A Model for Statistics of the Cell Division Process

BY A. L. KOCH AND M. SCHAECHTER

Departments of Microbiology and Biochemistry, College of Medicine, University of Florida, Gainesville, Florida, U.S.A.

(Received 27 December 1961)

SUMMARY

A model for the statistics of cell division is proposed. The model assumes:
(1) that growth at the cellular level is deterministic;
(2) that the mean size of a cell at division is under cellular and environmental control;
(3) that the distribution of sizes of cells at division has a small coefficient of variation, and is independent of the size at previous divisions;
(4) that the cell divides nearly into equal halves. The observed coefficient of variation of the life-length distribution \( f(\tau) \) results from a twofold smaller coefficient of variation in the distribution of cellular mass at division, \( g(c) \). The magnification of the coefficient of variation results from the fact that the mass variable enters twice, once in determining the size of cell formed at division, and, secondly, in determining the size of the cell when it in turn divides. In the expression for \( \tau \), these sizes enter as the logarithm of their ratio. These mathematical operations contribute to the increased coefficient of variation of the \( f(\tau) \) distribution over the \( g(c) \) distribution. The skewed nature of the \( f(\tau) \) distribution is attributable to a number of causes. We feel that an important source of the skewness is due to deviations from equi-partition of cell constituents at division. In this respect the present model is in disagreement with previous models which presume that the skewness results from the statistics of a small number of molecular events taking place inside each cell. The experimentally observed positive sister-sister life-length correlation and the negative mother-daughter life-length correlation are explained by the model. Deviations from the predicted values of +0.5 and -0.5, respectively may be explained in part by deviation from equal distribution of cell contents at division. Deviations of the correlations in excess of these, observed in some cases, imply the existence of the other biological processes. The size distribution of cells in balanced growth cultures based on this model is given.

INTRODUCTION

Since the work of Kelly & Rahn (1932) many workers have attempted to interpret the distribution of the times between successive bacterial cell divisions. This has usually been done in terms of the statistics of intracellular molecular events. Thus, Rahn (1931–1932) postulated that cell division ensues when all of a group of cellular elements, pictured as genes, has replicated. He proposed that the replication of each genetic element takes place by a first-order reaction and, therefore, is random in time. This theory was statistically refined by Finney & Martin (1951). Kendall (1948) viewed the process of cell division as resulting from the completion of a succession of necessary steps. He pictured each step to be a random Markov process, i.e. a stochastic process, and that the ‘convolution’ or concatenation of
these steps determines the distribution of ages that cells will achieve before dividing. Each model thus generates a cell interdivision time or life-length distribution of a particular mathematical form that may be fitted to experimental data. Kendall in 1952 combined both his previous model and Rahn's model in a general theory. For all models we shall refer to the interdivision time distribution as \( f(t) \), where the argument \( t \) is the span of time between two successive cell divisions.

We should like to explore a radically different type of model, one which postulates that the basic process of biosynthesis of cellular protoplasm is deterministic, i.e. precisely predictable at the level of the single cell. Such a view is much more probable today than in the past because of our increased knowledge in the regulatory biology of micro-organisms. As examples, it appears that deoxyribonucleic acid synthesis at the level of individual bacterial cells is essentially continuous and that ribosomes participate in protein synthesis at a constant rate (e.g. Schaechter, Bentzon & Maalee, 1959; Neidhardt & Magasanik, 1960).

It is well known that the growth rate of micro-organisms can be controlled by the amount of nutrients supplied in the medium. It is also clear that a large number of enzyme molecules in each cell deal with the nutrient substrates. Therefore, if stochastic processes limited by nutrients are considered to control cell division, it is necessary to postulate that the critical steps involved must be those controlling the synthesis of the limiting enzyme. This, however, is unlikely because enzyme synthesis (Jacob & Monod, 1961) is under the genetic and physiological control of both positive and negative repression mechanisms. These tend to regulate the enzyme synthesis very precisely. Thus, one postulates that, in a culture in balanced growth, the bulk of the constituents of each cell increases with the same growth constant that characterizes the cultures as a whole. Our measurements (Schaechter, Williamson, Hood & Koch, 1962) of elongation of individual cells through the growth cycle are in accord with this notion. At the level of the single cell it is difficult to distinguish between an exponential curve and certain other functions (in particular, the linear function). None the less, all cells have nearly the same growth rate constant. There are certainly no measurable major discontinuities in the growth process.

Thus, we take as our fundamental relation that

\[
\frac{dx}{dt} = \lambda x, \tag{1}
\]

where \( \lambda \) is the growth rate constant of the culture and \( x \) is the amount of any cell constituent or some combination of constituents in a cell. In particular, we shall consider the total protoplasmic mass, \( m \).

The second postulate to be introduced is that the mean size of a cell at division is under physiological control; that is to say, on the average, a cell grows until it has reached a critical size \( \bar{c} \) dependent on physiological and biochemical parameters; division then ensues. The notion of critical size is inherent in the ideas of Rashevsky (1938), although little credence may be given to the particular mechanism he proposes. This model has been considered and rejected out of hand as applying to higher organisms such as: Colpoda (Adolph, 1929), sea urchin eggs (Swann, 1954) and Tetrahymena pyriformis GL (Scherbaum & Rasch, 1957). On the other hand, this idea was tacitly assumed by Hershey (1939) to apply to bacteria. We found that in growing bacteria the size at division is critical, in the sense of having a smaller
Statistics of cell division

coefficient of variation than does the age at division (Schaechter et al. 1962). In
fact, Hood & Schaechter (unpublished observation) have found that the critical
size of bacteria at division changes monotonically, and without increase in its
coefficient of variation, throughout the reorganization that accompanies a shift of a
growing culture from one medium to another medium in which the growth rate and
bacterial size are different. Thus, we propose that the average mass of a cell at
division is a well-controlled physiological parameter.

We propose as a third postulate that although the control is effective, the size at
division may vary slightly due to random causes. For the present discussion it is
immaterial whether one assumes (1) that each cell divides precisely when it reaches
its critical size and that the critical size varies from individual to individual, or
(2) that every cell of the same genotype in the same environment has precisely the
same critical size but that for secondary reasons, the appearance of visible evidence
division may be premature or delayed. In either event we shall assume as an
important part of this postulate that this variation for certain cases of interest is
small, Gaussian, and uncorrelated with events in previous cell cycles or in sister
cells. In analogy with the life-length distribution we shall designate this distribution
$g(c)$, where $c$ is the critical mass of the cell at division. Its coefficient of variation
will be designated $q$.

Our last postulate is that cell division results in an equal or nearly equal division
of the cytoplasm. We thus introduce the distribution $h(p)$, where $p$ is the proportion
of the cytoplasm of the mother, incorporated into one daughter. Obviously, the other
daughter has a proportion of $1 - p$. In most of the present discussion we shall
assume that division is precise, i.e. mean $= 0.5$, variance $= 0$. In the remainder
we shall assume that the mean is $0.5$ but that it has a small coefficient of variation,
$q_h$.

Distribution in balanced growth

We shall first consider the relationships between distributions of certain pro-
properties at division with the distribution of these properties in whole cultures. We do
this because certain data in the literature must be accounted for by our theory. In
balanced growth situations, in particular, it is of interest to be able to calculate the
distribution of a property, such as age or mass, for the cells of the entire culture.
For the simplest case in which all cells reach the same age $\bar{a}$, before dividing, the
age distribution $\phi(a)$ is given by

$$\phi(a) = 2\lambda e^{-a}; \quad 0 \leq a \leq \frac{\ln 2}{\lambda} = \bar{a}$$

and

$$\phi(a) = 0; \quad a < 0, \quad a > \bar{a}$$

where $a$ is the age of the cell, $\phi(a)da$ the probability that a cell is between $a$
and $a + da$ of age. This distribution has been derived previously by Scherbaum
& Rasch (1957), Edwards et al. (1960), James (1960), and Stanners & Till
(1960).

The situation for mass or length distribution has not been so often studied. For
the simplest case, where it is assumed that all cells grow exponentially to a critical
mass or length, $\bar{c}$, and then divide precisely in half, the distribution of cell mass in
A. L. KOC\m AND M. SCHAECHTER

the culture is readily shown to be (Maclean & Munson, 1961, and Appendix note 1)

$$\theta(m) = \frac{\bar{c} - m}{m^2}; \quad \frac{\bar{c}}{2} \leq m \leq \bar{c}$$

and

$$\theta(m) = 0; \quad m < \frac{\bar{c}}{2}, \quad m > \bar{c}$$

where $\theta(m)dm$ is the probability that a cell has mass between $m$ and $m + dm$. Several interesting comparisons between these limiting forms of the $\phi(a)$ and $\theta(m)$ distributions should be noted; these are listed in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Comparison of the simplest cases for the $\phi(a)$ and $\theta(m)$ distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>distribution</td>
</tr>
<tr>
<td>$\phi(a)$</td>
</tr>
<tr>
<td>Formula:</td>
</tr>
<tr>
<td>Range:</td>
</tr>
<tr>
<td>Ratio-maximum to minimum frequency:</td>
</tr>
<tr>
<td>Mean value:</td>
</tr>
<tr>
<td>Percent further growth to complete cycle:</td>
</tr>
<tr>
<td>Coefficient of variation:</td>
</tr>
</tbody>
</table>

The limiting form of the mass distribution, equation (3), is more useful than the limiting form of the age distribution, because the assumptions on which the former is based are more nearly met for normally growing bacteria. Pertinent data are available in the literature. For example, it would be readily possible to fit equation (3) to the data of Kubitschek (1962), Maclean & Munson (1961) and to our own data with Escherichia coli and Salmonella typhimurium (Schaechter et al. 1962). It must be recognized that data fitting equation (3) have not been obtained with the electronic resistance counter: e.g. Coulter Counter. With this instrument a significant proportion of particles many times larger than average is found (Kubitschek, 1958; Lark & Lark, 1960). These results are probably due to coincident entrance of several smaller particles into the sensitive volume (which is not merely that of the orifice) of the counter, to the formation of elongated forms, i.e. filaments, or to the formation of aggregates. Each aggregate would be counted as a single large particle instead of several small particles.

Powell (1956) and Harris (1959) have derived the general solution for the age distribution. The probability distribution is given by

$$\phi(a) = 2\lambda e^{-\lambda a} \left[ 1 - \int_0^a f(\tau) d\tau \right], \quad (4)$$

where $f(\tau) d\tau$ is the probability that a cell will divide between age $\tau$ and $\tau + d\tau$.

A general solution for the mass distribution has not previously been presented, although an empirical calculation has been given (Scherbaum & Rasch, 1957). We have been unable to derive a completely general formula; a somewhat more general distribution is derived in Appendix, note 2 and is

$$\theta(m) = \frac{C}{m^2} \left[ \int_0^{2m} g(c) dc - \int_0^m g(c) dc \right], \quad (5)$$
Statistics of cell division

where \( g(c) \, dc \) is the probability that a cell will divide when its mass lies between \( c \) and \( c + dc \). This distribution, applies only if the distribution \( g(c) \) is not abroad. Since for balanced growth cultures of certain bacteria, the \( g(c) \) distribution does have a small coefficient of variation, it may be hoped that this formula will have a quite broad range of application. It has also been assumed that cell division results in precisely equivalent daughter cells. If this is not true, the distribution in the first integral can be replaced with a broader distribution to correct for inequality of division.

**Relationship of mass distribution at division \( g(c) \) to life-length distribution \( f(\tau) \)**

We will first consider a case based on the four assumptions given above, with the further restriction that the cell divides evenly in half, i.e. mean of \( h(p) = 0.5 \), variance = 0. Designate the mass of the mother cell at division as \( c_1 \), and the mass of the daughter cell at division as \( c_2 \). Since the growth in mass is deterministic and exponential, we may immediately write

\[
c_2 = \frac{1}{2} c_1 e^{+\lambda \tau}.
\]

Solving this equation for the life-length, \( \tau \), we obtain

\[
\tau = \frac{1}{\lambda} \ln \frac{2c_2}{c_1}.
\]

The form of this equation suggests two things. First, that antilog of \( \tau \lambda \) be taken as the fundamental variable for statistical analysis. This is considered below. Secondly, that the form of the distribution of \( f(\tau) \) will not only depend on the form of \( g(c) \), from which both \( c_1 \) and \( c_2 \) are drawn, but also on the influence of forming a ratio and of taking a logarithm.

In order to see how these mathematical operations influence the distribution, let us consider a variable that is normally distributed and has a mean value of unity. We may write this variable as \( 1 + qx \), where \( q \) is the coefficient of variation and \( x \) is the normal deviate; i.e. \( x \) is a normally distributed variable with zero mean and unit variance. Then, from the statistical principle that 68-26% of the cases lie between one standard deviation below and one standard deviation above the mean, it follows that 15.87% of the cases will be below \( 1 - q \) and 15.87% of the cases will lie above \( 1 + q \). Now the same relationship will hold for the distribution of the reciprocals of those values, i.e. 15.87% of the cases will lie above \( 1/(1 + q) \), etc. The new distribution is no longer symmetric since \( 1/(1 - q) \) is larger than one by a greater amount than \( 1/(1 + q) \) is smaller than one. Thus the effect of taking the reciprocal tends to skew the distribution by extending the right tail (i.e. positive skewness). For any particular value of \( q \) the new distribution is readily calculated by choosing \( x \) corresponding to other percentage points on the normal distribution curve and plotting these against the corresponding values of \( 1/(1 + qx) \). In this way a cumulative distribution is obtained. The ordinary distribution can then be computed by substraction.

Accordingly, the ratio of two uncorrelated normal variables will also be positively skewed, but the calculation is more complex. The general solution for the fiducial limits of the ratio of variables has been given previously (Finney, 1952) and is known as Fieller's theorem (1940). We are interested in a very restricted sub-case of this theorem. If we take the numerator to be \( A = A(1 + xq_a) \), and the
denominator to be \( B = B(1 + xq_b) \), the ratio, \( R \), corresponding to given values of the normal deviate, \( x \), may then be written

\[
R = \frac{A}{B} \left(1 + \frac{\sqrt{(q_a^2 + q_b^2) - x^2 q_a^2 q_b^2}}{1 - x^2 q_a^2}\right).
\]  

(8)

This equation is derived from general solution by assuming that variations in \( A \) are completely uncorrelated with variations in \( B \), and that a large number of single pairs of observations is made. The latter restriction is imposed so that the distribution of ratio of observations and not the ratio of parameters is obtained.

For our immediate purposes we shall assume that the coefficients of variation of both the numerator, \( q_a \), and the denominator, \( q_b \), are the same, since they both are equal to the coefficient of variation, \( q \), of the distribution \( g(c) \). From equation 8 and a table of the error function one can readily compute the resultant cumulative distributions. When this is done for a coefficient of variation of \( g(c) \) chosen to be 10 or 20%, the resultant cumulative distributions are markedly skewed in the positive sense (Fig. 1).

Fig. 1. Cumulative distribution curves for ratio of two variables having the same value of \( q \).

Fig. 2. Cumulative life-length distribution for cells dividing into two equal halves. \( \tau / \bar{\tau} \) is equal to \( \ln (A/B) \). See eq. 7, p. 439.

However, as shown in Fig. 2, the logarithm of the ratio is nearly normally distributed. Increasing \( q \) increases the deviations at the extremes alone, without distorting the central portion. It is not unexpected that the logarithm should be less positively skewed than the ratio since the logarithm function compresses large values. However, the close fit to the normal curve is certainly not obvious from inspection of the equations or their series expansions.

If the coefficient of \( g(c) \), \( q \), is small, the coefficient of variation of the ratio is \( 1.414 q \). This follows from the statistical principle that the coefficient of variation of
Statistics of cell division

the quotient is the square root of the sums of squares of coefficients of variation. It may also be deduced from equation (8) by neglecting the third term under the square root sign and the second term in the denominator. However, the empirically calculated value of the coefficient of variation of the logarithm of the ratio is considerably larger. It is almost precisely twice that of the distribution $g(c)$, being 2.09 times larger for $q = 10\%$ and 2.07 times larger for $q = 20\%$ (see Fig. 2).

Thus, the theory derived on the basis of these assumptions can explain the observed coefficients of variation in the age at division (20%) in our experiments (Schaechter et al. 1962) entirely on the basis of the observed coefficients of variation in the critical mass (10%) in the same experiments. However, it has been repeatedly stated that $f(\tau)$ has positive skewness. This feature has been emphasized and explained in various ways by previous authors. It does not simply follow from our theory as developed so far. Neither was it found in our experiments (Schaechter et al. 1962). In those cases where skew distributions are experimentally observed, they may be explained on the basis of the present model with the following three reasons, alone or in combination. First, $g(c)$ may not be a normal distribution. However, since the same distribution appears in the numerator and the denominator of the ratio $c_2/c_1$, the quotient tends to have a more normal distribution than that of $g(c)$ (see Appendix, note 3). If this factor be the sole cause of the skewness of the $f(\tau)$ distribution observed in some cases, then $g(c)$ would have to be negatively skewed to quite a high degree. Although this supposition is improbable, it is experimentally difficult to rule out completely in our experiments owing to the small size of the bacteria employed.

The second explanation is simply that environmental fluctuation and experimental vicissitudes are more frequently deleterious than beneficial. Thus a deleterious fluctuation would increase the life length much more than the equivalent beneficial fluctuation would decrease it. This would then distort the distribution in the observed direction. In fact, looking at published work it is quite clear that the skewed character of the curve has decreased as experimental technique has advanced.

The third explanation, which we believe is valid in certain cases, is that the skewness of the $f(\tau)$ distribution arises directly from inequalities of the cell division process. If the division process is not quite precise, the denominator (but not the numerator) of the mass ratio $c_2/c_1$ contains an additional fluctuation, i.e. the factor $1 + x q_h$. We have seen that the logarithm of the ratio of two quantities with the same coefficient of variation is nearly normally distributed (Fig. 2). It follows that an increased coefficient of variation in the denominator over that of the numerator will lead to a positive skewness in the distribution of the logarithm of the ratio. This is shown by numerical calculation in Fig. 3 (see Appendix, note 4). On this explanation, the skewness of the $f(\tau)$ distribution is a measure of $q_h$. On this basis our experimental observations (Schaechter et al. 1962) indicate that the division process in the cases and under the conditions employed is quite even, since as mentioned above the $f(\tau)$ distributions in these cases were not significantly skew.
Correlation between sisters

The life-lengths of sister bacteria have repeatedly been found to be positively correlated. A number of explanations may be given: (i) this may represent some common 'trigger' or clock setting in their history (see Appendix, note 5); (ii) it might be caused by a common local micro-environment experienced by both sisters; (iii) it might result from a division synchronization mediated by some chemical stimulus from one cell to the other sister; (iv) it may be the consequence of improperly pooling data from different parts of the same experiment or from different experiments. If the means of the various groups of data are different, a more positive coefficient of intraclass correlation is obtained. However, it may result mainly from the fact that sisters have a common mother, whose critical mass may have been smaller than average, and thus giving both daughters a handicap, or larger than average, thereby giving both daughters a head start.

Let us consider the cell interdivision times for the two sisters resulting from the even division of a mother cell. It is seen from equation (7) that the cellular interdivision times will be correlated to the degree that $c_{2a}/c_1$ and $c_{2b}/c_1$ are correlated. These ratios will be correlated even if $c_1$, $c_{2a}$ and $c_{2b}$ are completely uncorrelated. This phenomenon is well known, and was termed by Karl Pearson (1897) 'spurious correlation'. It can be shown (see Appendix note 6), that if the corresponding coefficients of variation, $q_1$, $q_{2a}$, $q_{2b}$ are all small, the correlation of these ratios is given by

$$
\rho = \frac{q_{1}^2}{\sqrt{(q_{2a}^2 + q_1^2)}} \sqrt{(q_{2b}^2 + q_1^2)}.
$$

![Diagram of cumulative distribution for cell division process](image)
Thus, if cell division results in an even distribution of cytoplasmic constituents and \( g(c) \) has the same dispersion for both generations the three coefficients will be equal and \( \rho \) will be +0.5.

The sister-sister correlation will decrease if the cell division process is not even (see Appendix, note 6). In fact, if \( g(c) \) is constant it may be shown (see Appendix, note 8) that

\[
\rho = \frac{q^2 - q_h^2}{2q^2 + q_h^2}.
\]

Thus, \( \rho \) varies from +0.5 to \(-1\) as \( q_h \) becomes large relative to \( q \).

A final possible cause for the correlation between sisters need be considered. If instead of dividing by a common variable, we subtract a common variable from each of two uncorrelated variables, another type of spurious correlation arises. The two cases are similar, and the ratio case becomes the difference case as the coefficient of variation decreases, since then \((1+y)/(1+x)\) becomes \(1+y-x\). As in the ratio case, if the three variables are uncorrelated and have equal variance, the correlation coefficient of the derived quantities is +0.5. Powell (1958) has discussed this difference type of spurious correlation. He has pointed out that it tends to make the intraclass correlation between sisters more positive since a common measurement of time, i.e. the time of the maternal division, is involved in the estimation of the life-lengths and is subject to random experimental error.

**Correlation between successive generations**

A second kind of correlation has been sought experimentally, namely, that between the life-lengths of mothers and daughters. A formula applicable to this case similar to equation (9) is derived in Appendix, note 6. If \( c_1, c_2 \) and \( c_3 \) are the critical mass of the grandmother, mother and daughter at division and the coefficient of variations of the three relevant divisions are \( q_1, q_2 \) and \( q_3 \) respectively, the correlation of the ratios is given by

\[
\rho = \frac{-q_2^2}{\sqrt{(q_2^2 + q_h^2)} \sqrt{(q_2^2 + q_3^2)}}.
\]

This formula is based on the assumption that division is precisely even. If in addition, it is assumed that the dispersion of the \( g(c) \) distribution is the same for the three generations, the correlation coefficient becomes \(-0.5\). If \( q_h \) is not zero although \( g(c) \) is constant the relationship becomes (see Appendix, note 7)

\[
\rho = \frac{-q^2}{2q^2 + q_h^2}.
\]

Thus \( \rho \) in this case can vary from \(-0.5\) to zero, as \( q_h \) becomes large relative to \( q \). The need for a negative mother–daughter correlation is self evident in our model. Since we assume mass to increase monotonically in a deterministic manner, a premature division of a cell in one generation will necessarily require a fluctuation.
in later generations toward longer life-length. In the present model this recovery is immediate and automatic.

Observed values of the correlation coefficient of mother and daughter bacteria are either zero or slightly negative, although our recent experiments have yielded more negative correlation coefficients than those of other workers. Random experimental error in assessing the division time of the mother would introduce an error in the opposite direction in the life-length of mother and daughter and, for reasons similar to that outlined above, lead to a negative mother–daughter correlation. Combining data from cells with different growth rates will tend to make the correlation coefficient more positive. This can come about in several ways. First, data from different experiments may be improperly pooled. This is not serious because all workers in the field are cognizant of this possibility. Secondly, there may be microheterogeneity in the cultural conditions so that a particular region of the agar is endowed with more or less favourable conditions for bacterial growth. The critical test for this would be the consistency of the mean time for cell division to go through several cycles from one microclone to another; the variance should not be larger than the variance for a single cell division.

This test should also serve to measure fluctuations in growth rate due to physiological changes transmitted through a few generations or to permanent genetic changes. It should be noted that although our observed logarithmic rates of elongation of cells were quite constant from cell to cell and seemingly independent of the division size of the cell, they did fluctuate slightly and the altered rate did continue into the next generation. This phenomenon, regardless of source, will tend to make both the sister–sister intraclass correlation coefficient and the mother–daughter correlation coefficient more positive.

To compare actual measurements to our model, here again it is necessary to perform the transformation, antilog $\tau$, and recalculate the intraclass coefficient. When this is done, the numerical values are altered slightly. In one case drawn from data of Schaechter et al. (1962) the coefficient of the untransformed data was $-0.370$ and that of the transformed data, $-0.366$. On making the transformation the change in a number of other examples drawn from the data at hand was always less than 0.1. The magnitude of the correction becomes smaller as the magnitude of the correlation coefficient becomes larger. Therefore, we have felt that extensive transformation and recalculation was not necessary.

**DISCUSSION**

The model proposed here is qualitatively different from those presented by Rahn and by Kendall. Their models are purely statistical in that they presuppose that the elementary events controlling cell divisions are completely random, i.e. as unpredictable as the decay of a single radioactive atom. These models do however explain the fact that the interdivision time of a bacterium is characterized by life-span and not by a half-life. In addition, they interpret the reported magnitudes of the variance and skewness of the interdivision time distribution in terms of the sequential or simultaneous completion of a small number of such random events necessary for the cell to divide. These models predict that the sister–sister and mother–daughter life-length correlations be zero.
Statistics of cell division

The present model assumes four postulates which have been experimentally verified in our laboratory for a few enteric bacteria under particular conditions. Two of these postulates concern source of fluctuations in interdivision time. One of these assumes that quite small, and in some cases very small, random variations take place in determining the amount of cytoplasm in a cell at the time of division. The other postulate makes similar assumptions about the proportion of the cytoplasm distributed to each daughter cell. The sources of variation of these two distributions may arise from the statistics of a small number of fundamental molecular events or from manifold sets of accidental and historical reasons (such as might be imagined to determine the adult weight of a man). For the present, a distinction between these two sources of error is immaterial. This is in part because the variance of the two distributions, \( g(c) \) and \( h(p) \), is small for the cases we have investigated, so that it is difficult to establish the shape of the distribution by direct observation. The shape of the distribution is, of course, more indicative of the source of variation than is its coefficient of variation. More important, moderate deviation from normality for either distribution would not influence very much the observed parameters of cell division, e.g. the variance of the \( f(\tau) \) distribution, the skewness of the \( f(\tau) \) distribution, the sister–sister intraclass correlation coefficient, and the mother–daughter correlation coefficient. On the other hand, we believe that the magnitude of these four parameters is directly attributable to the magnitude of the coefficient of variation of the \( g(c) \) and \( h(p) \) distributions. Thus, we feel that the observable distributions and correlations of \( \tau \) values are dependent on the second moments and not particularly on the third moments of these two distributions.

Deviations from the predictions of the model must be understood in terms of experimental artifact or in the failure of one or more of the postulates. Only in the case of the experimental systems we have used (Schaechter et al. 1962) are enough different kinds of data available to allow testing of the model. In general, the enteric bacteria studied fit the model quite well. Unfortunately, at the present time, not all of the assumptions of the model have been checked experimentally. As noted below, the constancy of \( \lambda \) is a good first approximation, but possibly not exactly correct. Likewise, the assumption that \( q_h \) be zero is consistent with most observations, but will be difficult to verify experimentally. Finally, the assumption of the independence of the size attained at division with those at any other division await direct test, though it fits the facts and our expectation.

Next we may consider the generality of the model. The first postulate that protoplasmic growth is continuous and deterministic at the cellular level, appears to apply as well for Paramecium (Kimball et al. 1959). It does not appear to be true for a coccus (Mitchison, 1961), a fission yeast (Mitchison, 1957), or an amoeba (Prescott, 1955). Possibly its range of validity only extends to those organisms which have no protein turnover during their growth (Koch & Levy, 1955). As we have mentioned, it can not apply to Colpoda or to sea urchin eggs.

A better reason for employing this postulate is certainly the recent developments in the cellular biochemistry and control mechanisms of bacteria. On this basis it need not and, in fact, would not be expected to hold for higher forms, with more complex control mechanisms.

The even stronger part of the postulate needed to arrive at the conclusions presented here is that the substance of each cell grows with exactly the same growth
rate constant as that of every other cell. As mentioned above, this is very nearly true for the enteric bacteria we have studied. Persistent deviations contribute in a small degree to increasing the sister–sister and the mother–daughter correlation coefficient. They would also tend to broaden the \( f(\tau) \) distribution and skew the distribution positively.

The second postulate, that genotype and environment acting through physiological mechanisms regulate the average size of the cell at division, implies the existence of control mechanisms that somehow ‘sense’ the cell size or some closely related property. This would appear to be an excellent first approximation, since many cell components are synthesized continuously so that the amount of any one is proportional to the amount of any other substance or group of substances more specifically associated with cell or nuclear division. Note, also, that the amount of substance at division is proportional to the amount at an earlier time. Thus, lengths, for example, will serve to describe the statistics as well as quantity or quantities unknown but more intimately involved with division processes.

That cell division is strongly correlated with cell size is an obvious conclusion from the small variation in the size at division. This dependence is emphasized by the unpublished observations from our laboratory showing that the critical size varies monotonically during a shift of growth rate. The same conclusion follows logically from the published finding of Schaechter et al. (1958) that the average cell size is precisely dependent on growth rate, together with equations (3) and (5) of the present paper, which requires that the average size in balanced populations be proportional to the mean critical size.

The third postulate, that size of cell at division has only a small coefficient of variation, is valid for the bacteria we have studied. Adolph & Bayne-Jones (1932) observed that the adult size of Bacillus megaterium had a coefficient of variation of only 6.5%. Similar coefficients of variation in the range 6–13% have been quoted for certain coliform organisms (Bayne-Jones & Adolph, 1933) and for Tetrahymena pyriformis GL by Scherbaum & Rasch (1957). In other instances also the coefficient of variation may be very small. For example, this is evident from the fact that the coccus Lampropedia hyalina grows in two dimensional sheets which are almost perfectly rectangular (Pringsheim, 1955). In some cocci which grow in chains, all the cells of a chain appear to be in the same stage of cell division. It is not known whether such instances represent precise timing of cell division or an exceptionally high sister–sister correlation as was reported for a yeast by Burns (1956) who found that in yeast cells of certain ploidies and grown at certain temperatures, there was complete positive correlation between the time of division of a cell and the next division of the mother cell from which it was derived. Since cells not attached to each other were not dividing in phase, it seems likely that this synchronization was mediated by chemical or physical interaction. On the other hand, there are some bacteria which tend to grow in filaments which break down almost at random. For these organisms the theory would not apply, since the coefficient of variation would not be small as required by the mathematical development presented here, and the second and fourth postulates would not be obeyed.

Not enough is known about the equality of division of the cytoplasm in bacteria; it has been discussed in reference to yeast (Spiegelman, De Lorenzo & Campbell 1951). From the \( f(\tau) \) distribution, the sister–sister correlation, and mother–
Statistics of cell division

daughter correlation, we infer that in the cases studied (Schaechter et al. 1962) cell division must divide the cytoplasm quite evenly.

The present model assumes nothing about ‘triggers’ or a period of time in which cytokinesis is irreversibly preordained, as is the case with sea urchin eggs (Swann, 1954). However, in our proposed model one cell division determined the next. It certainly does not preordain the time of the next division event. There might be more specific triggers, but as such they have not yet been found in bacteria. Of course, finding this type of mechanism would not invalidate the considerations presented here. They would lead to a refined model in which the prime variable(s) had even smaller dispersion than the size at division. The idea that division size is well controlled may be a fundamental biological notion. A search for the mechanism controlling it should be a fruitful area of investigation. Obvious approaches, such as a study of how the critical size changes with the environment are under way and will be extended. The search for genetic, chemical, and environmental agents which affect the control of cell division, rather than the process itself, may be fruitful.

We wish to thank the following persons for helpful discussions or comments: A. Campbell, R. E. Ecker, R. Hoffman, H. Kubitschek, O. Seherbaum, and J. P. Williamson. Support for the experimental work in the authors’ laboratory is acknowledged in the accompanying experimental paper (Schaechter, et al. 1962).

APPENDIX

Note 1. It appears worthwhile to derive this distribution in a different manner. To do so, we note that the number of cells now of mass \( m \) is related to the number of divisions taking place at a time when these cells had mass \( \frac{1}{2}m \). If the number of cells at that time \(-t\), was \( N_0 \), it follows from our first assumption that

\[
\frac{N_0}{N} = e^{-\lambda t} = \frac{\bar{c}}{2m} \quad \text{or} \quad N_0 = \frac{\bar{c}N}{2m}.
\]

The growth rate was then

\[
\lambda N_0 = \frac{\lambda \bar{c}N}{2m} = \frac{dN}{dt}.
\]

The rate of production of new cells is twice the growth rate, \( dN/dt \), since a division produces two new cells, but only one additional cell. Therefore, we have for \( dn/dt \), the rate of production of new cells

\[
\frac{dn}{dt} = \frac{\lambda \bar{c}N}{m}.
\]

On eliminating \( dt \) by use of the expression \( dm/dt = \lambda m \), we obtain

\[
dn = \frac{\lambda \bar{c}N dt}{m} = \frac{\lambda \bar{c}N dm}{m \lambda m} = \frac{N \bar{c}dm}{m^2}.
\]

and consequently

\[
\theta(m) = \frac{1}{N \bar{c}} \frac{dn}{dm} = \frac{\bar{c}}{m^2}.
\]
448

A. L. KOCH AND M. SCHAECHTER

Note 2. The following derivation is patterned after that given by Powell (1956) for the derivation of the age distribution. Define

\[ G(m) = \int_m^\infty g(c)dc. \]

Then an organism of mass \( m \) has the following chance of achieving a mass of \( m(1+k) \) without dividing,

\[ \frac{G(m(1+k))}{G(m)}. \]

Hence, of the original \( N\theta(m)dm \) organisms, \( \frac{N\theta(m)G(m(1+k))}{G(m)} \) organisms are still present. At a later time \( t = (1/\lambda) \ln (1+k) \) the number of organisms have grown to \( Ne^{\lambda t} \) and the survivors form a fraction of the whole, equal to

\[ \frac{\theta(m)G(m(1+k))e^{-\lambda t}}{G(m)} \]

This must hold all values of \( k \) and the corresponding values of \( t \).

For small \( k \) and \( t \),

\[ G(m(1+k)) = G(m) + \frac{dG(m)}{dm} (mk), \]

\[ \theta(m(1+k)) = \theta(m) + \frac{d\theta(m)}{dm} (mk), \]

and

\[ e^{-\lambda t} = 1 - \lambda t = 1 - k. \]

Therefore

\[ \frac{(G(m) + mk(dG(m)/dm))(1-k)}{G(m)} = \frac{(\theta(m) + mk(d\theta(m)/dm))(1+k)}{\theta(m)}. \]

Expanding and neglecting higher powers of \( k \) we obtain

\[ \frac{mk}{G(m)} \frac{dG(m)}{dm} - k = \frac{mk}{\theta(m)} \frac{d\theta(m)}{dm} + k, \]

or

\[ \frac{d \ln G(m)}{d \ln m} - 2 = \frac{d \ln \theta}{d \ln m}. \]

Multiplying by \( d \ln m \) and integrating, we obtain

\[ \ln G(m) - 2 \ln m = \ln \theta(m) - C \]

\[ \theta(m) = \frac{CG(m)}{m^2}, \]

\[ \theta(m) = \frac{C}{m^2} \left[ 1 - \int_0^m g(c)dc \right]. \]

Since the first integral in equation (5) becomes unity for values of \( m \) somewhat larger than \( m = \frac{1}{2} \varepsilon \), equation (5) is thus derived in part. This method can not be extended to yield the lower end of the distribution. We shall give here a different type of derivation that allows us to calculate the lower as well as the upper ends of the \( \theta(m) \) distribution, under conditions where the coefficient of variation of the \( g(c) \) distribution is small. We have not yet solved the general case.

From our definition, the number of cells of mass between \( m \) and \( m + dm \) is \( N\theta(m)dm \), when \( N \) is the total number of organisms present. At an earlier time
these cells had mass \( m' = m - \Delta m \). The total number of cells in the population then was (see note 1) \( N(m'/m) = N(m - \Delta m)/m \), and the interval of mass was

\[
dm' = \frac{m'}{m} \, dm = \frac{(m - \Delta m) \, dm}{m}.
\]

Therefore, these cells contributed to the present class \( m \) by an amount equal to

\[
N\left(\frac{m - \Delta m}{m}\right) \, \theta(m - \Delta m) \left(\frac{m - \Delta m}{m}\right) \, dm
\]

or

\[
N\left(\frac{m - \Delta m}{m}\right)^2 \left(\theta(m) - \frac{d\theta(m)}{dm} \, \Delta m\right) \, dm.
\]

Expanding, and neglecting second and higher powers of \( \Delta \) we obtain

\[
N\left(\theta(m) - \frac{2\Delta m}{m} \frac{d\theta(m)}{dm} \, \Delta m\right) \, dm.
\]

Therefore, the net gain or loss to the class of mass \( m \) to \( m + dm \) from the corresponding class \( m' \) to \( m' + dm' \) will be

\[
N \Delta m \, dm \left(\frac{-2\theta(m)}{m} - \frac{d\theta(m)}{dm}\right).
\]

In balanced growth this need be counterbalanced by the attrition due to cell division in the same interval

\[
\frac{N \theta(m) \, dm \, g(m) \, \Delta m}{\int_m^\infty g(c) \, dc},
\]

and to augmentation by division of cells of double size

\[
-2N \theta(2m) \, d2m \, g(2m) \, \Delta 2m \, g(2m) \, dc.
\]

Equating these sources and cancelling out \( N \Delta m \, dm \) we obtain

\[
-2 \frac{\theta(m)}{m} - \frac{d\theta(m)}{dm} = \theta(m) \frac{g(m)}{\int_m^\infty g(c) \, dc} - 8\theta(2m) \frac{g(2m)}{\int_{2m}^\infty g(c) \, dc}.
\]

It is readily noted if the right side is set to equal to zero, equation (3) may be obtained. If just the first term on the right is retained (high end of the distribution), the solution may be obtained by defining \( G(m) = \int_m^\infty g(c) \, dc \) and noting that \( dG(m)/dm = -g(m) \). The the equation becomes

\[
2 \frac{\theta(m)}{m} + \frac{d\theta(m)}{dm} = \theta(m) \frac{dG(m)}{G \, dm},
\]

\[
2 \frac{dm}{m} + \frac{d\theta(m)}{\theta(m)} = \frac{dG(m)}{G(m)}.
\]

Integration yields

\[
2 \ln m + \ln \theta(m) = \ln G(m) + \ln C,
\]

or

\[
m^2 \theta(m) = CG(m),
\]

\[
\theta(m) = \frac{C}{m^2} \int_m^\infty g(c) \, dc = \frac{C}{m^2} \left[ 1 - \int_0^m g(c) \, dc \right],
\]
and is thus identical to the results derived above. The lower end of the distribution is derived by substituting this value into the general equation and neglecting the first term on the right hand side of equation. We then obtain

\[ 2 \theta(m) \frac{d \theta(m)}{dm} = \frac{8 \theta(2m) g(2m)}{\int_{2m}^{\infty} g(c) dc} = \frac{8 C \int_{2m}^{\infty} g(c) dc}{4m^2 \int_{2m}^{\infty} g(c) dc} = \frac{2C}{m^2 g(2m)}, \]

which on clearing yields

\[ 2m \theta(m) \frac{d \theta(m)}{dm} + m^2 \theta(m) = 2C g(2m) \frac{d m}{dm}, \]

or

\[ m^2 \theta(m) = 2C \int_0^m g(c) dc, \]

which on integration yields

\[ m^2 \theta(m) = C \int_0^m g(c) dc. \]

Therefore, the entire distribution is given equation (5) where \( C \) will have slightly different value than for either end alone, but may be determined from the relationship that

\[ \int_0^m \theta(m) = 1. \]

\( C \) approaches \( \delta \) as the coefficient of variation of \( g(c) \) approaches zero.

Note that the effects of imprecise division of cell constituents between daughter cells can readily be taken into account by simply employing the correct \( g(c) \) distribution for the second integral in equation (5) and an effective broader distribution \( g'(c) \) for the first integral. This broader distribution then takes into account both the variability of the critical size as well as the cell distribution variability. Further theoretical work along these lines is in progress.

Note 3. If the variation in the numerator and denominator are small, we may write

\[ \frac{\delta_2}{\delta_1} = \left( \frac{1 \pm y}{1 \pm x} \right) = \frac{\delta_2}{\delta_1} \left( 1 \pm y \pm x \ldots \right), \]

where \( y \) and \( x \) are the fractional deviations from the mean. We then simply note that the central limit theorem requires that this ratio will tend towards normality.

Note 4. Experimentally obtained \( \tau \) distributions obtained previously fit well a number of possible distribution curves, such as the Pearson type III, the log normal, and recently the distribution of the reciprocal of a normally distributed variable with relatively small coefficient of variation (see Kubitschek, 1962). From the numerical calculation presented above it is clear that, over certain ranges of variability, the logarithmic normal curve and the reciprocal are equivalent transformations. Kubitschek’s ideas would then be equivalent to the assertion that \( \mu \) for each cell is normally distributed.

Note 5. Instead of the positive correlation found with bacteria, a negative sister-sister correlation is found with Tetrahymena geleii (Prescott, 1959). This suggests that extreme dissymmetry of some cell substance takes place at division of this organism.
Note 6. Both equations (9) and (11) can be derived as limiting forms for the general solution given by Pearson (1897). We shall derive the solution for the mother–daughter case, and refer the reader to Yule & Kendall (1944; p. 300) for the sister–sister case. The proofs for the two cases are, of course, very similar. Call $R_{12} = c_2/c_1$ and $R_{23} = c_3/c_2$ the ratios whose correlation coefficient $\rho$ we desire. The standard deviation and mean of these ratios are readily derived from equation (8) on the assumption that the errors are small; these are

$$
\frac{\bar{c}_2}{c_1} = \frac{\bar{c}_2}{c_1} (1 + q_1^2); \quad \frac{\bar{c}_3}{c_2} = \frac{\bar{c}_3}{c_2} (1 + q_2^2); \quad s_{12} = \frac{\bar{c}_3}{c_1} \sqrt{(q_1^2 + q_2^2)};
$$

and

$$
s_{23} = \frac{\bar{c}_3}{c_2} \sqrt{(q_2^2 + q_3^2)}.
$$

The correlation coefficient is by definition

$$
\rho = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{c_2}{c_1} \bar{c}_2 \bar{c}_3 \bar{c}_3 \right) - \frac{1}{n} \sum_{i=1}^{n} \left( \frac{c_2}{c_1} \bar{c}_2 \bar{c}_3 \bar{c}_3 \right).
$$

The numerator may be expanded to yield

$$
\frac{1}{n} \sum_{i=1}^{n} c_2 - \left( \frac{1}{n} \sum_{i=1}^{n} c_1 \right) \frac{\bar{c}_2}{c_1} - \left( \frac{1}{n} \sum_{i=1}^{n} c_2 \right) \frac{\bar{c}_2}{c_1} + \bar{c}_2 \bar{c}_3 \bar{c}_3.
$$

Note that the last three terms are identical except for sign. Cancellation yields

$$
\frac{1}{n} \sum_{i=1}^{n} c_2 - \frac{1}{n} \sum_{i=1}^{n} c_1 \frac{\bar{c}_2}{c_1}.
$$

Substituting $c_1 = \bar{c}_1 (1 + xq_1)$ and $c_3 = \bar{c}_3 (1 + xq_3)$, where $x$ is the normal deviate, in the first term of this expression yields for this term

$$
\frac{1}{n} \sum_{i=1}^{n} c_2 = \frac{1}{n} \sum_{i=1}^{n} (1 + xq_3) = \frac{1}{n} \sum_{i=1}^{n} 1 + xq_3.
$$

Expanding the argument and neglecting higher powers than second, yields:

$$
\frac{1}{n} \sum_{i=1}^{n} 1 + xq_3 = \frac{xq_3}{\bar{c}_1}.
$$

The second term may be written

$$
\frac{\bar{c}_2}{c_1} \bar{c}_3 \bar{c}_3 = \frac{\bar{c}_2}{c_1} (1 + q_1^2) \frac{\bar{c}_3}{c_2} (1 + q_2^2) = \frac{\bar{c}_3}{c_1} (1 + q_1^2 + q_2^2 + q_3^2).
$$

Therefore, the numerator becomes (neglecting $q_1^2 q_2^2$)

$$
\frac{\bar{c}_3}{c_1} q_2^2 - \frac{\bar{c}_3}{c_1} (1 + q_1^2 + q_2^2) = \frac{\bar{c}_3}{c_1} q_2^2.
$$

And the correlation coefficient is then given by equation (11).

Note 7. As in the previous footnote we shall derive the equation for the mother–daughter correlation, and note that the sister–sister equation may be derived similarly. It is simpler to start with the general formulation of Pearson (1897). As before we may designate the ratios $R_{12} = c_2/c_1$ and $R_{23} = c_3/c_2$. It will be necessary to distinguish between $c_2$ and $c_3$ because the coefficients of variation of the masses at division and after division in this case are not equal, and the correlation between
the two, \( \rho_{22'} \), is not unity as in the previous case. Pearson’s equations with the condition that the other variables are uncorrelated may be written

\[
\rho = \frac{-\rho_{22'} q_2 q_{2'}}{\sqrt{(q_2^2 + q_{2'}^2)} \sqrt{(q_{2'}^2 + q_2^2)}}
\]

which is, of course, extremely similar to the equation 11.

The assumptions for this case, give immediately

\[
q_2^2 = q_{2'}^2 = q^2
\]

and

\[
q_1^2 = q_{2'}^2 = q^2 + q_2^2.
\]

The value of \( \rho_{22'} \) is readily computed from the correlation of the product \( cp \), with one of the variables, \( c \). \( p \) was defined above as the proportion of protoplasm appearing in one daughter cell.

As in the note 6 we write

\[
\rho_{22'} = \rho c, cp = \frac{\sum c (cp - \bar{cp}) (c - \bar{c})}{\sum c^2 p - n \bar{cp} \bar{c}}
\]

\[
= \frac{\sum c^2 p - n \bar{cp} \bar{c}}{\sum c^2 p - n \bar{cp} \bar{c}}
\]

\[
= \frac{\bar{cp} \bar{c} \left( \sum (1 + qx^2) (1 + q_h x') \right) - n \bar{cp} \bar{c}}{\sum (1 + qx^2) (1 + q_h x')}
\]

\[
= \frac{\bar{cp} \bar{c} q^2}{q_{2'} \cdot q} = \frac{q^2}{\sqrt{(q^2 + q_h^2)}}.
\]

Therefore, we have for the correlation of mass ratio, equation (12)

\[
\rho = \frac{q}{\sqrt{(q^2 + q_h^2)}} \frac{q}{\sqrt{(q^2 + q_h^2)}} = \frac{q^2}{2q^2 + q_h^2}.
\]

REFERENCES


Statistics of cell division


PEARSON, K. (1897). On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. roy. Soc.* 60, 489.


G. Microb. xxix


