THE DISTRIBUTION OF THE NUMBERS OF MUTANTS IN BACTERIAL POPULATIONS

By D. E. LEA,* *Department of Radiotherapeutics and Strangeways Laboratory, Cambridge* and C. A. COULSON, *Wheatstone Physics Laboratory, King's College, London*

Introduction

Luria & Delbruck (1943) have shown that if a culture of some hundreds or thousands of millions of *Bacterium coli*, grown from a single cell, is plated out on a nutrient medium impregnated with a bacteriophage to which the strain of *coli* is sensitive, the vast majority of the bacteria are lysed, but a few give rise to colonies. These colonies contain only bacteria resistant to the bacteriophage, and give rise only to resistant bacteria on further subcultivation. Evidently hereditary variations or mutations can occur in bacteria. Numerous other examples are known of mutations in bacteria, affecting fermentation reactions (e.g. Lewis, 1934), resistance to chemicals (e.g. Stewart, 1947), to antibiotics (e.g. Demerec, 1945), or to radiation (Witkin, 1946).

The demonstration of phage-resistant mutants necessarily involves the exposing of the bacteria to the phage, and it is not immediately obvious whether the mutation to phage resistance occurs spontaneously during the growth of the culture, and is merely made apparent by subsequently testing with phage, or whether the mutation is induced by the phage and does not occur until the bacteria are brought into contact with phage. Most experiments on bacterial variation have left open the two alternatives of spontaneous mutation on the one hand, and induced mutation or adaptation on the other, and the interpretation adopted has usually been determined by the previous training of the individual worker rather than by any compelling evidence provided by the experiments.

Luria and Delbruck, however, in their paper, described a method by which a decision between the two alternative explanations may be reached, and concluded that the acquirement of resistance to phage is a spontaneous mutation which occurs during the growth of the culture and prior to its treatment with phage. Demerec (1945) and Witkin (1946) have applied the same method to mutants resistant to penicillin and to X-rays respectively, and have concluded that these changes also are spontaneous mutations occurring independently of the penicillin or of the radiation respectively.

The principle of Luria and Delbruck's test is as follows. A culture of (say) 10^9 bacteria is divided into (say) ten equal portions which are separately tested for phage-resistant organisms by plating out on a phage-impregnated medium. A small number is found in each of the ten portions, and the numbers are found experimentally to be distributed with a variance approximately equal to the mean. This result is not surprising on either hypothesis. On the spontaneous mutation theory, we suppose that mutations to phage resistance occurred from time to time during the growth of the culture. All the bacteria produced by subsequent divisions of a mutant bacterium were similarly phage resistant. Thus the culture of 10^9 bacteria contained a certain number of phage-resistant bacteria,

* [Note by C. A. C. A few days before Dr Lea's untimely death in June 1947, the manuscript and the calculations reported here had just been completed. It was Dr Lea's intention to make further experiments more suitable to a test of the theory outlined in this paper. These experiments cannot now be made, but it has been thought wise to publish the theory and numerical tables because of their value to other investigators.]

being either bacteria which had recently undergone mutation, or bacteria derived from the division of mutants which arose earlier in the growth of the culture. When the culture was divided into ten equal portions the phage-resistant organisms were distributed at random between the ten portions. We may expect, therefore, the numbers in the different portions to fall in a multinomial distribution with variance nearly equal to the mean.

On the adaptation or induced mutation theory, it is supposed that no phage-resistant bacteria arose during the growth of the culture. The ten portions, at the time of plating out, each contained 10^8 normal bacteria and no resistant bacteria. On being brought into contact with the phage most were lysed, but a few were able to adapt themselves to the phage (or the phage-induced mutations in them). The probability of this process is very small, but was presumably the same for all the bacteria. On this theory, therefore, we expect the number of resistant colonies on the ten parallel plates to be distributed in a Poisson distribution with variance equal to the mean. Either theory is thus capable of accounting for the experimental variance, and this experiment alone does not make possible a decision between the two theories.

A second experiment is now made in which (say) ten cultures, of (say) 10^8 bacteria are tested for phage-resistant organisms. On the adaptation or induced-mutation theory this experiment is not essentially different from the preceding one, and we again expect the numbers of phage-resistant colonies on the ten test plates to be distributed in a Poisson distribution with variance equal to the mean. For, on this theory, the phage-resistant mutants do not appear until the bacteria are plated out on the phage-impregnated medium, and there can be no relevant difference between a culture of 10^9 bacteria divided into ten equal portions, and ten separately grown cultures of 10^8 bacteria.

In practice a very different result is obtained: the distribution obtained is much wider than in the former experiment, and has a variance many times—perhaps fifty times—the mean.

On the spontaneous mutation hypothesis a very wide distribution of the number of phage-resistant bacteria in parallel cultures is to be expected. The reason is that not only do the parallel cultures differ in the numbers of mutations which have occurred, but also, and much more importantly, they differ in the stages at which the mutations occurred. If a mutation occurs towards the end of the growth of a culture, it will give rise to one phage-resistant organism, but if it occurs early in the growth, say when the culture is only one-hundredth of its final size, it will give rise to a large number of phage-resistant organisms. Thus even in cultures in which equal numbers of mutations have occurred, the numbers of phage-resistant organisms will usually be widely different.

It is evident, therefore, that the hypothesis that spontaneous mutation to phage resistance occurs during the growth of the culture before it is brought into contact with the phage is in qualitative agreement with the experimental result, while the alternative hypothesis of mutation induced by the phage, or adaptation of the bacterium to the phage, is not. Luria and Delbruck's method thus provides, for the first time, a clear means of distinguishing between the two hypotheses.

As left by Luria and Delbruck, the method is a qualitative one, since they do not derive the shape of the distribution to be expected on the spontaneous mutation theory. They do derive expressions for the mean and variance of the distribution, but as they point out, on account of the extreme skewness of the distribution, the mean and variance are very

inefficient statistics for estimating the parameters of the distribution from experimental results, or for testing the agreement of experiment and theory.

The purpose of the present paper is to extend Luria and Delbruck's method by calculating the form of the distribution of numbers of mutants in parallel cultures to be expected on the spontaneous mutation theory, so making the test of the applicability of the spontaneous mutation theory a quantitative test. Statistically efficient methods of deducing the mutation rate from experimental observations are also discussed.

The distribution

First method

During the active growth of a culture, the number of organisms increases as an exponential function of the time, and may be represented as

$$n = e^{\beta t}, \tag{1}$$

there being one organism at time $t = 0$.

Thus
$$dn = \beta n \, dt. \tag{2}$$

If α is the mutation rate, defined by the relation that $\alpha \, dt$ is the probability that an individual phage-sensitive organism shall undergo mutation to phage resistance in time dt, $n\alpha \, dt$ is the mean number of mutations which occur in time dt. (Strictly, since n is the total number of organisms, we should subtract from n in this formula the number of resistant organisms, but in practice the number of mutant organisms in a culture is a minute fraction of the total number.) Hence the mean number (m) of mutations which will have occurred in the culture by the time it has grown to size n at time t is

$$\int_0^t n\alpha \, dt = \frac{\alpha}{\beta} \int_1^n dn = \frac{\alpha}{\beta}(n-1).$$

Since at all relevant times n much exceeds unity, we may write

$$m = \frac{\alpha}{\beta} n \tag{3}$$

for the mean number of mutations in the culture by the time it has attained size n. The mean number of mutations which occur while the culture grows from n_1 to n_2 organisms is evidently

$$m\left(\frac{n_2 - n_1}{n}\right). \tag{4}$$

If it should happen that the mutation rate α and the growth rate β are equally affected by factors such as nutritional conditions and density of population which affect β, then the mutation rate *per generation*, though not *per unit time*, will be independent of these factors, and α/β, which may be regarded as the probability of mutation per division, will be constant even though α and β are not. Under these conditions equations (3) and (4) can be derived without the assumption of exponential growth.

We must distinguish between the number of *mutations* which have occurred, and the number of *mutants*, the latter being derived not only by mutation of the normal bacteria but also by division of bacteria which have suffered mutation earlier. If r ($r \geqslant m$) is the number of mutants in the culture at a given time, and if we assume that the division rate of the mutant is the same as that of the normal bacteria, then $r\beta \, dt = r \, dn/n$ is the probability

that one of the mutants shall divide in time dt. Similarly, $(r-1)\,dn/n$ is the probability that one of the mutants shall divide in time dt in a culture containing $(r-1)$ mutants.

Let p_r (a function of n) be the probability that a culture of n bacteria grown from a single bacterium at $t=0$ shall have r mutants (i.e. of a large number of cultures of n bacteria, a proportion p_r of cultures will have r mutants).

Consider the proportion $p_r+\frac{dp_r}{dn}\,dn$ of cultures which, at time $t+dt$, when the culture size is $n+dn$, have r mutants. These will be derived from:

(*a*) cultures which, at time t, had $r-1$ mutants and in which a mutation occurred in the interval dt;

(*b*) cultures which, at time t, had $r-1$ mutants and in which a mutant divided in the interval dt;

(*c*) cultures which, at time t, had r mutants and in which neither mutation nor division of a mutant occurred in the interval dt;

providing that the interval dt is small enough. For $(\alpha/\beta)\,dn$ and $r\,dn/n$ to be much less than unity, the possibility of more than one of the rare events mutation and division of a mutant in the interval dt can be neglected. We see, therefore, that

$$p_r+\frac{dp_r}{dn}\,dn=p_{r-1}\left\{\frac{\alpha}{\beta}\,dn+(r-1)\frac{dn}{n}\right\}+p_r\left\{1-\frac{\alpha}{\beta}\,dn-r\frac{dn}{n}\right\},$$

so that

$$\frac{dp_r}{dn}+\frac{\alpha}{\beta}p_r+\frac{r}{n}p_r=p_{r-1}\left(\frac{\alpha}{\beta}+\frac{r-1}{n}\right).$$

Making, from (3), the substitution $m=(\alpha/\beta)\,n$, we have

$$\frac{dp_r}{dm}+p_r+\frac{r}{m}p_r=p_{r-1}\left(1+\frac{r-1}{m}\right). \tag{5}$$

Multiplying by the integrating factor e^m we have

$$\frac{dq_r}{dm}+\frac{r}{m}q_r=q_{r-1}\left(1+\frac{r-1}{m}\right), \quad \text{where } q_r=e^m p_r. \tag{6}$$

Now m is the mean number of mutations which have occurred in a culture by the time it contains n bacteria. Therefore e^{-m}, the first term of the Poisson distribution, is the probability p_0 that no mutation shall have occurred. Thus $q_0=e^m p_0=1$ for all values of m. Evidently initially, when $m=0$, $q_r=0$ for all $r>0$.

Starting from $q_0=1$ we can calculate q_1, q_2, q_3, etc. in succession from the differential equation (6). Thus:

$$\frac{dq_1}{dm}+\frac{1}{m}q_1=1, \qquad \text{whence } q_1=\tfrac{1}{2}m;$$

$$\frac{dq_2}{dm}+\frac{2}{m}q_2=\tfrac{1}{2}m\left(1+\frac{1}{m}\right), \qquad \text{whence } q_2=\tfrac{1}{6}m+\tfrac{1}{8}m^2;$$

$$\frac{dq_3}{dm}+\frac{3}{m}q_3=(\tfrac{1}{6}m+\tfrac{1}{8}m^2)\left(1+\frac{2}{m}\right), \qquad \text{whence } q_3=\tfrac{1}{12}m+\tfrac{1}{12}m^2+\tfrac{1}{48}m^3.$$

Evidently q_r is a polynomial in m, with powers ranging from 1 to r.

Writing

$$q_r=\sum_{j=1}^{r}C_{j,r}\frac{m^j}{j!}=C_{1,r}\,m+C_{2,r}\frac{m^2}{2!}+\ldots+C_{r,r}\frac{m^r}{r!}, \tag{7}$$

we have
$$\frac{dq_r}{dm}=\sum_{j=1}^{r} C_{j,r}\frac{m^{j-1}}{(j-1)!}. \tag{8}$$

Inserting (7) and (8) in (6), and equating coefficients of $m^{j-1}/j!$, we have
$$(j+r)\,C_{j,r}=jC_{j-1,r-1}+(r-1)\,C_{j,r-1}. \tag{9}$$
With this recurrence relation a table of $C_{j,r}$ may be drawn up. Such a table, for values of $r\leqslant 10$, is given in the Appendix.

Evidently
$$C_{1,r}=\frac{1}{r\,(r+1)};\quad C_{r,r}=2^{-r}. \tag{10}$$

From (6) and (7), $p_0=e^{-m}$, and for $r\geqslant 1$
$$p_r=\sum_{j=1}^{r} C_{j,r}\left(e^{-m}\frac{m^j}{j!}\right)=C_{1,r}\,(e^{-m}m)+C_{2,r}\left(e^{-m}\frac{m^2}{2!}\right)+\ldots+C_{r,r}\left(e^{-m}\frac{m^r}{r!}\right). \tag{11}$$

A generating function for p_r

Define a function
$$f(x,m)=q_0+q_1x+q_2x^2+\ldots=\sum_{r=0}^{\infty} q_rx^r. \tag{12}$$

We have
$$\frac{\partial f}{\partial x}=\Sigma rx^{r-1}q_r \quad\text{and}\quad \frac{\partial f}{\partial m}=\Sigma x^r\frac{dq_r}{dm}. \tag{13}$$

Multiplying equation (6) by x^r and summing for all r we have
$$\Sigma x^r\frac{dq_r}{dm}+\frac{x}{m}\Sigma rx^{r-1}q_r=x\,\Sigma x^{r-1}q_{r-1}+\frac{x^2}{m}\Sigma(r-1)\,x^{r-2}q_{r-1},$$

or, using (13),
$$\frac{\partial f}{\partial m}+\frac{x}{m}\frac{\partial f}{\partial x}=xf+\frac{x^2}{m}\frac{\partial f}{\partial x},$$

whence
$$\frac{\partial\phi}{\partial m}+\frac{x}{m}(1-x)\frac{\partial\phi}{\partial x}=x, \tag{14}$$

where* $\phi=\log f$. This equation is satisfied by
$$\phi(x,m)=m\psi(x),$$

providing
$$\psi+x(1-x)\,\psi'=x,$$

i.e.
$$\psi'+\frac{\psi}{x(1-x)}=\frac{1}{1-x}.$$

Multiplying by $x/(1-x)$ and integrating
$$\frac{x}{1-x}\psi=\frac{1}{1-x}+\log(1-x)-1,$$

the integration constant -1 being introduced since when $x=0$, $f=q_0=1$, so that $\phi=0$ and so $\psi=0$. Thus
$$\psi=1+\frac{1-x}{x}\log(1-x)=\frac{x}{1.2}+\frac{x^2}{2.3}+\frac{x^3}{3.4}+\ldots,$$

whence
$$f=e^{m\psi}=e^m\,(1-x)^{m\,(1-x)/x},$$

so that $p_r=e^{-m}q_r$ is the coefficient of x^r in the expansion in ascending powers of x of
$$(1-x)^{m\,(1-x)/x} \quad\text{or of}\quad e^{-m}\exp\left[m\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)\right], \tag{15}$$

* log means natural logarithm to base $e=2{\cdot}718\ldots$ throughout this paper.

i.e. of $$e^{-m}+\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)e^{-m}\frac{m}{1!}+\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)^2 e^{-m}\frac{m^2}{2!}+\ldots. \quad (16)$$

Comparing with (11), it is evident that $C_{j,r}$ is the coefficient of x^r in the expansion of

$$\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)^j. \quad (17)$$

It will be shown later that $C_{j,r}$ is the probability that a culture in which exactly j mutations have occurred shall contain r mutants. If we define

$$D_{j,r}=\sum_{r=j}^{r} C_{j,r},$$

$D_{j,r}$ is the probability that a culture in which exactly j mutations have occurred shall have $\leqslant r$ mutants. Summing equation (9) over all r between j and r leads to the following recurrence relation between the $D_{j,r}$,

$$(r+j)\,C_{j,r}=j\,(D_{j-1,r-1}-D_{j,r-1}), \quad (18)$$

or

$$(r+j)\,D_{j,r}=rD_{j,r-1}+jD_{j-1,r-1}. \quad (19)$$

Equation (18) is useful as an arithmetical check during the computation of the $C_{j,r}$ by means of equation (9).

It is sometimes convenient to discuss P_r defined as

$$P_r=\sum_{r=0}^{r} p_r, \quad (20)$$

P_r being the probability that a culture shall contain 0, 1, 2, ..., or r mutants, i.e. any number of mutants up to r. $P_0=e^{-m}$; $P_r=1$ for all r at $m=0$. The following equations can be readily deduced by summing (5) and (11) over all r between 0 and r:

$$\frac{dP_r}{dm}+P_r\left(1+\frac{r}{m}\right)=P_{r-1}\left(1+\frac{r}{m}\right), \quad (21)$$

$$P_r=e^{-m}+\sum_{j=1}^{r} D_{j,r}\left(e^{-m}\frac{m^j}{j!}\right). \quad (22)$$

Second method

An alternative method of calculating p_r, the probability of a culture having r mutants, is instructive. A mutant appearing by mutation any time after the culture has passed the size $\frac{1}{2}n$ will not have time to divide by the time the culture size reaches n, and will therefore contribute 1 to the final complement of mutants. A mutant appearing by mutation during the period in which the culture grows from $\frac{1}{4}n$ to $\frac{1}{2}n$ will have time to divide once only and will therefore contribute 2 mutants to the final complement. A mutant appearing by mutation in the period in which the culture grows from $\frac{1}{8}n$ to $\frac{1}{4}n$ will have time to divide twice only and will contribute 4 mutants to the final complement, and so on.

Confining attention for the moment to all those cultures in which exactly one mutation occurs during the growth of the culture from one organism to n organisms, in one-half of the cultures the mutation will occur while the culture is growing from $\frac{1}{2}n$ to n, in one-quarter of the cultures the mutation will occur while the culture is growing from $\frac{1}{4}n$ to $\frac{1}{2}n$, and so on (compare equation (4)). Or, in other words, in one-half of these cultures the mutation occurs at such a time that it gives rise finally to 1 mutant, in one-quarter of the cultures

the mutation occurs at such a time that it gives rise finally to 2 mutants, in one-eighth at such a time that it gives rise to 4 mutants, and so on.

According to this argument, a mutation is necessarily represented, by the time the culture has reached size n, by a clone of 1 or 2 or 4 or 8, etc., organisms, there being no intermediate numbers. This would be so if divisions in a clone were synchronous. It is probably true that clones of 3, 5 or 7 cells will be less common than clones of 2, 4 or 8 cells (cp. Adolph & Bayne-Jones, 1932), but rather than make the extreme assumption that only integral powers of 2 are to be considered it is probably preferable to neglect this fact and to assume that the frequency of clones of different sizes is a smooth function of clone size.

A clone which, by time t, contains ν mutants will have originated when the number of bacteria in the culture was about n/ν. We thus replace the subdivision of the growth of the culture into generations by subdivision into intervals in which the population increased from $\frac{1}{2}n$ to n, from $\frac{1}{3}n$ to $\frac{1}{2}n$, from $\frac{1}{4}n$ to $\frac{1}{3}n$, and so on, and suppose that a mutation which occurred while the population increased from $n/(\nu+1)$ to n/ν is, by the time the population has grown to n, represented by a clone of ν mutants. Now, of those cultures in which exactly one mutation has occurred, the proportion in which the mutation occurred while the culture grew from $n/(\nu+1)$ to n/ν is $\left(\frac{1}{\nu}-\frac{1}{\nu+1}\right)=\frac{1}{\nu(\nu+1)}$ (cp. equation (4)). This may be represented as the coefficient of x^ν in the generating function

$$\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots+\frac{x^\nu}{\nu(\nu+1)}+\ldots. \tag{23}$$

Considering now all those cultures in which exactly j independent mutations occurred, the fraction of cultures in which the final number of mutants is ν is, from (23), evidently the coefficient of x^ν in the expansion of

$$\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\frac{x^3}{3.4}+\ldots\right)^j.$$

Now if m is the mean number of mutations per culture, the proportion of cultures in which exactly j mutations occurs is $e^{-m}\frac{m^j}{j!}$. Thus the proportion of all cultures in which the final number of mutants is r is the coefficient of x^r in the expansion of

$$\sum_{j=0}^{\infty}\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\frac{x^3}{3.4}+\ldots\right)^j e^{-m}.\frac{m^j}{j!}.$$

Thus p_r, the probability of a culture having r mutants, m being the mean number of mutations per culture, is the coefficient of x^r in the expansion of

$$e^{-m}\exp\left[m\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)\right]$$

in agreement with equation (15).

Arithmetical procedure

By means of the recurrence relation (9) and the boundary values (10) a table of values of $C_{j,r}$ has been computed for all (integral) values of j from 1 to 36 and of r from 1 to 64, subject to $r \geqslant j$. Equation (18) was employed as a check on the arithmetic at $j=63$. $C_{j,r}$

is the probability that a culture in which exactly j mutations have occurred shall contain r mutants. In practice when comparing the theoretical and experimental distributions the distributions will always be grouped. For economy of space, therefore, we do not publish the full table of $C_{j,r}$ but give in Table 1 grouped values for $r=1$, 2, 3, 4, 5–8, 9–16, 17–32, 33–64 and >64.

Table 1. $C_{j,r}$

$C_{j,r}$ is the probability that a culture in which exactly j mutations have occurred shall have r mutants ($r \geqslant j$). For values of r greater than 2, the values of $C_{j,r}$ have been grouped. Thus the numbers in the column headed '17–32' are values of $\sum_{r=17}^{r=32} C_{j,r}$. See also the Appendix for certain other values of $C_{j,r}$.

$j \backslash r$	1	2	3–4	5–8	9–16	17–32	33–64	>64
1	0·5000	0·1667	0·1333	0·0889	0·0523	0·0285	0·0149	0·0154
2	—	0·2500	0·2778	0·2118	0·1272	0·0671	0·0335	0·0325
3	—	—	0·2500	0·3100	0·2161	0·1161	0·0563	0·0515
4	—	—	0·0625	0·3093	0·2986	0·1735	0·0834	0·0726
5	—	—	—	0·2060	0·3479	0·2353	0·1149	0·0960
6	—	—	—	0·0885	0·3445	0·2949	0·1504	0·1216
7	—	—	—	0·0260	0·2907	0·3441	0·1894	0·1497
8	—	—	—	0·0039	0·2100	0·3753	0·2305	0·1804
9	—	—	—	—	0·1312	0·3830	0·2721	0·2137
10	—	—	—	—	0·0718	0·3665	0·3122	0·2495
11	—	—	—	—	0·0342	0·3294	0·3486	0·2878
12	—	—	—	—	0·0138	0·2790	0·3789	0·3284
13	—	—	—	—	0·0045	0·2234	0·4010	0·3711
14	—	—	—	—	0·0011	0·1698	0·4135	0·4157
15	—	—	—	—	0·0002	0·1228	0·4155	0·4616
16	—	—	—	—	0·0000	0·0846	0·4071	0·5084
17	—	—	—	—	—	0·0553	0·3891	0·5555
18	—	—	—	—	—	0·0343	0·3633	0·6024
19	—	—	—	—	—	0·0200	0·3315	0·6485
20	—	—	—	—	—	0·0109	0·2961	0·6930
21	—	—	—	—	—	0·0055	0·2591	0·7354
22	—	—	—	—	—	0·0026	0·2222	0·7752
23	—	—	—	—	—	0·0011	0·1871	0·8118
24	—	—	—	—	—	0·0004	0·1546	0·8450
25	—	—	—	—	—	0·0001	0·1254	0·8745
26	—	—	—	—	—	0·0000	0·0998	0·9001
27	—	—	—	—	—	—	0·0779	0·9221
28	—	—	—	—	—	—	0·0596	0·9404
29	—	—	—	—	—	—	0·0447	0·9553
30	—	—	—	—	—	—	0·0327	0·9673
31	—	—	—	—	—	—	0·0234	0·9766
32	—	—	—	—	—	—	0·0164	0·9836
33	—	—	—	—	—	—	0·0111	0·9889
34	—	—	—	—	—	—	0·0074	0·9926
35	—	—	—	—	—	—	0·0048	0·9952
36	—	—	—	—	—	—	0·0030	0·9970

p_r is the probability that a culture of such a size that the *mean* number of mutations is m shall contain r mutants. $p_0 = e^{-m}$ and for $r \geqslant 1$, $p_r = \sum_{j=1}^{r} C_{j,r} \left(e^{-m} \frac{m^j}{j!} \right)$. For any given value of m, p_r can be calculated with the aid of the table of $C_{j,r}$, for the same groupings of r. The calculation is facilitated if a table of Poisson coefficients $e^{-m} \frac{m^j}{j!}$ is available (Molina, 1942). In Table 2 values of p_r are given for a number of values of m from 0·05 to 15.

The figures in Tables 1, 2 and 4 are liable to occasional rounding errors of one unit in the last decimal place.

Limiting form of distribution for large numbers

Table 2 provides the means of testing the agreement between theoretical and experimental distributions in experiments in which the mean number of mutations per culture is 15 or fewer, and in which a minority of the cultures have more than 64 mutants. To extend Tables 1 and 2 by use of the recurrence relation (9) to cover experiments in which the mean number of mutations per culture considerably exceeds 15, and to subdivide the class

Table 2. p_r

p_r is the probability that a culture shall have r mutants, the average number of mutations which have occurred per culture being m. For values of r greater than 2, the values of p_r have been grouped. Thus the numbers in the column headed '17–32' are values of $\sum_{r=17}^{r=32} p_r$.

$m \backslash r$	0	1	2	3–4	5–8	9–16	17–32	33–64	>64
0·05	0·9512	0·0283	0·0082	0·0067	0·0045	0·0026	0·0014	0·0008	0·0008
0·10	0·9048	0·0452	0·0162	0·0134	0·0090	0·0053	0·0029	0·0015	0·0015
0·15	0·8607	0·0646	0·0239	0·0200	0·0137	0·0081	0·0044	0·0023	0·0023
0·20	0·8187	0·0819	0·0314	0·0267	0·0184	0·0109	0·0059	0·0031	0·0031
0·25	0·7788	0·0974	0·0385	0·0332	0·0231	0·0138	0·0074	0·0038	0·0039
0·30	0·7408	0·1111	0·0454	0·0397	0·0279	0·0167	0·0090	0·0046	0·0047
0·35	0·7047	0·1233	0·0519	0·0462	0·0328	0·0196	0·0106	0·0055	0·0055
0·40	0·6703	0·1341	0·0581	0·0525	0·0376	0·0226	0·0122	0·0063	0·0063
0·45	0·6376	0·1435	0·0640	0·0587	0·0425	0·0257	0·0139	0·0071	0·0071
0·50	0·6065	0·1516	0·0695	0·0648	0·0475	0·0288	0·0155	0·0079	0·0079
0·55	0·5769	0·1587	0·0747	0·0707	0·0524	0·0319	0·0172	0·0088	0·0087
0·60	0·5488	0·1646	0·0796	0·0765	0·0573	0·0351	0·0189	0·0096	0·0096
0·65	0·5220	0·1697	0·0841	0·0821	0·0622	0·0383	0·0207	0·0105	0·0104
0·70	0·4966	0·1738	0·0884	0·0876	0·0672	0·0415	0·0224	0·0114	0·0112
0·75	0·4724	0·1771	0·0923	0·0928	0·0721	0·0448	0·0242	0·0123	0·0120
0·80	0·4493	0·1797	0·0959	0·0979	0·0769	0·0482	0·0260	0·0132	0·0129
0·85	0·4274	0·1817	0·0992	0·1028	0·0818	0·0515	0·0279	0·0141	0·0137
0·90	0·4066	0·1830	0·1022	0·1076	0·0866	0·0549	0·0297	0·0150	0·0146
0·95	0·3867	0·1837	0·1049	0·1121	0·0914	0·0583	0·0316	0·0159	0·0154
1·0	0·3679	0·1839	0·1073	0·1164	0·0961	0·0617	0·0335	0·0168	0·0163
1·2	0·3012	0·1807	0·1145	0·1317	0·1144	0·0757	0·0414	0·0207	0·0198
1·4	0·2466	0·1726	0·1180	0·1438	0·1316	0·0899	0·0496	0·0247	0·0233
1·6	0·2019	0·1615	0·1184	0·1528	0·1473	0·1041	0·0581	0·0288	0·0270
1·8	0·1653	0·1488	0·1165	0·1587	0·1615	0·1184	0·0670	0 0332	0·0307
2·0	0·1353	0·1353	0·1128	0·1620	0·1738	0·1324	0·0761	0·0377	0·0345
2·2	0·1108	0·1219	0·1077	0·1629	0·1844	0·1462	0·0855	0·0423	0·0383
2·4	0·0907	0·1089	0·1016	0·1617	0·1930	0·1595	0·0951	0·0471	0·0423
2·6	0·0743	0·0966	0·0949	0·1587	0·1998	0·1723	0·1049	0·0521	0·0464
2·8	0·0608	0·0851	0·0880	0·1543	0·2048	0·1844	0·1149	0·0573	0·0505
3·0	0·0498	0·0747	0·0809	0·1487	0·2080	0·1958	0·1249	0·0626	0·0547
3·2	0·0408	0·0652	0·0739	0·1421	0·2095	0·2063	0·1351	0·0680	0·0590
3·4	0·0334	0·0567	0·0671	0·1350	0·2095	0·2159	0·1452	0·0736	0·0634
3·6	0·0273	0·0492	0·0607	0·1274	0·2081	0·2246	0·1554	0·0794	0·0679
3·8	0·0224	0·0425	0·0545	0·1195	0·2054	0·2324	0·1655	0·0853	0·0725
4·0	0·0183	0·0366	0·0488	0·1115	0·2016	0·2391	0·1755	0·0913	0·0771
4·2	0·0150	0·0315	0·0436	0·1036	0·1969	0·2447	0·1854	0·0975	0·0819
4·4	0·0123	0·0270	0·0387	0·0958	0·1912	0·2493	0·1951	0·1038	0·0867
4·6	0·0101	0·0231	0·0343	0·0882	0·1849	0·2529	0·2046	0·1102	0·0917
4·8	0·0082	0·0198	0·0303	0·0809	0·1780	0·2555	0·2139	0·1167	0·0967
5	0·0067	0·0168	0·0267	0·0739	0·1707	0·2571	0·2228	0·1234	0·1018
6	0·0025	0·0074	0·0136	0·0451	0·1311	0·2516	0·2620	0·1577	0·1289
7	0·0009	0·0032	0·0066	0·0258	0·0936	0·2285	0·2898	0·1932	0·1584
8	0·0003	0·0013	0·0031	0·0141	0·0630	0·1953	0·3043	0·2283	0·1903
9	0·0001	0·0006	0·0014	0·0074	0·0404	0·1586	0·3056	0·2615	0·2244
10	0·0000	0·0002	0·0006	0·0038	0·0249	0·1233	0·2950	0·2913	0·2608
11	0·0000	0·0001	0·0003	0·0019	0·0148	0·0923	0·2751	0·3165	0·2991
12	0·0000	0·0000	0·0001	0·0009	0·0086	0·0668	0·2486	0·3359	0·3391
13	0·0000	0·0000	0·0001	0·0004	0·0048	0·0469	0·2185	0·3489	0·3805
14	0·0000	0·0000	0·0000	0·0002	0·0027	0·0320	0·1871	0·3552	0·4228
15	0·0000	0·0000	0·0000	0·0001	0·0014	0·0214	0·1564	0·3549	0·4657

>64 mutants into further classes, e.g. 65–128, 129–256, 257–512 mutants, etc., would involve an impracticable amount of arithmetic. An attempt was therefore made to find asymptotic formulae for p_r or P_r valid for large values of m. We have not succeeded in finding explicit formulae, but have obtained some information on the form of the function.

If we consider P_r as a continuous function of the two variables r and m, then for values of $r \gg 1$ we have approximately $P_r - P_{r-1} = \partial P_r/\partial r$. Thus equation (21) approximates to

$$\frac{\partial P_r}{\partial m} + \left(1 + \frac{r}{m}\right)\frac{\partial P_r}{\partial r} = 0, \tag{24}$$

which is satisfied by

$$P_r = F\left(\frac{r}{m} - \log m\right), \tag{25}$$

where F is any function.

In Fig. 1 we have plotted P_r (derived from Table 2, i.e. based on the recurrence relation) against $(r/m - \log m)$ for $r = 8$, 16, 32 and 64, using five values of m (viz. 4, 6, 8, 13, 15)

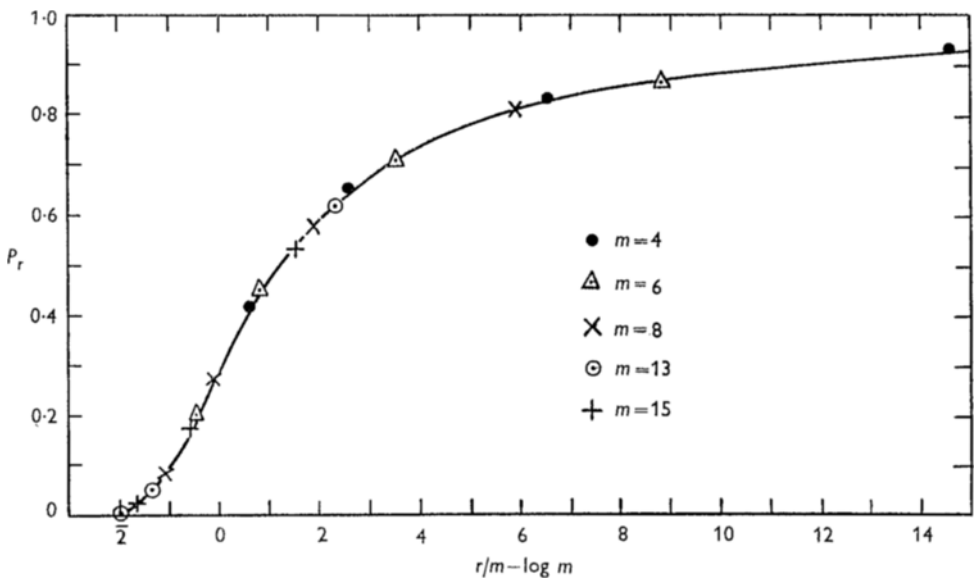

Fig. 1. P_r, for different r and m, is a function of $r/m - \log m$. The points are plotted for $r = 8$, 16, 32 and 64, and with $m = 4$, 6, 8, 13 and 15.

selected so that the twenty points are conveniently spaced. It is seen that the points lie quite well on a single curve, showing that these values of r are large enough for equation (24) to be a satisfactorily close approximation to equation (21). The smooth curve in Fig. 1 is thus a graph of the function F which enters into equation (25).

For any given value of m, $(r/m - \log m)$ is evidently distributed in a skew distribution about a median 1·24. We have found by trial that the derived variate $(r/m - \log m + 4{\cdot}5)^{-1}$ is distributed in a distribution rather closely approximating to a Gaussian distribution of standard deviation 0·086. This is shown by the closeness with which the points in Fig. 2 lie on a straight line. The points in Fig. 2 are derived from those of Fig. 1 by transforming the ordinates to probits, defined by the relation $\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{y-5} e^{-\frac{1}{2}y^2}dy = P_r$, where y is the probit corresponding to P_r. (Tables of probits are given in Fisher & Yates, 1938.) Also,

the abscissae are transformed to values of $(r/m-\log m+4{\cdot}5)^{-1}$. Fig. 2 shows that, approximately,

$$x=\left(\frac{1}{r/m-\log m+4{\cdot}5}-0{\cdot}174\right)\Big/0{\cdot}086=\left(\frac{11{\cdot}6}{r/m-\log m+4{\cdot}5}-2{\cdot}02\right) \tag{26}$$

is a normal deviate.

We conclude, in this semi-empirical manner, that when the spontaneous mutation theory is to be compared with experiments falling outside the scope of Table 2 (i.e. experiments in which cultures containing more than 64 mutants are frequent), it will be satisfactory for practical purposes to suppose $x=\left(\frac{11{\cdot}6}{r/m-\log m+4{\cdot}5}-2{\cdot}02\right)$ to be normally

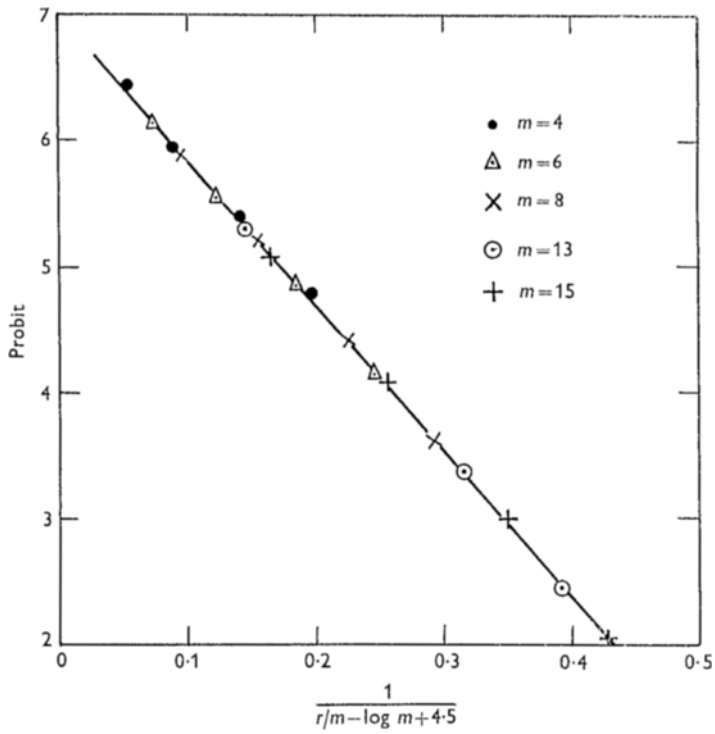

Fig. 2. $1/(r/m-\log m+4{\cdot}5)$ is distributed in an approximately normal distribution. The points are plotted for $r=8$, 16, 32 and 64, and with $m=4$, 6, 8, 13 and 15.

distributed with unit variance about the value 0. r is the number of mutants in an individual culture, m is the mean number of mutations per culture in the batch of parallel cultures.

The median of the distribution satisfies the relation

$$\frac{11{\cdot}6}{r/m-\log m+4{\cdot}5}-2{\cdot}02=0, \quad \text{i.e. at the median} \quad r/m-\log m=1{\cdot}24. \tag{27}$$

This equation provides a means of making a first estimate of m from the count of the number (r) of mutants in the median culture of the batch.

The quartiles of the distribution (i.e. values of r making $P_r=0{\cdot}25$ or $0{\cdot}75$) satisfy the relations:

$$\text{at quartiles:} \qquad r/m-\log m=-0{\cdot}2 \quad \text{and} \quad r/m-\log m=4{\cdot}1, \tag{28}$$

which relations may be used as a first test of whether the spread of an experimental distribution is comparable with the theoretical spread.

All the relations in this section are approximations, to be used only when dealing with experiments which lie outside the scope of Table 2. The approximation should not be used for the extreme ends of the distribution, e.g. for values of P_r exceeding 0·95 or less than 0·05.

THE ESTIMATION OF MUTATION RATE FROM EXPERIMENTAL OBSERVATIONS

m from the mean number of mutants per culture

As shown by Luria and Delbruck, the mean and variance of the distribution can be simply calculated, without knowing the distribution p_r, as follows:

While the culture grows from n_1 to n_1+dn_1, the mean number of mutations will be $(m/n)\,dn_1$ (cp. equation (4)), the actual number being distributed in a Poisson distribution about this mean with variance also $(m/n)\,dn_1$ (since the variance of a Poisson distribution is equal to the mean). The contribution to the final number of mutants (when the culture size is n) will be n/n_1 mutants for each mutation. Thus the contribution to the final number of mutants will be distributed about a mean $\frac{n}{n_1}\frac{m}{n}\,dn_1$ with a variance $\left(\frac{n}{n_1}\right)^2\frac{m}{n}\,dn_1$. Thus the mean of the required distribution is

$$\bar{r}=\int_1^n \frac{n}{n_1}\frac{m}{n}dn_1 = m\log n, \tag{29}$$

and the variance of an individual determination of r will be

$$\sigma^2=\int_1^n \left(\frac{n}{n_1}\right)^2\frac{m}{n}\,dn_1 \doteqdot mn.* \tag{30}$$

We can confirm that our distribution p_r yields the same mean and variance. The mean is

$$\bar{r}=\Sigma r p_r.$$

Since p_r is the coefficient of x^r in the expansion of $e^{-m}\exp\left[m\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)\right]$ (equation (15)), and since $q_r=e^m p_r$, we have

$$\Sigma\, q_r x^r = \exp\left[m\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)\right]. \tag{31}$$

Differentiating

$$\Sigma r x^{r-1} q_r = m\left(\frac{1}{2}+\frac{x}{3}+\frac{x^2}{4}+\ldots\right)\exp\left[m\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)\right]. \tag{32}$$

Inserting $x=1$, and putting $\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}\doteqdot\log n$ and $\frac{1}{1.2}+\frac{1}{2.3}+\ldots\doteqdot 1$, we have

$$\bar{r}=\Sigma r p_r = e^{-m}\,\Sigma\, r q_r \doteqdot m\log n.$$

Again, multiplying (32) by x^2 and differentiating,

$$\Sigma r(r+1)x^r q_r = m\left\{(x+x^2+x^3+\ldots)+m\left(\frac{x^2}{2}+\frac{x^3}{3}+\ldots\right)\left(\frac{1}{2}+\frac{x}{3}+\ldots\right)\right\}\exp\left[m\left(\frac{x}{1.2}+\frac{x^2}{2.3}+\ldots\right)\right].$$

Inserting $x=1$,

$$\Sigma\, r(r+1)q_r \doteqdot m e^m (n+m\log^2 n)$$

$$\Sigma\, r(r+1)p_r \doteqdot mn+m^2\log^2 n.$$

* But see Appendix: the correct value is $\sigma^2\doteqdot 2mn$.

Now the variance $\sigma^2 = \Sigma(r-\bar{r})^2 p_r = \Sigma r^2 p_r - 2\bar{r}\Sigma r p_r + \bar{r}^2 \Sigma p_r$

$= \Sigma r\,(r+1)\,p_r - \bar{r} - \bar{r}^2.$

Thus $\sigma^2 \doteqdot mn - m\log n$, i.e. $\sigma^2 \doteqdot mn$ since $n \gg 1$.

Since $\bar{r} = m\log n$, a possible method of determining m (and hence the mutation rate) experimentally would be to divide by $\log n$ the mean number of mutants per culture experimentally determined in a batch of N parallel cultures. However, on examination it appears that the precision of the estimate of m given by this method does not increase with increase of N. For it is evident that the total numbers of mutants in batches of N parallel cultures each of size n will be distributed (from batch to batch) in much the same way as the numbers of mutants in parallel cultures of size nN. The mean number of mutations in a culture of size nN will be mN (since the mean number of mutations is proportional to the size of the culture, cp. equation (3)), and hence by application of (30) the variance of the number of mutations in cultures of size nN is $mN.nN = mnN^2$. Thus the total number of mutants in a batch of N cultures of size n is distributed from batch to batch with a variance mnN^2. A fraction $1/N$ of this total number (i.e. the mean number per culture derived from a count of N cultures) is therefore distributed with variance mn.

Thus we see that the variance of the mean number $\bar{r}$ of mutants in N cultures is no smaller than the variance of the number of mutants in an individual culture, which shows that however many cultures are averaged, no improvement in precision is obtained over the use of a single culture selected at random. Consequently, the mean number of mutants per culture is an extremely inefficient statistic from which to calculate the mutation rate. If, nevertheless, this method of estimating m is employed, the variance (σ_m^2) of the estimate of m will (from (29) and (30)) be

$$\frac{mn}{(\log n)^2}, \tag{33}$$

independent of the number N of cultures averaged.

m from proportion of cultures without mutants

In view of the unsuitability of $\bar{r}$ as a means of estimating m from numerical data, Luria and Delbruck proposed its estimation by equating e^{-m} to the proportion of cultures experimentally determined to be without mutants. In a batch of N parallel cultures, in which the mean number of mutations is m per culture, the expected number of cultures without mutation is Ne^{-m}, the actual number being distributed about this mean in a binomial distribution having a variance $Ne^{-m}(1-e^{-m})$. Thus the variance of the estimate e^{-m} is $e^{-m}(1-e^{-m})/N$. Since $\dfrac{dm}{d(e^{-m})} = -e^m$, the corresponding estimate of m has a variance (σ_m^2) which is e^{2m} times as great, i.e.

$$\sigma_m^2 = \frac{e^m - 1}{N}. \tag{34}$$

Thus the standard error (σ_m) in the estimate of m is given by

$$\left(\frac{\sigma_m}{m}\right)^2 = \frac{1}{N}\left(\frac{e^m-1}{m^2}\right). \tag{35}$$

σ_m/m thus varies with m. It has a minimal value when $m = 1{\cdot}594$, when a fraction $0{\cdot}2032$ of cultures have no mutants. At $m = 1{\cdot}594$, $(\sigma_m/m)^2$ takes the value $1{\cdot}544/N$. At small or

large values of m (e.g. when the fraction of cultures without mutants exceeds 0·9 or is less than 0·01), the value of (σ_m/m) is much increased. Fig. 3*A* shows graphically $\frac{\sigma_m}{m}\sqrt{N}$ as a function of m.

The low precision at small values of m is to be attributed simply to the fact that an experiment in which the great majority of the cultures have no mutants does not provide much precise information about the mutation rate. The reduced precision at high values of m is, however, to be ascribed to the fact that this method of determining m does not

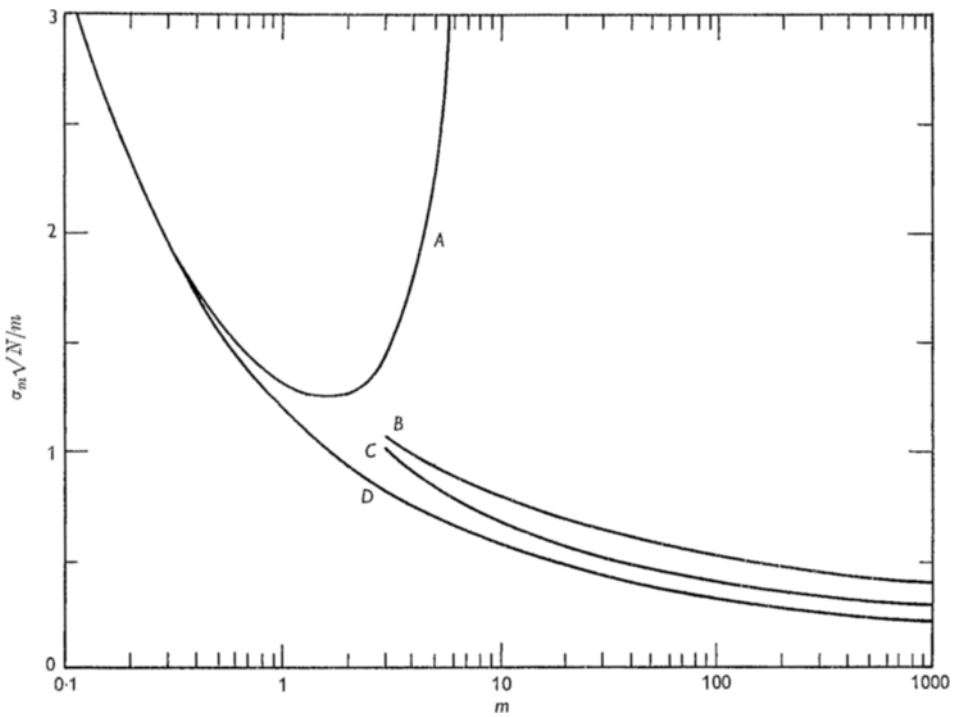

Fig. 3. The precision of the estimate of m derived by various methods: *A*, the method of the proportion of cultures without mutants; *B*, the method of the median; *C*, the method of $S[x]=0$; *D*, the method of maximal likelihood.

make full use of the experimental data, and in these cases more suitable methods, which we shall describe, enable a more precise estimate of m to be made from the same data.

m from the median

When the mutation rate is to be deduced from an experiment in which all, or nearly all, the cultures had mutants, so that the method just discussed is inapplicable, a very convenient method is to deduce m from the median of the distribution. The counts of the numbers of mutants in N parallel cultures are arranged in ascending order, and the middle one selected. The count in this culture is an estimate of r_0, the median of the distribution of r. Since we know that (approximately) the derived variate

$$x=\left(\frac{a}{r/m-\log m+b}-c\right) \quad \text{with } a=11{\cdot}6,\ b=4{\cdot}5,\ c=2{\cdot}02 \tag{36}$$

is normally distributed about median 0, it follows that

$$\frac{r_0}{m} - \log m = \frac{a}{c} - b = 1{\cdot}24. \tag{37}$$

This equation enables an estimate of m to be made from an experimentally determined value of r_0. With its aid Table 3 has been constructed, which enables m to be obtained for any value of r_0 up to 4400. While the derivation of m from the median is not the most efficient way of utilizing the experimental data from a statistical standpoint, it is the quickest satisfactory method, and is useful for making a preliminary estimate even if a more elaborate method is to be employed in making the final estimate.

Table 3. *Preliminary estimation of m from median value of r*

Thus if the middle culture of the series has 50 mutants, interpolation in the table between $r_0 = 49{\cdot}2$ and $r_0 = 55{\cdot}8$ gives $r_0/m = 3{\cdot}81$, so that $m = 50/3{\cdot}81 = 13{\cdot}1$. This is the mean number of mutations per culture.

r_0	r_0/m	r_0	r_0/m	r_0	r_0/m	r_0	r_0/m
1·4	1·3	15·3	2·9	117	4·5	787	6·1
1·6	1·4	17·4	3·0	132	4·6	884	6·2
1·9	1·5	19·9	3·1	150	4·7	993	6·3
2·3	1·6	22·7	3·2	169	4·8	1115	6·4
2·7	1·7	25·9	3·3	190	4·9	1251	6·5
3·2	1·8	29·5	3·4	215	5·0	1404	6·6
3·7	1·9	33·5	3·5	242	5·1	1575	6·7
4·3	2·0	38·1	3·6	273	5·2	1767	6·8
5·0	2·1	43·3	3·7	307	5·3	1981	6·9
5·7	2·2	49·2	3·8	346	5·4	2221	7·0
6·6	2·3	55·8	3·9	389	5·5	2490	7·1
7·7	2·4	63·2	4·0	438	5·6	2791	7·2
8·8	2·5	71·6	4·1	493	5·7	3127	7·3
10·1	2·6	81·1	4·2	554	5·8	3503	7·4
11·6	2·7	91·7	4·3	623	5·9	3924	7·5
13·3	2·8	104	4·4	700	6·0	4395	7·6

The precision of an estimate of m made in this way from counts of N cultures may be determined by calculating σ_m/m. We shall make use of the approximate result that x is distributed in a normal distribution with unit variance. The probability of x lying between x and $x+dx$ is $\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\,dx$. The probability of its lying between 0 and x is $\frac{1}{\sqrt{(2\pi)}}\int_0^x e^{-\frac{1}{2}x^2}\,dx$, and for observations in the neighbourhood of the median ($x^2 \ll 1$) we may write this as $\frac{x}{\sqrt{(2\pi)}}$. Thus $\left(\frac{1}{2}+\frac{x}{\sqrt{(2\pi)}}\right)$ is the probability of getting an observation $\leqslant x$, and $\left(\frac{1}{2}-\frac{x}{\sqrt{(2\pi)}}\right)$ is the probability of getting an observation $\geqslant x$. Thus the probability that, of $N = 2s+1$ observations, s shall be $\leqslant x$, s shall be $\geqslant x$, and one shall be between x and $x+dx$ is (providing x is in the neighbourhood of the median)

$$\frac{(2s+1)!}{s!\,s!}\left(\frac{1}{2}+\frac{x}{\sqrt{(2\pi)}}\right)^s\left(\frac{1}{2}-\frac{x}{\sqrt{(2\pi)}}\right)^s\frac{dx}{\sqrt{(2\pi)}} = \frac{(2s+1)!}{(2^s\,s!)^2}\left(1-\frac{2x^2}{\pi}\right)^s\frac{dx}{\sqrt{(2\pi)}}$$

$$\doteq \frac{(2s+1)!}{\sqrt{(2\pi)}\,(2^s\,s!)^2}\exp\left(-\frac{2sx^2}{\pi}\right)dx \doteq \left[2\pi\left(\frac{\pi}{4s}\right)\right]^{-\frac{1}{2}}\exp\left(\frac{-\frac{1}{2}x^2}{\pi/4s}\right) \tag{38}$$

for $s \gg 1$. Thus the median value of a set of $N = 2s+1$ values of x is distributed about $x = 0$ with variance $\frac{\pi}{4s} \doteq \frac{\pi}{2N}$.

It follows that if equation (37) is used to deduce from an experimentally determined median value r_0 an estimate of m, then this estimate will be subject to a variance $\frac{\pi}{2N}\Big/\left(\frac{\partial x}{\partial m}\right)_0^2$, the suffix 0 denoting evaluation at the median. Differentiating (36),

$$\frac{\partial x}{\partial m}=\frac{(x+c)^2\,(1-b+\log m)}{am}+\frac{x+c}{m}=(x+c)^2\,\frac{d}{m}+\frac{x+c}{m}, \tag{39}$$

where $d=(1-b+\log m)/a$. Hence at the median $x=0$, $\left(\frac{\partial x}{\partial m}\right)_0=\frac{c}{m}\,(cd+1)$. Thus the variance σ_m^2 of the estimate of m derived from the median is given by

$$\left(\frac{\sigma_m}{m}\right)^2=\frac{1}{N}\,\frac{\frac{1}{2}\pi\,(a/c^2)^2}{(1+a/c-b+\log m)^2}=\frac{1}{N}\,\frac{\frac{1}{2}\pi}{c^2\,(cd+1)^2}, \tag{40}$$

or, inserting the values of a, b, c from (36),

$$\left(\frac{\sigma_m}{m}\right)^2=\frac{1}{N}\,\frac{12{\cdot}70}{(2{\cdot}24+\log m)^2}. \tag{41}$$

Fig. 3B is a plot of $\left(\frac{\sigma_m}{m}\right)\sqrt{N}$ against m as given by (41). Having used Table 3 to make an estimate of m from the observation of the median value of r, Fig. 3B is consulted to obtain the standard deviation to be ascribed to the estimate of m.

m from $S\,[x]=0$

An alternative method of estimating m from experiments in which all or nearly all of the cultures have mutants is the following. Since x is distributed approximately normally about the value $x=0$, the mean value of x is zero. An estimate of m from a set of N observations can therefore be made by finding that value of m which makes

$$S\,[x]=S\left[\frac{a}{r/m-\log m+b}-c\right]=0, \tag{42}$$

the summation being over the N experimental observations. In using this method a first estimate of m is made by the median method. Inserting this value of m into (36), each experimental value of r is converted into a value of x, and the sum $S\,[x]$ formed. A series of adjacent values of m are then tried, and the value of m which makes $S\,[x]=0$ found (e.g. by plotting $S\,[x]$ against m).

The estimate of m obtained in this way is a little more precise than that based on the median. The mean $S\,[x]/N$ of a batch of N independent values of x will be distributed (from batch to batch) with variance $1/N$ about a mean zero. Suppose that its value for a particular batch is δ, so that

$$S\,[x]=N\delta.$$

If $m+\delta_m$ is the estimate of m derived from this particular batch (δ_m being the deviation between the estimated and true values of m),

$$S\left[x+\frac{\partial x}{\partial m}\,\delta_m\right]=0.$$

Thus $\delta_m \frac{S\,[\partial x/\partial m]}{N} = -\delta$ or approximately $\delta_m E\left[\frac{\partial x}{\partial m}\right] = -\delta$, where we have replaced the mean value of $\partial x/\partial m$ for the set of N observations by the *expectation* $E\left[\frac{\partial x}{\partial m}\right]$ of $\frac{\partial x}{\partial m}$.

Now from (39) we have

$$\frac{\partial x}{\partial m} = \frac{1}{m}\{x^2 d + x\,(2cd+1) + c^2 d + c\},$$

and x being normally distributed with unit variance about mean zero,

$$E\,[x^2] = 1 \quad \text{and} \quad E\,[x] = 0.$$

Thus

$$E\left[\frac{\partial x}{\partial m}\right] = \frac{1}{m}\{d\,(1+c^2)+c\},$$

and so

$$\delta_m = \frac{-m\delta}{d\,(1+c^2)+c}.$$

The variance of δ from batch to batch being $1/N$, we obtain for the variance$(\sigma_m{}^2)$ of m the relation

$$\left(\frac{\sigma_m}{m}\right)^2 = \frac{1}{N}\,\frac{1}{\{d\,(1+c^2)+c\}^2}, \tag{43}$$

with $a = 11{\cdot}6$, $b = 4{\cdot}5$, $c = 2{\cdot}02$, $d = (1-b+\log m)/a$.

A plot of $\frac{\sigma_m}{m}\sqrt{N}$ against m as computed by this formula is given in Fig. 3*C*. Having derived m by the method described in this section, the standard deviation to be ascribed to it is read from Fig. 3*C*.

Maximal likelihood method: large counts

None of the methods we have so far described is fully efficient statistically. At the expense of somewhat more laborious computation a fully efficient estimate of the mutation rate may be made by employing the method of maximal likelihood. We give two solutions: one for experiments which fall within the range of Tables 1 and 2, i.e. in which most of the cultures have fewer than 64 mutants, which is set out in the next section, and one for experiments falling outside the range of Tables 1 and 2, and for which the approximation that x is a normal deviate is employed, which is set out in the present section.

The probability that the number of mutants shall lie between r and $r+dr$ is given approximately (for r not too small) as

$$\frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}x^2}dx = \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}x^2}\frac{\partial x}{\partial r}dr = f\,dr \quad \text{(say)}. \tag{44}$$

Thus

$$\frac{d}{dm}\log f = \frac{1}{f}\frac{df}{dm} = -x\frac{\partial x}{\partial m} + \frac{\partial^2 x}{\partial r\,\partial m}\Big/\frac{\partial x}{\partial r}. \tag{45}$$

Now

$$x = \left(\frac{a}{r/m - \log m + b} - c\right) \quad \text{with } a = 11{\cdot}6,\ b = 4{\cdot}5,\ c = 2{\cdot}02, \tag{46}$$

and by differentiating we find

$$\frac{\partial x}{\partial m} = (x+c)^2\frac{d}{m} + \frac{x+c}{m}; \quad \frac{\partial^2 x}{\partial m\,\partial r}\Big/\frac{\partial x}{\partial r} = 2\,(x+c)\,\frac{d}{m} + \frac{1}{m}, \tag{47}$$

where
$$d=(1-b+\log m)/a. \tag{48}$$

Thus
$$\frac{1}{f}\frac{df}{dm}=\frac{1}{m}\{-x\,(x+c)^2\,d-x\,(x+c)+2\,(x+c)\,d+1\}. \tag{49}$$

Now L, the log likelihood, is (apart from irrelevant terms) $S\,[\log f]$, the summation being for the N observations of r, and the maximal likelihood condition is

$$0=\frac{dL}{dm}=S\left[\frac{1}{f}\frac{df}{dm}\right],$$

i.e.
$$S\,[x\,(x+c)^2\,d+x\,(x+c)-2\,(x+c)\,d-1]=0. \tag{50}$$

The routine for applying this method is as follows. Employing the preliminary estimate of m given by the median method, (48) is used to calculate d, and then (46) is used to calculate a value of x from each of the N experimental observations of r. For each of these N values of x the expression

$$x\,(x+c)^2\,d+x\,(x+c)-2\,(x+c)\,d-1$$

is evaluated and the N quantities added. The sum is similarly evaluated for several adjacent values of m, and by plotting against m (or otherwise) the value of m which satisfies (50) is deduced.

The variance to be attached to the maximal likelihood estimate of a parameter m is given by Fisher's formula (cp. e.g. Fisher, 1938)

$$\sigma_m{}^2=\frac{1}{Ni}, \tag{51}$$

where $i=E\left[\left(\frac{1}{f}\frac{df}{dm}\right)^2\right]$ is the expectation of $\left(\frac{1}{f}\frac{df}{dm}\right)^2$. Hence, using (49),

$$\begin{aligned} im^2 &= E\,[\{x^3d+x^2\,(2cd+1)+x\,(c^2d+c-2d)-(2cd+1)\}^2] \\ &= d^2\,E\,[x^6]+(6c^2d^2+6cd-4d^2+1)\,E\,[x^4] \\ &\quad +(c^4d^2+2c^3d-12c^2d^2-12cd+4d^2+c^2-2)\,E\,[x^2] \\ &\quad +(4c^2d^2+4cd+1) \\ &\quad +\text{terms involving odd powers of } x. \end{aligned}$$

Now it is readily shown that x being distributed normally about zero with unit variance, $E\,[x^n]$ vanishes for odd n, and

$$E\,[x^0]=E\,[x^2]=1, \quad E\,[x^4]=3, \quad E\,[x^6]=15.$$

Inserting these values in (52) we obtain

$$im^2=d^2\,(c^4+10c^2+7)+d\,(2c^3+10c)+(c^2+2),$$

so that
$$\left(\frac{\sigma_m}{m}\right)^2=\frac{1}{N}\frac{1}{d^2\,(c^4+10c^2+7)+d\,(2c^3+10c)+(c^2+2)}, \tag{53}$$

with $a=11{\cdot}6$, $b=4{\cdot}5$, $c=2{\cdot}02$, $d=(1-b+\log m)/a$.

The part of Fig. 3D to the right of $m=10$ is a plot of $\left(\frac{\sigma_m}{m}\right)\sqrt{N}$ against m. Having determined the maximal likelihood estimate of m, as described in this section, the standard deviation to attach to it is read off from Fig. 3D.

Maximal likelihood method: smaller counts

In this section we describe the method of arriving at the maximal likelihood estimate of m from an experiment falling within the scope of Tables 1 and 2; i.e. one in which the majority of cultures have fewer than 64 mutants.

p_r is the probability of a culture having r mutants. The log likelihood of a set of N values of r is (apart from irrelevant terms)

$$L = S\,[\log p_r],$$

S denoting summation over the N experimental values of r.

The maximal likelihood value of m is that satisfying

$$0 = \frac{dL}{dm} = S\left[\frac{1}{p_r}\frac{dp_r}{dm}\right].$$

Now from equation (11)

$$p_r = \sum_{j=1}^{\infty} C_{j,r}\, e^{-m}\frac{m^j}{j!}, \quad \frac{dp_r}{dm} = \sum_{j=1}^{\infty} C_{j,r}\, e^{-m}\left(\frac{m^{j-1}}{(j-1)!} - \frac{m^j}{j!}\right),$$

so that

$$\frac{1}{p_r}\frac{dp_r}{dm} = \frac{t_r - p_r}{p_r}, \tag{54}$$

where

$$t_r = \sum_{j=1}^{\infty} C_{j,r}\left(e^{-m}\frac{m^{j-1}}{(j-1)!}\right). \tag{55}$$

Thus the maximal likelihood estimate of m is that satisfying

$$S\left[\frac{t_r - p_r}{p_r}\right] = 0, \tag{56}$$

t_r has been computed for a range of values of m exactly as described earlier for p_r, and in Table 4 values of $(t_r - p_r)/p_r$ are listed for a range of values of m and for the same grouped ranges of r as were used previously.

The method of estimating m is therefore the following. A preliminary estimate of m is obtained either by the median method or by equating e^{-m} to the proportion of cultures without mutants. Table 4 is entered at the value of m nearest to this preliminary estimate, and a value of $(t_r - p_r)/p_r$ read off for each of the N experimental values of r. The N values are summed. The procedure is repeated for several adjacent values of m, and thence (graphically or otherwise) the value of m inferred which would make $S\left[\frac{t_r - p_r}{p_r}\right] = 0$.

The variance of this maximal likelihood estimate of m is given by the relation

$$\sigma_m^2 = \frac{1}{Ni}, \quad \text{where} \quad i = \sum_{r=0}^{\infty}\frac{1}{p_r}\left(\frac{dp_r}{dm}\right)^2 = \Sigma\frac{(t_r - p_r)^2}{p_r}.$$

Thus

$$\left(\frac{\sigma_m}{m}\right)^2 = \frac{1}{N}\frac{1}{m^2\Sigma\dfrac{(t_r - p_r)^2}{p_r}}. \tag{57}$$

Σ here means summation over all values of r from 0 to infinity, and is to be distinguished from S, meaning summation over the N experimental observations.

Table 4. $\frac{t_r - p_r}{p_r}$ *and* $\frac{\sigma_m}{m}\sqrt{N}$

This table is used in estimating mutation rate by the maximal likelihood method.

$\frac{t_0 - p_0}{p_0} = -1$ for all values of m.

$m \backslash r$	1	2	3–4	5–8	9–16	17–32	33–64	>64	$\sigma_m\sqrt{N}/m$
0·05	19·000	19·723	20·019	20·179	20·211	20·176	20·124	20·057	4·527
0·10	9·000	9·698	9·998	10·166	10·206	10·174	10·124	10·057	3·239
0·15	5·667	6·341	6·644	6·821	6·868	6·839	6·790	6·723	2·674
0·20	4·000	4·652	4·957	5·142	5·196	5·171	5·123	5·057	2·341
0·25	3·000	3·632	3·939	4·130	4·191	4·170	4·123	4·057	2·116
0·30	2·333	2·946	3·254	3·452	3·519	3·501	3·456	3·390	1·951
0·35	1·857	2·451	2·760	2·964	3·038	3·023	2·979	2·914	1·824
0·40	1·500	2·077	2·386	2·596	2·676	2·665	2·622	2·557	1·722
0·45	1·222	1·783	2·092	2·307	2·393	2·386	2·344	2·279	1·638
0·50	1·000	1·545	1·855	2·075	2·166	2·162	2·121	2·057	1·568
0·55	0·818	1·349	1·658	1·882	1·979	1·978	1·939	1·875	1·507
0·60	0·667	1·184	1·492	1·720	1·822	1·825	1·787	1·723	1·455
0·65	0·538	1·043	1·350	1·582	1·689	1·695	1·658	1·595	1·409
0·70	0·429	0·920	1·227	1·462	1·574	1·584	1·548	1·485	1·368
0·75	0·333	0·813	1·119	1·357	1·474	1·487	1·452	1·390	1·332
0·80	0·250	0·719	1·023	1·264	1·386	1·402	1·369	1·307	1·299
0·85	0·176	0·634	0·937	1·181	1·308	1·327	1·295	1·233	1·269
0·90	0·111	0·559	0·860	1·107	1·237	1·259	1·229	1·168	1·242
0·95	0·053	0·491	0·790	1·039	1·174	1·199	1·170	1·109	1·217
1·0	0·000	0·429	0·727	0·978	1·117	1·145	1·117	1·057	1·194
1·2	−0·167	0·228	0·520	0·777	0·931	0·971	0·948	0·890	1·118
1·4	−0·286	0·080	0·364	0·627	0·793	0·845	0·828	0·771	1·059
1·6	−0·375	−0·034	0·243	0·508	0·686	0·749	0·736	0·682	1·012
1·8	−0·444	−0·125	0·144	0·411	0·599	0·672	0·665	0·612	0·973
2·0	−0·500	−0·200	0·063	0·330	0·526	0·609	0·608	0·557	0·940
2·2	−0·545	−0·262	−0·007	0·260	0·464	0·556	0·560	0·511	0·913
2·4	−0·583	−0·315	−0·066	0·200	0·409	0·510	0·520	0·473	0·888
2·6	−0·615	−0·361	−0·119	0·147	0·362	0·471	0·486	0·441	0·867
2·8	−0·643	−0·401	−0·164	0·099	0·319	0·436	0·456	0·413	0·848
3·0	−0·667	−0·436	−0·205	0·057	0·280	0·404	0·430	0·390	0·831
3·2	−0·688	−0·467	−0·242	0·018	0·245	0·376	0·407	0·369	0·816
3·4	−0·706	−0·495	−0·275	−0·017	0·213	0·350	0·386	0·350	0·802
3·6	−0·722	−0·520	−0·305	−0·050	0·183	0·326	0·367	0·334	0·789
3·8	−0·737	−0·542	−0·333	−0·080	0·155	0·304	0·350	0·319	0·777
4·0	−0·750	−0·563	−0·358	−0·107	0·129	0·284	0·334	0·305	0·766
4·2	−0·762	−0·581	−0·381	−0·133	0·105	0·264	0·320	0·293	0·756
4·4	−0·773	−0·598	−0·402	−0·157	0·082	0·246	0·306	0·282	0·747
4·6	−0·783	−0·614	−0·422	−0·179	0·061	0·229	0·293	0·272	0·738
4·8	−0·792	−0·629	−0·441	−0·200	0·041	0·213	0·282	0·263	0·730
5·0	−0·800	−0·642	−0·458	−0·220	0·021	0·197	0·271	0·254	0·722
6	−0·833	−0·697	−0·530	−0·304	−0·062	0·130	0·223	0·219	0·690
7	−0·857	−0·737	−0·584	−0·369	−0·128	0·073	0·184	0·194	0·665
8	−0·875	−0·768	−0·626	−0·422	−0·184	0·025	0·151	0·174	0·644
9	−0·889	−0·792	−0·660	−0·465	−0·231	−0·016	0·121	0·157	0·627
10	−0·900	−0·812	−0·689	−0·502	−0·272	−0·053	0·095	0·143	0·613
11	−0·909	−0·828	−0·712	−0·534	−0·307	−0·086	0·071	0·131	0·600
12	−0·917	−0·842	−0·733	−0·561	−0·339	−0·116	0·048	0·120	0·589
13	−0·923	−0·853	−0·750	−0·586	−0·367	−0·143	0·028	0·110	0·580
14	−0·929	−0·863	−0·765	−0·607	−0·393	−0·167	0·008	0·101	0·571
15	−0·933	−0·872	−0·779	−0·626	−0·416	−0·190	−0·010	0·092	0·564

In the final column of Table 4 we have tabulated

$$\frac{1}{m\sqrt{\left(\Sigma\frac{(t_r-p_r)^2}{p_r}\right)}}=\frac{\sigma_m}{m}\sqrt{N}. \tag{58}$$

Having determined the maximal likelihood estimate of m as just described, the value of $\sigma_m\sqrt{N}/m$ is read off from the last column of Table 4. These values of $\sigma_m\sqrt{N}/m$ have been used in plotting the part of Fig. 3*D* to the left of $m=10$. Between $m=3$ and $m=15$, the values of $\sigma_m\sqrt{N}/m$ calculated from (58) and from (53) agree satisfactorily.

Summary

Statistical calculations are made of the distribution numbers of mutants in a culture of bacteria in which the number of mutants increases on account both of new mutations and of division of old mutants. In this way the largely qualitative conclusions of Luria and Delbruck are extended and placed on a firm quantitative basis. The results of these calculations, which enable the mutation rate to be inferred from experiments with parallel cultures, are presented in the form of tables. Statistically efficient methods of using these tables are discussed.

REFERENCES

Adolph, E. F. & Bayne-Jones, S. (1932). *J. Cell. Comp. Physiol.* **1**, 409.

Demerec, M. (1945). *Proc. Nat. Acad. Sci., Wash.*, **31**, 16.

Fisher, R. A. (1938). *Statistical Theory of Estimations.* Calcutta University Readership lectures.

Fisher, R. A. & Yates, F. (1938). *Statistical Tables for Biologists, Agriculturalists and Medical Research.* Edinburgh: Oliver and Boyd.

Lewis, I. M. (1934). *J. Bact.* **28**, 619.

Luria, S. E. & Delbruck, M. (1943). *Genetics*, **28**, 491.

Molina, E. C. D. (1942). *Poisson's Exponential Binomial Limit Tables.* New York: Van Nostrand.

Stewart, F. M. (1947). *J. Hyg., Camb.*, **45**, 28.

Witkin, E. M. (1946). *Proc. Nat. Acad. Sci., Wash.*, **32**, 59.

APPENDIX. (By C.A.C.)

(1) It has been suggested that a table of the individual coefficients $C_{j,r}$ introduced in equation (7), and which give the expansion of q_r in powers of m, might be useful. Such a table, for $r \leqslant 10$, is shown below.

Table of $C_{j,r}$

According to equation (7), $q_r \sum_{j=1}^{r} C_{j,r} \frac{m^j}{j!}$

$r \backslash j$	1	2	3	4	5	6	7	8	9	10
1	$\frac{1}{2}$									
2	$\frac{1}{6}$	$\frac{1}{4}$								
3	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{8}$							
4	$\frac{1}{20}$	$\frac{1}{9}$	$\frac{1}{8}$	$\frac{1}{16}$						
5	$\frac{1}{30}$	$\frac{7}{90}$	$\frac{5}{48}$	$\frac{1}{12}$	$\frac{1}{32}$					
6	$\frac{1}{42}$	$\frac{41}{720}$	$\frac{181}{2160}$	$\frac{1}{12}$	$\frac{5}{96}$	$\frac{1}{64}$				
7	$\frac{1}{56}$	$\frac{109}{2520}$	$\frac{97}{1440}$	$\frac{41}{540}$	$\frac{35}{576}$	$\frac{1}{32}$	$\frac{1}{128}$			
8	$\frac{1}{72}$	$\frac{853}{25200}$	$\frac{551}{10080}$	$\frac{173}{2592}$	$\frac{107}{1728}$	$\frac{1}{24}$	$\frac{7}{384}$	$\frac{1}{256}$		
9	$\frac{1}{90}$	$\frac{10}{700}$	$\frac{13570}{302400}$	$\frac{1311}{22680}$	$\frac{307}{5184}$	$\frac{203}{4320}$	$\frac{7}{256}$	$\frac{1}{96}$	$\frac{1}{512}$	
10	$\frac{1}{110}$	$\frac{1070}{75600}$	$\frac{251}{6720}$	$\frac{7567}{151200}$	$\frac{5900}{108864}$	$\frac{1081}{34600}$	$\frac{1100}{34560}$	$\frac{5}{288}$	$\frac{3}{512}$	$\frac{1}{1024}$

(2) It should perhaps be pointed out that the replacement in (3) of $n-1$ by n is an approximation whose effect is quite negligible provided that $r \ll n$, as occurs in all experiments. In fact, even for r of the order of $n^{\frac{1}{2}}$, the values of q_r are seriously in error. As a result of this, and of the fact that it allows r to exceed n (which is manifestly impossible since r is the number of mutants and n is the total number of bacteria), the generating function (15) actually gives an infinite value for all the moments. These two difficulties have been removed in a development of this theory, to be published by Mr D. G. Kendall, of Oxford. But unfortunately his more strictly correct generating function cannot be expanded with any ease to determine the q_r. Except for large r or small n, however, it differs insignificantly from our (15).

(3) Mr Kendall has kindly pointed out to me that the argument in (31) and (32), which was copied from Luria and Delbruck, is not quite valid. For in (32) the complete series $\frac{1}{2}+\frac{x}{3}+\frac{x^2}{4}+\ldots$ is not convergent when $x=1$, and in order to get an expression for the mean value and the variance it was necessary artificially to curtail this series by truncating it at its term x^n/n. This device is not a valid procedure, and it appears that although there is no change in the mean $\bar{r}$, the variance σ^2 of an individual determination of r requires to be multiplied by 2, so that the correct relation $\sigma^2 \doteqdot 2mn$.