APPROACHES TO MODELING KINETICS OF COLONY FORMING EXPERIMENTS

by

Thomas Kuczek
Purdue University

Technical Report #89-35

Department of Statistics
Purdue University

October 1989

◇

1. Introduction

The purpose of this communication is to outline the heuristics of a Bayesian approach to building statistical models for the analysis of the kinetics of cloning and subcloning experiments in microbiology. As a preface to this, in section 2, we will discuss the microstructure, i.e. detailed structure such as IDT and family tree, of cloning experiments, briefly discuss the models they have generated and point out some limitations of such studies. In section 3 we discuss the macrostructure of cloning experiments, i.e. ignoring the fine structure of colony formation and instead looking at the large structure of cloning experiments such as colony size distribution. We then discuss subcloning experiments and their uses and point out the need for statistical models to analyze the kinetics of cloning and subcloning experiments. In section 4, we outline the heuristics of a Bayesian approach to modeling the kinetics of cloning and subcloning experiments.

2. Microstructure of Cloning Experiments

In many experiments whose point is to study cell kinetics of colony formation, heretofore referred to as cloning experiments, a sample from a population of cells is plated out in media and the formation of colonies is observed through a fixed period of time. In one class of experiments of this sort, the objective is to study the *microstructure* of colony formation, i.e. to record complete information on interdivision times and pedigrees. These provide data on the kinetics of individual cells in pedigree. Experiments of this type include those of Kelly and Rahn (1932), Powell (1955, 1958) Powell and Errington (1958), Kubitschek (1962) and Schaecter et al (1962).

The data on interdivisional times of cells from these experiments motivated a number of efforts to model the kinetics of individual cells. In particular, there were a number of efforts to find a theoretical form for the generation time distribution, denoted $f(\tau)$, which would be consistent with the data. A variety of models were proposed based upon various hypothetical mechanisms, and usually fell into one of three classes: compartment type models, size-dependent models or multitype branching processes. (This is not to say other models don't exist which are suggested by other forms of experimentation. The reader should keep this in mind). The models of Rahn (1932) and Kendall (1948, 1952) were compartment type kinetic models, while the later models of Koch and Schaecter (1962) and Clifford and Sudbury (1973) was based upon a size control mechanism. The latter two models also attempted to explain the observed positive correlation in life lengths generally observed in the data. This was also the motivation in the queueing model of Kuczek (1983) which was based upon the work of Cooper and Helmstetter (1968). Finally, very general branching process models were developed, such as the Crump-Mode-Jagers (or CMJ) model, but which were somewhat divorced from the biology. (For more details and references, see Kuczek (1984)).

While the study of the microstructure of colony formation generated much data and many models, their contribution to the understanding of cell kinetics was limited for several reasons including:

1) The number of cells one could observe in an experiment were relatively small owing to the difficulty in making observations. Also, difficulty of observation caused censoring. This raises a question of how representative the observed cells were of the population

of cells in culture.

2) Identifiability. A variety of models for some characteristic, such as $f(\tau)$, fit equally well to the data. If the data couldn't differentiate between models, it couldn't differentiate between mechanisms.

3) There is typically a significant component of clonal heterogeneity and interexperiment variability in interdivisional time data which cause difficulty in building and interpreting models (see Kuczek and Axelrod (1986), Axelrod and Kuczek (1989) and Powell (1958)).

3. Macrostructure of Cloning Experiments

In cloning experiments whose point was to study the microstructure of colony formation, a subset of colonies were chosen for detailed study. The macrostructure of the experiment consists of the entire population of colonies formed. In particular, the entire distribution of colony sizes in the experiment is of interest (or some function thereof). Such experiments often take on a comparative form. For example, in the human tumor cloning assay (HTCA), biopsy tissue is plated out on either soft agar or agar plus chemotherapeutic agent. The drug is assumed effective if the number of colonies of more than fifty cells for the treated group is less than some percent of control. In this case only part of the cumulative frequency distribution is of interest. i.e. only a part of the macrostructure. Another type of comparative study is to compare cloning of two cell lines whose only difference is the presence or absence of a particular gene. For example, in Kuczek and Axelrod (1987), one finds cumulative frequency histograms of cells/colony for NIH3T3

mousefibroblast cells with and without the H-ras oncogene.

While the focus of microstructure studies was kinetics of individual cells in pedigree, one could envision studies where kinetics of colonies in pedigree were the focus of interest. This is in fact one purpose of subcloning (or respreading) experiments where colonies from cloning experiments are subcloned (respread) and colony sizes of the subclones or daughter colonies are recorded as well as sizes of corresponding mother colonies. Examples of these go back to Hughes (1955) and are as recent as Kuczek and Axelrod (1987). In these papers it was observed that a definitive tendency exists for colony size to persist. In particular, it was noted in Kuczek and Axelrod (1987) that this tendency could be affected by the presence of an oncogene. To generate a broad colony size distribution, however, a tendency to diversify must also exist. One can see this intuitively from the data, but for more detailed analysis of the data, statistical methods must be developed. A Bayesian aproach to fulfilling this goal will be outlined in the next section as it applies to subcloning experiments.

4. Bayesian Paradigm for Modeling the Kinetics of the Macrostructure.

For cells in steady state growth, we assume that there exists a distribution of growth rates which are heritable in the short term. This means that when cells are plated out, that each colony formed will grow at some rate α, where α has a prior distribution $F(\cdot)$. One may think of α in the following sense. If we let $Z(t)$ = colony size at time t, and if this colony is growing at rate α, then

$$Z(t+\tau)/Z(t) \approx \exp(\alpha\tau).$$

This idea of growth rate is consistent with the Malthusian parameter for the Bellman-Harris process (see Kuczek (1982a)).

It is the case, however, that two different cells, even if they grew at the same rate α, need not produce colonies of the same size after a fixed interval of time owing to:

1) asynchrony of the founder cells

2) random fluctuation in generation times in the first few generations.

Again there are analogues for the Bellman-Harris process. The analogue to 1) is that

$$A_t(x) \xrightarrow{\text{a.s.}} A(x)$$

where $A_t(x)$ is the empiric age distribution for the super critical Bellman-Harris process where $A(x)$ has a particular form (Kuczek (1982b)). The analogue to 2) is that

$$Z(t)/\exp(\alpha t) \xrightarrow{\text{a.s.}} W$$

where W satisfies a certain integral equation (see Athreya and Ney (1972)). We will therefore assume that

$$Z(t) = X\exp(\alpha t)$$

where X is a random variable which accounts for 1) and 2) above, and α is the (randomly distributed) growth rate.

With this in mind, one can now approach the problem of modeling the distribution of colony sizes in the following fashion. Suppose that we are given a sample of colony sizes

$$Z_1(t),\ Z_2(t),\ldots,Z_n(t)$$

which were observed at the end of a cloning experiment, and assuming that

$$Z_j(t) = X_j \cdot \exp(\alpha_j t),$$

then we may transform the data by taking logs. This yields

$$\log(Z_j(t)) = \log(X_j) + \alpha_j \cdot t.$$

Now making a further assumption that G is the distribution of log (X_j), F is the distribution of $\alpha_j \cdot t$ and that elements of these sequences are independent of each other, then the sequence $\log Z_j(t)$ are i.i.d. with distribution function $F * G(\cdot)$, i.e. the convolution of $F(\cdot)$ and $G(\cdot)$.

Now assuming that the data have this statistical structure, it is necessary to find actual forms for $F(\cdot)$ and $G(\cdot)$ in order to describe either the colony size distribution or the log-colony size distribution. There are two possible approaches.

Approach 1. One approach would be to estimate the generation time distribution from the data for a particular growth rate α denoted $f_\alpha(\cdot)$ from existing data on interdivisional times (summaries of such data have been previously published in Kuczek and Axelrod (1986)). This data can also be used to obtain information on the variation of growth rates α. Then one can use existing branching process theory to derive a form of $G(\cdot)$, i.e. by deriving the age distribution and approximating distribution of W. The behavior of $F(\cdot)$ could be approximated by estimating the observed variability of interdivisional times between small families.

Approach 2. This approach is more intuitive and is based upon the observation by Kubitschek (1962) that generation time rates, i.e. the reciprocals of interdivision times,

when plotted on normal probability paper, produce straight lines for a wide variety of organisms. He had demonstrated this for several previously published data sets. It is known that clonal heterogeneity of growth rate exists, so an alternative explanation of Kubitschek's results could be suggested, i.e., that α is distributed (approximately) normally and this gave the observed straight line normality plots. In addition, the subcloning data presented in Kuczek and Axelrod (1987) show that growth rate differences of subcolonies compared to mother colonies also exhibit approximate normality. So then one could, as a starting point, approximate $F(\cdot)$ and $G(\cdot)$ by normal distributions.

Up to now, there has been nothing Bayesian except an attempt to specify a prior distribution on growth rate, $F(\cdot)$. The application will come in attempting to explain subcloning experiments in the following fashion.

Step 1. Observe, in cloning experiment, primary colony of size x.

Step 2. Compute posterior distribution of growth rate α for this colony, given that it is of size x.

In step 2, the colony size distribution is given by $F * G(\cdot)$. Since F and G are normal, one can use standard calculations to get the posterior distribution of α given x which may be found in Press (1982).

Step 3. Predict distribution of subcolony sizes.

In order to carry out step 3, one must specify a transition kernel which will give a probability distribution of the growth rate of the daughter colonies. That is to say we must specify a stochastic mechanism for diversification of rate. Let α denote the growth

rate of the mother colony and α' denote the common growth rate of the daughter colonies, and $P(\alpha, \alpha')$ the transition kernel. Then if Post $(\alpha|x)$ denotes the posterior distribution of growth rate of a mother colony given that it is of size x ne could use this to predict the distribution of growth rate of subcolony which is given by

$$\int P(\alpha, \alpha') \; Post(d\alpha|x).$$

If subcolony sizes are distributed as

$$X \exp(\alpha' t),$$

then one can again calculate, at least in principle, the distribution of

$$\log(X) + \alpha' \cdot t$$

and compare it to the data.

What is of interest in the application is that in choosing F and G we are guided by the data, but the choice of $P(\alpha, \alpha')$ is left to the researcher who may wish to test various potential patterns of diversification over time.

In this way kinetics of colony growth over time could be elucidated. The actual analysis with existing data will be treated in a later manuscript where we will address the following questions among others:

1) How heritable is growth rate?

2) How can particular genes affect this heritability?

References

1. Athreya, K. B. and Ney, P. (1972). Branching Processes. Springer Verlag. New York.

2. Axelrod, D. E. and Kuczek, T. (1989). Clonal heterogeneity in populations of normal cells and tumor cells. *Comp. Math. Appl.* **18**, 871–881.

3. Clifford, P. and Sudbury, A. (1973). The linear cell-size dependent branching process. *J. Appl. Probl.* **9**, 687–696.

4. Cooper, S. and Helmstetter, C. E. (1968). Chromosome replication and the division cycle of E. Col. Blr. *J. Mol. Biol.* **31**, 519–540.

5. Hughes, W. H. (1955). The inheritance of growth rate in *Escherichia coli*. *J. Gen. Micro.* **12**, 265–268.

6. Kelly, C. D. and Rahn, O. (1932). The growth rate of individual bacterial cells. *J. Bact.* **23**, 147–153.

7. Kendall, D. G. (1948). On the role of variable generation time in the development of a stochastic birth process. *Biometrika* **35**, 316–330.

8. Kendall, D. G. (1952). On the choice of a mathematical model to represent normal bacterial growth. *J. Roy. Statist. Soc. Ser. B* **14**, 41–44.

9. Koch, A. L. and Schaecter, M. (1962). A model for the statistics of the cell division process. *J. Gen. Micro.* **29**, 435–454.

10. Kubitschek, H. E. (1962). Normal distribution of cell generation rate. *Exp. Cell Res.* **26**, 439–450.

11. Kuczek, T. (1982a). Almost sure limit results for the supercritical Bellman-Harris process. *J. Appl. Prob.* **19**, 668–674.

12. Kuczek, T. (1982b). On the convergence of the age distribution for the supercritical age dependent Bellman-Harris process. *Ann. Prob.* **10**, 252–258.

13. Kuczek, T. (1983). On the $G_a/G/\alpha$ queue. *Adv. in Appl. Prob.* **15**, 444–459.

14. Kuczek, T. (1984). Stochastic modeling for the bacterial life cycle. *Math. Biosci.* **69**, 159–169.

15. Kuczek, T. and Axelrod, D. E. (1986). The importance of clonal heterogeneity and interexperiment variability in modeling the eukaryotic cell cycle. *Math. Biosci.* **79** 87–96.

16. Kuczek, T. and Axelrod, D. E. (1987). Tumor cell heterogeneity: Divided colony assay for measuring drug response. *Proc. Nat. Acad. Sci.* USA **84**, 4490–4494.

17. Powell, E. O. (1955). Some features of the generation time of individual bacteria. *Biometrika* **42**, 16–45.

18. Powell, E. O. (1958). The pattern of bacterial generation times. *J. Gen. Micro.* **18**, 382–417.

19. Powell, E. O. and Errington, F. P. (1963). Generation times of individual bacteria: Some corroborative measurements. *J. Gen. Micro.* **31**, 315–327.

20. Press, S. J. (1982). Applied Multivariate Analysis. Krieger Publishing. Malabar, Florida.

21. Rahn, O. (1932). A chemical explanation of the variability of growth rate. *J. Gen. Physiol.* **15**, 257–267.

22. Schaecter, M., Williamson, J. P., Hood, J. R., and Koch, A. L. (1962). Growth, cell and nuclear division in some bacteria. *J. Gen. Micro.* **29**, 421–434.