpha x^2 n\} . \end{aligned}$$

PROOF. For $x \geq 2Bn^{-1}$, (3.1.27) follows from the foregoing lemma. Clearly, for $A > \exp\{4\alpha B^2\}$ the corollary also holds for $0 \leq x < 2Bn^{-1}$.

□

LEMMA 3.1.6 If X is the solution of (3.1.13) on $[0,\infty)$, then there exist positive numbers A and α such that for $n = 1,2,\ldots$ and $x \geq 0$,

$$(3.1.28) \quad P\,(|\lim_{t\to\infty} X_n(t) - \lim_{t\to\infty} X(t)| \geq x) \leq A\exp\{-\alpha x^2 n\} \quad .$$

PROOF. The lemma follows from (2.2.9) , LEMMA 3.1.4 and COROLLARY 3.1.1 .

□

For convenience we introduce a shorter notation for the quantities associated with the newly born A-cells . Let Y_n and Y be defined by

$$(3.1.29) \quad \begin{aligned} Y_n(t) &= n^{-1}Z_n(t) = n^{-1}\sum_{j=1}^{nX_n(t)} Z_{j,n} \ , & t \geq 0 \ . \\ Y(t) &= 2\bar{P}(X(t)) = 2\int_0^{X(t)} P(u)du = 2\int_0^t P(X(s))X'(s)ds \ , & t \geq 0 \ . \end{aligned}$$

Then, obviously

$$(3.1.30) \quad \begin{aligned} |Y_n(t) - Y(t)| &\leq n^{-1}|\sum_{j=1}^{nX_n(t)} (Z_{j,n} - 2P_{j,n})| \\ &+ 2|n^{-1}\sum_{j=1}^{nX_n(t)} P_{j,n} - \bar{P}(X_n(t))| + 2|\int_{X(t)}^{X_n(t)} P(u)du| \quad . \end{aligned}$$

Therefore, (3.1.3) , (3.1.12) , the boundedness of P , THEOREM 3.1.1 and LEMMA 3.1.6 immediately yield

CORROLARY 3.1.2 For each $M \geq 0$ there exist positive numbers A_0, α_0 depending on M, and A_1, α_1 such that for $n = 1,2,\ldots$ and $x \geq 0$,

$$\text{(i)}\quad P\,(\sup_{t\leq M} |Y_n(t) - Y(t)| \geq x) \leq A_0\exp\{-\alpha_0 x^2 n\} \quad , \tag{3.1.31}$$

$$\text{(ii)}\quad P\,(|\lim_{t\to\infty} Y_n(t) - \lim_{t\to\infty} Y(t)| \geq x) \leq A_1\exp\{-\alpha_1 x^2 n\} \quad . \tag{3.1.32}$$

□

Having proved the probability bounds (3.1.17) and (3.1.28) for bounded t and $t \to \infty$, respectively, we can now derive a similar result which holds uniformly for all $t \geq 0$. We write

$$\alpha_n = n^{-1}A_{z,n} \quad ,$$

$$X_A^0(t-c) = 1 - X(t) + Y(t-c) \geq 0 \quad , \qquad t \geq c \quad ,$$

$$X_{A,n}^0(t-c) = 1 - X_n(t) + Y_n(t-c) \geq 0 \quad , \qquad t \geq c \quad ,$$

$$p = P(X(\infty)) \geq 0 \quad , \qquad q = Q(X(\infty)) \geq 0 \quad , \qquad q' = Q'(X(\infty)) < 0 \quad .$$

Recall that for $t > c$, we have (cf. (3.1.5), (3.1.13), (2.2.10), (2.2.11) and (3.1.7))

$$\begin{aligned} X'(t) &= F(t,X) = X_A^0(t-c)Q(X(t)) \quad , \\ \alpha_n'(t) &= n^{-1}A_{z,n}(t) = X_{A,n}^0(t-c)Q(X_n(t))1_{[0,n^{-1}n_{s,n})}(X_n(t)) \quad , \end{aligned} \tag{3.1.33}$$

where α_n' denotes the derivative of α_n except on a countable set of jumps of X_n. Also, note that for $0 \leq u \leq X(\infty)$,

$$\begin{aligned} &0 \geq P'(u) > -(a_h b_h)^{-1} \quad , \qquad 0 \geq P''(u) > -2(a_h b_h)^{-2} \quad , \\ &0 > Q'(u) > -(d a_s b_s)^{-1} \quad , \qquad 0 \geq Q''(u) > -2d^{-1}(a_s b_s)^{-2} \quad . \end{aligned} \tag{3.1.34}$$

We shall need the following lemma.

<u>LEMMA</u> <u>3.1.7</u> Suppose that $\varepsilon > 0$ and define

$$(3.1.35) \quad t_0 = \inf\{\, t \geq c : X(\infty) - X(t-c) \leq \tfrac{1}{2}\varepsilon a_h b_h \,\} \ .$$

Then there exist positive numbers A_0 , A_1 , α_0 and α_1 depending on ε , such that for $n = 1,2,\ldots$ and $x \geq 0$,

$$(3.1.36) \quad P\,(\sup_{t\geq t_0} |X(\infty) - X_n(t-c)| \geq \varepsilon a_h b_h) \leq A_0 \exp\{-\alpha_0 n\} \ ,$$

$$(3.1.37) \quad P\,(\sup_{t\geq t_0} \bigl[\,|Y_n(t-c) - Y(t-c)| - 2(p+\varepsilon)|X_n(t-c) - X(t-c)|\,\bigr] \geq x)$$

$$\leq A_1 \exp\{-\alpha_1 x^2 n\} \ .$$

<u>PROOF.</u> Since X_n is monotone, we have for $t \geq t_0$,

$$|X(\infty) - X_n(t-c)| \leq |X(\infty) - X_n(T_n)| \vee |X(\infty) - X_n(t_0-c)|$$

$$\leq |X_n(T_n) - X(\infty)| + |X_n(t_0-c) - X(t_0-c)| + \tfrac{1}{2}\varepsilon a_h b_h$$

and in view of LEMMA 3.1.6 and THEOREM 3.1.1

$$P\,(\sup_{t\geq t_0} |X(\infty) - X_n(t-c)| \geq \varepsilon a_h b_h) \leq P\,(|X_n(T_n) - X(\infty)| \geq \tfrac{1}{4}\varepsilon a_h b_h)$$

$$+ P\,(|X_n(t_0-c) - X(t_0-c)| \geq \tfrac{1}{4}\varepsilon a_h b_h) \leq A_0 \exp\{-\alpha_0 n\} \ .$$

for positive A_0 , α_0 depending on ε . This proves (3.1.36) .

For $0 \leq x \leq \frac{8}{n}$, (3.1.37) trivially holds for an appropriate choice of A_1 and α_1 . Since $Y_n(t) \leq 2X_n(t) \leq 2b_s$ and $Y(t) \leq 2X(t) \leq 2b_s$ for all t , (3.1.37) is also satisfied for $x > 4b_s$. Assume therefore that $\frac{8}{n} < x \leq 4b_s$. By LEMMA 3.1.2

$$P\,(\sup_{t\geq c} n^{-1}\Bigl|\sum_{j=1}^{nX_n(t-c)} (Z_{j,n} - 2P_{j,n})\Bigr| \geq \frac{x}{2} - \frac{2}{n})$$

$$\leq P\,(\max_m n^{-1}\Bigl|\sum_{j=1}^{m} (Z_{j,n} - 2P_{j,n})\Bigr| \geq \frac{x}{4}) \leq A_2 \exp\{-\alpha_2 x^2 n\}$$

for positive constants A_2 and α_2 . From (3.1.12) we know that with probability 1

$$2|n^{-1}\sum_{j=1}^{nX_n(t-c)} P_{j,n} - \bar{P}(X_n(t-c))| \le \frac{2}{n} \qquad \text{for all } t \ge c \;.$$

Finally, since $P(X(\infty)) = p$ and $|P'(u)| < (a_hb_h)^{-1}$ for all u , we find that if $t \ge t_0$ and $|X(\infty) - X_n(t-c)| < \varepsilon a_hb_h$,

$$\begin{aligned}
|2\bar{P}(X_n(t-c)) - Y(t-c)| &= 2|\int_{X(t-c)}^{X_n(t-c)} P(u)du| \\
&\le 2[p + (a_hb_h)^{-1}\{(X(\infty) - X(t-c)) \vee |X(\infty) - X_n(t-c)|\}] \\
&\qquad \cdot |X_n(t-c) - X(t-c)| \\
&\le 2(p + \varepsilon)|X_n(t-c) - X(t-c)| \;.
\end{aligned}$$

Hence, in view of (3.1.35) and (3.1.36) , and since $\frac{x}{2} > \frac{4}{n}$,

$$\begin{aligned}
&P(\sup_{t\ge t_0} [|Y_n(t-c) - Y(t-c)| - 2(p + \varepsilon)|X_n(t-c) - X(t-c)|] > x) \\
&\le P(\sup_{t\ge t_0} [n^{-1}|\sum_{j=1}^{nX_n(t-c)} (Z_{j,n} - 2P_{j,n})| \\
&\qquad + 2|n^{-1}\sum_{j=1}^{nX_n(t-c)} P_{j,n} - \bar{P}(X_n(t-c))|] \ge \frac{x}{2}) \\
&+ P(\sup_{t\ge t_0} [|2\bar{P}(X_n(t-c)) - Y(t-c)| \\
&\qquad - 2(p + \varepsilon)|X_n(t-c) - X(t-c)|)] \ge \frac{4}{n}) \\
&\le A_2\exp\{-\alpha_2x^2n\} + P(\sup_{t\ge t_0} |X(\infty) - X_n(t-c)| \ge \varepsilon a_hb_h)
\end{aligned}$$

$$\leq A_2\exp\{-\alpha_2 x^2 n\} + A_0\exp\{-\alpha_0 n\}$$

$$\leq A_2\exp\{-\alpha_2 x^2 n\} + A_0\exp\{-\alpha_0 x^2 n/(4b_s)^2\}$$

$$\leq A_1\exp\{-\alpha_1 x^2 n\}$$

for appropriate positive A_1 and α_1 depending on ε .

□

Now we are ready to prove the final result.

<u>THEOREM 3.1.2</u> Let X be the solution of (3.1.13) . Then there exist positive numbers A and α such that for $n = 1,2,\ldots$ and $x \geq 0$,

(3.1.38) $$P\,(\sup_{t\geq 0} |X_n(t) - X(t)| \geq x) \leq A\exp\{-\alpha x^2 n\} \quad .$$

<u>PROOF.</u> Let T_n be the time X_n stops as defined by (2.2.8) . Since $\lim_{t\uparrow T_n} X_n(t) = X_n(T_n) - \frac{1}{n}$ and X_n and X are monotone, we have

$$\sup_{t\geq 0} |X_n(t) - X(t)| \leq \sup_{0\leq t<T_n} |X_n(t) - X(t)|$$

$$+ |X_n(T_n) - \lim_{t\to\infty} X(t)| + \frac{1}{n} \quad .$$

Therefore, in view of LEMMA 3.1.6 it suffices to show that

(3.1.39) $$P\,(\sup_{0\leq t<T_n} |X_n(t) - X(t)| \geq x) \leq A\exp\{-\alpha x^2 n\} \quad .$$

The proof consists of two parts. First suppose that the following assumption holds.

(A) There exist $t_0 \geq c$, positive constants $\tilde{A}$, $\tilde{\alpha}$ and $\tilde{C}$, and non-negative functions M_n, U_n, V_n with

$$U_n(t) + V_n(t) - M_n(t) < 0 , \qquad t_0 < t < T_n ,$$

such that for $n = 1,2,\ldots$, and $0 \leq x \leq \tilde{C}$,

$$(3.1.40) \quad P\,(\sup_{t \leq t_0} |\alpha_n(t) - X(t)| \geq x) \leq \tilde{A}\exp\{-\tilde{\alpha}x^2 n\} ,$$

$$(3.1.41) \quad \begin{cases} \alpha_n'(t) - X'(t) = -M_n(t)(\alpha_n(t) - X(t)) + R_n(t) , \quad t_0 < t < T_n , \\ \qquad \text{with } R_n \text{ such that} \\ P\,(|R_n(t)| \geq U_n(t)x + V_n(t)|\alpha_n(t-c) - X(t-c)| \\ \qquad \text{for some } t \in (t_0,T_n)\,) \leq \tilde{A}\exp\{-\tilde{\alpha}x^2 n\} . \end{cases}$$

Fix n and $x \in (0,\tilde{C}]$. Suppose that there exists $t_1 \in (t_0,T_n)$ such that $|\alpha_n(t) - X(t)| < x$ for $t < t_1$ and $|\alpha_n(t_1) - X(t_1)| = x$, and that $|R_n(t_1)| < U_n(t_1)x + V_n(t_1)|\alpha_n(t_1-c) - X(t_1-c)|$. Then

$$|R_n(t_1)| < U_n(t_1)x + V_n(t_1)|\alpha_n(t_1-c) - X(t_1-c)|$$

$$\leq (U_n(t_1) + V_n(t_1))|\alpha_n(t_1) - X(t_1)| .$$

Hence

$$(\alpha_n(t_1) - X(t_1))^{-1}(\alpha_n'(t_1) - X'(t_1))$$

$$\leq -M_n(t_1) + U_n(t_1) + V_n(t_1) < 0$$

which contradicts the assumption that t_1 is the first time when $\alpha_n(t) - X(t) = \pm x$. It follows from (3.1.40) and (3.1.41) that for $x \in (0,\tilde{C}]$,

$$P\,(\sup_{0\le t<T_n} |\alpha_n(t) - X(t)| \ge x) \le P\,(\sup_{t\le t_0} |\alpha_n(t) - X(t)| \ge x)$$

$$+ P\,(|R_n(t)| \ge U_n(t)x + V_n(t)|\alpha_n(t-c) - X(t-c)|$$

$$\text{for some } t \in (t_0,T_n)\,)$$

$$\le 2\tilde{A}\exp\{-\tilde{\alpha}x^2n\} \quad .$$

Combining this with LEMMA 3.1.1 we obtain (3.1.39) for $x \in (0,C]$ for some positive C . If $A \ge 1$, then (3.1.39) continues to hold for $x = 0$. Since $|X_n(t) - X(t)| \le 2b_s$ for all t , we see that for $x > C$, the left-hand side of (3.1.39) is bounded by $A\exp\{-\alpha x^2nC^2/(2b_s)^2\}$. Thus the theorem is proved under assumption (A) .

We now turn to the proof of Assumption (A) . We shall distinguish two cases.

Case (i). $p < \frac{1}{2}$. Choose $\varepsilon = \frac{1}{10}(1 - 2p)$ and let t_0 be as in (3.1.35) . From LEMMA 3.1.1 , THEOREM 3.1.1 and LEMMA 3.1.7 it follows that there exist positive numbers $\tilde{A}$, $\tilde{\alpha}$ and $\tilde{C}$ depending on ε , such that for $n = 1,2,\ldots$ and $0 \le x \le \tilde{C}$,

$$P\,(\sup_{t\ge 0} |\alpha_n(t) - X_n(t)| \ge \varepsilon x) \le \tfrac{1}{3}\tilde{A}\exp\{-\tilde{\alpha}x^2n\} \quad ,$$

$$P\,(\sup_{t\le t_0} |\alpha_n(t) - X(t)| \ge x) \le \tfrac{1}{3}\tilde{A}\exp\{-\tilde{\alpha}x^2n\} \quad ,$$

(3.1.42)

$$P\,(\sup_{t\ge t_0} \big[|Y_n(t-c) - Y(t-c)| - 2(p + \varepsilon)|X_n(t-c) - X(t-c)|\big] \ge \varepsilon x)$$

$$\le \tfrac{1}{3}\tilde{A}\exp\{-\tilde{\alpha}x^2n\} \quad .$$

For $c < t < T_n$ we have $\alpha_n'(t) = X^0_{A,n}(t-c)Q(X_n(t))$ (cf. (3.1.33)). Hence, except on a set of probability not larger than $\tilde{A}\exp\{-\tilde{\alpha}x^2n\}$, we have for all $t \in (t_0,T_n)$,

$$\begin{aligned}&\alpha_n'(t) - X'(t)\\ &= -Q(X(t))(X_n(t) - X(t)) + X^0_{A,n}(t-c)(Q(X_n(t)) - Q(X(t)))\\ &\qquad + (Y_n(t-c) - Y(t-c))Q(X(t))\\ &= -M_n(t)(X_n(t) - X(t)) + \tilde{R}_n(t) \quad ,\end{aligned}$$

where $M_n(t) > Q(X(t)) > 0$ and

$$|\tilde{R}_n(t)| < \{\varepsilon x + 2(p + \varepsilon)|X_n(t-c) - X(t-c)|\}Q(X(t)) \quad .$$

This follows from (3.1.42) and because

$$Q(X_n(t)) - Q(X(t)) = Q'(\theta X(t) + (1 - \theta)X_n(t))(X_n(t) - X(t)) \quad ,$$

for some $\theta \in (0,1)$, and $Q'(x) < 0$ for all $x < X(t) \vee X_n(t)$ and $t < T_n$. But then

$$\alpha_n'(t) - X'(t) = -M_n(t)(\alpha_n(t) - X(t)) + R_n(t) \quad ,$$

where,

$$\begin{aligned}&|R_n(t)| = |M_n(t)(\alpha_n(t) - X_n(t)) + \tilde{R}_n(t)|\\ &< M_n(t)\varepsilon x + \{\varepsilon x + 2(p + \varepsilon)\varepsilon x\\ &\qquad + 2(p + \varepsilon)|\alpha_n(t-c) - X(t-c)|\}Q(X(t))\\ &\le \{\varepsilon M_n(t) + 2\varepsilon Q(X(t))\}x + 2(p + \varepsilon)Q(X(t))|\alpha_n(t-c) - X(t-c)|\\ &\equiv U_n(t)x + V_n(t)|\alpha_n(t-c) - X(t-c)| \quad ,\end{aligned}$$

for all $t \in (t_0,T_n)$, except on a set of probability at most $\tilde{A}\exp\{-\tilde{\alpha}x^2n\}$. Note that $U_n(t)$ and $V_n(t)$ are positive. Moreover, we have

$$U_n(t) + V_n(t) - M_n(t) = -(1 - \varepsilon)M_n(t) + (4\varepsilon + 2p)Q(X(t))$$

$$< (5\varepsilon + 2p - 1)Q(X(t)) = (-\frac{1}{2} + p)Q(X(t)) < 0 \quad .$$

Case (ii). $p \ge \frac{1}{2}$. This implies that

$$X_A^0(\infty) = 1 - X(\infty) + Y(\infty) > 1 + X(\infty)(2p - 1) \ge 1 \quad ,$$

so that we must have $Q(X(\infty)) = 0$ and hence $X(\infty) = b_s$. Let $t_0^* > c$ be so large that $Q(X(t_0^*-c)) \le \frac{1}{2c}$. Since for $t - c \le s \le t$,

$$X'(s) \le (1 - X(t-c) + 2\int_0^{X(t-c)} P(u)du)Q(X(t-c)) \quad ,$$

we have for $t \ge t_0^*$,

$$X(t) - X(t-c) \le \frac{1}{2}(1 - X(t-c) + 2\int_0^{X(t-c)} P(u)du) \quad .$$

Hence, for $t \ge t_0^*$,

$$X_A^0(t-c) = (1 - X(t) + 2\int_0^{X(t-c)} P(u)du)$$

(3.1.43)

$$\ge \frac{1}{2}(1 - X(t-c) + 2\int_0^{X(t-c)} P(u)du)) \ge \frac{1}{2}(1 + X(t-c)(2p - 1)) \ge \frac{1}{2} \quad .$$

By (3.1.34) we know that $q' = Q'(b_s) < 0$. Take $0 \le \varepsilon \le \frac{1}{2}$ so small that $Q(b_s - \varepsilon a_h b_h) < \frac{1}{54}|q'|$ and $2\varepsilon a_h b_h / d(a_s b_s)^2 < \frac{1}{2}|q'|$. Define t_0 as in (3.1.35) and $\tilde{t}_0 = t_0 \vee t_0^*$. It follows from LEMMA 3.1.1 , THEOREM 3.1.1 and LEMMA 3.1.7 that there exist positive numbers $\tilde{A}$, $\tilde{\alpha}$ and $\tilde{C}$ depending on ε , such that for $n = 1,2,...$ and $0 \le x \le \tilde{C}$,

$$P\,(\sup_{t\geq 0}|\alpha_n(t) - X_n(t)| \geq \tfrac{1}{6}x) \leq \tfrac{1}{4}\tilde{A}\exp\{-\tilde{\alpha}x^2n\} \quad ,$$

$$P\,(\sup_{t\leq \tilde{t}_0}|\alpha_n(t) - X(t)| \geq x) \leq \tfrac{1}{4}\tilde{A}\exp\{-\tilde{\alpha}x^2n\} \quad ,$$

$$(3.1.44) \quad P\,(\sup_{t\geq \tilde{t}_0}|X(\infty) - X_n(t-c)| \geq \varepsilon a_h b_h) \leq \tfrac{1}{4}\tilde{A}\exp\{-\tilde{\alpha}x^2n\} \quad ,$$

$$P\,(\sup_{t\geq \tilde{t}_0}\left[|Y_n(t-c) - Y(t-c)| - 2(p+\varepsilon)|X_n(t-c) - X(t-c)|\right] \geq \tfrac{3}{4}x)$$

$$\leq \tfrac{1}{4}\tilde{A}\exp\{-\tilde{\alpha}x^2n\} \quad .$$

Then, except on set $\tilde{G}$ of probability not larger than $\tilde{A}\exp\{-\tilde{\alpha}x^2n\}$, we have for $\tilde{t}_0 < t < T_n$,

$$\alpha_n'(t) - X'(t)$$

$$= -Q(X_n(t))(X_n(t) - X(t)) + X_A^0(t-c)(Q(X_n(t)) - Q(X(t)))$$

$$+ (Y_n(t-c) - Y(t-c))Q(X_n(t))$$

$$= -M_n(t)(X_n(t) - X(t)) + \tilde{R}_n(t) \quad ,$$

where

$$M_n(t) = Q(X_n(t)) - Q'(\theta X_n(t) + (1-\theta)X(t))X_A^0(t-c) \geq \tfrac{1}{4}|q'| > 0 \quad ,$$

$$|\tilde{R}_n(t)| \leq \{\tfrac{3}{4}x + 2(p+\varepsilon)|X_n(t-c) - X(t-c)|\}Q(X_n(t))$$

$$\leq \tfrac{1}{72}|q'|x + \tfrac{1}{18}|q'||X_n(t-c) - X(t-c)| \quad .$$

This follows from (3.1.43) , (3.1.44) and because (cf. (3.1.34) and (3.1.44))

$$Q'(X_n(t)) = q' + Q''(\eta X_n(t) + (1-\eta)X(\infty))(X_n(t) - X(\infty))$$

$$\leq -|q'| + 2\varepsilon a_h b_h/d(a_s b_s)^2 \leq -\tfrac{1}{2}|q'| \quad .$$

But then, except on $\tilde{G}$, we have for $\tilde{t}_0 < t < T_n$,

$$\alpha_n'(t) - X'(t) = -M_n(t)((\alpha_n(t) - X(t)) + R_n(t) \quad ,$$

where

$$\begin{aligned} |R_n(t)| &= |M_n(t)(\alpha_n(t) - X_n(t)) + \tilde{R}_n(t)| \\ &< \{M_n(t) + \tfrac{1}{12}|q'| + \tfrac{1}{18}|q'|\}\tfrac{1}{6}x + \tfrac{1}{18}|q'||\alpha_n(t-c) + X(t-c)| \\ &\le U_n(t)x + V_n(t)|\alpha_n(t-c) - X(t-c)| \end{aligned}$$

with $U_n(t) = \frac{1}{3}M_n(t)$, $V_n(t) = \frac{1}{3}M_n(t)$. Hence $U_n(t) + V_n(t) - M_n(t) < 0$. This proves Assumption (A) and the proof of the theorem is complete.

□

<u>REMARK</u> <u>3.1.1</u> Note that, when $p < \frac{1}{2}$, then A and α in THEOREM 3.1.2 depend on p ; in case $p \ge \frac{1}{2}$, these numbers depend on q' .

<u>COROLLARY</u> <u>3.1.3</u> There exist positive numbers A and α such that for $n = 1,2,...$ and $x \ge 0$,

(3.1.45) $$P\,(\sup_{t\ge 0} |Y_n(t) - Y(t)| \ge x) \le A\exp\{-\alpha x^2 n\} \quad .$$

<u>PROOF.</u> From (3.1.30) it can be seen that combination of (3.1.3) , (3.1.12) , the boundedness of P and THEOREM 3.1.2 yields the desired result.

□

In the end we can say something about the value of $X(\infty)$. Recall that $X(\infty) = \inf\{\, z : 2\bar{P}(z) - z \le -1 \,\} \wedge b_s$ (see (3.1.14)) and that

$$\bar{P}(z) = \int_0^z P(s)ds = (z \wedge b_h) - a_h b_h \log\left[\frac{b_h + a_h b_h}{b_h + a_h b_h - (z \wedge b_h)} \right] \quad .$$

But $2\bar{P}(z) - z$ is a concave function of z , which equals zero for $z = 0$. Hence, if $1 + b_h - 2a_h b_h \log(\, 1 + a_h^{-1}) < 0$, then for some $z_0 < b_h$,

$$1 + z_0 - 2a_h b_h \log((a_h b_h + b_h)/(a_h b_h + b_h - z_0)) = 0$$

If not then,

$$\inf\{ z : 2\bar{P}(z) - z \le -1 \} = 1 + 2b_h - 2a_h b_h \log(1 + a_h^{-1}) \ge b_h .$$

Therefore we conclude

<u>LEMMA</u> <u>3.1.8</u>

(i) If $1 + b_h - 2a_h b_h \log(1 + a_h^{-1}) < 0$ or $b_s < b_h$, then

$$p = P(X(\infty)) > 0 .$$

(ii) If $1 + b_h - 2a_h b_h \log(1 + a_h^{-1}) < 0$, then

$$X(\infty) = \inf\left\{ z : 1 + z - 2a_h b_h \log\left[\frac{a_h b_h + b_h}{a_h b_h + b_h - z} \right] = 0 \right\} \wedge b_s$$
$$< b_h ;$$

if $1 + b_h - 2a_h b_h \log(1 + a_h^{-1}) \ge 0$, then

$$X(\infty) = \{1 + 2b_h - 2a_h b_h \log(1 + a_h^{-1})\} \wedge b_s \ge b_h \wedge b_s .$$

□

3.2. CONVERGENCE OF X_n , $n \to \infty$

Having performed the preliminary work in the foregoing section, we can easily prove the uniform convergence of $X_n(t)$ to the solution $X(t)$ of (3.1.13) for all $t \ge 0$, and determine its rate.

THEOREM 3.2.1

(3.2.1) $$\sup_{t\geq 0} |X_n(t) - X(t)| = \mathcal{O}_P(n^{-1/2}) \quad ,$$

(3.2.2) $$\sup_{t\geq 0} |X_n(t) - X(t)| = \mathcal{O}((n^{-1}\log n)^{1/2}) \quad \text{a.s.}$$

PROOF. The choice of $x = Bn^{-1/2}$ in THEOREM 3.1.2 , where B is a sufficiently large positive constant, immediately proves (3.2.1) .

When, again in THEOREM 3.1.2 , $x = B(n^{-1}\log n)^{1/2}$, B positive and large enough, is taken, then application of the Borel-Cantelli lemma yields (3.2.2) .

□

In view of COROLLARY 3.1.3 we can prove similar results for Y_n and Y :

COROLLARY 3.2.1

(3.2.3) $$\sup_{t\geq 0} |Y_n(t) - Y(t)| = \mathcal{O}_P(n^{-1/2}) \quad ,$$

(3.2.4) $$\sup_{t\geq 0} |Y_n(t) - Y(t)| = \mathcal{O}((n^{-1}\log n)^{1/2}) \quad \text{a.s.}$$

□

Hence we have

CORROLLARY 3.2.2

(3.2.5) $$\lim_{n\to\infty} \sup_{t\geq 0} |X_n(t) - X(t)| = 0 \quad \text{a.s.} \quad ,$$

(3.2.6) $$\lim_{n\to\infty} \sup_{t\geq 0} |Y_n(t) - Y(t)| = 0 \quad \text{a.s.}$$

□

Finally, we formulate the analogous results which follow from LEMMA 3.1.2 .

CORROLLARY 3.2.3

(3.2.7) $$\max_m n^{-1}\left|\sum_{j=1}^{m}(Z_{j,n} - 2P_{j,n})\right| = \mathcal{O}_P(n^{-1/2}) \quad ,$$

(3.2.8) $$\max_m n^{-1}\left|\sum_{j=1}^{m}(Z_{j,n} - 2P_{j,n})\right| = \mathcal{O}((n^{-1}\log n)^{1/2}) \quad \text{a.s.} \quad ,$$

(3.2.9) $$\sup_{t\geq c} n^{-1}|\Lambda_{z,n}(t) - \Lambda_{p,n}(t)| = \mathcal{O}_P(n^{-1/2}) \quad ,$$

(3.2.10) $$\sup_{t\geq c} n^{-1}|\Lambda_{z,n}(t) - \Lambda_{p,n}(t)| = \mathcal{O}((n^{-1}\log n)^{1/2}) \quad \text{a.s.}$$

□

3.3. A CENTRAL LIMIT THEOREM

This section is devoted to the derivation of a central limit theorem for the deviation of X_n from X . Because of the results in the foregoing section, we guess that $n^{1/2}$ should be the appropriate scaling factor. We introduce the following notation. For $t \geq 0$,

$$V_n(t) = n^{1/2}(X_n(t) - X(t)) \quad ,$$

$$V_n^*(t) = n^{1/2}(\alpha_n(t) - X(t)) \quad ,$$

$$U_n(t) = n^{1/2}(Y_n(t) - Y(t)) \quad ,$$

where $\alpha_n = n^{-1}\Lambda_{z,n}$ as before. In this section convergence in distribution is in $D[0,\infty)$ with respect to the Skorohod metric.

LEMMA 3.3.1

(3.3.1) $$n^{1/2}(X_n(t) - \alpha_n(t)) \overset{\mathcal{D}}{\rightarrow} W_0(X(t)) \quad , \text{ as } \quad n \rightarrow \infty \quad ,$$

where W_0 is a standard Wiener process.

PROOF. Since N_n^* in (2.2.13) is a unit Poisson process, we have

$$(3.3.2) \qquad n^{-1/2}(N_n^*(nu) - nu) \xrightarrow{\mathcal{D}} W_0(u) \quad ,$$

where W_0 is a standard Wiener process. Let $A_{z,n}(t) = nu_n$, then

$$\begin{aligned} n^{-1/2}(N_n^*(nu_n) - nu_n) &= n^{-1/2}(N_n(A_{z,n}^{-1}(nu_n)) - n - nu_n) \\ &= n^{-1/2}(N_n(t) - n - A_{z,n}(t)) = n^{1/2}(X_n(t) - \alpha_n(t)) \quad . \end{aligned}$$

Moreover, since with probability 1 ,

$$u_n = n^{-1}A_{z,n}(t) = \alpha_n(t) \to X(t) \quad , \text{ as } \quad n \to \infty \quad ,$$

the continuity of the paths of W_0 guarantees (3.3.1) .

□

LEMMA 3.3.2

$$(3.3.3) \qquad \begin{aligned} &(V_n(t) - V_n^*(t), U_n(t) - 2P(X(t))V_n(t)) \\ &\xrightarrow{\mathcal{D}} (W_0(X(t)), W_1(2\int_c^{X(t)} P(s)(1 - P(s))ds) \quad , \text{ as } \quad n \to \infty \quad , \end{aligned}$$

where W_0 and W_1 are independent standard Wiener processes.

PROOF. Since $V_n(t) - V_n^*(t) = n^{1/2}(X_n(t) - \alpha_n(t))$, the weak convergence of the first component was established in LEMMA 3.3.1 .

Consider $S_n(u) = n^{-1/2} \sum_{j=1}^{[nu]} (Z_{j,n} - 2P_{j,n})$ for $u \in [0,b_h]$ and $n = 1,2,\ldots$. Analogously to (3.1.12) we can prove

$$\max_m \left| \sum_{j=1}^m P_{j,n}(1 - P_{j,n}) - n\int_0^{m/n} P(s)(1 - P(s))ds \right| \le A$$

for some positive constant A . Hence, it follows from the extension of

Donsker's Theorem to sums of independent but not identically distributed random variables (cf. PROHOROV (1956)) that

$$(3.3.4)\qquad S_n(u) \xrightarrow{\mathcal{D}} W_1\left(2\int_0^u P(s)(1-P(s))ds\right) \quad , \text{ as } \quad n \to \infty \quad ,$$

where W_1 is a standard Wiener process. In view of (3.1.12) this means that

$$n^{-1/2}\left(\sum_{j=1}^{[nu]} (Z_{j,n} - 2n\bar{P}(u)\right) \xrightarrow{\mathcal{D}} W_1\left(2\int_0^u P(s)(1-P(s))ds\right) \quad .$$

Since $Y_n(t) = n^{-1} \sum_{j=1}^{nX_n(t)} Z_{j,n}$, $X_n \le b_s$ for all n , and $\sup_{t\ge 0} |X_n(t) - X(t)| = \mathcal{O}(n^{-1}(\log n)^{1/2}))$ a.s. , the continuity of the sample paths of W_1 ensures that

$$n^{1/2}(Y_n(t) - 2\bar{P}(X_n(t)) \xrightarrow{\mathcal{D}} W_1\left(2\int_0^{X(t)} P(s)(1-P(s))ds\right) \quad .$$

Now $U_n(t) = n^{1/2}(Y_n(t) - Y(t))$, $V_n(t) = n^{1/2}(X_n(t) - X(t))$, $Y(t) = 2\bar{P}(X(t))$ and $\bar{P}'(u) = P(u)$. In view of (3.1.34) and THEOREM 3.1.2 this implies

$$(3.3.5)\qquad U_n(t) - 2P(X(t))V_n(t) \xrightarrow{\mathcal{D}} W_1\left(2\int_c^{X(t)} P(s)(1-P(s))ds \quad , \text{ as } \quad n \to \infty \quad ,\right.$$

which proves the convergence of the second component in (3.3.3) .

As we pointed out below (2.2.13) , the processes N_n^* and S_n are independent, so that we have joint convergence in (3.3.2) and (3.3.4) and hence in (3.3.3) if W_0 and W_1 are chosen to be independent.

□

LEMMA 3.3.3

$$\sup_{t\geq c} \{V_n^{*\prime}(t) - [V_n(t)\{-Q(X(t)) + X_A^0(t-c)Q'(X(t))\} + U_n(t-c)Q(X(t))]\} = o(1) \quad , \quad \text{a.s.} \quad \text{as} \quad n \to \infty \quad .$$

PROOF. In view of (3.1.34), (3.2.2) and (3.2.4), we have with probability 1, uniformly in $t \geq c$,

$$\begin{aligned} V_n^{*\prime}(t) &= n^{1/2}(-(X_n(t) - X(t)) + Y_n(t-c) - Y(t-c))Q(X(t)) \\ &\qquad + n^{1/2}(1 - X_n(t) + Y_n(t-c))(Q(X_n(t)) - Q(X(t))) \end{aligned} \tag{3.3.6}$$

$$\begin{aligned} &= -V_n(t)Q(X(t)) + U_n(t-c)Q(X(t)) \\ &\qquad + (1 - X_n(t) + Y_n(t-c))Q'(X(t))V_n(t) + o(1) \\ &= -V_n(t)Q(X(t)) + U_n(t-c)Q(X(t)) \\ &\qquad + (1 - X(t) + Y(t-c))Q'(X(t))V_n(t) + o(1) \\ &= V_n(t)\{-Q(X(t)) + X_A^0(t-c)Q'(X(t))\} + U_n(t-c)Q(X(t)) + o(1) \quad , \\ &\qquad \text{as} \quad n \to \infty \quad . \end{aligned}$$

□

LEMMA 3.3.4 The sequence V_n is tight and every weak limit point is in $C[0,\infty)$ a.s.

PROOF. In view of Lemma 3.3.1, we only have to show that V_n^* is tight and that every weak limit point of V_n^* is in $C[0,\infty)$ a.s. From (3.3.6) and the boundedness of Q' it follows that

$$\sup_{t\geq c} |V_n^{*\prime}(t)| \leq A\{\sup_{t\geq c} |V_n(t)| + \sup_{t\geq c} |U_n(t)|\}$$

for some positive constant A . Therefore, we see from THEOREM 3.2.1 and COROLLARY 3.1.3 that positive $\tilde{A}$ and $\tilde{\alpha}$ exist such that

$$P\,(\sup_{t\geq c} |V_n^{*\prime}(t)| \geq x) \leq \tilde{A}\exp\{-\tilde{\alpha}x^2\} \leq \frac{\tilde{A}}{\tilde{\alpha}ex^2} \quad .$$

Hence, for $\varepsilon > 0$, $\eta > 0$ and $\delta = (\tilde{\alpha}e\eta^2\varepsilon/\tilde{A})^{1/2}$,

$$P\,(\sup\{\ |V_n^*(t) - V_n^*(s)|\ :\ c \leq s < t\ ,\ \ |t - s| \leq \delta\ \} \geq \eta)$$

$$\leq P\,(\sup_{t\geq c} |V_n^{*\prime}(t)| \geq \frac{\eta}{\delta}\ ) \leq \frac{\tilde{A}\delta^2}{\tilde{\alpha}e\eta^2} = \varepsilon \quad .$$

In view of POLLARD (1984), p.131 , this ensures that V_n^* is tight in $D[0,\infty)$ with respect to the Skorohod metric. It also implies, however, that V_n^* is tight in $C[0,\infty)$ with respect to the supremum norm on bounded intervals (cf. WHITT (1970)). The standard subsequence argument shows that every weak limit point of V_n^* is in $C[0,\infty)$ a.s.

□

Suppose that V and $\tilde{V}$ are weak limit points of V_n . Assuming our probability space to be sufficiently rich, LEMMAS 3.3.2 , 3.3.3 and a Skorohod construction allow us to redefine V_n , V_n^* , U_n , V , $\tilde{V}$, W_0 and W_1 such that for some subsequences $\{n_k\}$ and $\{\tilde{n}_k\}$, and all t ,

$$\lim_{k\to\infty} V_{n_k}(t) = V(t) \quad , \qquad \lim_{k\to\infty} V_{\tilde{n}_k}(t) = \tilde{V}(t) \quad ,$$

$$\lim_{n\to\infty} (V_n(t) - V_n^*(t)) = W_0(X(t)) \quad ,$$

$$(3.3.7) \qquad \lim_{n\to\infty} \big(U_n(t) - 2P(X(t))V_n(t)\big) = W_1(2\int_0^{X(t)} P(s)(1 - P(s))ds) \quad ,$$

$$\lim_{n\to\infty} \Big(V_n^*(t) - \int_c^t \big[V_n(s)\{-Q(X(s)) + X_A^0(s-c)Q'(X(s))\}$$

$$+ U_n(s-c)Q(X(s))\big]\,ds\Big) = 0 \quad ,$$

with probability one. Here we have used the continuity of the sample paths of W_0, W_1, V, $\tilde{V}$ (cf. LEMMA 3.3.4) to conclude that almost sure convergence of V_{n_k}, $V_{\tilde{n}_k}$ and $(V_n - V^*_n)$ in the Skorohod metric ensures almost sure convergence in the supremum metric on finite intervals, and hence for all t a.s. Hence

$$V(t) = W_0(X(t)) + \lim_{k\to\infty} V^*_{n_k}(t)$$

$$= W_0(X(t)) + \int_c^t V(s)\{-Q(X(s)) + X^0_A(s-c)Q'(X(s))\}$$

$$+ \int_c^t \Big[W_1\Big(2\int_0^{X(s-c)} P(u)(1 - P(u))du\Big) + 2P(X(s-c))V(s-c)\Big] Q(X(s))ds$$

with probability one, and the same holds with n_k replaced by $\tilde{n}_k$ and V replaced by $\tilde{V}$ throughout. Here the integrals are defined to be zero for $0 \le t < c$.

Consider $F^* : [0,\infty) \times D[0,\infty) \to \mathbb{R}$ defined by

$$F^*(t,v) = W_1\Big(2\int_c^{X(t-c)} P(s)(1 - P(s))ds\Big)Q(X(t))$$

$$+ v(t)\{-Q(X(t) + X^0_A(t-c)Q'(X(t))\} \tag{3.3.8}$$

$$+ 2v(t-c)P(X(t-c))Q(X(t)) \quad ,$$

and the system

$$v(t) = \begin{cases} W_0(X(t)) + \int_c^t F^*(s,v)ds \quad , & t \ge c \quad , \\ 0 \quad , & t < c \quad . \end{cases} \tag{3.3.9}$$

Then V and $\tilde{V}$ are solutions of (3.3.9) almost surely. Since for all t, $X(t) \le b_s$, P-almost every realization of $W_0(X(t))$ and

$W_1(2\int_0^{X(t)} P(s)(1 - P(s))ds)$ are bounded. Therefore, because of the boundedness of X, P, $\bar{P}$, Q and Q', and the continuity of the sample paths of W_0 and W_1, the system (3.3.9) satisfies the conditions of LEMMA A2.2 in Appendix A2. Hence, with probability 1, (3.3.9) has a unique solution on $[0,\infty)$, so that $V = \tilde{V}$ a.s. Since V_n is tight, it follows that $V_n \xrightarrow{\mathcal{D}} V$, where V is the unique solution of (3.3.9) and W_0 and W_1 are independent.

We finally remark that V is a Gaussian process. According to the proof of LEMMA A2.2 in Appendix A2, we have

$$\lim_{n\to\infty} v_n = V \quad \text{a.s.} \quad ,$$

where

$$v_0(t) = W_0(X(t)) \quad ,$$

$$v_{n+1}(t) = W_0(X(t)) + \int_c^t F^*(s,v_n)ds \quad \text{a.s.}$$

By induction v_n is a Gaussian process for every $n = 1,2,\ldots$, and as a result this is also true for V. Obviously, $EV = 0$, but its covariance structure is not immediately clear. We thus have proved

<u>THEOREM 3.3.1</u> Let W_0 and W_1 be independent standard Wiener processes, let F^* be defined by (3.3.8), and let V be the unique solution of (3.3.9). Then V is a Gaussian process with mean zero and

$$V_n \xrightarrow{\mathcal{D}} V \quad , \quad \text{as } n \to \infty \quad ,$$

where the convergence is in the sense of the Skorohod metric.

□

3.4. THE DURATION OF N_n

Now we shall investigate what can be said about the duration T_n of the process N_n. Let $\tau_n(\varepsilon)$ and $\tau(\varepsilon)$ be defined by

$$(3.4.1) \qquad \tau_n(\varepsilon) = \inf\{\ t : X_n(T_n) - X_n(t) \leq \varepsilon\ \} = \tau_{nX_n(T_n)-[\varepsilon n],n} \ ,$$

$$(3.4.2) \qquad X(\infty) - X(\tau(\varepsilon)) = \varepsilon \ .$$

Then the following holds.

LEMMA 3.4.1 There exist positive constants $\tilde{A}$ and $\tilde{\alpha}$ such that for $n = 1,2,\ldots$ and $\varepsilon > 0$,

$$P\ (\tau_n(\varepsilon) > \tau(\varepsilon/2)) \leq \tilde{A}\exp\{-\tilde{\alpha}\varepsilon^2 n\}$$

and hence $\tau_n(\varepsilon) = O(1)$ a.s.

PROOF. We have

$$|X_n(T_n) - X_n(\tau(\varepsilon/2))| \leq 2\sup_{t\geq 0} |X_n(t) - X(t)| + |X(\infty) - X(\tau(\varepsilon/2))|$$

and in view of THEOREM 3.1.2,

$$P\ (\tau_n(\varepsilon) > \tau(\varepsilon/2)) \leq P\ (X_n(T_n) - X_n(\tau(\varepsilon/2)) > \varepsilon)$$

$$\leq P\ (\sup_{t\geq 0} |X_n(t) - X(t)| > \varepsilon/4) \leq A\exp\{-\alpha\varepsilon^2 n/16\} \ .$$

□

However, the total duration of the process N_n is not bounded. To find bounds for T_n, we distinguish two cases.

3.4.1. THE CASE $X_A^0(\infty) \neq 0$ OR $q \neq 0$

Recall that $X_A^0(\infty) = 1 - X(\infty) + Y(\infty)$, $q = Q(X(\infty))$.

LEMMA 3.4.2 If $X_A^0(\infty) \neq 0$ or $q \neq 0$, then there exist positive constants C_0 , C_1 , ε_0 , ε_1 such that for all $t > c$,

$$(3.4.3) \qquad C_0\exp\{-\varepsilon_0 t\} \leq X(\infty) - X(t) \leq C_1\exp\{-\varepsilon_1 t\} \quad .$$

PROOF. First, note that the left-hand inequality in (3.4.3) is already proved in Section 3.1 (cf. (3.1.15)). Next, to prove the right-hand inequality, we distinguish two cases.

(i) $X_A^0(\infty) \neq 0$. In this case $X(\infty) = b_s$. It follows from LEMMA 3.1.4 that $X'(t) = X_A^0(t-c)Q(X(t))$ and both its factors are continuous and positive on (c,∞) . Thus there exists $\eta > 0$ such that for all $t \geq c$,

$$(3.4.4) \qquad X_A^0(t-c) \geq \eta$$

and hence $X'(t) \geq \eta Q(X(t))$. Since

$$Q'(X(t)) \leq Q'(0) < 0 \quad ,$$

we have for $t \geq c$,

$$X'(t) \geq \eta(Q(X(t)) - Q(X(\infty))) \geq -\eta Q'(0)(X(\infty) - X(t)) \quad ,$$

so that

$$-\eta Q'(0)(t - c) \leq \int_c^t \frac{X'(s)}{X(\infty) - X(s)} ds = \log\left[\frac{X(\infty)}{X(\infty) - X(t)} \right] \quad ,$$

or

$$X(\infty) - X(t) \leq X(\infty)\exp\{\eta Q'(0)(t - c)\} \quad .$$

Since for $t < c$, $X(t) = X(c) = 0$, this inequality holds for $t \geq 0$.

(ii) $q \neq 0$. This implies that $X(\infty) < b_s$ and by (3.1.14) , $2\bar{P}(X(\infty)) = X(\infty) - 1$ and for $t \geq c$,

$$X'(t) = \left[X(\infty) - X(t) - 2\int_{X(t-c)}^{X(\infty)} P(u)du\right] Q(X(t)) \quad . \tag{3.4.5}$$

Also $X(\infty) - 1 = 2\bar{P}(X(\infty)) > 2pX(\infty)$ ensures that $p < \frac{1}{2}$, so that $0 < \delta < \frac{1}{2}$ and $t_0 \geq 2c$ exist such that

$$P(X(t)) < \frac{1}{2} - \delta \qquad \text{for } t \geq t_0 \quad . \tag{3.4.6}$$

Hence for $s \geq t_0 + c$,

$$X'(s) > \left[X(\infty) - X(s) - (1 - 2\delta)(X(\infty) - X(s-c))\right] q$$

and for $t \geq t_0$, we may integrate over $(t + c, t + 2c)$ to obtain

$$X(t+2c) - X(t+c) > qc\left[X(\infty) - X(t+2c) - (1 - 2\delta)(X(\infty) - X(t))\right] \quad .$$

Replacing the left-hand side by $X(t+2c) - X(t)$, we find

$$\left[1 + qc(1 - 2\delta)\right](X(t+2c) - X(t)) > 2\delta qc(X(\infty) - X(t+2c))$$

or

$$\frac{X(\infty) - X(t)}{X(\infty) - X(t+2c)} > 1 + \frac{2\delta qc}{1 + qc(1 - 2\delta)} = e^{2\varepsilon c}$$

for an appropriately chosen $\varepsilon > 0$. But this means for $k = 1,2,\ldots$, and $t_0 + 2kc \leq t < t_0 + 2(k + 1)c$,

$$\frac{X(\infty) - X(t)}{X(\infty) - X(t_0)} < e^{-2\varepsilon ck} \leq \exp\{-\tfrac{1}{2}\varepsilon(t - t_0)\} \quad ,$$

which completes the proof.

□

Having proved LEMMA 3.4.2 , we can say more about the duration T_n of N_n . Indeed, since $|X(\infty) - X(T_n)| = \mathcal{O}((n^{-1}\log n)^{1/2})$ a.s. , it follows that for $\mathcal{P}$-almost all ω , for some constant $C > 0$ and $n > n_0(\omega)$,

$$C(n^{-1}\log n)^{1/2} \geq |X(\infty) - X(T_n)| \geq C_0 \exp\{-\varepsilon_0 T_n\}$$

and hence there exists a positive constant $\tilde{C}_0$ such that

(3.4.7) $T_n \geq \tilde{C}_0 \log n$.

To establish an upper bound for T_n , we again distinguish two cases.

First, if $q = 0$ and $X_A^0(\infty) > 0$, then we already concluded (see (3.4.4)) that for some $\eta > 0$,

$$\inf_{t \geq c} X_A^0(t-c) \geq \eta \quad .$$

Hence in view of (3.2.2) and (3.2.4) , we have with probability 1 for sufficiently large n ,

$$\inf_{t \geq c} X_{A,n}^0(t-c) > \tfrac{1}{2}\eta = \delta \quad .$$

Therefore $X_n(\infty) = n^{-1} n_{s,n}$. Let $0 < \varepsilon < 1$ and $\tau_n(\varepsilon)$ as defined in (3.4.1) . Then (3.1.6) and (3.1.33) imply that for sufficiently large n , $T_n - \tau_n(\varepsilon)$ is stochastically smaller than the sum S of $[\varepsilon n]$ independent exponential random variables with means A/j , $j = 1,2,\ldots,[\varepsilon n]$, where $A = db_s(1 + a_s)/\delta$. We have

$$Ee^{tS} = \prod_{j=1}^{[\varepsilon n]} \frac{j/A}{j/A - t} \leq 2 \prod_{j=2}^{[\varepsilon n]} \frac{j}{j-1} \leq 2\varepsilon n \qquad \text{for } t \leq (2A)^{-1} \quad .$$

and hence for $t = (2A)^{-1}$ and $x = 6A\log n$,

$$P\,(T_n - \tau_n(\varepsilon) \geq x) \leq P\,(S \geq x) \leq \frac{Ee^{tS}}{e^{tx}} \leq 2\varepsilon n^{-2} \quad .$$

The Borel-Cantelli lemma and LEMMA 3.4.1 yield

(3.4.8) if $q = 0$ and $X_A^0(\infty) > 0$, then $T_n = \mathcal{O}(\log n)$ a.s.

Next, suppose $q > 0$, $X_A^0(\infty) = 0$. Consider a process which starts with a single A-cell. Cells divide independently after an exponential waiting time with mean q^{-1} plus a constant time c . With each division cells independently become A-cells with probability $p \in [0,\frac{1}{2})$. Let $T(c)$ denote the time until the final division.

<u>LEMMA</u> <u>3.4.3</u> For every $q > 0$, $c > 0$ and $0 \le p < \frac{1}{2}$, there exists $\varepsilon > 0$ such that

$$(3.4.9) \quad P\,(T(c) \ge t) \le 2e^{-\varepsilon(t - 2c)} \qquad \text{for all } t > 2c \ .$$

<u>PROOF.</u> $T(c)$ is stochastically smaller than $T(0) + (N + 1)c$, where $T(0)$ is the time until the final division for $c = 0$, and N is the random number of generations the initial cell will ultimately produce, which is independent of c . Let G_j be the number of cells in the j-th generation, $j = 0,1,2,\ldots$. Then $P\,(G_0 = 1) = 1$ and the Markov inequality yields

$$P\,(G_j \ge x) \le x^{-1}(2p)^j \ , \qquad x > 0 \ , \quad j = 1,2,\ldots \ .$$

Hence

$$(3.4.10) \quad P\,(N > k) = P\,(G_k \ge 1) \le (2p)^k = e^{-k\log(2p)^{-1}} \ , \qquad k = 1,2,\ldots \ .$$

Next, let $M_0(t)$ denote the number of A-cells at time t for $c = 0$. Then, for $h \downarrow 0$,

$$E(h^{-1}(M_0(t+h) - M_0(t)) \mid M_0(t) = k)$$

$$= \{p^2 - (1 - p)^2\}qk + o(1)$$

$$= (2p - 1)qk + o(1) \ .$$

Hence, if $m(t) = EM_0(t)$, we get

$$m'(t) = (2p - 1)qm(t) \ ,$$

or

$$m(t) = m(0)e^{-(1 - 2p)qt} .$$

Since $m(0) = 1$, the Markov inequality gives

(3.4.11) $\mathcal{P}(T(0) > t) = \mathcal{P}(M_0(t) \geq 1) \leq \exp\{-(1 - 2p)qt\}$.

Now for $t > 2c$, (3.4.10) and (3.4.11) imply

$$\mathcal{P}(T(c) \geq t) \leq \mathcal{P}(T(0) + (N + 1)c \geq t) = \mathcal{P}(T(0) + Nc \geq t - c)$$

$$\leq \mathcal{P}(T(0) \geq 2^{-1}(t - 2c)) + \mathcal{P}(N \geq (2c)^{-1}t)$$

$$\leq \exp\{-(1 - 2p)q2^{-1}(t - 2c)\} + \exp\{-[(2c)^{-1}t - 1]\log(2p)^{-1}\}$$

$$\leq 2\exp\{-\varepsilon(t - 2c)\} ,$$

where $\varepsilon = (2^{-1}(1 - 2p)q) \wedge ((2c)^{-1}\log(2p)^{-1}) > 0$.

□

Let us return to the original problem. Since $X_A^0(\infty) = 0$, there exists t_0 such that $p_0 = P(X_n(t_0)) < \frac{1}{2}$ for $n \geq n(\omega)$. Then $T_n - t_0$ is stochastically smaller than the maximum of $n(b_s + 1)$ independent copies of $T(c)$ with $p = p_0$. It follows from LEMMA 3.4.3 that there exists $\varepsilon > 0$ so that for $t > 2c$,

$$\mathcal{P}(T_n - t_0 \geq t) \leq 2n(b_s + 1)\exp\{-\varepsilon(t - 2c)\} ,$$

so

$$\mathcal{P}(T_n \geq t_0 + 2c + 3\varepsilon^{-1}\log n) \leq 2(b_s + 1)n^{-2} .$$

The Borel-Cantelli lemma yields

(3.4.12) if $q > 0$ and $X_A^0(\infty) = 0$, then $T_n = \mathcal{O}(\log n)$ a.s. .

3.4.2. THE CASE $X_A^0(\infty) = 0$ AND $q = 0$

LEMMA 3.4.4 If $X_A^0(\infty) = 0$ and $q = 0$, then there exist positive constants C_0, C_1, ε_0, ε_1 such that for all t,

$$(3.4.13)\quad C_0(1 + \varepsilon_0 t)^{-1} \le X(\infty) - X(t) \le C_1(1 + \varepsilon_1 t)^{-1} .$$

PROOF. We have $X_A^0(\infty) = 1 - X(\infty) + Y(\infty) = 0$, and hence $p = P(X(\infty)) < \frac{1}{2}$ and (3.4.5) holds. Choose $0 < \delta < \frac{1}{2}$ and $t_0 \ge c$ so that for $t \ge t_0$,

$$P(X(t)) < \frac{1}{2} - \delta ,$$

$$X(\infty) - X(t) < [c(1 - 2\delta)|Q'(0)|]^{-1} .$$

Since $X_A^0(t) > 0$ and $Q(X(t)) = Q(X(t)) - Q(X(\infty)) \ge |Q'(0)|(X(\infty) - X(t))$ for all $t > c$, (3.4.5) ensures that for $t \ge t_0 + c$,

$$X'(t) \ge [X(\infty) - X(t) - 2\int_{X(t-c)}^{X(\infty)} P(u)du]\,|Q'(0)|(X(\infty) - X(t))$$

$$\ge [X(\infty) - X(t) - (1 - 2\delta)(X(\infty) - X(t-c))]\,|Q'(0)|(X(\infty) - X(t)) .$$

It follows that for $t \ge t_0$ and $t + c \le s \le t + 2c$,

$$\frac{X'(s)}{(X(\infty) - X(s))^2} \ge |Q'(0)|[2\delta - (1 - 2\delta)\frac{X(t+2c) - X(t)}{X(\infty) - X(t+2c)}]$$

and integrating over $(t + c, t + 2c)$, we obtain

$$(X(\infty) - X(t+2c))^{-1} - (X(\infty) - X(t))^{-1}$$

$$> (X(\infty) - X(t+2c))^{-1} - (X(\infty) - X(t+c))^{-1}$$

$$\ge c|Q'(0)|[2\delta - (1 - 2\delta)\frac{X(t+2c) - X(t)}{(X(\infty) - X(t+2c))(X(\infty) - X(t))} \cdot \frac{1}{c(1 - 2\delta)|Q'(0)|}]$$

or

$$(X(\infty) - X(t+2c))^{-1} - (X(\infty) - X(t))^{-1} > \delta c|Q'(0)|$$

for $t \geq t_0$. If $t_0 + 2kc \leq t \leq t_0 + 2(k + 1)c$ for $k = 1,2,\ldots$, then

$$\begin{aligned}(X(\infty) - X(t))^{-1} &\geq (X(\infty) - X(t_0+2kc))^{-1} \\ &\geq (X(\infty) - X(t_0))^{-1} + \tfrac{1}{4}\delta|Q'(0)|(t - t_0)\end{aligned}$$

which proves the right-hand inequality in (3.4.13) for $t \geq 2t_0$, and by an appropriate choice of C_1 , for all t .

To prove the left-hand inequality we merely remark that $Q(X(t)) \leq (da_s b_s)^{-1}(X(\infty) - X(t))$ and hence for all t ,

$$X'(t) \leq (da_s b_s)^{-1}(X(\infty) - X(t))^2 .$$

It follows that for all t ,

$$(X(\infty) - X(t))^{-1} - X(\infty)^{-1} = \int_c^t \frac{X'(s)}{(X(\infty) - X(s))^2}\, ds \leq (da_s b_s)^{-1} t$$

which implies the inequality we need.

□

From (3.2.2) and LEMMA 3.4.4 it now follows that, if $X_A^0(\infty) = 0$ and $q = 0$, then for some $C > 0$ and $n > n(\omega)$,

$$C(n^{-1}\log n)^{1/2} \geq X(\infty) - X(T_n) \geq C_0(1 + \varepsilon_0 T_n)^{-1} .$$

This means: there exists $\tilde{C} > 0$ such that for $n > n(\omega)$,

(3.4.14) $T_n \geq \tilde{C}(n(\log n)^{-1})^{1/2}$ a.s.

To find an upper bound for T_n , let $0 < \varepsilon < 1$. The last $[\varepsilon n]$ stimuli occur at $\tau_n(j/n) - c$ for $j = 1,2,\ldots,[\varepsilon n]$. The corresponding Q's at these times are at least as large as $j(dn)^{-1}(a_s b_s + \varepsilon)^{-1}$ for $j = 1,2,\ldots,[\varepsilon n]$. Since at least one A-cell is waiting for a stimulus just before a stimulus occurs, the corresponding rates exceed

(3.4.15) $(An)^{-1}$, $2(An)^{-1}$, ... , $[\varepsilon n](An)^{-1}$,

respectively, where $A = d(a_s b_s + 1)$. We remark that it may happen that after the previous stimulus there were no waiting A-cells present for some time. This time period is at most of length c . Hence $T_n - \tau_n(\varepsilon)$ is stochastically smaller than εnc plus the sum S of $[\varepsilon n]$ independent exponential random variables with means An/j , $j = 1,2,\ldots,[\varepsilon n]$. We have

$$Ee^{tS} = \prod_{j=1}^{[\varepsilon n]} \frac{j/An}{j/An - t} \le 2 \prod_{j=2}^{[\varepsilon n]} \frac{j}{j-1} \le 2\varepsilon n \qquad \text{for } t \le (2An)^{-1} .$$

Therefore, we get for $t = (2An)^{-1}$,

$$\mathcal{P}\,(T_n - \tau_n(\varepsilon) \ge \varepsilon nc + 6An\log n) \le \mathcal{P}\,(S \ge 6An\log n)$$

$$\le \exp\{-6tAn\log n\}Ee^{tS} \le 2\varepsilon n^{-2} .$$

Invoking the Borel-Cantelli lemma we get

(3.4.16) $T_n = \mathcal{O}(n\log n)$ a.s.

Note that the quantities in (3.4.15) possibly crudely underestimate the actual rates, since the number of waiting A-cells may vary from 1 to $\varepsilon n + \mathcal{O}((n\log n)^{1/2})$. The gap between this upper bound and the lower bound in (3.4.14) probably is the result of this underestimation. We conjecture that (3.4.16) can be sharpened.

3.4.3. SUMMARY

Summarizing the results of this section, we have for the total duration T_n of the process N_n :

THEOREM 3.4.1

(i) If $X_A^0(\infty) \neq 0$ or $q \neq 0$, then

$$0 < a \leq \liminf T_n(\log n)^{-1} \leq \limsup T_n(\log n)^{-1} \leq A < \infty \quad \text{a.s.}$$

(ii) If $X_A^0(\infty) = 0$ and $q = 0$, then

$$\liminf T_n(n^{-1}\log n)^{1/2} \geq b > 0 \quad \text{a.s.} \quad ,$$

$$\limsup T_n(n\log n)^{-1} \leq B < \infty \quad \text{a.s.}$$

Here a , A , b and B are constants.

□

3.5. GRAPHICAL COMPARISON OF X_n and X

In Chapter 1 we remarked that plant cell population growth is usually described by a deterministic model given by a set of differential equations. Such a deterministic model obeying the same mechanics as our stochastic model, would be given by

$$(3.5.1) \quad \begin{cases} S'(t) = -y_s^{-1}N_A^0(t)q(t) \quad , \\ H'(t) = -y_h^{-1}N_A^0(t-c)q(t-c)1_{\{t:H(t) > 0\}}(t) \quad , \\ N_A^{0\,\prime}(t) = -N_A^0(t)q(t) + N_A^0(t-c)q(t-c)2p(t) \quad , \\ N_A'(t) = N_A^0(t-c)q(t-c)(2p(t) - 1) \quad , \\ N'(t) = N_A^0(t-c)q(t-c) \quad , \end{cases}$$

where

$$q(t) = \frac{S(t)}{d(S(t) + N(0)k_s)} \quad , \quad p(t) = \frac{H(t)}{H(t) + N(0)k_h} \quad ,$$

k_s and k_h are constants, and the initial conditions are given by

$$(3.5.2) \quad \begin{cases} S(t) = \begin{cases} 0 \quad , & t \in [-c,0) \quad , \\ N(0)b_s y_s^{-1} \quad , & t = 0 \quad , \end{cases} \\ H(t) = \begin{cases} 0 \quad , & t \in [-c,0) \quad , \\ N(0)b_h y_h^{-1} \quad , & t = 0 \quad , \end{cases} \\ N_A^0(t) = N_A(t) = N(t) = N(0) \quad , \qquad t \in [-c,0] \quad , \end{cases}$$

where b_s , b_h , y_s , y_h are constants. Then obviously for $t \geq 0$,

$$S'(t-c) = -y_s^{-1} N'(t) \quad , \qquad H'(t) = -y_h^{-1} N'(t) \quad .$$

Using these relations it is not hard to see that the relative growth $X(t) = N(0)^{-1}(N(t) - N(0))$, $t \geq 0$, satisfies (3.1.13) . Hence, (3.5.1) together with (3.5.2) describe the deterministic model which corresponds to the limit as $n \to \infty$ of the stochastic model introduced in Chapter 2 : here S corresponds with S_n , etc.

To get an idea what the predicted cell number looks like for the two models, we have simulated the stochastic growth curve and computed the solution of (3.1.13) numerically for several sets of parameter values. In Fig. 3.1 four computer simulations of the stochastic process $N_n(t)$ of our model are shown. The parameter values were taken the same for all curves. In Fig. 3.2 three pairs of curves are shown. For the two curves of a pair the same parameter values were taken. The dotted curve of a pair was obtained by computer simulation of the stochastic growth process; the solid one is a numerically computed solution of (3.1.13) . In all cases the initial cell number was 100 . A detailed description of the simulation and computing techniques can be found in VAL (1987).

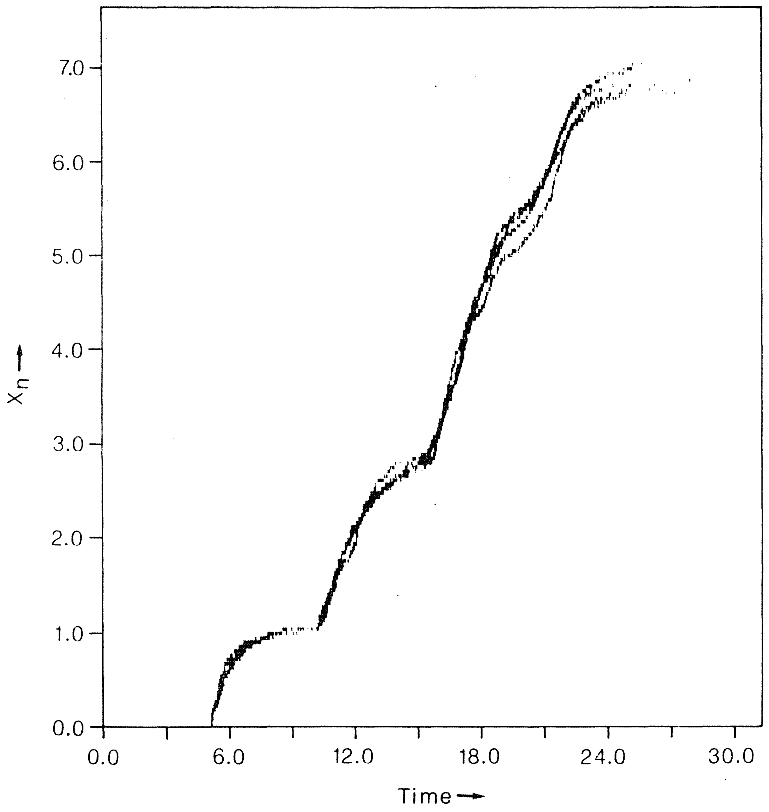
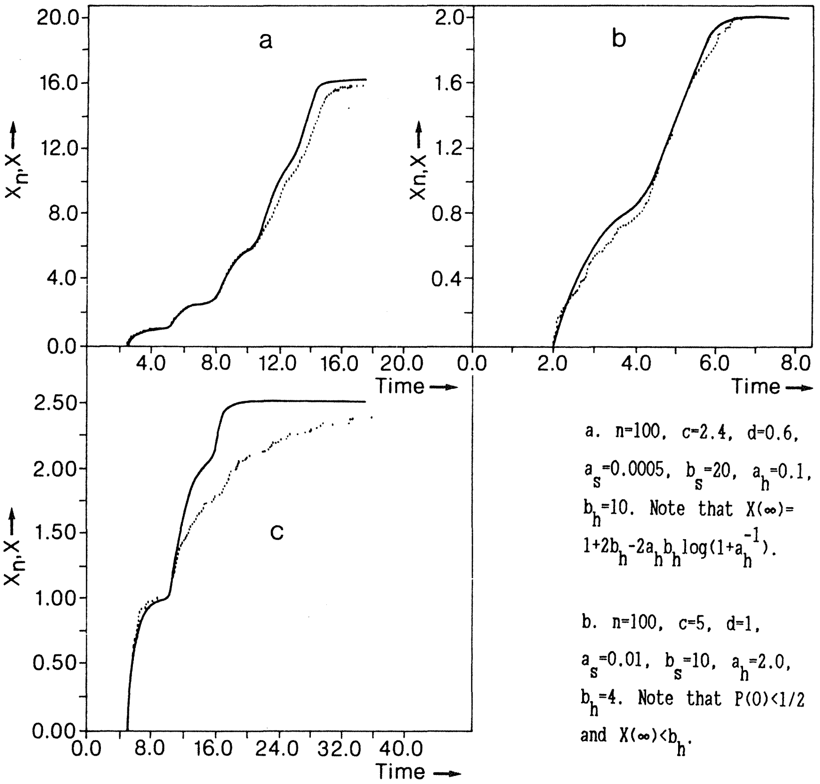


Fig.3.1. Four simulations of the stochastic growth process $X_n(t)$ with the same parameter values: $n=100$, $c=5$, $d=1$, $a_s=0.01$, $b_s=10$, $a_h=0.125$, $b_h=4$.



a. n=100, c=2.4, d=0.6, a_s=0.0005, b_s=20, a_h=0.1, b_h=10. Note that $X(\infty)=1+2b_h-2a_hb_h\log(1+a_h^{-1})$.

b. n=100, c=5, d=1, a_s=0.01, b_s=10, a_h=2.0, b_h=4. Note that $P(0)<1/2$ and $X(\infty)<b_h$.

c. n=100, c=2, d=1, a_s=0.05, b_s=2, a_h=0.025, b_h=4. Note that $X(\infty)=b_s$.

Fig.3.2. Three pairs of curves. The dotted curves are computer simulations of $X_n(t)$, the solid ones are numerically computed solutions of (3.1.13). Curves of a pair have the same parameter values.

CHAPTER 4

AN ALTERNATIVE APPROACH

In this chapter we show that some of the results derived in Chapter 3 can be arrived at by using known theory for population processes. The merits of both approaches are compared. The method applied in this chapter is due to Kurtz and can be found in KURTZ (1981; 1983) or ETHIER & KURTZ (1986).

The basic idea of this method is to represent counting processes having complicated dependence properties in terms of counting processes having strong independence properties. This results in a law of large numbers (e.g. Theorem 2.1 in KURTZ (1983)) and a central limit theorem (e.g. Theorem 2.2 in KURTZ (1983)) for the processes with dependence structure. To prove the latter, Gaussian approximation results for the processes with the independence properties are used.

To apply this theory directly to our counting process N_n as defined in Chapter 2 , we need an explicit expression in terms of $\{N_n(s)\}_{s\le t}$ for its unconditional intensity process $\Lambda_n(t)$ as defined in (2.2.14) . However, such an expression is not available. To obtain it, one would have to compute the conditional expectation with respect to $\hat{Z}_n$ of the conditional intensity process $\Lambda_{Z,n}(t)$ given $\{N_n(s)\}_{s\le t}$, but this would not yield a very manageable expression. In view of this, we could think of another approach by applying the theory to the bivariate process $\{N_n, Z_n\}$, where $Z_n(t) = \sum_{j=1}^{N_n(t)-n} Z_{j,n}$ and use $\Lambda_{z,n}(t)$ as the intensity of the first component. However, $\{N_n, Z_n\}$ is not a counting process and this does not work.

To overcome these problems, we construct a trivariate counting process $\tilde{N}_n$ which is closely related to N_n . For its intensity process we have a simple expression in terms of $\{\tilde{N}_n(s)\}_{s\le t}$. We apply Kurtz's theory to $\tilde{N}_n$ and from the results obtained we derive analogous results for N_n .

4.1. CONVERGENCE

Let $N_{AA,n}(t)$, $N_{AB,n}(t)$ and $N_{BB,n}(t)$ be the number of stimuli before or at time t resulting in two A-cells, one A-cell and one B-cell and two B-cells, respectively. Let

$$\bar{N}_n(t) = N_{AA,n}(t) + N_{AB,n}(t) + N_{BB,n}(t)$$

denote the total number of stimuli at or before time t. Recall that the number of type-A cells waiting for a stimulus at time t is denoted by $N^0_{A,n}(t)$. Then, according to the definitions in Chapter 2, for $t \geq 0$,

$$(4.1.1)\quad \begin{aligned} N_n(t) &= n + \bar{N}_n(t-c)1_{[c,\infty)}(t) \quad , \\ N^0_{A,n}(t) &= n - \bar{N}_n(t) + (2N_{AA,n}(t-c) + N_{AB,n}(t-c))1_{[c,\infty)}(t) \quad , \end{aligned}$$

and for the number $Z_n(t)$ of newly born A-cells before or at time t, $t \geq c$, we have

$$(4.1.2)\quad Z_n(t) = \sum_{j=1}^{N_n(t)-n} Z_{j,n} = 2N_{AA,n}(t-c) + N_{AB,n}(t-c) \quad .$$

Let $D_{\mathbb{R}^3}[0,\infty)$ denote the space of left-continuous functions with right-hand limits on $[0,\infty)$. Define for $t \geq 0$ and $x \in D_{\mathbb{R}^3}[0,\infty)$, the scalar-valued functions $\tilde{x}^0_A(t,x)$ and $\lambda_{i,n}(t,x)$, $i = 1,2,3$, by

$$(4.1.3)\quad \begin{aligned} \tilde{x}^0_A(t,x) &= 1 - a^Tx(t) + b^Tx(t-c)1_{[c,\infty)}(t) \quad , \\ \lambda_{1,n}(t,x) &= \tilde{x}^0_A(t,x)(\tilde{P}_n(a^Tx(t)))^2\tilde{Q}_n(a^Tx(t)) \quad , \\ \lambda_{2,n}(t,x) &= \tilde{x}^0_A(t,x)2\tilde{P}_n(a^Tx(t))(1 - \tilde{P}_n(a^Tx(t)))\tilde{Q}_n(a^Tx(t)) \quad , \\ \lambda_{3,n}(t,x) &= \tilde{x}^0_A(t,x)(1 - \tilde{P}_n(a^Tx(t)))^2\tilde{Q}_n(a^Tx(t)) \quad , \end{aligned}$$

where y^T denotes the transpose of $y \in \mathbb{R}^3$, $a = (1,1,1)^T$, $b = (2,1,0)^T$ and -

$$\tilde{P}_n(x) = \frac{b_h - x + n^{-1}}{b_h + a_h b_h - x + n^{-1}} \, 1_{[0,n^{-1}n_{h,n}]}(x) \quad ,$$

$$\tilde{Q}_n(x) = \frac{b_s - x}{d(b_s + a_s b_s - x)} \, 1_{[0,n^{-1}n_{s,n})}(x) \quad .$$

Let $\mathcal{F}_{t,n}$ be the σ-algebra generated by $\{\tilde{N}_n(s),\ s \le t\}$, where $\tilde{N}_n = \{N_{AA,n} , N_{AB,n} , N_{BB,n}\}$. Then $\tilde{N}_n$ is a trivariate counting process with respect to $\{\mathcal{F}_{t,n}\}_{t\ge 0}$. It is not hard to see that for $i = 1,2,3$, the intensity process of the i-th component process is given by the left-continuous version of $n\lambda_{i,n}(t,\tilde{X}_n)$, where

$$\begin{aligned} \tilde{X}_n(t) &= (\tilde{X}_{1,n}(t),\tilde{X}_{2,n}(t),\tilde{X}_{3,n}(t))^T \\ &= \tfrac{1}{n}(N_{AA,n}(t),N_{AB,n}(t),N_{BB,n}(t))^T \quad . \end{aligned}$$

Following KURTZ (1983), we can represent $\tilde{X}_n(t)$ as the solution of

$$\begin{aligned} \tilde{X}_n(t) = \tfrac{1}{n}(\Pi_1(n\int_0^t \lambda_{1,n}(s,\tilde{X}_n)ds),\Pi_2(n\int_0^t \lambda_{2,n}(s,\tilde{X}_n)ds), \\ \Pi_3(n\int_0^t \lambda_{3,n}(s,\tilde{X}_n)ds))^T \quad , \end{aligned} \tag{4.1.4}$$

where Π_1 , Π_2 and Π_3 are independent standard Poisson processes. Unfortunately, Theorem 2.1 in KURTZ (1983) on the a.s. convergence of $\tilde{X}_n$ cannot be immediately applied, since $\lambda_{i,n}$, $i = 1,2,3$, depends on n . Define $\lambda_i(t,x)$ as $\lambda_{i,n}(t,x)$ with $\tilde{P}_n$ and $\tilde{Q}_n$ replaced by P and Q (cf. (3.1.6)), and note that

$$\sup_{x\ge 0} |\tilde{P}_n(x) - P(x)| = \mathcal{O}(n^{-1}) \quad , \qquad \sup_{x\ge 0} |\tilde{Q}_n(x) - Q(x)| = \mathcal{O}(n^{-1}) \quad .$$

It follows that there exists a constant $C > 0$ such that for $x = (x_1,x_2,x_3)^T \in D_{\mathbb{R}^3}[0,\infty)$ and $t > 0$, (4.1.5) $$\sup_{s\le t} |\lambda_{i,n}(s,x) - \lambda_i(s,x)| \le C(1 + 3\sum_{i=1}^{3} \sup_{s\le t} |x_i(s)|)n^{-1}$$

and this will enable us to approximate $\lambda_{i,n}$ by λ_i and carry through Kurtz's proof.

Define $\tilde{F} : [0,\infty) \times D_{\mathbb{R}^3}[0,\infty) \to \mathbb{R}^3$ by

(4.1.6) $$\tilde{F}(t,x) = (\lambda_1(t,x),\lambda_2(t,x),\lambda_3(t,x))^T \ .$$

Obviously, there exist positive constants C_1, C_2 and C_3 such that for all x and y in $D_{\mathbb{R}^3}[0,\infty)$ and $t > 0$,

(4.1.7) $$\sup_{s\le t} |\tilde{F}(s,x)| \le C_1 + C_2 \sup_{s\le t} |x(s)| \ ,$$

(4.1.8) $$|\tilde{F}(t,x) - \tilde{F}(t,y)| \le C_3 \sup_{s\le t} |x(s) - y(s)| \ ,$$

where for $a = (a_1,a_2,a_3)^T \in \mathbb{R}^3$, $|a| = \max_i |a_i|$. Consider the equation

(4.1.9) $$x(t) = \int_0^t \tilde{F}(s,x)ds \ , \qquad x \in D_{\mathbb{R}^3}[0,\infty) \ .$$

The inequalities (4.1.7) and (4.1.8) and LEMMA 3.1.4 ensure the existence of a unique solution $\tilde{X} = (\tilde{X}_1,\tilde{X}_2,\tilde{X}_3)^T$ of (4.1.9). Now the analogue of Theorem 2.1 in KURTZ (1983) can be proved.

<u>THEOREM</u> <u>4.1.1</u> Suppose $\tilde{X}_n$, $n = 1,2,\ldots$, satisfies (4.1.4) and $\tilde{X}$ is the solution of (4.1.9). Then for all $M \ge 0$,

(4.1.10) $$\lim_{n\to\infty} \sup_{t\le M} |\tilde{X}_n(t) - \tilde{X}(t)| = 0 \quad \text{a.s.}$$

<u>PROOF.</u> We have

$$\tilde{X}_n(t) - \tilde{X}(t) = \varepsilon_n(t) + \eta_n(t) + \int_0^t (\tilde{F}_n(s,\tilde{X}_n) - \tilde{F}(s,\tilde{X}))ds \ ,$$

where

$$\varepsilon_n(t) = \tilde{X}_n(t) - \left(\int_0^t \lambda_{1,n}(s,\tilde{X}_n)ds, \int_0^t \lambda_{2,n}(s,\tilde{X}_n)ds, \int_0^t \lambda_{3,n}(s,\tilde{X}_n)ds\right)^T ,$$

$$\eta_n(t) = \left(\int_0^t (\lambda_{1,n}(s,\tilde{X}_n) - \lambda_1(s,\tilde{X}_n))ds, \int_0^t (\lambda_{2,n}(s,\tilde{X}_n) - \lambda_2(s,\tilde{X}_n))ds,\right.$$

$$\left.\int_0^t (\lambda_{3,n}(s,\tilde{X}_n) - \lambda_3(s,\tilde{X}_n))ds\right)^T .$$

Following the argument in KURTZ (1983), it only remains to prove

$$\lim_{n\to\infty} \sup_{t\le M} |\eta_n(t)| = 0 \quad \text{a.s.}$$

But this follows immediately from (4.1.5) and the fact that, for $i = 1,2,3$, $\lambda_{i,n}(t,\tilde{X}_n) = 0$ if $a^T\tilde{X}_n(t) > b_s$, so that $\sum_{i=1}^{3} \sup_{s\le t} |\tilde{X}_{i,n}(s)| \le b_s$.

□

From (4.1.1) and (4.1.2) we see that for $t \ge c$,

$$X_n(t) = a^T\tilde{X}_n(t-c) \quad , \qquad Y_n(t) = b^T\tilde{X}_n(t-c) \quad ,$$

where X_n and Y_n are as defined in Chapter 3 . Hence the following corollary is obvious.

<u>COROLLARY</u> <u>4.1.2</u> For $t \ge 0$, let X and Y be defined by

(4.1.11) $X(t) = a^T\tilde{X}(t-c)1_{[c,\infty)}$, $\qquad Y(t) = b^T\tilde{X}(t-c)1_{[c,\infty)}$.

Then for each $M \ge 0$,

$$\lim_{n\to\infty} \sup_{t\le M} |X_n(t) - X(t)| = 0 \quad \text{a.s.} \quad ,$$

$$\lim_{n\to\infty} \sup_{t\le M} |Y_n(t) - Y(t)| = 0 \quad \text{a.s.}$$

□

Writing out the expressions for $X(t)$ and $Y(t)$ we get

$$X(t) = \int_0^{t-c} \sum_{i=1}^{3} \lambda_i(s,\tilde{X})ds = \int_0^{t-c} \tilde{x}_A^0(s,\tilde{X})Q(a^T\tilde{X}(s))ds$$

$$= \int_0^{t-c} \{1 - a^T\tilde{X}(s) + b^T\tilde{X}(s-c)\}Q(a^T\tilde{X}(s))ds$$

$$= \int_0^{t-c} \{1 - X(s+c) + Y(s)\}Q(X(s+c))ds$$

$$= \int_c^{t} \{1 - X(s) + Y(s-c)\}Q(X(s))ds$$

and similarly,

$$Y(t) = 2\int_c^t P(X(s))X'(s)ds \quad ,$$

which are the same expressions for $X(t)$ and $Y(t)$ as we had in Chapter 3. Note that convergence is proved here only on bounded intervals as in THEOREM 3.1.1 .

4.2. THE CENTRAL LIMIT THEOREM

Following KURTZ (1983) again, the central limit theorem for the deviation of X_n from X can be derived. Here too we have to modify the theory slightly to supply our needs.

Define $\hat{F} : [0,\infty) \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ by

$$(4.2.1) \quad \begin{aligned} \hat{F}(t,y,z) &= (\hat{F}_1(t,y,z),\hat{F}_2(t,y,z),\hat{F}_3(t,y,z))^T \\ &= [1 - a^Ty + b^Tz\ 1_{[c,\infty)}(t)]Q(a^Ty)\begin{bmatrix} P(a^Ty)^2 \\ 2P(a^Ty)(1 - P(a^Ty)) \\ (1 - P(a^Ty))^2 \end{bmatrix} \end{aligned} \quad .$$

By (4.1.3) and (4.1.6) we have

$$(4.2.2) \qquad \tilde{F}(t,x) = \hat{F}(t,x(t),x(t-c)) \quad .$$

Let $\frac{\partial \hat{F}}{\partial y}$ and $\frac{\partial \hat{F}}{\partial z}$ be the 3×3 matrix valued functions with $\left[\frac{\partial \hat{F}}{\partial y}\right]_{ij} = \frac{\partial \hat{F}_i}{\partial y_j}$ and $\left[\frac{\partial \hat{F}}{\partial z}\right]_{ij} = \frac{\partial \hat{F}_i}{\partial z_j}$. We find for $j = 1,2,3$,

$$\begin{aligned}
\frac{\partial \hat{F}_1}{\partial y_j}(t,y,z) &= -P(a^Ty)^2Q(a^Ty) + \{1 - a^Ty + b^Tz\, 1_{[c,\infty)}(t)\} \\
&\quad \cdot (2P(a^Ty)P'(a^Ty)Q(a^Ty) + P(a^Ty)^2Q'(a^Ty)) \quad , \\
\frac{\partial \hat{F}_2}{\partial y_j}(t,y,z) &= -2P(a^Ty)(1 - P(a^Ty))Q(a^Ty) + 2\{1 - a^Ty \\
(4.2.3) \qquad &\quad + b^Tz\, 1_{[c,\infty)}(t)\}(P'(a^Ty)(1 - 2P(a^Ty))Q(a^Ty) \\
&\quad + P(a^Ty)(1 - P(a^Ty)Q'(a^Ty)) \quad , \\
\frac{\partial \hat{F}_3}{\partial y_j}(t,y,z) &= -(1 - P(a^Ty))^2Q(a^Ty) + \{1 - a^Ty + b^Tz\, 1_{[c,\infty)}(t)\} \\
&\quad \cdot (-2(1 - P(a^Ty))P'(a^Ty)Q(a^Ty) \\
&\quad + (1 - P(a^Ty))^2Q'(a^Ty)) \quad ,
\end{aligned}$$

where

$$P'(x) = \frac{-a_hb_h}{(b_h + a_hb_h - x)^2}\, 1_{[0,b_h]}(x) \quad ,$$

$$Q'(x) = \frac{-a_sb_s}{d(b_s + a_sb_s - x)^2}\, 1_{[0,b_s]}(x) \quad .$$

Also, $\dfrac{\partial\hat{F}}{\partial z}(t,y,z)$ equals

$$(4.2.4)\qquad \begin{bmatrix} 2P(a^Ty)^2 & P(a^Ty)^2 & 0 \\ 4P(a^Ty)(1-P(a^Ty)) & 2P(a^Ty)(1-P(a^Ty)) & 0 \\ 2(1-P(a^Ty))^2 & (1-P(a^Ty))^2 & 0 \end{bmatrix} Q(a^Ty)\,1_{[c,\infty)}(t)\ .$$

Let $K : [0,\infty) \times [0,\infty) \times D_{\mathbb{R}^3}[0,\infty) \to M_{3\times3}$, where $M_{3\times3}$ is the space of real valued 3×3-matrices , be defined by

$$(4.2.5)\qquad K(t,u,x) = \frac{\partial\hat{F}}{\partial y}(u,x(u),x(u-c)) + \frac{\partial\hat{F}}{\partial z}(u+c,x(u+c),x(u))\cdot 1_{[0,t-c]}(u)\ .$$

Since we use the Skorohod topology on $D_{\mathbb{R}^3}[0,\infty)$, we have for $A \subset D_{\mathbb{R}^3}[0,\infty)$, A compact and $t > 0$ (cf. ETHIER & KURTZ (1986), p.122),

$$(4.2.6)\qquad \sup_{v\in A}\sup_{s\le t} |v(s)| < \infty\ .$$

Hence, the following holds.

LEMMA 4.2.1 For $t > 0$, $x \in D_{\mathbb{R}^3}[0,\infty)$ and compact $A \subset D_{\mathbb{R}^3}[0,\infty)$,

$$(4.2.7)\qquad \lim_{n\to\infty}\sup_{v\in A}\sup_{s\le t}\Big| \int_0^s n^{1/2}(\tilde{F}(u,x+n^{-1/2}v) - \tilde{F}(u,x))du - \int_0^s K(s,u,x)v(u)du\Big| = 0\ .$$

PROOF. By (4.2.1) and (3.1.34) all second order partial derivatives of $\hat{F}$ are bounded, and hence by (4.2.2) there exists $C > 0$ such that for all $t > 0$ and $x,\ v \in D_{\mathbb{R}^3}[0,\infty)$,

$$\sup_{u\le t} |\tilde{F}(u,x + n^{-1/2}v) - \tilde{F}(u,x) - n^{-1/2} \frac{\partial \hat{F}}{\partial y}(u,x(u),x(u-c))v(u)$$

$$- n^{-1/2}\frac{\partial \hat{F}}{\partial z}(u,x(u),x(u-c))v(u-c)| \le Cn^{-1}\sup_{u\le t} |v(u)|^2 \ .$$

Integrating over $u \in (0,s)$ for $0 < s \le t$, changing variables in the third integral and recalling that $\frac{\partial \hat{F}}{\partial z}(u,y,z) = 0$ for $u < c$ by (4.2.4) , we obtain

$$\sup_{s\le t} | \int_0^s \{\tilde{F}(u,x + n^{-1/2}v) - \tilde{F}(u,x)\}du$$

$$- n^{-1/2} \int_0^s \frac{\partial \hat{F}}{\partial y}(u,x(u),x(u-c))v(u)du$$

$$- n^{-1/2} \int_0^{s-c} \frac{\partial \hat{F}}{\partial z}(u+c,x(u+c),x(u))v(u)\}du|$$

$$\le Cn^{-1}t \sup_{s\le t} |v(s)|^2 \ .$$

Because the second term on the right in (4.2.5) equals zero if $u > t - c$, we see that

$$\sup_{v\in A} \sup_{s\le t} | \int_0^s n^{1/2}(\tilde{F}(u,x+n^{-1/2}v) - \tilde{F}(u,x))du$$

$$- \int_0^s K(s,u,x)v(u)du| \le Cn^{-1}t \sup_{v\in A} \sup_{s\le t} |v(s)|^2$$

and application of (4.2.6) completes the proof.

□

Let for $t \geq 0$,

$$
\begin{aligned}
&\tilde{V}_n(t) = n^{1/2}(\tilde{X}_n(t) - \tilde{X}(t)) \quad , \\
(4.2.8) \qquad & \\
&\tilde{W}_{i,n}(t) = n^{-1/2}(\Pi_i(nu) - nu) \quad , \qquad\qquad i = 1,2,3 \quad ,
\end{aligned}
$$

with Π_i as in Section 4.1 . Then we have

LEMMA 4.2.2 For every $t \geq 0$,

$$
\begin{aligned}
&\sup_{s \leq t} |\tilde{V}_n(s) - \Big[(\tilde{W}_{1,n}(\int_0^s \lambda_1(u,\tilde{X}_n)du), \tilde{W}_{2,n}(\int_0^s \lambda_2(u,\tilde{X}_n)du), \\
(4.2.9) \qquad & \\
&\tilde{W}_{3,n}(\int_0^s \lambda_3(u,\tilde{X}_n)du))^T + \int_0^s n^{1/2}(\tilde{F}(u,\tilde{X}_n) - \tilde{F}(u,\tilde{X}))du\Big] | \to 0 \quad \text{a.s.} \quad .
\end{aligned}
$$

PROOF. Let Π be a unit Poisson process. Then it follows from the monotonicity of Π that for all positive constants A and a ,

$$
\begin{aligned}
&P\,(\sup_{\substack{0 \leq u \leq A \\ 0 \leq h \leq a}} |\Pi(nu+h) - \Pi(nu)| \geq 6\log n) \\
&\leq P\,(\max_{0 \leq k \leq [a^{-1}An]+1} (|\Pi((k+1)a) - \Pi(ka)| \geq 3\log n) \\
&\leq \sum_{k=0}^{[a^{-1}An]+1} P\,(\Pi((k+1)a) - \Pi(ka) \geq 3\log n) \\
&= ([a^{-1} An] + 1)P\,(\Pi(a) \geq 3\log n) \\
&\leq (a^{-1}An + 1)\exp\{2a - 3\log n\} \leq Bn^{-2}
\end{aligned}
$$

for some positive constant B . By definitions (4.1.4) , (4.1.6) , (4.1.9) and (4.2.8) , we have

$$
\begin{aligned}
&|\tilde{V}_n(s) - \Big[(\tilde{W}_{1,n}(\int_0^s \lambda_1(u,\tilde{X}_n)du),\tilde{W}_{2,n}(\int_0^s \lambda_2(u,\tilde{X}_n)du),\\
&\quad \tilde{W}_{3,n}(\int_0^s \lambda_3(u,\tilde{X}_n)du))^T + \int_0^s n^{1/2}(\tilde{F}(u,\tilde{X}_n) - \tilde{F}(u,\tilde{X}))du\Big]|\\
&= |n^{-1/2}(\Pi_1(n\int_0^s \lambda_{1,n}(u,\tilde{X}_n)du),\Pi_2(n\int_0^s \lambda_{2,n}(u,\tilde{X}_n)du),\\
&\quad \Pi_3(n\int_0^s \lambda_{3,n}(u,\tilde{X}_n)du))^T - n^{-1/2}(\Pi_1(n\int_0^s \lambda_1(u,\tilde{X}_n)du),\\
&\quad \Pi_2(n\int_0^s \lambda_2(u,\tilde{X}_n)du),\Pi_3(n\int_0^s \lambda_3(u,\tilde{X}_n)du))^T| \quad .
\end{aligned}
$$

Because $\sum_{i=1}^{3} \tilde{X}_{i,n} \le b_s$ (cf. the proof of THEOREM 4.1.1),

$$\sup_{s\le t} n\int_0^s \lambda_{i,n}(u,\tilde{X}_n)du \le A_i n \quad ,$$

and in view of (4.1.5) ,

$$\sup_{s\le t} n\int_0^s |\lambda_{i,n}(u,\tilde{X}_n) - \lambda_i(u,\tilde{X}_n)|du \le a_i$$

for $i = 1,2,3$ and positive constants A_i and a_i . The lemma now follows from the Borel-Cantelli lemma.

□

Having proved LEMMAS 4.2.1 and 4.2.2 , we have verified the conditions of Kurtz's central limit theorem (KURTZ, 1983), which in our case yields the following.

THEOREM 4.2.1 Let $\tilde{V}_n$ be defined as in (4.2.8) , then

$$\tilde{V}_n \overset{\mathcal{D}}{\to} \tilde{V} \quad ,$$

where $\tilde{V}$ is the Gaussian process satisfying

$$\text{(4.2.10)} \qquad \begin{aligned} \tilde{V}(t) = (\tilde{W}_1(\int_0^t \lambda_1(s,\tilde{X})ds),\tilde{W}_2(\int_0^t \lambda_2(s,\tilde{X})ds),\\ \tilde{W}_3(\int_0^t \lambda_3(s,\tilde{X})ds))^T + \int_0^t K(t,s,\tilde{X})\tilde{V}(s)ds \quad , \end{aligned}$$

and $\tilde{W}_1$, $\tilde{W}_2$ and $\tilde{W}_3$ are independent standard Wiener processes.

□

As in Section 4.1 we can use this to obtain similar results for V_n and U_n . Indeed, remembering that $X_n(t) = a^T\tilde{X}_n(t-c)$, $X(t) = a^T\tilde{X}(t-c)$, $Y_n(t) = b^T\tilde{X}_n(t-c)$ and $Y(t) = b^T\tilde{X}(t-c)$, we immediately have the following corollary.

COROLLARY 4.2.1 For $t \geq c$, let

$$\text{(4.2.11)} \quad V(t) = a^T\tilde{V}(t-c) \quad , \qquad U(t) = b^T\tilde{V}(t-c) \quad .$$

Then

$$V_n(t) \overset{\mathcal{D}}{\to} V(t) \quad , \qquad U_n(t) \overset{\mathcal{D}}{\to} U(t) \quad ,$$

and V and U are Gaussian processes with mean zero.

□

To obtain THEOREM 3.3.1 by the methods employed in the present chapter, it remains to be shown that V satisfies (3.3.8) and (3.3.9) for independent standard Wiener processes W_0 and W_1 . Straightforward calculation shows that (4.2.3) - (4.2.5) and THEOREM 4.2.1 imply that V satisfies

$$V(t) = \sum_{i=1}^{3} \tilde{W}_i\Big(\int_0^{t-c} \lambda_i(s,\tilde{X})ds\Big) + \int_0^{t-c} \Big[\{-Q(a^T\tilde{X}(s)) + \tilde{x}_A^0(s,\tilde{X})Q'(a^T\tilde{X}(s))\}a^T\tilde{V}(s) + Q(a^T\tilde{X}(s))b^T\tilde{V}(s-c)\Big] ds \quad . \tag{4.2.12}$$

Define a Gaussian process W_0 by

$$W_0(X(t)) = \tilde{W}_1\Big(\int_0^{X(t)} P(s)^2 ds\Big) + \tilde{W}_2\Big(\int_0^{X(t)} 2P(s)(1 - P(s))ds\Big) + \tilde{W}_3\Big(\int_0^{X(t)} (1 - P(s))^2 ds\Big) \quad . \tag{4.2.13}$$

Since $\tilde{W}_1$, $\tilde{W}_2$ and $\tilde{W}_3$ are independent standard Wiener processes, W_0 is also a standard Wiener process. Hence, after a change of integration variables and in view of (4.1.3) , (4.1.11) and (4.2.11) , (4.2.12) reduces to

$$V(t) = W_0(X(t)) + \int_c^t V(s)\{-Q(X(s)) + X_A^0(s-c)Q'(X(s))\}ds + \int_c^t Q(X(s))U(s-c)ds \quad . \tag{4.2.14}$$

Similarly,

$$U(t) = 2\tilde{W}_1\Big(\int_0^{t-c} \lambda_1(s,\tilde{X})ds\Big) + \tilde{W}_2\Big(\int_0^{t-c} \lambda_2(s,\tilde{X})ds\Big) + \int_0^{t-c} \Big[\{-2P(a^T\tilde{X}(s)) \cdot Q(a^T\tilde{X}(s)) + 2\tilde{x}_A^0(s,\tilde{X})(P'a^T\tilde{X}(s))Q(a^T\tilde{X}(s)) + P(a^T\tilde{X}(s)) \cdot Q'(a^T\tilde{X}(s)))\}a^T\tilde{V}(s) + 2P(a^T\tilde{X}(s))Q(a^T\tilde{X}(s))b^T\tilde{V}(s-c)\Big] ds \quad . \tag{4.2.15}$$

Changing the integration variables in (4.2.15) and using (4.2.12) and (4.2.13) , we obtain for $t \geq c$,

$$U(t) = 2\tilde{W}_1(\int_0^{X(t)} P(s)^2 ds) + \tilde{W}_2(\int_0^{X(t)} 2P(s)(1 - P(s))ds)$$

$$+ \int_c^t \Big[2P'(X(s))X'(s))V(s) + 2P(X(s))\{-Q(X(s))$$

$$+ X_A^0(s-c)Q'(X(s))\}V(s) + 2P(X(s))Q(X(s))U(s-c)\Big] ds$$

$$= 2\tilde{W}_1(\int_0^{X(t)} P(s)^2 ds) + \tilde{W}_2(\int_0^{X(t)} 2P(s)(1 - P(s))ds) \tag{4.2.16}$$

$$+ \int_c^t \Big[2P(X(s))\Big]'V(s)ds + \int_c^t 2P(X(s))\Big[V(s) - W_0(X(s))\Big]'ds$$

$$= 2\tilde{W}_1(\int_0^{X(t)} P(s)^2 ds) + \tilde{W}_2(\int_0^{X(t)} 2P(s)(1 - P(s))ds)$$

$$+ \int_c^t 2P(X(s))dW_0(X(s)) + 2P(X(t))V(t)$$

$$= W_1(2\int_0^{X(t)} P(s)(1 - P(s))ds) + 2P(X(t)V(t) \quad ,$$

where W_1 is easily seen to be a standard Wiener process which is independent of W_0 . By (4.2.14) and (4.2.16) , V satisfies (3.3.8) and (3.3.9) and we have reproved THEOREM 3.3.1 .

4.3. CONCLUSION

In Chapters 3 and 4 we have outlined two different methods to approach the asymptotic theory for the process N_n . The approach in the present chapter is to consider a trivariate counting process for which the intensity process is easily available. Apart from a constant time shift, the process N_n is the sum of the components of this trivariate process and N_n inherits some of the asymptotic properties of this process. The approach adopted in Chapter 3 is to consider the process N_n conditionally on the sequence $\hat{Z}_n = (Z_{1,n}, Z_{2,n}, \ldots)$. The conditional intensity process of N_n has a structure very similar to the unconditional intensity of the trivariate process.

The uniform convergence of X_n to X on bounded intervals is shown here in essentially the same manner as in Chapter 3 . The proof of the convergence $V_n \overset{\mathcal{D}}{\to} V$ is also essentially the same. However, the representation of V in terms of two independent Wiener processes in (3.3.8) and (3.3.9) is almost impossible to guess without going into the special structure of the underlying processes as is done in LEMMA 3.3.2 .

To obtain further asymptotic properties of X_n , as we have done in Chapter 3 , the special character of the process must play a crucial role. The uniform convergence of X_n to X on $[0,\infty)$ appears to depend heavily on the boundedness of X_n and our investigation on the duration T_n is very specific for the process we consider.

CHAPTER 5

OTHER DISTRIBUTION RESULTS; A SECOND ORDER APPROXIMATION FOR THE UNCONDITIONAL INTENSITY PROCESS

Because the distribution of V, which is the limit distribution of $n^{1/2}(X_n - X)$ (see Section 3.3), is rather complicated, some statistics with a more tractable (limit) distribution are constructed in this chapter.

5.1. CONDITIONAL DISTRIBUTION RESULTS

When for each division the number of newly born A-cells is known, that is, conditionally on $\hat{Z}_n = (Z_{1,n}, Z_{2,n}, \ldots)$, N_n can be represented as (cf. (4.1.4))

$$(5.1.1) \qquad N_n(t) - n \cong \Pi(A_{z,n}(t)) \quad , \qquad t \geq 0 \quad ,$$

where Π is a unit Poisson process and $\cong$ denotes equality in distribution. Moreover, we already saw in Section 2.2 that the process N_n^* as defined in (2.2.13) is a Poisson process. In other words,

$$(5.1.2) \qquad N_n(A_{z,n}^{-1}(u)) - n \cong \Pi_0(u) \quad , \qquad 0 \leq u \leq A_{z,n}(T_n) \quad ,$$

where Π_0 is a unit Poisson process stopped at the random time $A_{z,n}(T_n)$ and as a result we proved in LEMMA 3.3.1 ,

$$(5.1.3) \qquad n^{-1/2}(N_n(t) - n - A_{z,n}(t)) \overset{\mathcal{D}}{\to} W_0(X(t)) \quad \text{as} \quad n \to \infty \quad , \qquad t \geq 0 \quad .$$

where W_0 is a standard Wiener process, X the solution of (3.1.13) and the convergence is in the sense of the Skorohod metric.

Also it was noted (Section 2.2) that given $\hat{Z}$, the process Π_0 stops at the non-random level $N_n(T_n) - n$. Denote this level by $K_n = K_n(\hat{Z}_n)$ and $A_{z,n}(T_n)$ by U_n. Then (see SHORACK & WELLNER (1986), p.335)

$$(K_n - 1)^{-1}\Pi_0(xU_n) \cong F_{K_n-1}(x) \quad , \qquad x \in (0,1) \quad ,$$

where F_k is the empirical distribution function of k independent random variables which are uniformly distributed on $(0,1)$. It follows that

$$(K_n - 1)^{1/2}\{(K_n - 1)^{-1}\Pi_0(xU_n) - x\} \cong \mathcal{E}_n(x) \quad , \qquad x \in (0,1) \quad ,$$

where $\mathcal{E}_n$ denotes the uniform empirical process. Since K_n is of exact order n , we thus have (BILLINGSLEY, 1968) for $x \in (0,1)$ and $n \to \infty$,

$$(K_n - 1)^{1/2}\{(K_n - 1)^{-1}\Pi_0(xU_n) - x\} \overset{\mathcal{D}}{\to} B_0(x) \quad ,$$

where B_0 is a Brownian bridge and the convergence is again in the sense of the Skorohod metric. Translating these results back to the original process N_n , we get

<u>THEOREM</u> <u>5.1.1</u> Conditionally on $\hat{Z}_n$, for $x \in (0,1)$ and $n \to \infty$,

$$(5.1.4) \qquad \frac{N_n(A_{z,n}^{-1}(xA_{z,n}(T_n))) - n}{N_n(T_n) - n - 1} \cong F_{N_n(T_n)-n-1}(x) \quad ,$$

$$(5.1.5) \qquad (N_n(T_n) - n - 1)^{1/2}\left[\frac{N_n(A_{z,n}^{-1}(xA_{z,n}(T_n))) - n}{N_n(T_n) - n - 1} - x\right] \cong \mathcal{E}_n \overset{\mathcal{D}}{\to} B_0(x) \ .$$

□

Note that (5.1.3) and the convergence result in (5.1.5) in principle enable us to test the model of Chapter 2 . Unfortunately, in practice $\hat{Z}_n$ and $A_{z,n}$ are usually unknown. However, there obviously also exists a unit Poisson process $\tilde{\Pi}$ such that (5.1.1) holds unconditionally with $\tilde{\Pi}$ and the unconditional integrated intensity process A_n instead of Π and $A_{z,n}$, respectively. Moreover, because the distributions of Π_0 , W_0 and B_0 do not depend on $\hat{Z}_n$, the results (5.1.2) , (5.1.3) and the convergence in (5.1.5) also hold unconditionally, and there too $A_{z,n}$ can be replaced by A_n . Hence, if A_n can be computed from the data, the results

$$(5.1.6) \qquad n^{-1/2}(N_n(t) - n - A_n(t)) \xrightarrow{\mathcal{D}} W_0(X(t)) \quad \text{as} \quad n \to \infty \quad , \qquad t \ge 0$$

and

$$(5.1.7) \qquad (N_n(T_n) - n\,)^{1/2}\left[\frac{N_n(A_n^{\;-1}(xA_n(T_n))) - n}{N_n(T_n) - n} - x\right] \xrightarrow{\mathcal{D}} B_0(x) \quad , \quad x \in (0,1)$$

allow us to test our model. This is why in the following sections of this chapter we derive an approximation with remainder $\mathcal{O}(n^{1/2})$ for the unconditional intensity process Λ_n , which yields, on bounded intervals, an approximation up to $\mathcal{O}(n^{1/2})$ for A_n .

For convenience the index n is omitted in the remainder of this chapter when no confusion is possible.

5.2. NOTATION AND PRELIMINARY RESULTS

Fix $t_0 < \infty$ and choose $t \in [c,t_0]$ so that $N(t) > n$ and $X(\infty) - X(t) \ge X(\infty) - X(t_0) > 0$. Suppose that

$$(5.2.1) \qquad \begin{aligned} &N(t) - n = k \quad , \qquad N(t-c) - n = \ell \quad , \\ &\tau_j = t_j \quad , \qquad N(t_j-c) - n = \ell(j) \quad , \qquad j = 1,2,\ldots,k \;. \end{aligned}$$

Define

$$(5.2.2) \qquad m(s) = 2n - N(s) + 2\sum_{j=1}^{nX_n(s-c)} P_j \quad ,$$

$$(5.2.3) \qquad \alpha_i = \alpha_i(t) = \left[\int_{[t_i+c,t]} m(s)^{-1}dN(s) - \int_{[t_i+c,t]} Q(X_n(s))ds\right]1_{[t_i+c,\infty)}(t) \quad ,$$

so that $\alpha_i = 0$ for $i = \ell + 1, \ell + 2, \ldots$. We have the following result.

LEMMA 5.2.1 There exist positive numbers δ , B , β and x_0 such that for $n = 1,2,\ldots$ and $0 \le x \le x_0$,

$$(5.2.4)\qquad P(\inf_{s\le t_0} Q(X_n(s)) \le \delta) + P(\inf_{s\le t_0} m(s) \le \delta n) \le B\exp\{-\beta n\} ,$$

$$(5.2.5)\qquad P(\sup_{c\le t\le t_0}\ \max_{1\le i\le \ell} |\alpha_i(t)| \ge x) \le B\exp\{-\beta x^2 n\} .$$

PROOF. Since $X(\infty) - X(t_0) > 0$, the first assertion follows from THEOREM 3.1.2 , COROLLARY 3.1.3 , (3.1.3) and LEMMA 3.1.4 .

For $t \in [c,t_1 + c)$, we have $\alpha_i(t) = 0$. Fix i and consider the processes $m(s)$ and $N(s) - n - A_z(s)$ for $t_i + c \le s \le t \le t_0$. Since $f(s) = m(s)^{-1}$ and $g(s) = N(s) - n - A_z(s)$ are functions of bounded variation over finite intervals, right-continuous and with left-hand limits, we have

$$f(t)g(t) - f((t_i+c)_-)g((t_i+c)_-)$$

$$= \int_{[t_i+c,t]} f(s)dg(s) + \int_{[t_i+c,t]} g(s_-)df(s) ,$$

and hence

$$|\int_{[t_i+c,t]} f(s)dg(s)| \le 2 \sup_{s\in[t_1,t_0]} |f(s)g(s)|$$

$$+ \sup_{s\in[t_1,t_0]} |g(s)| \int_{[t_1+c,t_0]} |df(s)| .$$

Because of LEMMA 3.1.1 and (5.2.4) ,

$$P(\sup_{s\in[t_1+c,t_0]} |f(s)g(s)| \ge x/8) \le B_1\exp\{-\beta_1 x^2 n\}$$

for $0 \le x \le x_1$. Here and in what follows B_ν , β_ν and x_ν are fixed positive numbers. Also because of (5.2.4) ,

$$\int_{[t_1+c,t_0]} |df(s)| \le (\delta n)^{-2} \int_{[t_1+c,t_0]} |dm(s)|$$

$$\le (\delta n)^{-2}\Big[N(t_0) - n + 2\sum_{j=1}^{n_{h,n}} P_j\Big] \le (b_s + 2b_h)\delta^{-2}n^{-1}$$

with probability at least $1 - B_2\exp\{-\beta_2 x_2^2 n\}$. Applying LEMMA 3.1.1 once more, we find that

$$(5.2.6)\qquad P\Big(\sup_{t_1+c\le t\le t_0} \max_{1\le i\le \ell} \Big| \int_{[t_i+c,t]} f(s)dg(s)\Big| \ge \tfrac{1}{2}x\Big) \le B_3\exp\{-\beta_3 x^2 n\}$$

for $0 \le x \le x_3$.

In view of (3.1.3) and (5.2.4)

$$P\Big(\sup_{t\ge 0} \Big|\Big(2n - N(t) + \sum_{j=1}^{nX_n(t-c)} Z_j\Big)^{-1} - m(t)^{-1}\Big| \ge xn^{-1}\Big)$$

$$\le B_4\exp\{-\beta_4 x^2 n\}$$

for $0 \le x \le x_4$. Hence

$$P\Big(\sup_{t\ge 0} |Q(X_n(t)) - m(t)^{-1}\Lambda_z(t)| \ge x(2t_0)^{-1}\Big) \le B_5\exp\{-\beta_5 x^2 n\}$$

for $0 \le x \le x_5$, and therefore (5.2.6) implies that with probability at least $1 - B\exp\{-\beta x^2 n\}$, for $0 \le x \le x_0$,

$$\sup_{t_1+c\le t\le t_0} \max_{1\le i\le \ell} |\alpha_i(t)| \le \sup_{t_1+c\le t\le t_0} \max_{1\le i\le \ell} \Big| \int_{[t_i+c,t]} m(s)^{-1}dN(s)$$

$$- \int_{[t_i+c,t]} m(s)^{-1}\Lambda_z(s)ds \Big| + \tfrac{1}{2}x$$

$$= \sup_{t_1+c\le t\le t_0} \max_{1\le i\le \ell} \Big| \int_{[t_i+c,t]} f(s)dg(s)\Big| + \tfrac{1}{2}x \le x \quad .$$

□

Consider the system of linear equations in $\beta_1 , \ldots , \beta_\ell$,

$$(5.2.7) \qquad \beta_i = \alpha_i(t) - 2 \sum_{j=1}^{\ell} \beta_j P_j(1 - P_j) \int_{[(t_i \vee t_j)+c,t]} m(s)^{-2} dN(s) 1_{[t_i+c,\infty)}(t) \quad ,$$

for $i = 1,2,\ldots,\ell$. We shall show that this system has a unique solution $\beta_i = \beta_i(t)$, $i = 1,2,\ldots,\ell$. For these $\beta_1 , \ldots , \beta_\ell$, define

$$(5.2.8) \qquad \tilde{P}_i = \tilde{P}_i(t) = \frac{P_i \exp\{\beta_i(t)\}}{P_i \exp\{\beta_i(t)\} + 1 - P_i} \quad , \qquad i = 1,2,\ldots,\ell \quad ,$$

$$(5.2.9) \qquad \tilde{m}(s) = 2n - N(s) + 2 \sum_{j=1}^{nX_n(s-c)} \tilde{P}_j(t) \quad ,$$

$$(5.2.10) \qquad \delta_i = \delta_i(t) = \left[\int_{[t_i+c,t]} \tilde{m}(s)^{-1} dN(s) - \int_{[t_i+c,t]} Q(X_n(s)) ds \right] \cdot 1_{[t_i+c,\infty)}(t) - \beta_i(t) \quad , \qquad i = 1,2,\ldots,\ell \quad .$$

<u>LEMMA</u> <u>5.2.2</u> The system of linear equations (5.2.7) has a unique solution $\beta_i = \beta_i(t)$, $i = 1,2,\ldots,\ell$. Moreover, there exist positive numbers B , β and x_0 such that for $n = 1,2,\ldots$ and $0 \le x \le x_0$,

$$(5.2.11) \qquad P \left(\sup_{c \le t \le t_0} \max_{1 \le i \le \ell} |\beta_i(t)| \ge x \right) \le B\exp\{-\beta x^2 n\} \quad ,$$

$$(5.2.12) \qquad P \left(\sup_{c \le t \le t_0} \max_{1 \le i \le \ell, P_i > 0} \left| \frac{\tilde{P}_i - P_i}{P_i(1 - P_i)} - \beta_i \right| \ge x \right) \le B\exp\{-\beta x n\} \quad ,$$

$$(5.2.13) \qquad P \left(\sup_{c \le t \le t_0} \max_{1 \le i \le \ell, P_i > 0} \left| \frac{\tilde{P}_i - P_i}{P_i(1 - P_i)} \right| \ge x \right) \le B\exp\{-\beta x^2 n\} \quad ,$$

$$(5.2.14) \qquad P \left(\inf_{s \le t_0} \tilde{m}(s) \le \delta n \right) \le B\exp\{-\beta n\} \quad ,$$

$$(5.2.15) \qquad P \left(\sup_{c \le t \le t_0} \max_{1 \le i \le \ell} |\delta_i(t)| \ge x \right) \le B\exp\{-\beta x n\} \quad .$$

PROOF. The system (5.2.7) may be written as

$$(5.2.16)\quad (C\Pi + I)\beta = \alpha \quad ,$$

where $\alpha = (\alpha_1,\dots,\alpha_\ell)^T$, $\beta = (\beta_1,\dots,\beta_\ell)^T$,

$$(5.2.17)\quad C = \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_\ell \\ c_2 & c_2 & c_3 & \cdots & c_\ell \\ c_3 & c_3 & c_3 & \cdots & c_\ell \\ \vdots & \vdots & \vdots & & \vdots \\ c_\ell & c_\ell & c_\ell & \cdots & c_\ell \end{bmatrix} \quad , \quad \Pi = \begin{bmatrix} \pi_1 & 0 & & \cdots & 0 \\ 0 & \pi_2 & \ddots & & \vdots \\ \vdots & \ddots & \pi_3 & \ddots & \\ & & \ddots & \ddots & 0 \\ 0 & \cdots & & 0 & \pi_\ell \end{bmatrix} \quad ,$$

$\pi_j = 2P_j(1 - P_j)$, $j = 1,\dots,\ell$, I denotes the $\ell \times \ell$ identity matrix, and

$$(5.2.18)\quad c_i = \int_{[t_i+c,t]} m(s)^{-2}dN(s) \quad , \qquad i = 1,\dots,\ell \quad .$$

Note that $c_1 , c_2 , \dots , c_\ell$ form a non-increasing and non-negative sequence.

A standard computation shows that

$$\det(C\Pi + I) = \sum_{r=0}^{\ell} \sum_{1\le i_1<i_2<\dots<i_r\le\ell} \cdots\cdots \sum c_{i_r} \prod_{j=1}^{r-1} (c_{i_j} - c_{i_{j+1}}) \prod_{j=1}^{r} \pi_{i_j}$$

$$\ge 1 \quad ,$$

so that $(C\Pi + I)$ is non-singular. Hence (5.2.16) determines $\beta = \beta(t)$ uniquely.

Next we note that C is non-negative definite since for $\lambda < 0$,

$$\det(C - \lambda I) = \sum_{r=0}^{\ell} (-\lambda)^{\ell-r} \sum_{1\le i_1<i_2<\dots<i_r\le\ell} \cdots\cdots \sum c_{i_r} \prod_{j=1}^{r-1} (c_{i_j} - c_{i_{j+1}})$$

$$\ge (-\lambda)^{\ell} > 0 \quad .$$

It follows that if Π is non-singular, then

$$(C\Pi + I)^T(C\Pi + I) - 4\Pi^2 = \Pi(C + \Pi^{-1})(C + \Pi^{-1})\Pi - 4\Pi^2$$

is non-negative definite, because the diagonal elements of Π^{-1} are all ≥ 2 and hence the smallest eigenvalue of $(C + \Pi^{-1})$ is ≥ 2 . By continuity, we find that $(C\Pi + I)^T(C\Pi + I) - 4\Pi^2$ is non-negative definite, even if Π is singular. Therefore

$$\sum_{j=1}^{\ell} \alpha_j^2 = \beta^T(C\Pi + I)^T(C\Pi + I)\beta \geq 4\beta^T\Pi^2\beta = 4\sum_{j=1}^{\ell}(\pi_j\beta_j)^2 \quad .$$

By (5.2.4) we have $\max_i |c_i| \leq nb_s(\delta n)^{-2} = b_s\delta^{-2}n^{-1}$ on a set of probability at least $1 - B_1\exp\{-\beta_1 n\}$ and on this set (5.2.7) ensures that

$$\begin{aligned}
\max_i |\beta_i| &\leq \max_i |\alpha_i| + b_s\delta^{-2}n^{-1}\sum_{j=1}^{\ell}\pi_j|\beta_j| \\
&\leq \max_i |\alpha_i| + b_s\delta^{-2}n^{-1/2}\Big(\sum_{j=1}^{\ell}(\pi_j\beta_j)^2\Big)^{1/2} \\
&\leq \max_i |\alpha_i| + \tfrac{1}{2}b_s\delta^{-2}n^{-1/2}\Big(\sum_{j=1}^{\ell}\alpha_j^2\Big)^{1/2} \\
&\leq (1 + \tfrac{1}{2}b_s^{3/2}\delta^{-2})\max_i |\alpha_i| \quad .
\end{aligned}$$

Hence (5.2.11) follows from (5.2.5) . A little arithmetic shows that for $|\beta_i| \leq \log 2$,

$$\left|\frac{\tilde{P}_i - P_i}{P_i(1 - P_i)} - \beta_i\right| \leq 4\beta_i^2 \tag{5.2.19}$$

and (5.2.11) produces (5.2.12) and (5.2.13) .

Finally, (5.2.7) may be written as

$$\beta_i = \alpha_i(t) - 2\int_{[t_i+c,t]} m(s)^{-2}\sum_{j=1}^{nX_n(s-c)} P_j(1 - P_j)\beta_j dN(s) \quad , \tag{5.2.20}$$

for $i = 1,\dots,\ell$ and together with (5.2.3) and (5.2.10) this yields

$$\delta_i(t) = \int_{[t_i+c,t]} \Big[\tilde{m}(s)^{-1} - m(s)^{-1} + 2m(s)^{-2} \sum_{j=1}^{nX_n(s-c)} P_j(1 - P_j)\beta_j\Big] dN(s) \quad ,$$

for $i = 1,\dots,\ell$. Using (5.2.2) and (5.2.9) we find

$$\delta_i(t) = \int_{[t_i+c,t]} \Big[-2m(s)^{-2} \sum_{j=1}^{nX_n(s-c)} (\tilde{P}_j - P_j - P_j(1 - P_j)\beta_j)$$

$$+ 4m(s)^{-2}\tilde{m}(s)^{-1} \Big(\sum_{j=1}^{nX_n(s-c)} (\tilde{P}_j - P_j)\Big)^2\Big] dN(s) \quad ,$$

Now (5.2.4) and (5.2.13) imply that $\inf_{s \le t_0} m(s) \ge \delta n$ and $\inf_{s \le t_0} \tilde{m}(s) \ge \frac{1}{2}\delta n$ with probability at least $1 - B_2\exp\{-\beta_2 n\}$, which proves (5.2.14) . Application of (5.2.12) and (5.2.13) establishes (5.2.15) .

□

Let Λ_z be the conditional intensity process process of the counting process N given $(Z_1 , Z_2 , \dots) = (z_1 , z_2 , \dots)$ as defined by (2.2.10) and (3.1.7) , i.e.

$$(5.2.21) \quad \Lambda_z(s) = (2n - N(s) + \sum_{j=1}^{nX_n(s-c)} z_j)Q(X_n(s))1_{[0,n^{-1}n_s)}(X_n(s)) \quad .$$

For $c \le s \le t \le t_0$, let

$$(5.2.22) \quad \Lambda_{\tilde{p}}(s) = (2n - N(s) + \sum_{j=1}^{nX_n(s-c)} \tilde{P}_j(t))Q(X_n(s))1_{[0,n^{-1}n_s)}(X_n(s)) \quad ,$$

in analogy to Λ_p as given in (3.1.2) . Note that in the notation we suppress the dependence of $\Lambda_{\tilde{p}}$ on t . It follows from (5.2.2) , (5.2.3) , (5.2.7) and (5.2.8) that the $\tilde{P}_i(t)$ and hence $\Lambda_{\tilde{p}}(t)$ are $\mathcal{F}_t$ - measurable, where $\mathcal{F}_t$ denotes the σ-algebra generated by $\{N(s)\}_{s \le t}$. In the next section we shall show that $\Lambda_{\tilde{p}}(t)$ is an asymptotic approximation with remainder $o(n^{1/2})$ for the unconditional intensity $\Lambda(t) = E(\Lambda_Z(t) | \mathcal{F}_t)$.

At this point it suffices to note that for $c \le s \le t \le t_0$, and provided that $Q(X_n(s)) > (da_s b_s n)^{-1}$,

$$
\begin{aligned}
(5.2.23) \qquad \log\{\Lambda_{\tilde{p}}(s)^{-1}\Lambda_z(s)\} &= \log(1 + M(z,s)) \\
&= M(z,s) - \tfrac{1}{2}M(z,s)^2 + \tfrac{1}{3}M(z,s)^3 + R(z,s) \quad ,
\end{aligned}
$$

where M is given by

$$
(5.2.24) \qquad M(z,s) = \tilde{m}(s)^{-1} \sum_{j=1}^{nX_n(s-c)} (z_j - 2\tilde{P}_j(t)) \quad .
$$

For fixed $t \in [c,t_0]$, let $\tilde{Z}(t) = (\tilde{Z}_1(t) , \tilde{Z}_2(t) , \ldots) = (\tilde{Z}_1 , \tilde{Z}_2 , \ldots)$ be a sequence of random variables. Given $\mathcal{F}_t$, $\tilde{Z}_1 , \tilde{Z}_2 , \ldots$ are independent and $\tilde{Z}_i$ has a binomial distribution with parameters 2 and $\tilde{P}_i = \tilde{P}_i(t)$. Note that $\tilde{P}_i = P_i = 0$ for $n > n_h$.

For $0 \le t \le t_0$, let $\hat{F}_t \in \mathcal{F}_t$ denote the set where $\inf_{s\le t} Q(X_n(s)) \ge \delta$, $\inf_{s\le t} m(s) \ge \delta n$, and $\inf_{s\le t} \tilde{m}(s) \ge \delta n$. By (5.2.4) and (5.2.14) we have $\mathcal{P}(\hat{F}_t) \ge 1 - B\exp\{-\beta n\}$ for $t \le t_0$.

<u>LEMMA 5.2.3</u> There exist positive numbers B , β and x_0 such that for $n = 1,2,\ldots$ and $x \ge 0$,

$$
(5.2.25) \qquad \sup_{c\le t\le t_0} \ \sup_{\hat{F}_t} \mathcal{P}\Big(\sup_{c\le s\le t} |M(\tilde{Z},s)| \ge x | \mathcal{F}_t\Big) \le B\exp\{-\beta x^2 n\} \quad ,
$$

whereas for $0 \le x \le x_0$,

$$
(5.2.26) \qquad \sup_{c\le t\le t_0} \ \sup_{\hat{F}_t} \mathcal{P}\Big(\sup_{c\le s\le t} |R(\tilde{Z},s)| \ge x | \mathcal{F}_t\Big) \le B\exp\{-\beta x^{1/2} n\} \quad .
$$

<u>PROOF.</u> Since $\tilde{P}_j = 0$ for $n > n_h$, LEMMA A1.2 yields (5.2.25) . To prove (5.2.26) , note that $|R(\tilde{Z},s)| < x$ if $|M(\tilde{Z},s)| \le \frac{1}{2}$ and $M(\tilde{Z},s)^4 < x$, and apply (5.2.25) .

□

5.3. AN APPROXIMATION OF Λ

The conditional joint density of τ_1 , τ_2 , ..., τ_k , $N(t-c)$, $N(t)$, given $\hat{Z} = z$, at t_1 , t_2 , ..., t_k , $\ell + n$, $k + n$, is

$$
\begin{aligned}
& f_{\hat{Z}=z}(t_1,t_2,\ldots,t_k,\ell+n,k+n) \\
&= \exp\{-A_z(t_1)\}\Lambda_z(t_1)\exp\{-(A_z(t_2) - A_z(t_1))\}\Lambda_z(t_2)\cdot\ldots \\
&\quad \cdot \exp\{-(A_z(t_\ell) - A_z(t_{\ell-1}))\}\Lambda_z(t_\ell)\exp\{-(A_z(t-c) - A_z(t_\ell))\} \\
&\quad \cdot \exp\{-(A_z(t_{\ell+1}) - A_z(t-c))\}\Lambda_z(t_{\ell+1}) \\
&\quad \cdot \exp\{-(A_z(t_{\ell+2}) - A_z(t_{\ell+1}))\}\Lambda_z(t_{\ell+2})\cdot\ldots \\
&\quad \cdot \exp\{-(A_z(t_k) - A_z(t_{k-1}))\}\Lambda_z(t_k)\exp\{-(A_z(t) - A_z(t_k))\} \\
&= \exp\{-A_z(t)\} \prod_{j=1}^{k} \Lambda_z(t_j)
\end{aligned}
$$

for $c \le t_1 < t_2 < \ldots < t_\ell \le t - c < t_{\ell+1} < t_{\ell+2} < \ldots < t_k \le t$. Recall that the Z_i have binomial distributions with parameters 2 and P_i . Hence the unconditional joint density of τ_1 , τ_2 , ..., τ_k , $N(t-c)$, $N(t)$, Z_1 , Z_2 , ..., Z_ℓ at t_1 , t_2 , ..., t_k , $\ell + n$, $k + n$, z_1 , z_2 , ..., z_ℓ , is

$$
\begin{aligned}
& f(t_1,t_2,\ldots,t_k,\ell+n,k+n,z_1,z_2,\ldots,z_\ell) \\
&= \prod_{i=1}^{\ell} \binom{2}{z_i} P_i^{z_i}(1 - P_i)^{2-z_i}\exp\{(-A_z(t)\} \prod_{j=1}^{k} \Lambda_z(t_j)
\end{aligned}
$$

for $c \le t_1 < t_2 < \ldots < t_\ell \le t - c < t_{\ell+1} < t_{\ell+2} < \ldots < t_k \le t$, and $z_i = 0, 1$ or 2 , $i = 1,2,\ldots,\ell$. Therefore the unconditional intensity Λ is given by

$$\Lambda(t) = \lim_{h \downarrow 0} h^{-1} P\,(N(t+h) - N(t) \geq 1 \mid F_t) = E(\Lambda_Z(t) \mid F_t)$$

$$= \Big[\sum_{z_1=0}^{2} \cdots \sum_{z_\ell=0}^{2} \prod_{i=1}^{\ell} \binom{2}{z_i} P_i^{z_i}(1 - P_i)^{2-z_i} \exp\{-A_z(t)\} \prod_{j=1}^{k} \Lambda_z(t_j) \Big]^{-1}$$

$$\cdot \sum_{z_1=0}^{2} \cdots \sum_{z_\ell=0}^{2} \prod_{i=1}^{\ell} \binom{2}{z_i} P_i^{z_i}(1 - P_i)^{2-z_i}$$

$$\cdot \exp\{-A_z(t)\}\Lambda_z(t) \prod_{j=1}^{k} \Lambda_z(t_j)$$

$$= \Big[\sum_{z_1=0}^{2} \cdots \sum_{z_\ell=0}^{2} \prod_{i=1}^{\ell} \binom{2}{z_i} P_i^{z_i}(1 - P_i)^{2-z_i} \exp\{-A_z(t)\} \prod_{j=1}^{k} \Lambda_{\tilde{p}}(t_j)$$

$$\cdot \exp\{ \sum_{j=1}^{k} \big(M(z,t_j) - \tfrac{1}{2}M(z,t_j)^2 + \tfrac{1}{3}M(z,t_j)^3 + R(z,t_j)\big) \} \Big]^{-1}$$

$$\cdot \sum_{z_1=0}^{2} \cdots \sum_{z_\ell=0}^{2} \prod_{i=1}^{\ell} \binom{2}{z_i} P_i^{z_i}(1 - P_i)^{2-z_i} \exp\{-A_z(t)\}\Lambda_{\tilde{p}}(t)$$

$$\cdot \prod_{j=1}^{k} \Lambda_{\tilde{p}}(t_j) \exp\{ \sum_{j=1}^{k} \big(M(z,t_j) - \tfrac{1}{2}M(z,t_j)^2 + \tfrac{1}{3}M(z,t_j)^3$$

$$+ R(z,t_j)\big) + M(z,t) - \tfrac{1}{2}M(z,t)^2 + \tfrac{1}{3}M(z,t)^3 + R(z,t)\} \quad ,$$

where we have used (5.2.23) . Note that the following factors occur in the numerator as well as in the denominator of the last expression:

$$\prod_{j=1}^{k} \Lambda_{\tilde{p}}(t_j)\exp\{-A_z(t) + \sum_{j=1}^{k} M(z,t_j)\} \prod_{i=1}^{\ell} P_i^{z_i}(1 - P_i)^{2-z_i}$$

$$= \Big[\prod_{j=1}^{k} \Lambda_{\tilde{p}}(t_j)\exp\{- \int_c^t (2n - N(s))Q(X_n(s))ds - 2\sum_{j=1}^{k} \sum_{i=1}^{nX_n(t_j-c)} \tilde{m}(t_j)^{-1}\tilde{P}_i\}\Big]$$

$$\cdot \Big[\exp\{- \int_c^t \sum_{i=1}^{nX_n(s-c)} z_i Q(X_n(s))ds + \sum_{j=1}^{k} \sum_{i=1}^{nX_n(t_j-c)} \tilde{m}(t_j)^{-1} z_i\}\Big]$$

$$\cdot \prod_{i=1}^{\ell} \binom{2}{z_i} P_i^{z_i}(1 - P_i)^{2-z_i} \quad .$$

The first factor in square brackets doesn't depend on $z_1 , \ldots , z_\ell$ and may therefore be cancelled in numerator and denominator. In view of (5.2.10) the second factor in square brackets equals $\exp\{\sum_{i=1}^{\ell}(\beta_i + \delta_i)z_i\}$. The factor $\exp\{\sum_{i=1}^{\ell}\beta_i z_i\}$ may be combined with $\prod_{i=1}^{\ell} P_i^{z_i}$ to yield $\prod_{i=1}^{\ell}(P_i e^{\beta_i})^{z_i}$. Dividing the numerator as well as the denominator by $\prod_{i=1}^{\ell}(P_i e^{\beta_i} + 1 - P_i)^2$, this amounts to replacing P_i by $\tilde{P}_i = \tilde{P}_i(t)$ as given in (5.2.8) throughout. Finally, we may replace $\exp\{\sum_{i=1}^{\ell}\delta_i z_i\}$ by $\exp\{\sum_{i=1}^{\ell}\delta_i(z_i - 2\tilde{P}_i)\}$ again. As a result we find

$$\Lambda(t) = \Big[\sum_{z_1=0}^{2} \cdots \sum_{z_\ell=0}^{2} \prod_{i=1}^{\ell} \binom{2}{z_i} \tilde{P}_i^{z_i}(1 - \tilde{P}_i)^{2-z_i}\exp\{\sum_{i=1}^{\ell}\delta_i(z_i - 2\tilde{P}_i)$$

$$+ \sum_{j=1}^{k} \big(-\tfrac{1}{2}M(z,t_j)^2 + \tfrac{1}{3}M(z,t_j)^3 + R(z,t_j)\big)\}\Big]^{-1}$$

$$\cdot\ \Lambda_{\tilde{p}}(t) \sum_{z_1=0}^{2} \cdots \sum_{z_\ell=0}^{2} \prod_{i=1}^{\ell} \binom{2}{z_i} \tilde{P}_i^{z_i}(1 - \tilde{P}_i)^{2-z_i}$$

$$\cdot \exp\{\sum_{i=1}^{\ell}\delta_i(z_i - 2\tilde{P}_i) \sum_{j=1}^{k} \big(-\tfrac{1}{2}M(z,t_j)^2 + \tfrac{1}{3}M(z,t_j)^3 + R(z,t_j)\big)$$

$$+ M(z,t) - \tfrac{1}{2}M(z,t)^2 + \tfrac{1}{3}M(z,t)^3 + R(z,t)\} \quad .$$

Now let us introduce the random variables $\tilde{Z}_1$, $\tilde{Z}_2$, ... , which were discussed in Section 5.2 , once more. Recall that, conditionally on $\mathcal{F}_t$, these variables are independent, $\tilde{Z}_j$ having a binomial distribution with parameters 2 and $\tilde{P}_j$. Hence

$$\Lambda(t) = \Bigg[E\Bigg(\exp\{\sum_{i=1}^{\ell} \delta_i(\tilde{Z}_i - 2\tilde{P}_i) + \sum_{j=1}^{k} \big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2$$

$$+ \tfrac{1}{3}M(\tilde{Z},t_j)^3 + R(\tilde{Z},t_j)\big)\}\,|\,\mathcal{F}_t\Bigg)\Bigg]^{-1}$$

$$(5.3.1) \qquad \cdot\ \Lambda_{\tilde{p}}(t)E\Bigg(\exp\{\sum_{i=1}^{\ell} \delta_i(\tilde{Z}_i - 2\tilde{P}_i)$$

$$+ \sum_{j=1}^{k} \big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2 + \tfrac{1}{3}M(\tilde{Z},t_j)^3 + R(\tilde{Z},t_j)\big)$$

$$+ \tilde{M}(\tilde{Z},t) - \tfrac{1}{2}M(\tilde{Z},t)^2 + \tfrac{1}{3}M(\tilde{Z},t)^3 + R(\tilde{Z},t)\}\,|\,\mathcal{F}_t\Bigg) \quad .$$

From here on, we shall have to approximate $\Lambda(t)$ and we shall take care to ensure that the error is uniform on $[0,t_0]$. Of course this can't be achieved on the entire probability space and for an arbitrarily small $\varepsilon > 0$, we shall restrict attention to a set $F_t \in \mathcal{F}_t$ with

$$(5.3.2) \qquad \sup_{c \le t \le t_0} \mathcal{P}(F_t) \ge 1 - \varepsilon \quad .$$

First of all, we choose this set as in LEMMA 5.2.3 to ensure that on F_t we have

$$(5.3.3) \qquad \inf_{s \le t} Q(X_n(s)) \ge \delta \quad , \qquad \inf_{s \le t} m(s) \ge \delta n \quad , \qquad \inf_{s \le t} \tilde{m}(s) \ge \delta n \quad ,$$

and (5.2.25) and (5.2.26) hold. In fact we have already made use of the first inequality to ensure that $Q(X_n(s)) \ge (da_s b_s n)^{-1}$ (cf.

(5.2.21) - (5.2.23)). Furthermore we require that on F_t ,

$$(5.3.4) \qquad \max_{1\le i\le \ell} |\beta_i(t)| \le Dn^{-1/2} \quad ,$$

$$(5.3.5) \qquad \max_{\substack{1\le i\le \ell \\ P_i>0}} \left| \frac{\tilde{P}_i(t) - P_i}{P_i(1 - P_i)} \right| \le Dn^{-1/2} \quad ,$$

$$(5.3.6) \qquad \max_{1\le i\le \ell} |\delta_i(t)| \le Dn^{-1} \quad ,$$

for some positive constant D and all n . This is allowed by (5.2.11) , (5.2.13) and (5.2.15) .

Let us first consider the denominator in (5.3.1) . Define the sets of outcomes

$$C_t = \{ | \sum_{j=1}^{k} M(\tilde{Z},t_j)^3 | \le n^{-1/2}(2\beta^{-1}\log n)^{3/2} \} \quad ,$$

$$\tilde{C}_t = \{ | \sum_{j=1}^{k} R(\tilde{Z},t_j) | \le n^{-1}(2\beta^{-1}\log n)^{2} \} \quad .$$

For $c \le t \le t_0$, (5.2.25) and (5.2.26) ensure that on F_t ,

$$\mathcal{P}\,(C_t^c | \mathcal{F}_t) \le n\mathcal{P}\,(\sup_{c\le s\le t} |M(\tilde{Z},s)| \ge (2\beta^{-1}n^{-1}\log n)^{1/2} | \mathcal{F}_t) \le Bn^{-1} \quad ,$$

$$\mathcal{P}\,(\tilde{C}_t^c | \mathcal{F}_t) \le n\mathcal{P}\,(\sup_{c\le s\le t} |R(\tilde{Z},s)| \ge (2\beta^{-1}n^{-1}\log n)^{2} | \mathcal{F}_t) \le Bn^{-1} \quad .$$

Furthermore from (5.2.23) and (5.3.6) we see that on F_t , we have for $c \le s \le t$,

$$(5.3.7) \qquad -\tfrac{1}{2}M(\tilde{Z},s)^2 + \tfrac{1}{3}M(\tilde{Z},s)^3 + R(\tilde{Z},s) = \log(1 + M(\tilde{Z},s)) - M(\tilde{Z},s) \le 0 \quad ,$$

$$(5.3.8) \qquad | \sum_{i=1}^{\ell} \delta_i(\tilde{Z}_i - 2\tilde{P}_i) | \le 2Db_s \quad .$$

Hence, on F_t ,

$$E\Big[\exp\{\sum_{i=1}^{\ell}\delta_i(\tilde{Z}_i - 2\tilde{P}_i) + \sum_{j=1}^{k}\big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2$$

$$+ \tfrac{1}{3}M(\tilde{Z},t_j)^3 + R(\tilde{Z},t_j)\big)\}1_{C_t^c \cup \tilde{C}_t^c} | \mathcal{F}_t\Big]$$

$$\le 2B\exp\{2Db_s\}n^{-1} \quad ,$$

so that, uniformly for $c \le t \le t_0$,

$$E\Big[\exp\{\sum_{i=1}^{\ell}\delta_i(\tilde{Z}_i - 2\tilde{P}_i) + \sum_{j=1}^{k}\big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2$$

$$+ \tfrac{1}{3}M(\tilde{Z},t_j)^3 + R(\tilde{Z},t_j)\big)\} | \mathcal{F}_t\Big]$$

$$= E\Big[\exp\{\sum_{i=1}^{\ell}\delta_i(\tilde{Z}_i - 2\tilde{P}_i) + \sum_{j=1}^{k}\big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2$$

$$+ \tfrac{1}{3}M(\tilde{Z},t_j)^3\big)\}1_{C_t \cap \tilde{C}_t} | \mathcal{F}_t\Big] + O(n^{-1}(\log n)^2)$$

$$= E\Big[\exp\{\sum_{i=1}^{\ell}\delta_i(\tilde{Z}_i - 2\tilde{P}_i) + \sum_{j=1}^{k}\big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2$$

$$+ \tfrac{1}{3}M(\tilde{Z},t_j)^3\big)\}1_{C_t} | \mathcal{F}_t\Big] + O(n^{-1}(\log n)^2) \quad .$$

To deal with the numerator in a similar manner, we notice that on F_t ,

$$(5.3.9) \qquad |M(\tilde{Z},t)| \le (\delta n)^{-1}2nb_s = 2b_s\delta^{-1}$$

and in view of (5.2.25) , the terms $-\frac{1}{2}M(\tilde{Z},t)^2 + \frac{1}{3}M(\tilde{Z},t)^3 + R(\tilde{Z},t)$ in the exponent can be relegated to the remainder term. Thus, uniformly for $c \le t \le t_0$,

$$
\begin{aligned}
\Lambda(t) = \Lambda_{\tilde{p}}(t)\Big[E\Big(&\exp\{\sum_{i=1}^{\ell}\delta_i(\tilde{Z}_i - 2\tilde{P}_i) + \sum_{j=1}^{k}\big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2 \\
&+ \tfrac{1}{3}M(\tilde{Z},t_j)^3\big)\}1_{C_t}\,|\mathcal{F}_t\Big) + \mathcal{O}(n^{-1}(\log n)^2)\Big]^{-1} \\
\cdot\ \Big[E\Big(&\exp\{\sum_{i=1}^{\ell}\delta_i(\tilde{Z}_i - 2\tilde{P}_i) + M(\tilde{Z},t) \\
&+ \sum_{j=1}^{k}\big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2 + \tfrac{1}{3}M(\tilde{Z},t_j)^3\big)\}1_{C_t}\,|\mathcal{F}_t\Big) + \mathcal{O}(n^{-1}(\log n)^2)\Big] \\
= \Lambda_{\tilde{p}}(t)\Big\{1 + \Big[E\Big(&\exp\{\sum_{i=1}^{\ell}\delta_i(\tilde{Z}_i - 2\tilde{P}_i) + \sum_{j=1}^{k}\big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2 \\
&+ \tfrac{1}{3}M(\tilde{Z},t_j)^3\big)\}1_{C_t}\,|\mathcal{F}_t\Big) + \mathcal{O}(n^{-1}(\log n)^2)\Big]^{-1} \\
\cdot\ \Big[E\Big(&(\exp\{M(\tilde{Z},t)\} - 1)\exp\{\sum_{i=1}^{\ell}\delta_i(\tilde{Z}_i - 2\tilde{P}_i) \\
&+ \sum_{j=1}^{k}\big(-\tfrac{1}{2}M(\tilde{Z},t_j)^2 + \tfrac{1}{3}M(\tilde{Z},t_j)^3\big)\}1_{C_t}\,|\mathcal{F}_t\Big) + \mathcal{O}(n^{-1}(\log n)^2)\Big]\Big\} .
\end{aligned}
$$

What we achieved in the last step is, heuristically speaking, that we replaced a ratio $W^{-1}(V + W)$ where both W an $(V + W)$ are $\mathcal{O}(1)$ by $1 + W^{-1}V$, where $V = \mathcal{O}(n^{-1/2})$. As will be seen, this enables us to allow remainder terms of larger order than we have so far. Using (5.3.8) and the definition of C_t , we find that deleting $\exp\{\frac{1}{3}\sum_{j=1}^{k} M(\tilde{Z},t_j)^3\}$ in the denominator will raise the remainder term in the denominator to $\mathcal{O}(n^{-1/2}(\log n)^{3/2})$. Next, we may delete the factor 1_{C_t} in the denominator, since this gives rise to a remainder term of the lower order $\mathcal{P}\,(C_t^c|\mathcal{F}_t) = \mathcal{O}(n^{-1})$.

In the numerator, deleting the factor $\exp\{\frac{1}{3}\sum_{j=1}^{k} M(\tilde{Z},t_j)^3\}$ will add a remainder term

$$
\mathcal{O}(n^{-1/2}(\log n)^{3/2})E\big(|\exp\{M(\tilde{Z},t)\} - 1|\ |\mathcal{F}_t\big) ,
$$

and (5.3.9) and (5.2.25) ensure that this is $\mathcal{O}(n^{-1}(\log n)^{3/2})$. For the same reason, replacing $\exp\{M(\tilde{Z},t)\} - 1$ by $M(\tilde{Z},t)$ will only add a remainder term $\mathcal{O}(n^{-1})$ and deleting 1_{C_t} produces another $\mathcal{O}(n^{-1})$ term in the numerator. All of this holds on F_t uniformly for $c \le t \le t_0$, and we find

$$(5.3.10)\quad \Lambda(t) = \Lambda_{\tilde{p}}(t)\left\{1 + \frac{V + \mathcal{O}(n^{-1}(\log n)^2)}{W + \mathcal{O}(n^{-1/2}(\log n)^{3/2})}\right\},$$

where

$$V = E\left[M(\tilde{Z},t)\exp\{\sum_{i=1}^{\ell} \delta_i(\tilde{Z}_i - 2\tilde{P}_i) - \frac{1}{2}\sum_{j=1}^{k} M(\tilde{Z},t_j)^2\} \mid \mathcal{F}_t\right],$$

$$W = E\left[\exp\{\sum_{i=1}^{\ell} \delta_i(\tilde{Z}_i - 2\tilde{P}_i) - \frac{1}{2}\sum_{j=1}^{k} M(\tilde{Z},t_j)^2\} \mid \mathcal{F}_t\right].$$

We begin the final part of our analysis by noting that (5.2.25) implies that on F_t and for $x = (\beta b_s)^{-1}\log(2B)$,

$$\mathcal{P}\left(\sum_{j=1}^{k} M(\tilde{Z},t_j)^2 \le x \mid \mathcal{F}_t\right) \ge \mathcal{P}\left(\sup_{c\le s\le t} M(\tilde{Z},s)^2 \le x(b_s n)^{-1} \mid \mathcal{F}_t\right)$$

$$\ge 1 - B\exp\{-\beta b_s x\} = \frac{1}{2}$$

and together with (5.3.8) this implies that there exists $\varepsilon > 0$ such that

$$(5.3.11)\quad W \ge \varepsilon$$

on F_t.

Expanding $\exp\{\sum_{i=1}^{\ell} \delta_i(\tilde{Z}_i - 2\tilde{P}_i)\}$ in V and using (5.2.25), (5.3.3) and (5.3.6), we find that on F_t,

$$(5.3.12)\quad V = E\left[M(\tilde{Z},t)\exp\{-\frac{1}{2}\sum_{j=1}^{k} M(\tilde{Z},t_j)^2\} \mid \mathcal{F}_t\right] + \mathcal{O}(n^{-1})$$

uniformly for $c \le t \le t_0$.

At this point, the random variables $\tilde{Z}_i = \tilde{Z}_i(t)$ have served their purpose and the fact that their distribution given $\mathcal{F}_t$ depends on t through the $\tilde{P}_i(t)$, is going to be bothersome. We shall replace them by random variables $\bar{Z}_1, \bar{Z}_2, \ldots$ which, given $\mathcal{F}_t$, are conditionally independent, $\bar{Z}_i$ having a binomial distribution with parameters 2 and P_i. To simplify our calculations we assume that $\tilde{Z}_i$ and $\bar{Z}_i$ are coupled as follows:

$$\tilde{Z}_i = \bar{Z}_i + W_i \quad ,$$

where, given $\mathcal{F}_t$ and $\bar{Z} = (\bar{Z}_1, \bar{Z}_2, \ldots)$, $W_1, W_2, \ldots$ are independent and W_i has a binomial distribution with parameters $(2 - \bar{Z}_i)$ and $(\tilde{P}_i - P_i)/(1 - P_i)$ if $\tilde{P}_i \geq P_i$, and $-W_i$ has a binomial distribution with parameters $\bar{Z}_i$ and $(P_i - \tilde{P}_i)/P_i$ if $\tilde{P}_i < P_i$. It is easy to see that this yields the correct marginal distribution for $\tilde{Z}_i$, and that

$$\tilde{Z}_i - 2\tilde{P}_i = \bar{Z}_i - 2P_i + W_i - E(W_i \mid \mathcal{F}_t, \bar{Z}) + \zeta_i(\bar{Z}_i - 2P_i) \quad , \tag{5.3.13}$$

where

$$E(W_i \mid \mathcal{F}_t, \bar{Z}) = (2 - \bar{Z}_i)\,\frac{\tilde{P}_i - P_i}{1 - P_i} \quad , \qquad \zeta_i = \frac{P_i - \tilde{P}_i}{1 - P_i} \quad \text{if } \tilde{P}_i \geq P_i \quad ,$$

$$E(W_i \mid \mathcal{F}_t, \bar{Z}) = \bar{Z}_i\,\frac{\tilde{P}_i - P_i}{P_i} \quad , \qquad \zeta_i = \frac{\tilde{P}_i - P_i}{P_i} \quad \text{if } \tilde{P}_i < P_i \quad .$$

It follows from (5.3.5) that for $j = 1,2,\ldots,n_h$, we have on the set F_t

$$E\left(\left[\sum_{i=1}^{j}(W_i - E(W_i \mid \mathcal{F}_t, \bar{Z}))\right]^4 \middle| \mathcal{F}_t, \bar{Z}\right) \leq (nb_h)^2 2^4 2^5 Dn^{-1/2}$$

$$= 2^9 Db_h{}^2 n^{3/2} \quad , \tag{5.3.14}$$

$$E\left(\left[\sum_{i=1}^{j}\zeta_i(\bar{Z} - 2P_i)\right]^4 \middle| \mathcal{F}_t\right) \leq (nb_h)^2(Dn^{-1/2})^4 2^4 = 2^4 D^4 b_h{}^2 \quad .$$

Define, for $c \le s \le t_0$,

$$(5.3.15) \quad L(z,s) = \tilde{m}(s)^{-1} \sum_{i=1}^{nX_n(s-c)} (z_i - 2P_i) \quad .$$

By (5.2.24) , (5.3.13) and (5.3.15)

$$\begin{aligned} M(\tilde{Z},s) - L(\overline{Z},s) &= \tilde{m}(s)^{-1} \sum_{i=1}^{nX_n(s-c)} (\tilde{Z}_i - 2\tilde{P}_i - \overline{Z}_i + 2P_i) \\ (5.3.16) \qquad &= \tilde{m}(s)^{-1} \sum_{i=1}^{nX_n(s-c)} (W_i - E(W_i | \mathcal{F}_t, \overline{Z}) + \zeta_i(\overline{Z}_i - 2P_i)) \quad , \end{aligned}$$

and (5.3.3) and (5.3.14) ensure that on F_t ,

$$\begin{aligned} E\left(\left[M(\tilde{Z},t_j) - L(\overline{Z},t_j)\right]^4 | \mathcal{F}_t\right) &\le (\delta n)^{-4}\left[2^{12} D b_h{}^2 n^{3/2} + 2^7 D^4 b_h{}^2\right] \\ &= \mathcal{O}(n^{-5/2}) \end{aligned}$$

for $j = 1,2,\dots,k$. Also, on $\mathcal{F}_t$,

$$E\left(\left[M(\tilde{Z},t_j) + L(\overline{Z},t_j)\right]^4 | \mathcal{F}_t\right) \le (\delta n)^{-4}(nb_h)^2 2^8 = \mathcal{O}(n^{-2})$$

for $j = 1,2,\dots,k$. It follows that on F_t ,

$$\begin{aligned} &\left| E\left(M(\tilde{Z},t)\exp\{-\tfrac{1}{2}\sum_{j=1}^{k} M(\tilde{Z},t_j)^2\} | \mathcal{F}_t\right) - E\left(M(\tilde{Z},t)\exp\{-\tfrac{1}{2}\sum_{j=1}^{k} L(\overline{Z},t_j)^2\} | \mathcal{F}_t\right) \right| \\ &\le \tfrac{1}{2} E\left(|M(\tilde{Z},t)| \left| \sum_{j=1}^{k} \{M(\tilde{Z},t_j)^2 - L(\overline{Z},t_j)^2\} \right| \Big| \mathcal{F}_t\right) \\ &\le \tfrac{1}{2}\left[E\left(M(\tilde{Z},t)^2 | \mathcal{F}_t\right) E\left(\left(\sum_{j=1}^{k} \{M(\tilde{Z},t_j)^2 - L(\overline{Z},t_j)^2\}\right)^2 \Big| \mathcal{F}_t\right)\right]^{1/2} \\ &\le \tfrac{1}{2}\left[(\delta n)^{-2} \tfrac{1}{2} n b_h k^2 \max_{1 \le j \le k} E\left(\left(M(\tilde{Z},t_j)^2 - L(\overline{Z},t_j)^2\right)^2 \Big| \mathcal{F}_t\right)\right]^{1/2} \end{aligned}$$

$$\leq \delta^{-1} b_s b_h{}^{1/2} n^{1/2} \max_{1\leq j\leq k} \left[E\left[\left(M(\tilde{Z},t_j) - L(\bar{Z},t_j)\right)^4 \middle| \mathcal{F}_t\right]\right.$$

$$\left.\cdot\ E\left[\left(M(\tilde{Z},t_j) + L(\bar{Z},t_j)\right)^4 \middle| \mathcal{F}_t\right]\right]^{1/4} = \mathcal{O}(n^{-5/8})$$

and together with (5.3.12) this yields

$$V = E\left[M(\tilde{Z},t)\exp\{-\tfrac{1}{2}\sum_{j=1}^{k} L(\bar{Z},t_j)^2\} \mid \mathcal{F}_t\right] + \mathcal{O}(n^{-5/8})$$

uniformly for $c \leq t \leq t_0$. From (5.3.16) we see that

$$E(M(\tilde{Z},t) - L(\bar{Z},t) \mid \mathcal{F}_t\ ,\ \bar{Z}) = \tilde{m}(t)^{-1} \sum_{i=1}^{\ell} \zeta_i(\bar{Z}_i - 2P_i)$$

and since $\tilde{m}(t)^{-2}E(\sum_{i=1}^{\ell} \zeta_i(\bar{Z}_i - 2P_i))^2 \leq \frac{1}{2}D^2(\delta n)^{-2}$ on F_t by (5.3.3) and (5.3.5) , we have

$$(5.3.17)\quad V = E\left[L(\bar{Z},t)\exp\{-\tfrac{1}{2}\sum_{j=1}^{k} L(\bar{Z},t_j)^2\} \mid \mathcal{F}_t\right] + \mathcal{O}(n^{-5/8})$$

on F_t , uniformly for $c \leq t \leq t_0$.

Consider the partial sum process

$$S_n(u) = n^{-1/2} \sum_{i=1}^{[nu]} (\bar{Z}_i - 2P_i)$$

for $0 \leq u \leq b_h$. The process $S_n(u)$ converges in distribution with respect to the supremum metric to

$$W(n^{-1} \sum_{i=1}^{[nu]} P_i(1 - P_i))$$

on $[0,b_h]$, where W is a standard Wiener process. A Skorohod construction allows us to define W in such a way that

$$\lim_{n\to\infty} \sup_{0\leq u\leq b_h} |S_n(u) - W(n^{-1} \sum_{i=1}^{[nu]} P_i(1 - P_i))| = 0 \quad \text{a.s.}$$

Since the processes S_n are independent of $\mathcal{F}_t$, we may also choose W independent of $\mathcal{F}_t$.

By (5.3.17)

$$V = \tilde{m}(t)^{-1} n^{1/2} E\left[S_n(\ell/n)\exp\{-\frac{1}{2}\sum_{j=1}^{k} \tilde{m}(t_j)^{-2} n S_n(\ell(j)/n)^2\} \,|\, \mathcal{F}_t \right] + O(n^{-5/8}) .$$

On F_t , $\tilde{m}(s)^{-1} \leq (\delta n)^{-1}$, and hence

$$S_n(\ell/n) 1_{[-a,a]}(S_n(\ell/n)) \exp\{-\frac{1}{2}\sum_{j=1}^{k} \tilde{m}(t_j)^{-2} n S_n(\ell(j)/n)^2\}$$

is a continuous and bounded function of S_n with respect to the supremum distance, and hence its conditional expectation tends to that of the same function of the corresponding Wiener process. Letting a tend to infinity we find that on F_t

$$(5.3.18) \quad \begin{aligned} V = \tilde{m}(t)^{-1} n^{1/2} E\Big[& W(n^{-1} \sum_{i=1}^{\ell} P_i(1 - P_i)) \\ & \cdot \exp\{-\frac{1}{2}\sum_{j=1}^{k} \tilde{m}(t_j)^{-2} n W(n^{-1} \sum_{i=1}^{\ell(j)} P_i(1 - P_i))^2\} + o(1)\Big] . \end{aligned}$$

Since there is only a single weakly convergent sequence of processes involved for all $t \in [c,t_0]$, the $o(1)$ term is still uniform for $c \leq t \leq t_0$.

Finally we note that the conditional expectation in (5.3.18) vanishes because

$$(5.3.19) \quad E(W(t) \,|\, W(s)^2 , 0 \leq s \leq t) = 0 .$$

Combination of (5.3.10) , (5.3.11) , (5.3.18) and (5.3.19) yields

$$(5.3.20) \quad \Lambda(t) = \Lambda_{\tilde{p}}(t)(1 + o(n^{-1/2})) = \Lambda_{\tilde{p}}(t) + o(n^{1/2})$$

on F_t , uniformly for $c \leq t \leq t_0$. By choosing $\varepsilon > 0$ arbitrarily small in (5.3.2) , we see that (5.3.20) implies that

$$(5.3.21)\qquad \sup_{c\le t\le t_0} n^{-1/2}|\Lambda(t) - \Lambda_{\tilde{p}}(t)| = o_P(1) \quad .$$

As in Section 5.1 , let A_n denote the unconditional integrated intensity process

$$A_n(t) = \int_c^t \Lambda_n(s)ds$$

and, re-introducing the subscript n , let $\Lambda_{\tilde{p},n}$ be given by (5.2.22) , (5.2.3) , (5.2.7) and (5.2.8) . Define

$$A_{\tilde{p},n}(t) = \int_c^t \Lambda_{\tilde{p},n}(s)ds \quad .$$

We have proved

<u>THEOREM</u> <u>5.3.1</u> For every $t_0 > 0$,

$$(5.3.22)\qquad \sup_{c\le t\le t_0} n^{-1/2}|A_n(t) - A_{\tilde{p},n}(t)|$$

tends to zero in probability.

□

By Theorem VI.10 in POLLARD (1984) this implies that $n^{-1/2}(A_n - A_{\tilde{p},n}) \overset{\mathcal{D}}{\to} 0$ in $D[0,\infty)$ and combining this with (5.1.6) we obtain

<u>THEOREM</u> <u>5.3.2</u>

$$(5.3.23)\qquad n^{-1/2}(N_n(t) - n - A_{\tilde{p},n}(t)) \overset{\mathcal{D}}{\to} W_0(X(t)) \quad ,$$

where W_0 is a standard Wiener process and the convergence is in the sense of the Skorohod metric in $D[0,\infty)$.

□

We end this section with two remarks. The first is that we have shown in this section that $\Lambda_{\tilde{p},n}$ is a sufficiently accurate approximation of Λ_n , even though $\Lambda_{p,n}$ is not. What this amounts to, in fact, is that

$$E(Z_{i,n}|N_n(s)\ ,\ s \le t) = 2\tilde{P}_{i,n}(t) + o_p(n^{-1/2})\quad ,$$

whereas $(\tilde{P}_{i,n}(t) - P_{i,n})$ is typically of exact order $n^{-1/2}$.

The second remark is that, although THEOREM 5.3.2 is quite satisfactory from a mathematical point of view, it does not yet solve the practical problem of divising a test for our model. To apply THEOREM 5.2.3 one needs to compute $\Lambda_{\tilde{p},n}(t)$ for at least one - but preferably many more - values of t , and for this one needs to know $\Lambda_{\tilde{p},n}(s)$ and hence $N_n(s)$ for all $s \in (c,t)$. This can only approximately be the case in situations where the sampling frequency is high. Moreover, solving the ℓ linear equations (5.2.7) for β_1 , ... , β_ℓ is certainly a non-trivial problem, especially since ℓ is of the same order as n . In Section 5.4 we provide a solution to this problem, but this solution still involves two sequences of ℓ numbers which are defined iteratively. Thus, more work on this problem is needed in order to be able to compare this kind of model by means of the distribution results of this chapter or those of Section 3.3 with actual data. In those cases where the measurement error in N_n exceeds the order $n^{1/2}$ of the difference between N_n and $n(X + 1)$, however, we can resort to non-linear regression techniques. This will be dealt with in the next chapter.

5.4. SOME LINEAR ALGEBRA

To compute $\Lambda_{\tilde{p},n}(t)$, we notice that

$$\Lambda_{\tilde{p},n}(t) = \Lambda_{p,n}(t) + 2Q(X_n(t))1_{[c,n^{-1}n_{s,n}]}(X_n(t))\sum_{i=1}^{\ell}(\tilde{P}_{i,n}(t) - P_{i,n}) \tag{5.4.1}$$

where, as in (3.1.2) ,

$$\Lambda_{p,n}(t) = (2n - N(t) + 2\sum_{i=1}^{nX_n(t-c)} P_{i,n})Q(X_n(t))1_{[c,n^{-1}n_{s,n})}(X_n(t))\quad . \tag{5.4.2}$$

In this section we shall find an explicit expression for

$\sum_{i=1}^{\ell} (\tilde{P}_{i,n}(t) - P_{i,n})$.

Let us adopt the notation introduced in (5.2.16) - (5.2.18) . By (5.2.8) , (5.2.12) and (5.2.16) - (5.2.18) ,

$$
\begin{aligned}
2\sum_{i=1}^{\ell} (\tilde{P}_i - P_i) &= 2\sum_{i=1}^{\ell} P_i(1 - P_i)\beta_i + O_P(1) \\
&= 1^T\Pi\beta + O_P(1) = 1^T\Gamma\alpha + O_P(1) = \sum_{i=1}^{\ell}\sum_{i'=1}^{\ell} \gamma_{ii'}\alpha_{i'} + O_P(1) \quad ,
\end{aligned}
\tag{5.4.3}
$$

where $1^T = (1,\ldots,1)_{1\times\ell}$ and $\Gamma = (\gamma_{ii'})_{\ell\times\ell} = \Pi(C\Pi + I)^{-1}$. Note that Γ is symmetric.

Because of the special structure of $(C\Pi + I)$ we may obtain an expression for Γ after a bit of matrix manipulation. By means of direct computation the reader can verify that $\Gamma = (\gamma_{ii'})$ is given by

$$
\gamma_{ii'} = \begin{cases}
-\mu_i^{-1}\lambda_{i'} \prod_{k=i+1}^{i'-1} (1 - \lambda_k) + \lambda_i\lambda_{i'} \sum_{\sigma=1}^{i-1} \mu_\sigma^{-1} & \\
\quad \cdot \prod_{j=\sigma+1}^{i-1} (1 - \lambda_j) \prod_{k=\sigma+1}^{i'-1} (1 - \lambda_k) \quad , & i < i' \quad , \\
\mu_i^{-1} + \lambda_i^2 \sum_{\sigma=1}^{i-1} \mu_\sigma^{-1} \prod_{j=\sigma+1}^{i-1} (1 - \lambda_j)^2 \quad , & i = i' \quad , \\
-\mu_{i'}^{-1}\lambda_i \prod_{k=i'+1}^{i-1} (1 - \lambda_k) + \lambda_i\lambda_{i'} \sum_{\sigma=1}^{i'-1} \mu_\sigma^{-1} & \\
\quad \cdot \prod_{j=\sigma+1}^{i-1} (1 - \lambda_j) \prod_{k=\sigma+1}^{i'-1} (1 - \lambda_k) \quad , & i > i' \quad ,
\end{cases}
\tag{5.4.4}
$$

where empty sums and products have to be interpreted as 0 and 1 respectively, and λ_i and μ_i , $i = 1,\ldots,\ell$, are defined by

$$(5.4.5)\qquad \begin{cases} \lambda_\ell = c_\ell(c_\ell + b_\ell)^{-1} \;, \quad d_\ell = c_\ell^{2}(c_\ell + b_\ell)^{-1} \;, \quad \mu_\ell = c_\ell + b_\ell \;, \\ \lambda_i = (c_i - \sum_{j>i}^{\ell} d_j)(c_i + b_i - \sum_{j>i}^{\ell} d_j)^{-1} \;, \qquad i = 1,\ldots,\ell - 1 \;, \\ d_i = (c_i - \sum_{j>i}^{\ell} d_j)\lambda_i \;, \qquad i = 1,\ldots,\ell - 1 \;, \\ \mu_i = c_i + b_i - \sum_{j>i}^{\ell} d_j \;, \qquad i = 1,\ldots,\ell - 1 \;. \end{cases}$$

Here c_i is given by (5.2.18) and

$$b_i = [2P_i(1 - P_i)]^{-1} \;, \qquad i = 1,\ldots,\ell \;.$$

Note that uniformly for $c \le t \le t_0$,

$$(5.4.6)\qquad \lambda_i = \mathcal{O}_P(n^{-1}) \;, \quad \mu_i = \mathcal{O}_P(1) \;, \qquad i = 1,\ldots,\ell \;.$$

Laborious calculation yields

$$(5.4.7)\qquad \begin{aligned} \sum_{i'=1}^{\ell} \gamma_{ii'} = \mu_i^{-1} \sum_{\sigma=i+1}^{\ell} (1 - \lambda_\sigma) - \lambda_i \sum_{j=1}^{i-1} \mu_j^{-1} \prod_{\sigma=j+1}^{\ell} (1 - \lambda_\sigma) \\ \cdot \prod_{\tau=j+1}^{i-1} (1 - \lambda_\tau) \;. \end{aligned}$$

Hence

$$(5.4.8)\qquad \sup_{c \le t \le t_0} |\Lambda_{\tilde{p},n}(t) - \hat{\Lambda}_n(t)| = \mathcal{O}_P(1) \;,$$

where

$$\hat{\Lambda}_n(t) = \Lambda_{p,n}(t) + 2Q(X_n(t)) 1_{[c, n^{-1} n_{s,n}]}(X_n(t))$$

$$\cdot \sum_{i=1}^{\ell} \Big[\mu_{i,n}(t)^{-1} \sum_{\sigma=i+1}^{\ell} (1 - \lambda_{\sigma,n}(t)) - \lambda_{i,n}(t) \sum_{j=1}^{i-1} \mu_{j,n}(t)^{-1} \tag{5.4.9}$$

$$\cdot \prod_{\sigma=j+1}^{\ell} (1 - \lambda_{\sigma,n}(t)) \prod_{\tau=j+1}^{i-1} (1 - \lambda_{\tau,n}(t)) \Big] \alpha_{i,n}(t) \quad .$$

Define

$$\hat{A}_n(t) = \int_c^t \hat{\Lambda}_n(s) ds \quad .$$

THEOREM 5.3.2 and (5.4.8) yield

<u>THEOREM 5.4.1</u>

$$n^{-1/2}(N_n(t) - n - \hat{A}_n(t)) \overset{\mathcal{D}}{\to} W_0(X(t)) \quad , \tag{5.4.10}$$

where W_0 is a standard Wiener process.

□

CHAPTER 6

COMPARISON OF THE MODEL WITH DATA

In order to test the agreement between theory and actual data, experiments were performed in which a batch culture of tobacco cells was followed in time. Samples from the culture were taken from the time at which the culture was inoculated until growth by cell division had stopped. The cells in the samples were counted and from this the population growth was determined. Also the concentration of sugars in the medium was measured at various time points. The results are discussed and statistically analysed.

6.1. INTRODUCTION

In the previous chapters we have formulated a model for the behaviour of individual plant cells in terms of cell cycle events, and we have considered the consequences of this model for the growth of batch cultures of these cells. In order to compare the predictions of the model with experimental data on population growth of plant cells, we needed more detailed information than is present in available published and unpublished data. For this reason we designed a set of specific experiments with the primary aim of obtaining precise information about the growth in terms of cell number of batch cultures. In what follows by (population) growth we mean proliferation, that is growth by cell division, and by cell growth we mean growth by cell enlargement.

Since in our laboratory cell lines of tobacco (designated LT lines) are grown which are being used for molecular studies on cell division and cell differentation (cf. VAN DER ZAAL et al., (1987); VAN DER ZAAL et al., 1987), these cell lines were also chosen for our experiments. The LT lines are suitable for the experiments, because they have reasonable doubling times, to wit approximately 30 h under standard culture conditions, and because they only require the auxin analogue 2,4-dichlorophenoxyaceticacid (2,4-D) as a growth factor (see Section 6.2). Moreover, the characteristics of plant cell lines may change rather drastically in time from their initiation from callus. In many cases cell lines obtain stable characteristics only after years in culture. The LT lines have been grown in our laboratory for about ten years and apparently have stable features. For instance, the LT lines have stabilized at one ploidy class and probably accumulate at G_1 at the end of each 7-days growth cycle.

To obtain as much information as possible, the experiments were designed to allow frequent sampling from the cell cultures during the entire growth cycle. The routine culture system for maintenance of the cell lines, consisting of a relatively large number of small culture vessels (250 ml erlenmeyer flasks each containing 50 ml of culture medium), did not seem to be suitable for reliable long-term sampling procedures. A much more appropriate device was a 3 l turbine stirred round bottom fermentor equipped with an autosampling device which we could use at the Department of Biochemical Engineering of the University of Delft. Preliminary experiments showed that transfer of inocula of cells from the stock cultures to the fermentor probably has no significant influence on the basic growth characteristics of the LT lines. Thus, no adaptation cycles were required for obtaining undisturbed growth of the cells in the fermentor. The only precaution taken was inoculation of the fermentor with a higher cell density ($1.5 - 2.0 \cdot 10^5$ cells ml^{-1}) than the one generally used in the stock cultures ($0.5 - 1.0 \cdot 10^5$ cells ml^{-1}).

After the first trials two separate but essentially similar pilot experiments were performed to analyse the increase of cell density and substrate consumption in the fermentor. Due to practical limitations the sampling procedure consisted of drawing no more than three samples at every 4 - 8 h over a culture period of approximately 10 days. Two of them were used for cell counting and one for analysis of substrate consumption.

Another practical constraint was brought about by the very laborious and time consuming cell-counting method, which has to be employed for plant cells. Since plant cells grow as small cell aggregates of varying size, no sophisticated devices for particle counting can be used. The cells have to be counted by eye using a light microscope.

Thus far, the data of one of the pilot experiments have been analysed in sufficient detail. The results will be described and discussed in the following sections.

6.2. MATERIALS AND METHODS

Cell culture

In the experiments 2,4-D-dependent cell cultures of *Nicotiana tabacum* L. c.v. White Burley (designated LT) were used. These tobacco cultures were obtained from leaf callus and were grown in suspension for about 10 years. The various cell lines differed in the concentration of 2,4-D in which they were maintained.

For the experiments described in this chapter, the LT 0.05 line was chosen. This line was grown in LS-medium (LINSMAIER & SKOOG, 1965) containing 0.05 mg l^{-1} ($2.26 \cdot 10^{-7}$ M) 2,4-D at 25° C in a dark room on a gyrotary shaker (120 rpm, 5 cm displacement). For maintenance of the culture, once every week the cells were transferred to a fresh medium at an initial density of about $0.5 - 1.0 \cdot 10^{5}$ cells ml^{-1} in 250 ml erlenmeyer flasks containing 50 ml of the growth medium.

When the cells were used for the experiments, a sample of ca. 200 ml of suspension which was at the end of its 7-days growth cycle (early stationary phase), was transferred to a commercially available 3 l turbine stirred round bottom fermentor (Applicon) containing 1800 ml of LS-medium supplemented with 0.05 mg l^{-1} 2,4-D and 0.25 ml of Silicone antifoam agent (BHD). The fermentor was autoclave sterilized at 120° C, the medium at 110° C, both for 30 min. A schematic representation of the experimental system is given in Fig. 6.1 . The initial density in the fermentor was about $1.5 - 2.0 \cdot 10^{5}$ cells ml^{-1}.

To resuspend the cells, after inoculation the two six-blade turbine impellers were set at a rotation speed of 750 rpm for 2 min. Subsequently, the rotation speed was set at 128 rpm. The suspension was kept at a temperature of 24.5° C by pumping temperature controlled water through the water jacket of the vessel. The culture was aerated by means of a sintered steel sparger at the bottom of the vessel at a rate of 30 vvm. Predried influent air passed through a cotton wool filter and a 0.2 µm hydrofobe membrane filter (Millipore, Milex FG50). Effluent air left the fermentor through a glass condensor kept at 4° C by means of a cryostat (Lauda). The evaporation of water was measured by leading the effluent air through a pre-weighed silica column. From time to time, i.e. at day 1, 2 and 3, about 10 ml of a diluted antifoam solution was added. The initial pH was 5.6 and as a check for undisturbed growth it was continuously measured by a sterilized pH-electrode (Ingold).

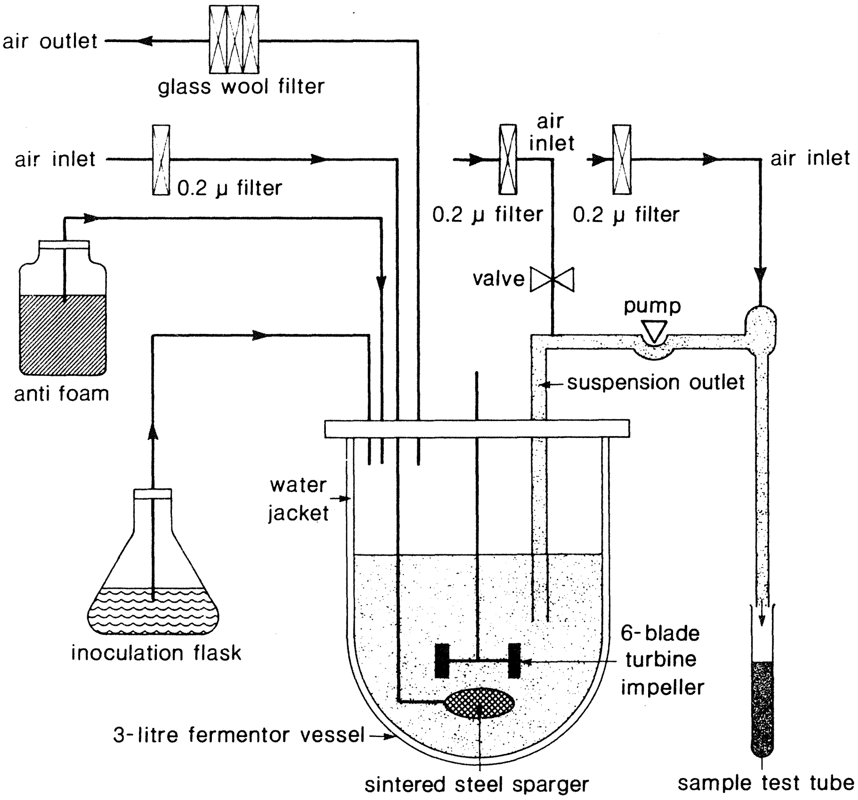


Fig.6.1. A schematic representation of the experimental system.

Sampling

The sampling of cells from the fermentor was conducted by an autosampling system ruled by a programmable sequential timer (Omron). Every four hours three samples of about 7.5 ml were pumped into different test tubes, which were weighed before. The true sample volume per tube was calculated from the increase of the weight of each tube. Of each set of three tubes, the first two contained 1.0 ml of a CrO_3 solution (60 %). These first two samples were used for cell counts. In the third sample the medium was separated from the cells by means of a cotton wool filter in the upper part of the tube. This medium was used for the analysis of several components of the medium. The three samples were drawn consecutively, the whole operation lasting about 10 min.

Cell counting

Dividing plant cells tend to stick together by means of their walls. Hence, plant cell cultures consist of cell aggregates of varying size. For cell counts it is necessary to separate the cells as much as possible. Since CrO_3 destroys the middle lamella which holds cell walls together, it is an agent for maceration of cell aggregates. This is why the tubes for the cell counts contained a CrO_3 solution. These tubes were incubated during 20 min at 70° C in order to optimize the cell separation. Next, the cells were pelleted by centrifugation at 3000 rpm for 2 - 5 min. The cellular pellet was resuspended in water, the final volume being 40 ml. Remaining cell aggregates were separated further by several passages through a hypodermal syringe. From each of the 40 ml solutions 5 samples of 1 ml were taken. Depending on the time at which the original sample was taken, a suitable dilution of the 1 ml samples was made such that 0.025 ml contained 100-200 cells. From each of the diluted solutions a sample of 0.025 ml was taken. In these final samples the number of cells was determined by means of cell counting using a light microscope. Each 0.025 ml sample was counted independently by two persons. The cell density in the fermentor at each time point was estimated by the mean of the 20 counts of the 0.025 samples corresponding to this time point.

Analysis of substrate consumption

Sucrose, fructose and glucose concentrations were determined by high performance liquid chromatography (Waters). An HPX 87P column (bio-rad) was used. Its temperature was 85° C. The eluent was doubly destilled deionized water; its flow rate was 0.6 ml $\min^{-1}$. Detection of the sugars was performed by means of a refractometer. The detection bound was 0.5 g l^{-1}.

Computation of the least squares estimates of the parameters
The parameter values which were used in Fig. 6.2, 6.4 and 6.5 were iteratively obtained by means of a least squares method. For several fixed values of the starting cell density n and the parameter c, the quantity $S^2 = \sum_{i=1}^{27} \frac{(\hat{N}(t_i) - n - nX(t_i))^2}{\hat{N}(t_i)^2}$ was minimized with respect to the parameters d, a_s, b_s, a_h, b_h. Here for $i = 1,\ldots,27$, $\hat{N}(t_i)$ denotes the mean of the 20 estimates of the cell density at the observation time t_i originating from the 20 0.025 ml samples corresponding to t_i, and $X(t_i)$ is the value of the limit function X, which is the solution of (3.1.13) at t_i.

For the values $n = 162500$ ml^{-1}, $c = 26$ h and several sets of different starting values of the other parameters, the parameter values which minimized S^2 were found to be $d = 4$ h, $a_s = 0.04$, $b_s = 29.9$, $a_h = 0.07$, $b_h = 4.7$. Note that this means for the parameters y_s, k_s, y_h, k_h which were defined in Chapter 2,: $y_s = 5.4 \cdot 10^{10}$ cel mol^{-1}, $k_s = 2.2 \cdot 10^{-11}$ mol cel^{-1}, $y_h = 3.4 \cdot 10^{15}$ cel mol^{-1}, $k_h = 9.7 \cdot 10^{-17}$ mol cel^{-1}. The minimal value of S^2 was 0.25.

The algorithm which was used for the iterative solution of the non-linear least squares problem was a Gauss-Newton method (see e.g. STOER & BURLISCH, 1980). In each iteration step the values of the function X at the time points $t_1, \ldots, t_{27}$ were computed numerically by means of a Runge-Kutta method, where Hermite interpolation of third order was used to approximate unknown function values. This method is described by IN'T HOUT (1988), who also proved the convergence of this procedure. A detailed description of all computational procedures is given in PIETERS (1988).

6.3. STATISTICAL ANALYSIS

To analyse the results of the experiment described in the previous section we first have to formulate a statistical model for the sampling procedure as used in the experiments. Let $t_1 < t_2 < \dots < t_r$ denote the observation times, and $N_1, N_2, \dots, N_r$ the number of cells present in the fermentor at these times. Let $n_{ijk\ell}$ denote the number of cells counted by investigator ℓ ($\ell = 1,2$) in 0.025 ml of the diluted k-th 1 ml subsample ($k = 1,2,\dots,5$) of the resuspended cells in the j-th 7.5 ml sample ($j = 1,2$) taken at time t_i $(i = 1,2,\dots,r)$. On the basis of $n_{ijk\ell}$ an estimate $\hat{N}_{ijk\ell}$ of N_i is computed by successive multiplication by the various measured dilution factors f_{ijk} and f_{ij} :

$$\hat{N}_{ijk\ell} = n_{ijk\ell} f_{ijk} f_{ij} \quad .$$

However, due to measurement errors these dilution factors differ from the true factors ϕ_{ijk} and ϕ_{ij} . In fact we have

$$\frac{f_{ijk}}{\phi_{ijk}} = 1 + \varepsilon_{ijk} \quad , \qquad \frac{f_{ij}}{\phi_{ij}} = 1 + \varepsilon_{ij} \quad ,$$

and since each of these errors is caused each time by an almost identical sequence of actions, it is reasonable to assume that the ε_{ijk} are identically distributed for each (i,j,k) , that the ε_{ij} are identically distributed for each (i,j) , and that all ε's are independent.

A second type of error is the error in the counted numbers of cells $n_{ijk\ell}$. If the true number of cells is denoted by $\nu_{ijk\ell}$, we may again model this as a multiplicative error, i.e. $n_{ijk\ell} = \nu_{ijk\ell}(1 + \varepsilon_{ijk\ell})$ with independent and identically distributed (i.i.d.) $\varepsilon_{ijk\ell}$, which are also independent of the ε's previously defined. Here the justification for a multiplicative error model is simply that approximately the same number of cells is counted each time.

A third source of errors are the repeated (sub-)sampling procedures. These sampling errors occur when taking the 2 original 7.5 ml samples, the 2×5 1 ml samples from the resuspended cells, and the 2×5 0.025 ml samples used for counting. Under ideal mixing conditions the cells in a vessel are spatially distributed as points in a three dimensional Poisson process, and sampling a fixed volume will give rise to a sampling error of order $n^{1/2}$, which is presumably negligible compared to the errors defined so far, whose absolute magnitude is of order n . Of course, such ideal mixing conditions do not prevail, but clustering would have to be excessive to produce an error of order n . In that case, however, these errors could be subsumed under the multiplicative errors we have introduced already.

Finally, there are errors caused by evaporation of water and addition of antifoam. A rough calculation shows that these errors may be neglected.

These considerations lead us to adopt the multiplicative model

$$\hat{N}_{ijk\ell} = (1 + \varepsilon_{ij})(1 + \varepsilon_{ijk})(1 + \varepsilon_{ijk\ell})\phi_{ij}\phi_{ijk}\nu_{ijk\ell}$$

$$= N_i(1 + \varepsilon_{ij})(1 + \varepsilon_{ijk})(1 + \varepsilon_{ijk\ell})(1 + \text{negligible terms}) \quad .$$

Taking logarithms we obtain the additive model

$$(6.3.1) \qquad Y_{ijk\ell} = \log\hat{N}_{ijk\ell} = \log N_i + e_{ij} + e_{ijk} + e_{ijk\ell} \ , \quad \begin{array}{l} i = 1,2,\dots,r \ , \\ j = 1,2 \ , \\ k = 1,2,\dots,5 \ , \\ \ell = 1,2 \ , \end{array}$$

where each of the groups of random variables $e_{ij} = \log(1 + \varepsilon_{ij})$, $e_{ijk} = \log(1 + \varepsilon_{ijk})$ and $e_{ijk\ell} = \log(1 + \varepsilon_{ijk\ell})$ are identically distributed and all e's are independent. In the absence of systematic errors it is reasonable to assume that the e's are all normally distributed with mean zero and variances

$$(6.3.2) \qquad \sigma^2(e_{ij}) = \sigma_1^2 \ , \quad \sigma^2(e_{ijk}) = \sigma_2^2 \ , \quad \sigma^2(e_{ijk\ell}) = \sigma_3^2 \ .$$

Note that the classical dilemma whether to assume normality of the ε's or the e's is of no great significance here since the $|\varepsilon|$'s are very much smaller than 1 .

According to THEOREM 3.2.1 we have

$$N_i = N(t_i) = n(1 + X(t_i)) + n\delta_i \ ,$$

where $\delta_1 , \dots , \delta_r$ are dependent errors which are $\mathcal{O}_P(n^{-1/2})$. Hence

(6.3.3) $$\log N_i = \log\{n(1 + X(t_i))\} + \frac{\delta_i}{1 + X(t_i)} + \mathcal{O}_p(n^{-1}) \quad .$$

To a first approximation our model reduces to

(6.3.4) $$Y_{ijk\ell} = \log\{n(1 + X(t_i))\} + e_{ij} + e_{ijk} + e_{ijk\ell} \quad .$$

Consider the more general model

(6.3.5) $$Y_{ijk\ell} = \mu_i + e_{ij} + e_{ijk} + e_{ijk\ell}$$

for arbitrary $\mu_1 , \ldots , \mu_r$. Define the sums of squares

$$S_1 = \sum_{i=1}^{r} \sum_{j=1}^{2} \sum_{k=1}^{5} \sum_{\ell=1}^{2} (Y_{ijk\ell} - Y_{ijk\cdot})^2 \quad ,$$

(6.3.6) $$S_2 = 2 \sum_{i=1}^{r} \sum_{j=1}^{2} \sum_{k=1}^{5} (Y_{ijk\cdot} - Y_{ij\cdot\cdot})^2 \quad ,$$

$$S_3 = 10 \sum_{i=1}^{r} \sum_{j=1}^{2} (Y_{ij\cdot\cdot} - Y_{i\cdot\cdot\cdot})^2 \quad ,$$

where a dot instead of an index indicates averaging over the possible values of that index. It is easy to see that under the general model (6.3.5) ,

$$S_1 \cong \sigma_3^2 \chi^2_{10r} \quad ,$$

(6.3.7) $$S_2 \cong (\sigma_3^2 + 2\sigma_2^2)\chi^2_{8r} \quad ,$$

$$S_3 \cong (\sigma_3^2 + 2\sigma_2^2 + 10\sigma_1^2)\chi^2_{r} \quad ,$$

where $\cong$ indicates equality in distribution and χ^2_ν denotes a random variable having a chi-squared distribution with ν degrees of freedom. Moreover, S_1 , S_2 , S_3 and $Y_{1\cdot\cdot\cdot}$, $Y_{2\cdot\cdot\cdot}$, $\ldots$, $Y_{r\cdot\cdot\cdot}$ are independent under (6.3.5) .

For the special model (6.3.4) we have $\mu_i = \log\{n(1 + X(t_i))\}$ and X depends on a vector of seven unknown parameters $\theta = (n,c,d,a_s,b_s,a_h,b_h)$. Since these parameters act in an independent and presumably locally linear manner, the model behaves asymptotically as a seven parameter linear model. Hence, under the special model (6.3.4) ,

$$(6.3.8) \qquad S_4 = 20 \inf_{\theta} \sum_{i=1}^{r} \{Y_{i\cdots} - \log\{n(1 + X_\theta(t_i))\}\}^2$$

is approximately distributed as $(\sigma_3^2 + 2\sigma_2^2 + 10\sigma_1^2)\chi^2_{r-7}$. As S_4 depends on the $Y_{ijk\ell}$ only through $Y_{1\cdots}$, ... , $Y_{r\cdots}$, S_1 , S_2 , S_3 and S_4 are independent even under the general model (6.3.5) . The random vector $\hat{\theta} = (\hat{n},\hat{c},\hat{d},\hat{a}_s,\hat{b}_s,\hat{a}_h,\hat{b}_h)$ minimizing the sum of squares in (6.3.8) is the vector of least squares estimators of θ , thus

$$(6.3.9) \qquad S_4 = 20 \sum_{i=1}^{r} \{Y_{i\cdots} - \log\{n(1 + X_{\hat{\theta}}(t_i))\}\}^2 \ .$$

Let $\chi_{\nu,\alpha}$ and $F_{\nu_1,\nu_2;\alpha}$ denote the upper α-point of the chi-squared distribution with ν degrees of freedom and the F-distribution with ν_1 and ν_2 degrees of freedom, respectively. We have the following (approximate) test and confidence intervals:

(i) Test for the model (6.3.4)

The model should be rejected if

$$(6.3.10) \qquad F = \frac{rS_4}{(r - 7)S_3} \geq F_{r-7,r;\alpha} \ .$$

Note that if the model is rejected, it is still possible that $N(t) = n(1 + X(t)) + \mathcal{O}_p(n^{1/2})$, but that the differences $N(t_i) - n(1 + X(t_i)) = n\delta_i$ are not negligible compared to the other errors. In a sense, this would also indicate a deviation from our theory. Another way of looking at this is the following. If $N(t_i) = n(1 + X(t_i)) + n\delta_i$, then

(6.3.11) $$D = \frac{(r - 7)}{20r}\left[\frac{S_4}{r - 7} - \frac{S_3}{r}\right]$$

estimates $\frac{1}{r}\sum_{i=1}^{r}\frac{E\delta_i^2}{(1 + X(t_i))^2}$, the average expected squared relative deviation of the $N(t_i)$ from $n(1 + X(t_i))$.

(ii) Confidence intervals

The following confidence statements have confidence coefficient $(1 - \alpha)$:

(6.3.12) $$\frac{S_1}{\chi_{10r,\alpha/2}} \le \sigma_3^2 \le \frac{S_1}{\chi_{10r,1-\alpha/2}} ,$$

(6.3.13) $$\frac{S_2}{\chi_{8r,\alpha/2}} \le \sigma_3^2 + 2\sigma_2^2 \le \frac{S_2}{\chi_{8r,1-\alpha/2}} ,$$

(6.3.14) $$\frac{S_3}{\chi_{r,\alpha/2}} \le \sigma_3^2 + 2\sigma_2^2 + 10\sigma_1^2 \le \frac{S_3}{\chi_{r,1-\alpha/2}} .$$

To carry out this analysis of variance on the actual experimental data is a non-trivial undertaking. The only presently available computer program for computing the vector of least squares estimates $\hat{\theta}$ minimizes

(6.3.15) $$\sum_{i=1}^{r}\frac{(\hat{N}(t_i) - n - nX_\theta(t_i))^2}{\hat{N}(t_i)^2} ,$$

where $\hat{N}(t_i) = \hat{N}_{i\cdots}$ is the average of the 20 estimates of $N(t_i)$. Since (6.3.15) is a first order approximation of

(6.3.16) $$\sum_{i=1}^{r}\{Y_{i\cdots} - \log\{n(1 + X_\theta(t_i))\}\}^2 ,$$

the minimum of (6.3.16) should be close to that of (6.3.15) , which equals 0.25 for our data (cf. the end of Section 6.2). Hence S_4

should be approximately equal to $20 \cdot 0.25 = 5.0$. By choosing values of θ close to the minimum of (6.3.15) we have indeed found $S_4 \approx 4.94$; strictly speaking, 4.94 is of course an upper bound for S_4 .

The complete results of the analysis are

$$r = 27 \quad ,$$
$$S_1 = 0.501 \quad ,$$
$$S_2 = 15.46 \quad ,$$
$$S_3 = 4.48 \quad ,$$
$$S_4 \leq 4.94 \quad .$$

The F-test (6.3.10) for the model yields

$$F \leq 1.49 \quad ,$$

which is not significant at the 0.10 level for 20 and 27 degrees of freedom. In view of this there is no reason to reject the model. By (6.3.11)

$$D = 0.0030$$

estimates $\frac{1}{27} \sum_{i=1}^{27} \frac{E\delta_i^2}{(1 + X(t_i))^2}$.

Choosing $\alpha = 0.10$, (6.3.12) - (6.3.14) produce the following (approximate) 0.90-confidence intervals:

$$0.0016 \leq \sigma_3^2 \leq 0.0021 \quad ,$$

$$0.062 \leq \sigma_3^2 + 2\sigma_2^2 \leq 0.085 \quad ,$$

$$0.11 \leq \sigma_3^2 + 2\sigma_2^2 + 10\sigma_1^2 \leq 0.28 \quad .$$

The corresponding point estimates are

$$\hat{\sigma}_3^2 = 0.0019 \quad ,$$

$$\hat{\sigma}_3^2 + 2\hat{\sigma}_2^2 = 0.0716 \quad ,$$

$$\hat{\sigma}_3^2 + 2\hat{\sigma}_2^2 + 10\hat{\sigma}_1^2 = 0.1659 \quad ,$$

so that

$$\hat{\sigma}_3^2 = 0.0019 \quad ,$$

$$\hat{\sigma}_2^2 = 0.0349 \quad ,$$

$$\hat{\sigma}_1^2 = 0.0094 \quad .$$

Since

$$\sigma^2(Y_{i\cdot\cdot\cdot}) = \frac{1}{20}\,(\sigma_3^2 + 2\sigma_2^2 + 10\sigma_1^2) \quad ,$$

we see that with the present design of the experiment, the errors made in counting the cells (variance σ_3^2) play no role in the ultimate accuracy of the experiment. The second step sampling and dilution errors e_{ijk} (variance σ_2^2) are approximately as important for the final result as the first step errors e_{ij} (variance σ_1^2).

6.4. RESULTS AND DISCUSSION

The growth of the LT 0.05 line during a fermentor run is shown in Fig. 6.2 . The curve through the data points was obtained by using numerical solutions of (3.1.13) and a fitting program as explained in Section 6.2 . The results show that the data do not violate the assumptions on cell cycle events made in the model (see Chapter 2). Assuming that the curve in Fig. 6.2 does indeed represent a realistic description of the growth of the LT cells, we may infer that a population doubling time of 30 h (sum of the parameters $c = 26$ h and $d = 4$ h) and a lag phase of 26 h (parameter $c = 26$ h) are fair estimates of the real values. These estimates are in good agreement with experimentally obtained values from apparently exponentially growing stock cultures (data not shown). In terms of the model this would imply that transfer of early stationary phase cells to a fresh medium -even if this transfer is from stock culture vessels to a fermentor- quickly reactivates all of the cells to become A-cells waiting for a division stimulus, and that most of the so-called lag phase in fact

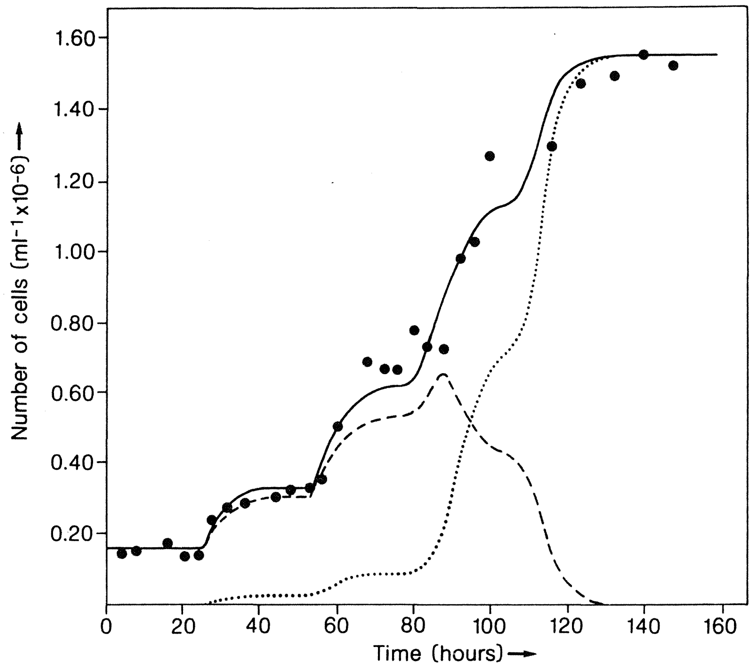


Fig.6.2. Experimental data of the growth of a batch culture of tobacco cells (• • •). The curve through the data points (———) was fitted using numerical solutions of (3.1.13). The parameter values were c=26h, d=4h, a_s=0.04ml^{-1}, b_s=29.9, a_h=0.07ml^{-1}, b_h=4.7. The two other curves represent the number of A-cells per ml (– – – – –) and the number of B-cells per ml (············), as predicted by the model with the same parameter values as used for the computation of the total cell number per ml.

represents part of the cell cycle, i.e. the fixed part c . The model also seems to give a fair description of the observed transient stepwise increase in cell number, which obviously follows after the massive reactivation of the transferred cells. This means that the often observed hump in the first data points (cf. Fig. 2.1) is not due to measurement error, but represents a real phenomenon.

According to the estimated values of the paramaters as used in the numerical solution represented in Fig. 6.2 , the hormone (2,4-D) was the growth limiting factor in the described experiment. This is not unlikely, since the concentration of the hormone at which the cells were grown ($2.26 \cdot 10^{-7}$ M), was relatively low. At present hormone metabolism by cells in the fermentor has not yet been analysed. Thus, verification of hormone consumption as predicted by the model (see Fig. 6.5) awaits further investigation.

It is unlikely, however, that the major substrate, i.e. sugar, was the growth limiting factor. An overall picture of the sugar consumption by the cells is shown in Fig. 6.3 . At about 120 h when population growth had almost stopped (see Fig. 6.2), only approximately 30 % of the initial total amount of sugar (0.09 M) had been consumed. The non-dividing cells, however, continued to consume sugar up till about 210 h. This sugar was probably mainly used for cell growth and storage in the form of starch and/or sucrose as will be discussed below.

The transport and storage sugar sucrose is commonly used as major substrate in plant cell culture practice, but as can be inferred from Fig. 6.3 , it is rapidly split by the cells into glucose and fructose, which can also serve as substrate. This has to be taken into account if one wants to analyse substrate consumption by plant cells. The determination of sucrose concentration alone would give a false picture of substrate consumption.

Sugar consumption by the cells over the first 116 h is shown in somewhat more detail in Fig. 6.4 . The lower curve in Fig. 6.4 represents the prediction of sugar consumption by the theoretical growth curve of Fig.6.2 . The curve through the data points was obtained by translation of the lower curve over plus $6.5 \cdot 10^{-3}$ M . The thus translated curve gives a reasonable description of the sugar consumption as observed between ca. 10 h and 96 h. Between 0 h and 10 h, however, the curves of the observed sugar consumption and the predicted substrate consumption roughly look like mirror images, i.e. for the first 10 h the model predicts a

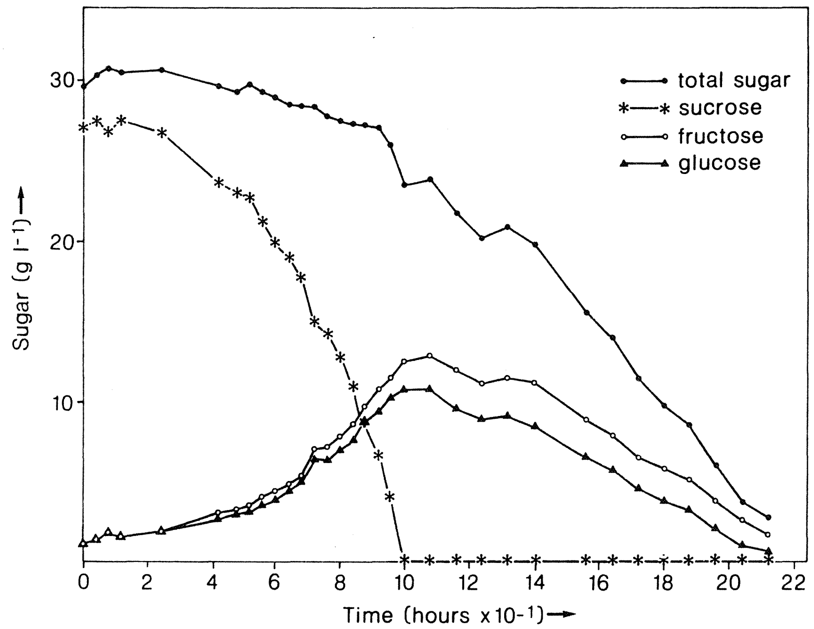


Fig.6.3. Experimental data repesenting the concentration of sugars in the medium at various time points.

progressive decrease of substrate concentration, whereas in the experiment a progressive increase of sugar concentration was found. A possible explanation for this discrepancy between model and data is that the transferred early-stationary phase cells contained stored sugars (starch and/or sucrose), part of which, after being degraded to the monomers glucose and fructose, was preferentially used by the cells as substrate upon receiving a division stimulus; at the same time another part of these stored sugars might have been excreted in the form of monomers into the medium. This is an interesting point and deserves further attention in future experiments. At present we must be content with the fact that if the sugar concentration at time $t = 10$ h is considered as a new starting value, then the theoretical substrate consumption does give a fair description of the total sugar consumption as found in the experiments up till about 96 h. Hereafter the model predicts that substrate consumption stops at ca. 96 h, while in the experiments sugar consumption continued up till approximately 210 h (Fig. 6.3). Since differentiation of cells from meristems of intact plants is usually accompanied by cell growth and sugar consumption, an obvious reason for this difference is that B-cells do consume a substantial amount of substrate, an aspect which we did not take into account in the model. Before discussing this, we shall first describe how B-cells appeared in the experiments.

The model predicts that cells are progressively leaving the cell cycle to become B-cells which start to differentiate. Theoretical curves of the progression of the densities of A- and B-cells as predicted by the model are also shown in Fig. 6.2 . At present we are not able to compare these curves with quantitative data on A- or B-cell densities. This is due to the fact that we have no early markers to unambiguously distinguish A-cells from B-cells. Nevertheless, a study of macerated cell samples by light microscopy revealed some features of the cell population which certainly do not contradict the A- versus B-cell concept.

In order to give the reader an impression of the morphology of the cells in the fermentor, we have taken a series of pictures of macerated cell samples drawn at various times after inoculation of cells to the fermentor (Fig. 6.6). Between 0 h and 26 h the cell population consisted of parenchymatous cells of varying size (Fig. 6.6.a,b,c). As time progressed between 0 h and 26 h, we observed morphological changes such as displacement of nuclei from a position near one of the cell walls to the centre of the cells, increase in volume of nuclei, a gradual increase in

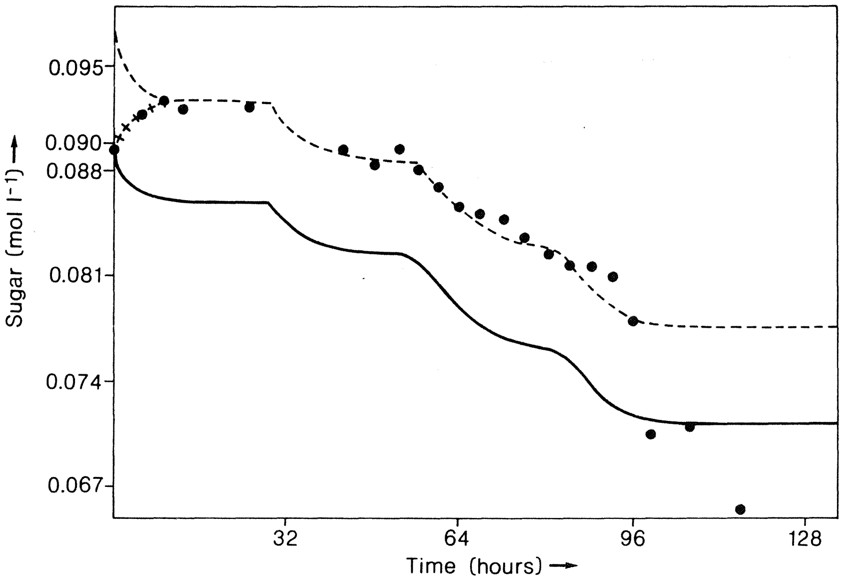


Fig.6.4. Concentration of total sugar in the medium during the first 116 h of the experiment (● ● ● ● ● ●). The solid curve represents the concentration of substrate as predicted by the model with the same parameter values as in Fig.6.2.; the dashed curve was obtained by translation of this theoretical curve over plus $6.5 \cdot 10^{-3}$M.

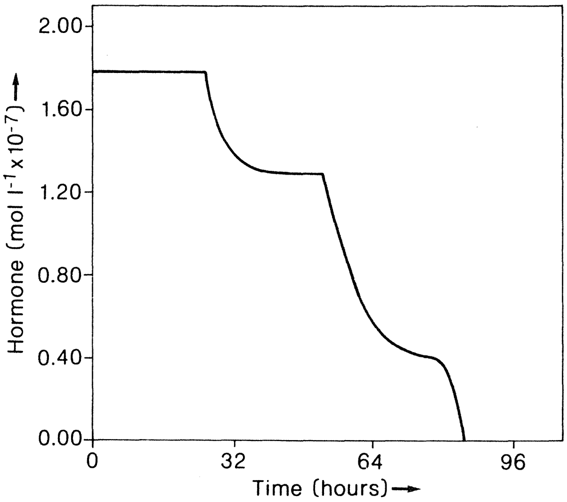


Fig.6.5. Concentration of hormone (————) as predicted by the model with the same parameter values as in Fig.6.2.

size of the nucleolus etc. (Fig. 6.6.c), which are all features of cells preparing for cell division (cf. VENVERLOO & LIBBENGA, (1987)). Within the next time interval (26 h - 52 h) a drastic increase in cell aggregates consisting of two cells was observed, i.e. the cells were split into two daughter cells held together within the confinement of the original wall of the mother cell (Fig. 6.6.d,e). At the end of this time interval many cells appeared to prepare for new cell divisions. From 52 h to 78 h we observed a progressive increase in original cells split into 3 to 4 daughter cells (Fig. 6.6.f). At this stage of the development of the culture, however, the relationship between daughter cells and the original mother cell became already less obvious due to separation of daughter cells by maceration. The interval between 78 h and 104 h was characterized by aggregates consisting of relatively many small cells (Fig. 6.6.g,h). After ca. 104 h the picture changed rapidly. By then we observed phenomena which were the reverse of those occurring during the first time interval: the average size of the cells increased, the nuclei became smaller and were displaced to positions near one of the cell walls, and the nucleoli became less conspicuous. At the end of the experiment the population was observed to consist only of parenchymatous cells of relatively large size (Fig.6.6.i,j).

The above observations are consistent with the idea that the proliferation of the cell population quite suddenly stopped because cells progressively left the cell cycle and started to differentiate into parenchymatous cells. In terms of the model: the population growth stopped because the A-cells have become extinct. Because we could recognize B-cells only after a delay when morphological changes had occurred, it is not unlikely that a rapid increase of B-cell density started at about 84 h after inoculation of cells into the fermentor, as predicted by the model (Fig. 6.2). Since the real B-cells most probably did consume sugar, we may expect a substantial and rapidly increasing contribution of B-cells to the total sugar consumption after this time. This might explain the deviation of the observed total sugar consumption from the predicted one over this time period (cf. Fig. 6.3 and Fig. 6.4).

6.5. CONCLUSIONS

The LT 0.05 cell line grown in a turbine stirred fermentor as described in this chapter provides an appropriate experimental system for further investigation of the mechanisms of the growth of plant cell cultures.

The results of the experiment we have described show that the model introduced in Chapter 2 accounts for the observed population growth reasonably well. In particular, the observations support the idea that transfer of early stationary phase cells to a fresh medium quickly reactivates all of the cells to become A-cells waiting for a stimulus to divide.

We have also seen that, as predicted by the model, the lag phase seems to constitute a fixed part of the cell cycle. Sugar consumption during the first step of population growth in the experiment was very low. In terms of the model this means that the value Q, which is the parameter of the exponential distribution of the waiting time for a stimulus of an individual cell, was nearly constant during this period. Note that this implies that the course of the population growth in the first step was approximately similar to the one predicted by the transition probability model (see Chapter 1), the mean cycle time being $c + Q^{-1}$ ($\approx$ 30 h in the present experiment). In order to verify this more rigorously, more precise data should be obtained by increasing the sampling frequency and the number of samples per observation epoch over this initial period.

Cytological observations of macerated cell samples provide good evidence that in the experiment population growth quite suddenly stopped because cells progressively left the cell cycle and started to differentiate into parenchymatous cells. For a better understanding of the way in which the hormone controls this phenomenon, experiments should be performed with different starting concentrations of the hormone in which uptake and conversion of the hormone by the cells is analysed at the same time. The results should be compared with the predictions made by the model.

Finally, although the model provides a fair description of the observed sugar consumption between about 10 h and 96 h, it has to be slightly adjusted in order to cover the entire range of sugar consumption. It should take into account the observed increase in sugar concentration over the first 10 h and the observed sugar consumption by non-dividing cells. At present we see no need, however, to take account of substantial maintenance energy requirement. For further investigation of the relationship between the sugar consumption and the division rate, made explicit in the model by means of the dependence of Q on the substrate concentration, continuous cell culture systems as for instance described in VAN GULIK et al. (submitted) may be more suitable than batch cultures. A first attempt to adjust the model to continuous cultures looks promising and will be published elsewhere.

Fig.6.6. The next five pages show a series of photographs of cell samples drawn at various times after inoculation of cells to the fermentor. The pictures were made from macerated cells using a Leitz Laborlux D light microscope equipped with a Wild photo automat MPS45. Magnification scales are placed in the pictures.

a,b. Cells at 4 h and 24 h after inoculation, respectively. The pictures give an impression of the shape of cells in the inoculum.

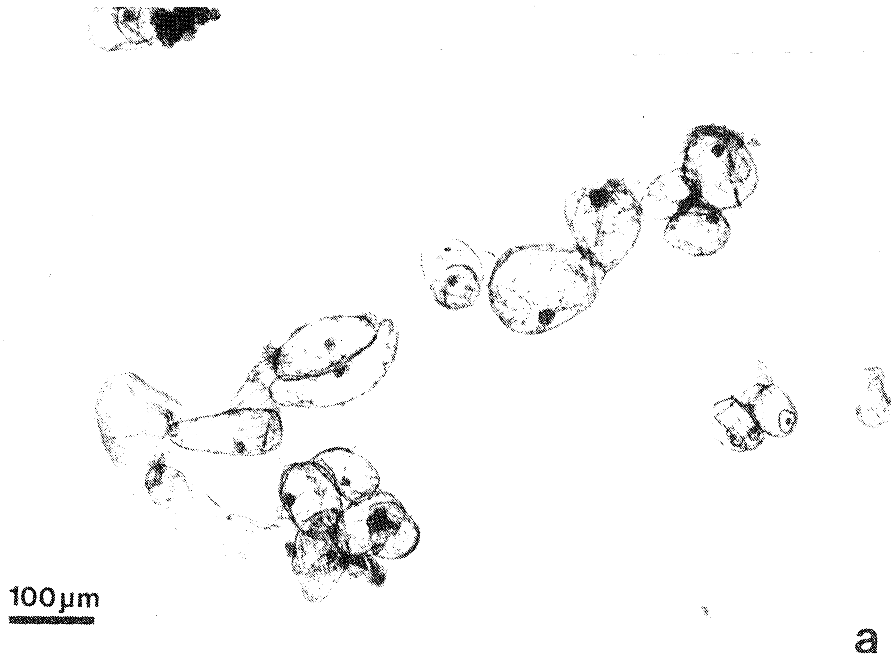
c. Detail of picture b, showing cells preparing for cell division. Note the enlarged nuclei with clearly visible nucleoli (arrow).

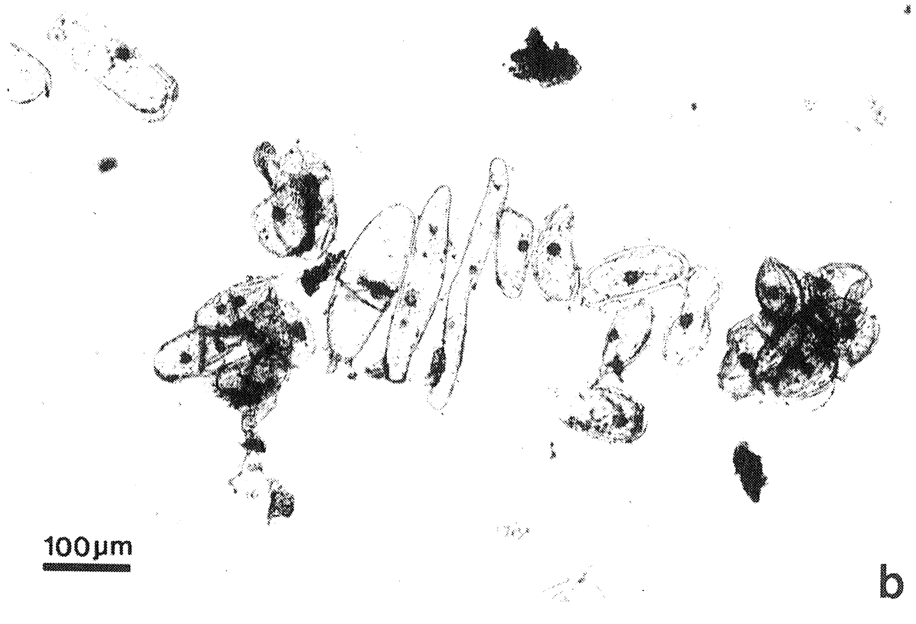
d. Recently divided cells. The picture was taken only 4 h after the pictures b and c were taken.

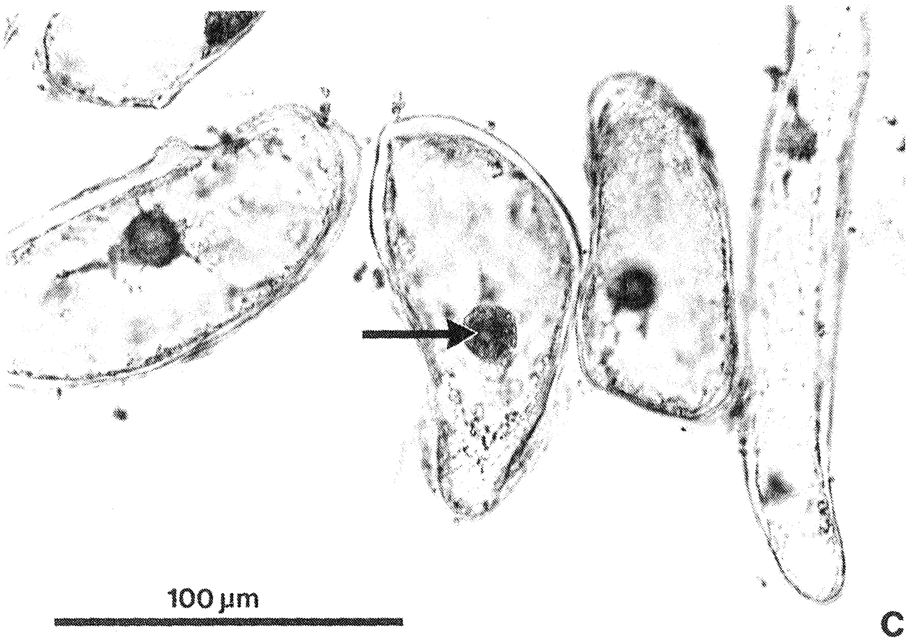
e,f. Arrows indicate predominant clusters of two cells at 36 h (e), and first clusters of 3 - 4 cells at 56 h (f).

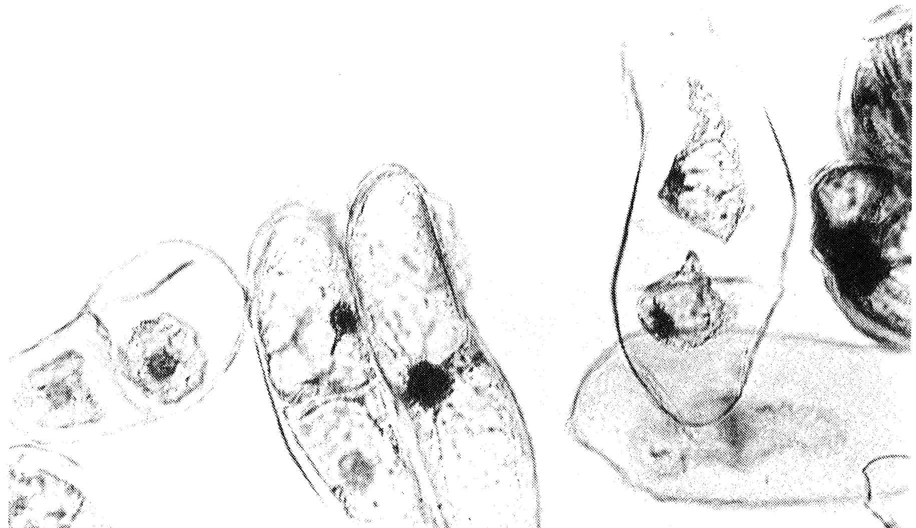
g,h. Cells in clusters of varying size at 92 h. At about this time B-cells could already be observed (arrows). The cluster of cells in picture h still shows the confinement of the original mother cell. This picture clearly demonstrates that during active population growth the cells progressively become smaller.

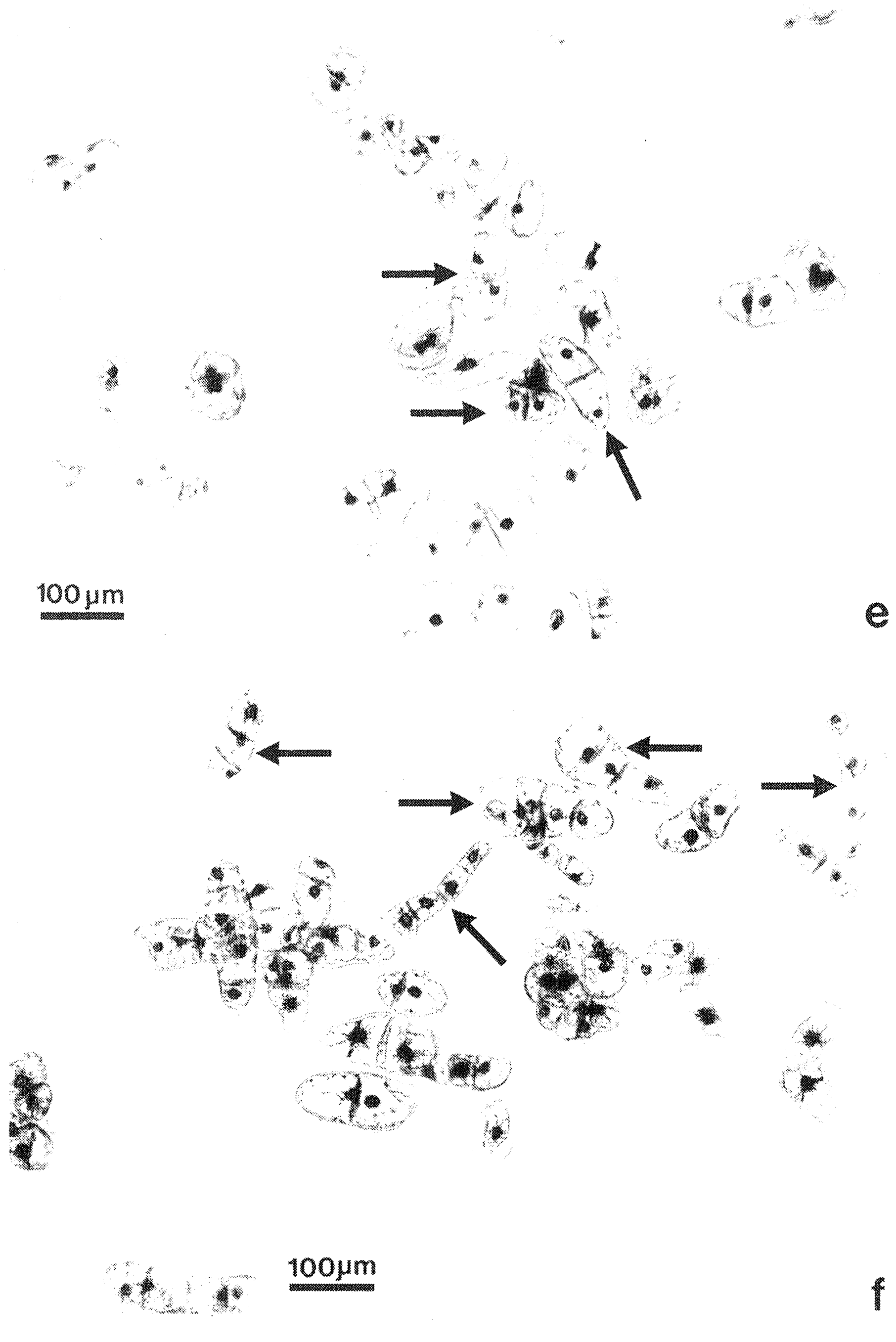
i,j. Cells at 124 h and 188 h, respectively. These photographs show growing B-cells. The cells in picture i have the same outward appearance as the cells used for inoculation (cf. a).

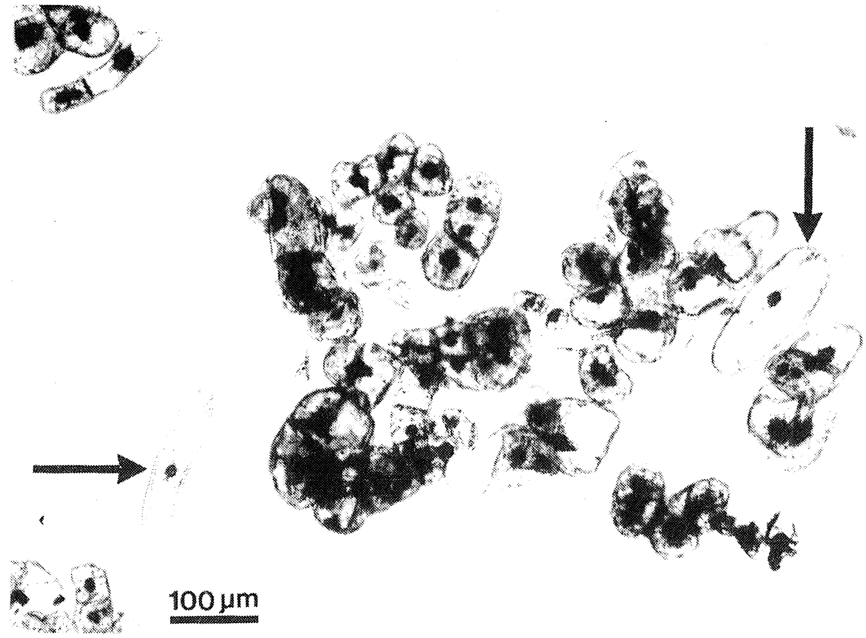

100µm
a


100µm
b

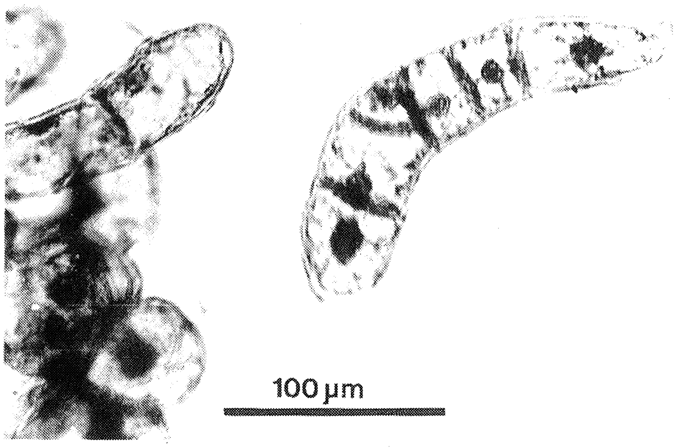

100 µm
c


100 µm
d

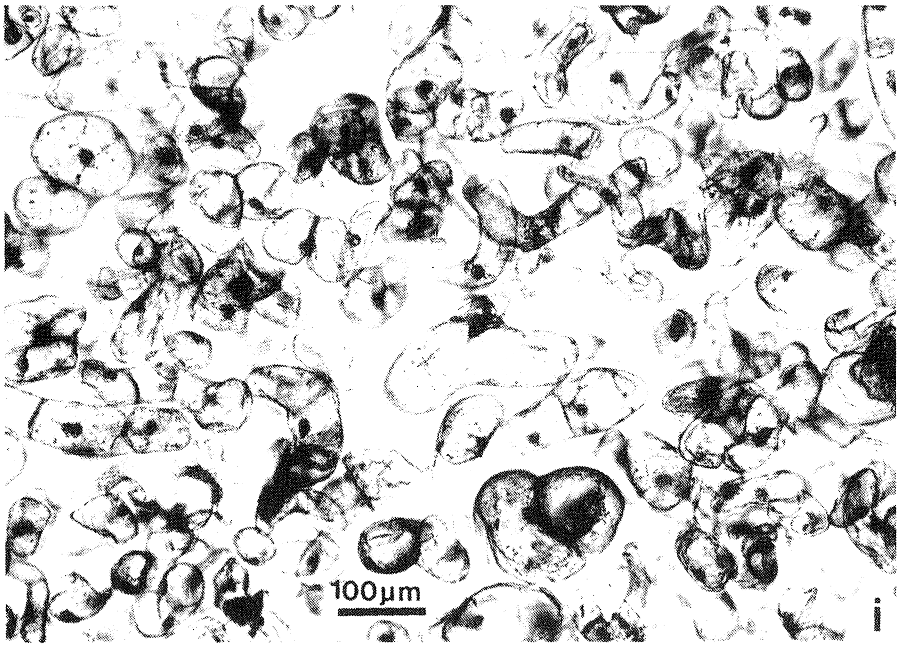

100 µm
e
100µm
f



g



h


100µm
i


100µm
j

CHAPTER 7

FINAL REMARKS

Although a large amount of work has been carried out within frameworks of plant cell biotechnology programs on biosynthesis of secondary metabolites in plant cell cultures, little information is available on growth, differentiation and product accumulation kinetics of different culture systems. However, in order to be able to manipulate the productivity of cell cultures in multi-litre vessels, detailed knowledge of the regulation of division, differentiation, growth and product formation of cells grown under different conditions is required (cf. for instance MORRIS, 1986). Our approach of this problem has been to start with a mathematical model based upon ideas about growth and cell differentiation kinetics in plant cell cultures as they have emerged in our laboratory over the past years, to consider the implications of the model, and to perform some specific experiments aimed at its testing. Although at present testing of the model is in its initial stage, the first results which were described in the previous chapter, show that this approach is very promising.

An interesting feature of our model is that it links population growth to cell differentiation, with as major controlling factor the - in batch cultures declining - hormone concentration in the medium. Since it is generally believed that secondary product formation in batch culture systems is related to cell differentiation, the model might in principle be able to link population growth also to product accumulation. It has occasionally been observed that the biosynthesis of certain secondary products seemed to be confined to dividing cells, whereas in the same system other, related products seemed to be synthesized only by differentiating cells. In this respect it is interesting to note that production curves of ajmalicine and serpentine in batch-cultured *Catharanthus roseus* cells look like the theoretical growth curves of A- and B-cells as shown in Fig. 6.2 , respectively, (MORRIS, 1986; BLOM, personal communication). This makes the *Catharanthus roseus* system, which is also studied in our laboratory, an interesting one to use for extending our model to other systems.

Thus far, we have ascribed to B-cells in our model no other features than that they do not divide. The above discussion and the experimental results described in Chapter 6 make clear, however, that the model will gain in value if we would be able to account for growth, sugar consumption and secondary product formation of these cells as well.

From a mathematical point of view this study is devoted to the analysis of a point process of a somewhat more complicated structure than is usually considered. In part the complications arise from the fact that the point process is non-Markovian. Another, perhaps more important difficulty is that one doesn't have an expression for the intensity process available. As a result, we can only appeal to standard theory for part of our investigation.

Moreover, our results go somewhat further than the usual ones. In particular the study of the process on the entire real line as opposed to bounded intervals, introduces additional problems. The study of the duration of the process and the asymptotic behaviour of the intensity process are both very special for the particular model considered. It would be of interest to know whether results of this type can be obtained more generally for bounded point processes.

APPENDIX A1

THREE MAXIMAL INEQUALITIES

We prove probability inequalities for the maximum of two stochastic processes, which we use in Chapter 3 ; their applicability is not restricted to the subject of this study though.

LEMMA A1.1 If $X(t)$, $t \geq 0$ is a unit Poisson process, then

$$(A1.1) \qquad P\,(\sup_{0\leq t\leq a} |X(t) - t| \geq x) \leq \tfrac{8}{3}\exp\{-x^2/(16a)\} \;, \qquad 0 \leq x \leq 4a \;,$$

$$(A1.2) \qquad P\,(\sup_{\{t:X(t)\leq a\}} |X(t) - t| \geq x) \leq \tfrac{11}{3}\exp\{-x^2/(32a)\} \;, \qquad 0 \leq x \leq 2a \;.$$

PROOF. If X has a Poisson distribution with mean μ , then for $|t| \leq \mu^{1/2}$,

$$\log(E\exp\{t\mu^{-1/2}(X - \mu)\}) = -t\mu^{1/2} + \mu(\exp\{t\mu^{-1/2}\} - 1)$$

$$= \tfrac{1}{2}t^2 + t^2 \sum_{k=3}^{\infty} \frac{(t\mu^{-1/2})^{k-2}}{k!} \leq t^2 \;.$$

For $0 \leq x \leq 2\mu^{1/2}$ with $t = \frac{x}{2}$ and $\tilde{t} = \frac{-x}{2}$ respectively, Chebyshev's inequality gives

$$(A1.3) \qquad \begin{aligned} &P\,(\mu^{-1/2}(X - \mu) \geq x) \leq \exp\{t^2 - tx\} = \exp\{-x^2/4\} \;, \\ &P\,(\mu^{-1/2}(X - \mu) \leq -x) \leq \exp\{\tilde{t}^2 + \tilde{t}x\} = \exp\{-x^2/4\} \;. \end{aligned}$$

Hence

$$(A1.4) \qquad P\,(\mu^{-1/2}|X - \mu| \geq x) \leq 2\exp\{-x^2/4\} \;, \qquad 0 \leq x \leq 2\mu^{1/2} \;.$$

Combining this with Inequality 5 in SHORACK & WELLNER (1986), p.844, we find that for $x \in [2(2a)^{1/2}, 4a]$,

$$P\left(\sup_{0\le t\le a}|X(t)-t|\ge x\right)\le\frac{4}{3}P\left(|X(a)-a|\ge x-(2a)^{1/2}\right)$$

$$\le\frac{4}{3}P\left(|X(a)-a|\ge\frac{x}{2}\right)\le\frac{8}{3}\exp\{-x^2/(16a)\}\quad .$$

Since (A1.1) is trivially true for $0\le x\le 2(2a)^{1/2}$, this proves (A1.1) .

Next we note that for $x\in[0,2a]$, (A1.3) and (A1.1) imply

$$P\left(\sup_{\{t:X(t)\le a\}}|X(t)-t|\ge x\right)$$

$$\le P\left(X(2a)\le a\right)+P\left(\sup_{0\le t\le 2a}|X(t)-t|\ge x\right)$$

$$\le P\left(X(2a)-2a\le -a\right)+\frac{8}{3}\exp\{-x^2/(32a)\}$$

$$\le\exp\{-a/8\}+\frac{8}{3}\exp\{-x^2/(32a)\}$$

$$\le\frac{11}{3}\exp\{-x^2/(32a)\}$$

and the proof is complete.

□

<u>LEMMA</u> <u>A1.2</u> If X_1 , X_2 , ... are independent, bounded random variables, $0\le X_j\le a$, $(j=1,2,\ldots)$, then for all $M\in\mathbb{N}$ and every $x\ge 0$,

$$\text{(A1.5)}\qquad P\left(\max_{1\le m\le M}\left|\sum_{j=1}^{m}(X_j-EX_j)\right|\ge x\right)\le 4\exp\{-2x^2/(9a^2M)\}\quad .$$

<u>PROOF.</u> Clearly it suffices to prove the lemma for $x\ge\frac{3}{2}a(2M)^{1/2}$. From Theorem 2 in HOEFFDING (1963) it follows that for all $m\in\mathbb{N}$ and every $x\ge 0$,

$$\text{(A1.6)}\qquad P\left(\sum_{j=1}^{m}(X_j-EX_j)\ge x\right)\le\exp\{-2x^2/(a^2m)\}\quad .$$

Combining this with Lévy's inequality (SHORACK & WELLNER (1986), p.844), we get

$$P\left(\max_{1\le m\le M}\sum_{j=1}^{m}(X_j - EX_j) \ge x\right)$$

$$\le 2P\left(\sum_{j=1}^{M}(X_j - EX_j) \ge x - \left(2\sum_{j=1}^{M}\sigma^2(X_j)\right)^{1/2}\right)$$

$$\le 2P\left(\sum_{j=1}^{M}(X_j - EX_j) \ge x - a(2M)^{1/2}\right)$$

$$\le 2P\left(\sum_{j=1}^{M}(X_j - EX_j) \ge \frac{x}{3}\right) \le 2\exp\{-2x^2/(9a^2M)\} \quad ,$$

provided $x \ge \frac{3}{2}a(2M)^{1/2}$. Replacing X_j by $a - X_j$ we see that the same bound holds for $P\left(\min_{1\le m\le M}\sum_{j=1}^{m}(X_j - EX_j) \le -x\right)$, and the lemma is proved.

□

APPENDIX A2

SOME RESULTS CONCERNING INTEGRAL EQUATIONS IN $D[0,\infty)$

LEMMA A2.1 (a version of Gronwall's lemma)

Let $u : [0,\alpha] \to \mathbb{R}$ be non-negative and right-continuous with left-hand limits. Suppose $A \geq 0$, $K \geq 0$ exist such that for all $t \in [0,\alpha]$,

$$u(t) \leq A + K \int_0^t \sup_{r\leq s} u(r)ds \quad ,$$

then for all $t \in [0,\alpha]$,

$$u(t) \leq Ae^{Kt} \quad .$$

PROOF. Define

$$v(t) = \int_0^t \sup_{r\leq s} u(r)ds$$

Then for $t \in [0,\alpha]$,

$$v'(t) \leq \sup_{s\leq t} u(s) \leq A + K \sup_{s\leq t} \int_0^s \sup_{\sigma\leq\tau} u(\sigma)d\tau$$

$$= A + K \int_0^t \sup_{\sigma\leq\tau} u(\sigma)d\tau = A + Kv(t) \quad .$$

From this we get

$$v'(t)e^{-Kt} - v(t)Ke^{-Kt} \leq Ae^{-Kt} \quad ,$$

and integration yields

$$v(t)e^{-Kt} \le \int_0^t Ae^{-Ks} ds \quad ,$$

or

$$v(t) \le \frac{A}{K} (e^{Kt} - 1) \quad .$$

Thus we have for $t \in [0,\alpha]$,

$$u(t) \le A + Kv(t) \le Ae^{Kt} \quad .$$

□

Let $D[0,\infty)$ denote the space of right-continuous functions with left-hand limits on $[0,\infty)$; $C[0,\infty)$ denotes the space of continuous functions on $[0,\infty)$. Consider the integral equation

$$\text{(A2.1)} \qquad x(t) = f(t) + \int_0^t F(s,x) ds \quad ,$$

where $f : [0,\infty) \to \mathbb{R}$ and $F : [0,\infty) \times D[0,\infty) \to \mathbb{R}$.

<u>LEMMA</u> <u>A2.2</u> Assume that f is continuous and bounded on $[0,\infty)$, and that there exist constants C_1 , C_2 and C_3 such that for all $t \in [0,\infty)$ and for every x , $y \in C[0,\infty)$ the following two conditions hold.

$$\text{(A2.2)} \qquad \sup_{s\le t} |F(s,x)| \le C_1 + C_2 \sup_{s\le t} |x(s)| \quad ,$$

$$\text{(A2.3)} \qquad |F(t,x) - F(t,y)| \le C_3 \sup_{s\le t} |x(s) - y(s)| \quad .$$

Then there exists a unique solution X of (A2.1) in $D[0,\infty)$ and this solution is continuous.

<u>PROOF.</u> (i) Existence. Fix $t > 0$ and define for $s \in [0,t]$ the sequence $\{x_n\}_{n\in\mathbb{N}_0}$ by

$$x_0(s) = f(s) \quad , \quad x_{n+1}(s) = f(s) + \int_0^s F(r,x_n) dr \quad , \quad n \ge 0 \quad .$$

Then x_n is continuous and for $n \geq 1$,

$$|x_{n+1}(s) - x_n(s)| = |\int_0^s (F(r,x_n) - F(r,x_{n-1}))dr|$$

$$\leq \int_0^s |F(r,x_n) - F(r,x_{n-1})|dr$$

$$\leq C_3 \int_0^s \sup_{u\leq r} |x_n(u) - x_{n-1}(u)|dr \quad .$$

Since f is bounded, say $|f| \leq C_4$ on $[0,\infty)$,

$$\sup_{s\leq t} |x_1(s) - x_0(s)| \leq (C_1 + C_2C_4)t \equiv C_5t \quad ,$$

and by induction we get for $n = 0,1,2,\ldots$,

$$\sup_{s\leq t} |x_{n+1}(s) - x_n(s)| \leq \frac{C_5C_3{}^n\, t^{n+1}}{(n + 1)!} \quad .$$

From this we conclude that for all $s \leq t$,

$$X(s) = \lim_{n\to\infty} x_n(s) = \lim_{n\to\infty} (f(s) + \sum_{k=1}^{n} (x_k(s) - x_{k-1}(s)))$$

$$= f(s) + \sum_{k=1}^{\infty} (x_k(s) - x_{k-1}(s))$$

exists and is continuous on $[0,\infty)$. Furthermore, from (A2.3) we obtain

$$\sup_{s\leq t} |F(s,x_n) - F(s,X)| \leq C_3 \sup_{r\leq t} |x_n(r) - X(r)| \to 0 \ , \quad (n \to \infty) \quad .$$

Hence

$$X(s) = \lim_{n\to\infty} x_{n+1}(s) = \lim_{n\to\infty} (f(s) + \int_0^s F(r,x_n)dr)$$

$$= f(s) + \int_0^s F(r,X)dr \quad .$$

Since the foregoing holds for all $t \in [0,\infty)$, X is a solution of (A2.1) on $[0,\infty)$.

(ii) Uniqueness. Let X_1 and X_2 be two solutions of (A2.1) on $[0,\infty)$ and define for $0 \le s \le t < \infty$,

$$u(s) = |X_1(s) - X_2(s)| \quad .$$

Then using (A2.3) we have

$$u(s) = | \int_0^s F(r,X_1) - F(r,X_2)dr|$$

$$\le C_3 \int_0^s \sup_{\sigma\le r} |X_1(\sigma) - X_2(\sigma)|dr$$

$$= C_3 \int_0^s \sup_{\sigma\le r} u(\sigma)dr \quad ,$$

and invoking LEMMA A2.1 we get $u(s) = 0$ for all $s \in [0,t]$. Since t was arbitrary, we conclude that for all $t \ge 0$,

$$X_1(t) = X_2(t) \quad .$$

□

REFERENCES

AALEN, O.O. & J.M. HOEM (1978), *Random time changes for multivariate counting processes*, Scand. Act. J. 5, p.81-101.

ALBERTS, B., BRAY, D., LEWIS, J., RAFF, M., ROBERTS, K. & J.D. WATSON (1983), *Molecular biology of the cell*, Garland Publishing, Inc., New York.

BARLOW, P.W. & P.D.M. MACDONALD (1973), *An analysis of the mitotic cycle in the root meristem of* Zea mays, Proc. R. Soc. Lond. B 183, p.385-398.

BAYLISS, M.W. (1975), *The duration of the cell cycle of* Daucus carota L. *in vivo and in vitro*, Exptl. Cell Res. 92, p.31-38.

BAYLISS, M.W. (1976), *Variation of cell cycle duration within suspension cultures of* Daucus carota *and its consequences for the induction of ploidy changes with colchicine*, Protoplasma 88, p.279-285.

BAYLISS, M.W. (1985), *The regulation of the cell division cycle in cultured plant cells*, in: BRYANT, J.A. & D. FRANCIS (eds.), *The cell division cycle in plants*, Cambridge University Press, Cambridge.

BERTUZZI, A., GANDOLFI, A. & M.A. GIOVENCO (1981), *Mathematical models of the cell cycle with a view to tumor studies*, Math. Biosc. 53, p.159 -188.

BERTUZZI, A. & A. GANDOLFI (1983), *Recent views on the cell cycle structure*, Bull. Math. Biol. 45, p.605-616.

BILLINGSLEY, P. (1968), *Convergence of probability measures*, Wiley, New York.

BROOKS, R.F. (1976), *Regulation of the cell cycle by serum*, Nature 260, p.248-250.

BROOKS, R.F. (1981), *Variability in the cell cycle and the control of proliferation*, in: JOHN, P.C.L. (ed.), *The cell cycle*, Cambridge University Press.

BROOKS, R.F., BENNETT, D.C. & J.A. SMITH (1980), *Mammalian cells need two random transitions*, Cell 19, p.493-504.

CASTOR, L.N. (1980), *A G_1 rate model accounts for cell-cycle kinetics attributed to 'transition probability'*, Nature 287, p.857-859.

CLOWES, F.A.L. (1961), *Duration of the mitotic cycle in a meristem*, J.Eptl. Bot. 12, p.283-293.

COOPER, S. (1982), *The continuum model: statistical implications*, J. Theor. Biol. 94, p.783-800.

EISEN, M. (1979), *Mathematical models in cell biology and cancer chemother-*

apy, Lecture Notes in Biomathematics 30, Springer-Verlag, Berlin.

ERIKSSON, T. (1967), *Duration of the mitotic cycle in cultures of* Haplopappus gracilis, Physiol. Plant. 20, p.348-354.

ETHIER, S.N. & T.G. KURTZ (1986), *Markov processes; characterization and convergence*, Wiley, New York.

EVANS, G.H. & H. REES (1971), *Mitotic cycles in dicotyledons and monocotyledons*, Nature 233, p.350-351.

EVANS, L.S. & J. VAN'T HOF (1975), *The age-distribution of cell cycle populations in plant root meristems*, Exptl. Cell Res. 90, p.401-410.

FOSKET, D.E. (1977), *The regulation of the plant cell cycle by cytokinin*, in: ROST, T.L. & E.M. GIFFORD (eds.), *Mechanisms and control of cell division*, Dowden, Hutchison & Ross, Inc. Stroudsburg, Pennsylvania.

FRIEDBERG, S.H. & D. DAVIDSON (1970), *Duration of S phase and cell cycles in diploid and tetraploid cells of mixoploid meristems*, Exptl. Cell Res. 61, p.216-218.

GOODWIN, B.C. (1976), *Analytical physiology of cells and developing organisms*, Academic Press, London.

GOULD, A.R. (1977), *Temperature response of the cell cycle of* Haplopappus gracilis *in suspension culture and its significance to the* G_1 *transition probability model*, Planta 137, p.29-36.

GOULD, A.R., BAYLISS, M.W. & H.E. STREET (1974), *Studies on the growth in culture of plant cells XVII. Analysis of the cell cycle of asynchronously dividing* Acer pseudoplatanus. L. *cells in suspension cultures*, J. Exptl. Bot. 25, p.468-478.

HARRIS, T.E. (1959), *A mathematical model for multiplication by binary fission*, in: STOHLMAN, F. (ed.), *The kinetics of cellular proliferation*, Grune & Stratton, Orlando, Florida.

HOEFFDING, W. (1963), *Probability inequalities for sums of bounded random variables*, JASA 53, p.13-30.

HOWARD, A. & S.R. PELC (1953), *Synthesis of desoxyribonucleic acid in normal and irradiated cells and its relation to chromosome breakage*, Heredity (Suppl.) 6, p.261-273.

IN 'T HOUT, K. (1988), *Een eenstapsmethode voor differentiaalvergelijkingen met achterlopend argument*, Masters Thesis, Department of Mathematics, University of Leiden.

JACOBSEN, M. (1982), *Statistical analysis of counting processes*, Lecture Notes in Statistics, 12, Springer-Verlag, New York.

JAGERS, P. (1983), *Stochastic models for cell kinetics*, Bull. Math. Biol.

45, p.507-519.

JIMENEZ DE ASUA, L., O'FARRELL, M., BENNETT, D., CLINGAN, D. & Ph. RUDLAND (1977), *Interaction of two hormones and their effect on observed rate of mitiation of DNA synthesis in 3T3 cells*, Nature 256, p.151-153.

KELLY, C.D. & O. RAHN (1932), *The growth rate of individual bacterial cells*, J. Bacteriol. 23, p.147-153.

KENDALL, D. (1948), *On the role of variable generation time in the development of a stochastic birth process*, Biometrika 35, p.316-329.

KENDALL, D. (1952), *On the choice of a mathematical model to represent normal bacterial growth*, J. R. Statist. Soc. B 14, p.41-44.

KILLANDER, D. & A. ZETTERBERG (1965), *Quantitative cytochemical studies on interphase growth, I. determination of DNA, RNA and mass content of age determined mouse fibroblasts in vitro and of intercellular variation in generation time*, Exptl. Cell Res. 38, p.272-284.

KING, P.J. (1980), *Plant tissue culture and the cell cycle*, in: FIECHTER, A. (ed.), *Plant cell cultures II* , Advances in biochemical engineering 18, Springer-Verlag, Berlin.

KING, P.J. & H.E. STREET (1973), *Growth patterns in cell cultures*, in: STREET, H.E. (ed.), *Plant tissue and cell culture*, Botanical Monography 11, Blackwell Scientific Publications, Oxford.

KUBITSCHEK, H.E. (1971), *The distribution of cell generation times*, Cell Tissue Kinet. 4, p.113-122.

KURTZ, T.G. (1981), *Approximation of population processes*, SIAM, Philadelphia.

KURTZ, T.G. (1983), *Gaussian approximations for Markov chains and counting processes*, Proceedings of the 44-th ISI Session, p.361-376.

LEGAY, J-J. & J. GUERN (1975), *Quantitative effects of 2,4-Dichlorophenoxy-acetic Acid on growth of suspension-cultured* Acer pseudoplatanus *cells*, Plant Physiol. 56, p.356-359.

LINSMAIER, E.M. & F. SKOOG (1965), *Organic growth factor requirements of tobacco tissue cultures*, Physiol. Plant. 18, p.100-127.

MACDONALD, P.D.M. (1970), *Statistical inference from the fraction labelled mitoses curves*, Biometrika 57, p.489-503.

MORRIS, P. (1986), *Kinetics of growth and alkaloid accumulation in* Catharanthus roseus *cell suspension cultures*, in: MORRIS, P., SCRAGG, A.H., STAFFORD, A. & M.W. FOWLER (eds.), *Secondary metabolism in plant cell cultures*, Cambridge University Press, Cambridge.

NELSON, S. & P.J. GREEN (1981), *The random transiton model of the cell* *cycle. A critical review*, Cancer Chemother. Pharmacol. 6, p.11-18.

PARDEE, A.B. (1974), *A restriction point for control of normal animal cell proliferation*, Proc. Nat. Acad. Sci. USA 71, p.1286-1290.

PIETERS, E. (1988), *Parameter identificatie toegepast op een groeimodel*, Masters Thesis, Department of Mathematics, University of Leiden.

POLLARD, D. (1984), *Convergence of stochastic processes*, Springer Series in Statistics, Springer-Verlag, New York.

POWELL, E.O. (1955), *Some features of the generation times of individual bacteria*, Biometrika 42, p.16-44.

PROHOROV, YU.V. (1956), *Convergence of random processes and limit theorems in probability theory*, Theor. Probability Appl. 1, p.157-214.

ROBINSON, J.H., SMITH, J.A., TOTTY, N.F. & P.N. RIDDLE (1976), *Transition probability and the hormonal and density-dependent regulation of cell proliferation*, Nature 262, p.298-300.

ROELS, J.A. (1983), *Energetics and kinetics in biotechnology*, Elsevier Biomedical Press, Amsterdam.

SHIELDS, R., BROOKS, R.F., RIDDLE, P.N., CAPELLARO, D.E. & D. DELIA (1978), *Cell size, cell cycle and transition probability in mouse fibroblasts*, Cell 15, p.469-474.

SHIELDS, R. & J.A.SMITH (1977), *Cells regulate their proliferation through alterations in transition probability*, J. Cell. Physiol. 91, p.345 -356.

SHORACK, G.R. & J.A. WELLNER (1986), *Empirical processes with applications to statistics*, Wiley, New York.

SISKEN, J.E. & L.M. MORASCA (1965), *Intrapopulation kinetics of the mitotic cycle*, J. Cell Biol. 25, p.179-189.

SMITH, J.A. & L. MARTIN (1973), *Do cells cycle?*, Proc. Nat. Acad. Sci. USA 70, p.1263-1267.

STOER J. & R. BULIRSCH (1980), *Introduction to numerical analysis*, Springer-Verlag, New York.

STREET, H.E. (1973), *Cell (suspension) cultures - techniques*, in: STREET, H.E. (ed.), *Plant tissue and cell culture*, Botanical Monography 11, Blackwell Scientific Publications, Oxford.

THOMPSON, J. & F.A.L. CLOWES (1968), *The quiescent centre and rates of mitosis in the root of* Allium sativum, Ann. Bot. 32, p.1-13.

TREWAVAS, A.J. (1985), *Growth substances, calcium and the regulation of cell division*, in: BRYANT, J.A. & D. FRANCIS (eds.), *The cell division cycle in plants*, Cambridge University Press, Cambridge.

VAL, J. (1987), *Population growth in cell suspensions, a theoretical approach*, Masters Thesis, Department of Plant Molecular Biology, University of Leiden.

VAN DER ZAAL, E.J., MENNES, A.M. & K.R. LIBBENGA (1987), *Auxin-induced rapid changes in translatable mRNAs in tobacco cell suspension*, Planta 172, p.514-519.

VAN DER ZAAL, E.J., MEMELINK, J., MENNES, A.M., QUINT A. & K.R. LIBBENGA (1987), *Auxin-induced mRNA species in tobacco cell cultures,* Plant Molec. Biol. 10, p.145-157.

VAN GULIK, W.M., MEIJER, J.J., TEN HOOPEN, H.J.G. & K.C.A.M. LUYBEN, *Growth of a* Catharanthus roseus *cell suspension culture in a chemostat under glucose limiting conditions*, submitted to Appl. Microbiol. and Biotechnol.

VAN'T HOF, J. (1965), *Cell population kinetics of excised roots of* Pisum sativum, J. Cell Biol. 27, p.179-189.

VAN'T HOF, J. (1985), *Control points within the cell cycle*, in: BRYANT, J.A. & D. FRANCIS (eds.), *The cell division cycle in plants*, Cambridge University Press, Cambridge.

VENVERLOO, C.J. & K.R. LIBBENGA (1987), *Regulation of the plane of cell division in vacuolated cells, I. the function of nuclear positioning and phragmosome formation*, J. Plant Physiol. 131, p.267-284.

WADA, M., HAYAMI, J. & A. KADOTA (1984), *Returning dark induced cell cycle to the beginning of G_1 phase by red light irradiation in* Fern adianthum protonemata, Plant & Cell Physiol. 25, p.1053-1058.

WEBSTER, P.L. (1979), *Variation in sister-cell cycle duration and loss of synchrony in cell lineages in root apical meristems*, Plant Sci. Letters 14, p.13-22.

WHITT, W. (1970), *Weak convergence of probability measures on the function space* $C[0,\infty)$, Ann. Math. Stat. 41, p.939-944.

WIMBER, D.E. (1960), *Duration of the nuclear cycle in* Tradescantia paludosa *root tips as measured with H^3-Thymidine*, Amer. J. Bot. 47, p.828-834.

SUBJECT INDEX

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.

36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.

37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.

38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.

39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.

40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.

41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.

42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.

43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.

44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.

45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.

46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.

47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.

48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.

49 F.W. Wubs. *Numerical solution of the shallow-water equations.* 1988.

50 F. de Kerf. *Asymptotic analysis of a class of perturbed Korteweg-de Vries initial value problems.* 1988.

51 P.J.M. van Laarhoven. *Theoretical and computational aspects of simulated annealing.* 1988.

52 P.M. van Loon. *Continuous decoupling transformations for linear boundary value problems.* 1988.

53 K.C.P. Machielsen. *Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space.* 1988.

54 L.C.R.J. Willenborg. *Computational aspects of survey data processing.* 1988.

55 G.J. van der Steen. *A program generator for recognition, parsing and transduction with syntactic patterns.* 1988.

56 J.C. Ebergen. *Translating programs into delay-insensitive circuits.* 1989.

57 S.M. Verduyn Lunel. *Exponential type calculus for linear delay equations.* 1989.

58 M.C.M. de Gunst. *A random model for plant cell population growth.* 1989.