FIRST ROCKY MOUNTAIN REGIONAL CONFERENCE ON MEDICAL APPLICATIONS OF STATISTICS. Preprint volume, 27-33, 1982.

Stochastic Models for Plasmid Copy Number

Eugene Seneta
Dept. of Math. Statistics
University of Sydney
NSW 2006
Australia

Simon Tavaré
Department of Statistics
Colorado State University
Fort Collins, Colorado
80524

ABSTRACT

Some stochastic models for the copy number of plasmids in a cell line are considered. We discuss plasmid copy number in both random lines of descent and the whole cell line. The models are compared by use of asymptotic cure rates and the asymptotic fraction of cells containing a given number of plasmids. The results are derived using the extensive theory of multitype branching processes. Numerical comparisons are also presented.

I. Introduction

In this paper we study the behavior of some stochastic models of plasmid reproduction proposed by Gardner et al.[2] and Novick and Hoppensteadt [6]. Consider a cell line initiated by a single cell at time $n = 0$, and growing by binary splitting at times $n = 1, 2, \ldots$. The initial cell contains a number of plasmids which replicate in the cell, and are then distributed at random to the two daughter cells at cell division. A cell is said to be cured if it contains no plasmids, and then its daughter cells are also cured. The main features we seek to analyse are the stability of plasmid copy number and the cure rate under different models of plasmid reproduction. We now describe briefly three such models developed in [2].

Doubling model. If the plasmid copy number in a cell is currently $i (\geq 0)$, then there are i birth events, each of which survives to partition with probability p. Denoting the number of survivors by Y, the number of plasmids in the cell

before partition is $i + Y$.

Additive model. There is a fixed number N of birth events, each of which is successful with probability p. If there are $i(> 0)$ plasmids before replication, there will be $i + Y$ before partition.

Equilibrium model. If there are currently $i(> 0)$ plasmids in a cell, there are $2N - i$ birth events, each of which is successful with probability p. The number of plasmids in the cell before partition is then $i + Y$.

We remark that in each case above, survival of a birth to partition occurs independently for each birth event. To avoid obscuring the replication mechanisms described above, we will be assuming a random partition mechanism so that the probability $P_{\ell j}$ that a cell now containing ℓ plasmids contributes j to a particular daughter cell is

$$P_{\ell j} = \binom{\ell}{j} 2^{-\ell} \quad , \quad j = 0, 1, \ldots, \ell \quad . \tag{1.1}$$

II. RANDOM LINES OF DESCENT

We first discuss the evolution of plasmid copy number along a random line of descent; that is, we choose at random a daughter cell of that cell chosen at the previous stage, so constructing a random line of descent from the initial cell. Let X_n be the number of plasmids before replication in the cell chosen at time n. It is clear that $\{X_n, n \geq 0\}$ is a Markov chain with one-step transition probabilities p_{ij} given by

$$p_{ij} = \Pr\{X_{n+1} = j \mid X_n = i\} = \sum_{\ell} R_{i\ell} \, P_{\ell j} \quad , \tag{2.1}$$

where $P_{\ell j}$ is given in (1.1), and $R_{i\ell}$ is the probability that a cell initially containing i plasmids produces ℓ before partition. The $R_{i\ell}$ are computed for the three models from the prescriptions given in the introduction.

For the doublind model, we can write $X_{n+1} = \xi_1 + \ldots + \xi_{X_n}$, where the ξ_i are i.i.d. random variables with probability generating function f(s) given by

$$f(s) = \frac{1}{2} - \frac{p}{4} + \frac{s}{2} + \frac{p s^2}{4} \quad , \quad 0 \leq s \leq 1 \tag{2.2}$$

It follows immediately that $\{X_n\}$ is a Galton-Watson branching process (GWBP) with offspring p.g.f. f(s).

A similar representation holds for the additive model. If $X_n > 0$, we have $X_{n+1} = \xi_1 + \ldots + \xi_{X_n} + I$, the ξ_i being i.i.d.

with p.g.f. f(s), and independent of I, which has p.g.f. b(s):

$$f(s) = \frac{1}{2} + \frac{1}{2}s\,, \quad b(s) = (1 - \frac{p}{2} + \frac{ps}{2})^N, \quad 0 \leq s \leq 1 \qquad (2.3)$$

The chain $\{X_n\}$ is now identified as a GWBP with immigration stopped at zero. Further details are contained in [10].

Finally, the equilibrium model leads to a finite Markov chain with

$$P_{ij} = \sum_{\ell=i \vee j} \binom{2N-i}{\ell-i} p^{\ell-i}(1-p)^{2N-\ell}\binom{\ell}{j}2^{-\ell} \qquad (2.4)$$

It is intuitively clear that eventually the line of descent is cured (so that $X_n = 0$ for some n) and interest then focusses on the curing time, T_0. In Markov chain terminology, state 0 is absorbing, and so we can write (p_{ij}) in the form

$$(p_{ij}) = \begin{pmatrix} 1 & 0' \\ \underline{0} & Q \end{pmatrix} \qquad (2.5)$$

The matrix Q, and its Perron-Frobenius eigenvalue ρ play a fundamental role in our analysis. While it is possible to find properties of the cure time numerically (cf. [1], [2] for the case $p = 1$), we concentrate on the asymptotic cure rate θ defined by

$$\theta = \lim_{n\to\infty} \Pr\{X_{n+1} = 0 \mid X_n > 0\} = 1 - \lim_{n\to\infty} \frac{\Pr\{T_0 > n+1\}}{\Pr\{T_0 > n\}}\,. \qquad (2.6)$$

In all cases, it can be shown that $\theta = 1 - \rho$. This is verified for the multipicative model (when $p < 1$) via Kolmogorov's theorem [3, p. 18], and for the additive model using results established in [9]. Using subscripts M, E, A to distinguish the models, we find in particular that $\rho_M = (1+p)/2$ (cf. [11]). Neither ρ_E nor ρ_A can be found explicitly (except when $p = 1$, $\rho_E = 1 - 2^{-2N}$), although a functional equation for ρ_A can be given [9]. Computation of ρ will be described in section IV.

III. COPY NUMBER IN THE WHOLE CELL LINE

The behavior of copy number in a random line of descent tells us little about copy number density in the whole cell line. To study this problem, we use a method due to Harris [3]. See also [6]. For $j \geq 0$, let $X_n^{(j)}$ be the number of cells in generation n containing j plasmids before replication. Since $\sum_{j\geq 0} X_n^{(j)} = 2^n$, we concentrate on the vector process

$\underline{X}_n' = (X_n^{(1)}, X_n^{(2)}, \ldots)$. It turns out [10] that $\underline{X}_n'$ is a multitype GWBP, with mean matrix $M = (m_{ij})$ (where m_{ij} is the expected number of cells containing j plasmids produced by a cell containing i before replication) given by $M = 2Q$, Q being determined in (2.5). Further analysis depends on the behavior of powers of M, which in view of the relationship $M = 2Q$ can be carried out by studying Q. In particular, let $\underline{v}'$, $\underline{w}$ be the left and right eigenvector of Q corresponding to eigenvalue ρ. We can normalise these so that

$$\sum_{i \geq 1} v_i = 1, \quad \sum_{i \geq 1} w_i v_i = 1 \ . \tag{3.1}$$

To describe the cure rate in the whole cell line, we look at the proportion of cells cured from the n th to (n + 1)st generation. This is given by

$$\theta_n = 1 - E \frac{\underline{X}_n' M \underline{1}}{2\underline{X}_n' \underline{1}} \tag{3.2}$$

and, using [4], [5] we have $\lim_{n \to \infty} \theta_n = 1 - \rho$. Thus $1 - \rho$ gives the asymptotic cure rate for the whole cell line; notice that this is the same as for the random line of descent in section 2.

To assess the copy number density we chose a measure which is relatively easy to compute. Using the observation that extinction is impossible in our models, we find that

$$\underline{X}_n' / (\underline{X}_n' \underline{1}) \to \underline{v}' \text{ as } n \to \infty \ . \tag{3.3}$$

(almost surely for the equilibrium model, in probability for the others). In any case, v_i can be viewed as the asymptotic fraction of cells containing i plasmids, and the mean and variance of the distribution $\{v_i\}$ provide a way of comparing the stability of plasmid copy number across models.

IV. NUMERICAL COMPARISONS AND DISCUSSION

Having established in (3.2), (3.3) our criteria for comparing these replication systems, we give some numerical results. These were derived using algorithms described in [10], [7], [8, Chapter 7].

TABLE I

p	Asymptotic cure rates $(1-\rho)$						
	Multi-plicative	Additive			Equilibrium		
		N = 5	10	15	N = 5	10	15
.1	.45	.3194	.2098	.1377	.2373	.1056	.0457
.3	.35	.1327	.0337	.0078	.0611	.0051	.0004
.5	.25	.0494	.0039	.0003	.0165	.0003	
.7	.15	.0157	.0003		.0049	3×10^{-5}	
.9	.05	.0041			.0016		

TABLE II

p	Means of copy number distribution ν'						
	Multi-plicative	Additive			Equilibrium		
		N = 5	10	15	N = 5	10	15
.1	1.104	1.384	1.723	2.070	1.599	2.250	2.974
.3	1.353	2.042	3.217	4.571	2.547	4.652	6.927
.5	1.729	2.774	5.039	7.505	3.409	6.669	10.000
.7	2.504	3.613	7.004		4.141	8.236	
.9	6.020	4.537			4.745	9.474	

TABLE III

p	Variance of copy number distribution						
	Multi-plicative	Additive			Equilibrium		
		N = 5	10	15	N = 5	10	15
.1	.108	.328	.766	1.178	.600	1.343	2.186
.3	.414	1.082	2.432	3.900	1.489	3.447	5.308
.5	1.045	1.812	4.071	6.227	2.059	4.432	6.668
.7	3.108	2.506	5.352		2.346	4.844	
.9	26.811	3.068			2.460	4.980	

V. CONCLUSIONS

We examined three models of plasmid replication proposed in [2] and [6] under the assumption of random partition from the standpoint of a random line of descent and of a whole cell line. The concept of asymptotic cure rate, defined for each approach, coincided. The second viewpoint also allowed consideration of the asymptotic fractions of cells containing a given number of plasmids, and seems useful in that the study of copy number by random line of descent [1] [2] is likely to be generally misleading. The study of the evolution of the whole cell line was achieved by the use, in a simple way, of the extensive theory of multitype branching processes.

The multiplicative model allows for "unrestricted" replication, and is unlikely to be accepted as plausible. Numerical results indicate that a decrease in cure rate with increasing p for a given type of model, or for fixed moderate p in proceeding from multiplicative to additive to equilibrium, as accompanied by an increase in the mean of the corresponding copy number distribution (as might have been expected). The variance of this distribution is somewhat higher for the equilibrium model than for the additive one, for the same values of the parameters N and p. For p close to unity the multiplicative model has both high mean and variance in this distribution, which highlights its unsuitability.

REFERENCES

[1] EMERSON, W. D., HOPPENSTEADT, F. (1981). On the stability of plasmid copy number. Research Report, Dept. of Mathematics, University of Utah.

[2] GARDNER, R. C., METCALF, P., BERGQUIST, P.L. (1977). Replication, Segregation and the hereditary stability of bacterial plasmids. Unpublished report, University of Auckland.

[3] HARRIS, T. E. (1963). The Theory of Branching Processes, Springer Verlag, New York.

[4] KESTEN, H., STIGUM, B. P. (1966). A limit theorem for multidimensional Galton-Watson processes. Ann. Math. Stat, 37, 1211-1223.

[5] MOY, S.-T. C. (1967). Extensions of a limit theorem of Everett, Ulam and Harris on multitype branching processes to a branching process with countably many types. Ann. Math. Stat., 38, 992-999.

[6] NOVICK, R. P., HOPPENSTEADT, F. (1978). On plasmid incompatibility. PLASMID, 1, 421-434.

[7] POLLAK, E. (1969). Bounds for certain branching processes. J. Appl. Prob., 6, 201-204.

[8] SENETA, E. (1981). Non-Negative Matrices and Markov Chains, Springer Series in Statistics, Springer-Verlag, New York.

[9] SENETA, E., TAVARÉ, S. (1982). A note on models using the branching process with immigration stopped at zero. Submitted to J. Appl. Prob.

[10] SENETA, E., TAVARÉ, S. (1982). Stochastic models for plasmid copy number. Submitted to Theor. Popn. Biol.

[11] SENETA, E., VERE-JONES D. (1966). On quasi-stationary distributions in discrete time Markov chains with a denumerable infinity of states. J. Appl. Prob., 3, 403-434.