THE DESIGN OF EXPERIMENTS ON MUTATION RATES

By J. B. S. HALDANE

Genetics and Biometry Laboratory, Government of Orissa, Bhubaneswar

On page 16 of her interesting book on Mutation, Auerbach (1962) writes, "In general, it is a good idea to plan experiments so that the numbers in each series are roughly in inverse proportion to the expected mutation frequencies". This implies that in each series, say a control and the progeny after irradiation or other treatment with two different doses of a mutagen, the number of mutants to be counted is about the same. I venture to question this statement.

Let us suppose that two treatments give frequencies p and q of mutants, where $p>q$. We may ask three rather different questions:

(1) Is it reasonably certain that p exceeds q ? Is our treatment mutagenic at all ? Or, again, does ·05% of a mutagenic compound give significantly more mutants than ·01% ?

(2) What is the value of $p-q$? If our final result is to be a graph in which mutation rate increases in a more or less linear manner with dose, this is the obvious question.

(3) What is the value of pq^{-1} or of $\log p - \log q$? This would arise if we wished to know by what percentage the standard rate was raised by a small mutagenic dosage.

In each case we suppose that a total of s organisms are counted, of which m are expected to include a fraction p of mutants, n a fraction q, where $m+n=s$. Our problem is how to divide up the series so as to get the best test of significance, the estimate of $p-q$ with the lowest sampling variance, or the estimate of $\log p - \log q$ with the lowest sampling variance.

In answering the first two questions there is no need to be sure of the value of the lower of the two mutation rates. If we have effectively examined 100,000 gametes from untreated parents for mutations of a certain type, and found none, we can say that if the mutation rate q were 5×10^{-5}, the probability of finding no mutants would be less than ·01. So q is probably under ·00005. If 20 mutants have been found among 10,000 treated gametes we are in no doubt of the efficacy of the mutagenic treatment, and have a fair idea of the mutation rate induced by it. This is about ·002. If we had tested 109,000 untreated gametes and 1,000 treated, the first group would probably have included one or no mutants, the second about two, though quite possibly none. We should have come nearer to satisfying Auerbach's criterion, but should know a good deal less about mutagenesis. On the other hand, if we are interested in the ratio of mutation frequencies, we certainly need information about the lower of the two mutation frequencies. It remains to be seen whether Auerbach's criterion, or some other, should be used.

In what follows, I shall suppose m and n to be large, that is to say over 100, so that m^{-1} and n^{-1} can be neglected in comparison with unity. I shall also often assume p and q to be small, i.e. $<\cdot 01$. I suppose that there were a mutants out of m, and b out of n, as in Table 1.

Table 1

	Normal	Mutants	Total
First treatment	$m-a$	a	m
Second treatment	$n-b$	b	n

The second question is very easily answered. Let x and y be our estimates of p and q. Then

$$x = am^{-1}, \qquad \text{Var}\,(x) = p(1-p)m^{-1}$$
$$y = bn^{-1}, \qquad \text{Var}\,(y) = q(1-q)n^{-1}.$$

The variance of the difference is the sum of these two variances, that is to say

$$\text{Var}\,(x-y) = p(1-p)m^{-1}+q(1-q)n^{-1}.$$

Now if $m+n=s$, we may increase m by unity, and if so must decrease n by unity. The variance becomes

$$p(1-p)(m+1)^{-1}+q(1-q)(n-1)^{-1}.$$

That is to say it is increased by $\dfrac{q(1-q)}{n(n-1)} - \dfrac{p(1-p)}{m(m+1)}$

This is zero when $\dfrac{m(m+1)}{n(n-1)} = \dfrac{p(1-p)}{q(1-q)}$,

or very nearly $\dfrac{m}{n} = \left[\dfrac{p(1-p)}{q(1-q)}\right]^{\frac{1}{2}}$

$$= \left(\frac{p}{q}\right)^{\frac{1}{2}} \quad \text{nearly, if } p \text{ and } q \text{ are small.} \qquad (1)$$

We can also see that the value of the variance is a minimum, for it increases when m or n becomes small. We could obtain these results a little quicker by treating m as a continuous variable, and differentiating with regard to it. This is however illegitimate, even though in this case it gives a correct result.

When p and q are small, we should choose

$$m = \frac{p^{\frac{1}{2}}s}{p^{\frac{1}{2}}+q^{\frac{1}{2}}}, \quad n = \frac{q^{\frac{1}{2}}s}{p^{\frac{1}{2}}+q^{\frac{1}{2}}}. \qquad (2)$$

The sampling variance is $\left(p^{\frac{1}{2}}+q^{\frac{1}{2}}\right)^2 s^{-1}$.

Thus the numbers in the two series should be proportional to the square roots of the mutation frequencies, and not inversely proportional, as Auerbach suggests. For example if $s=10{,}000$ and p and q are believed to be $\cdot 0081$ and $\cdot 0001$, we should choose $m=9{,}000$, and $n=1{,}000$. It may be remarked that a considerable deviation from

these values raises the variance very little. Thus for $m=9{,}000$, $n=1{,}000$ the variance is $0{\cdot}9927\times10^{-6}$, for $m=9{,}500$, $n=500$, it is $1{\cdot}0457\times10^{-6}$, for $m=8{,}000$, $n=2{,}000$ it is $1{\cdot}0543\times10^{-6}$. On the other hand Auerbach's suggestion is $m=122$, $n=9{,}878$, giving Var $(x-y)=6{\cdot}586\times10^{-5}$, so that the standard sampling error is increased eightfold.

In this case $p-q=\cdot0080$. The standard error of $x-y$, for a correct choice of m and n, is $\cdot0010$, but with the incorrect expression it becomes $\cdot0081$, so the method is valueless.

However (2) is inapplicable if q is believed to be zero or very small. Even if no mutants of the type under investigation have been found in a large number of organisms previously observed, it is desirable to verify that few or none occur under the conditions of the particular experiment in question. One may be dealing with a stock in which a recessive mutation has just occurred, an exceptionally mutable one, and so on.

If p is roughly known, but q believed to be very small, the following argument seems reasonable. If n is, say, 1,000, and includes no mutants, we cannot be sure that if we count 1,001 the next organism will not be a mutant, but we can bet heavily that the next two will not be mutants, or even that there will not be more than one mutant in the next ten. Let us choose n so that whether it includes one mutant or none, this will make a difference to $x-y$ less than the standard error of x.

That is to say $n^{-1}<\left(\frac{p}{m}\right)^{\frac{1}{2}}$ or $n^2>mp^{-1}$.

Since $m=s-n$, this implies

$$n>\frac{(4sp+1)^{\frac{1}{2}}-1}{2p} > \left(\frac{s}{p}\right)^{\frac{1}{2}}-\frac{1}{2p}. \tag{3}$$

For example if $s=10{,}000$, $p=\cdot01$, $n>950$.

Of course if in such a case the sample of $n=1{,}000$, say, includes 2 or 3 mutants, we can be reasonably sure that our previous information was incorrect, and q exceeds $\cdot001$.

Next consider the third case. We wish to estimate $\log p-\log q$, but it is more convenient to consider the Napierian logarithm of the ratio, $\ln p-\ln q$. If we call the estimate of this difference r, $r=\ln a-\ln m-\ln b+\ln n$.

The sampling variance of $\ln a$ is $\frac{1-p}{mp}$, so

$$\mathrm{Var}(r)=\frac{1-p}{mp}+\frac{1-q}{nq}$$

When m is increased and n decreased by unity this increases by

$$\frac{1-q}{qn(n-1)}-\frac{1-p}{pm(m+1)}.$$

The variance, as before, is minimal when this is zero, or

$$\frac{m}{n}=\left[\frac{q(1-p)}{p(1-q)}\right]^{\frac{1}{2}} \tag{4}$$
$$=\left(\frac{q}{p}\right)^{\frac{1}{2}} \quad \text{nearly.}$$

$$\text{Thus } m = \frac{q^{\frac{1}{2}}s}{p^{\frac{1}{2}}+q^{\frac{1}{2}}} \quad , \quad n = \frac{p^{\frac{1}{2}}s}{p^{\frac{1}{2}}+q^{\frac{1}{2}}} \quad \text{nearly.} \tag{5}$$

In fact the sample number is nearly inversely as the square root of the mutation frequency. There is no danger in this case of n being too small unless, of course, s is too small to give a clear result. The minimum value of the variance is

$$\frac{\left(p^{\frac{1}{2}}+q^{\frac{1}{2}}\right)^2 \left(1-p^{\frac{1}{2}}q^{\frac{1}{2}}\right)}{pqs}$$

$$= \frac{\left(p^{\frac{1}{2}}+q^{\frac{1}{2}}\right)^2}{pqs} \quad \text{nearly.}$$

Taking the same example as before, namely $p=\cdot0081$, $q=\cdot0001$, $s=10{,}000$, we have $m=1{,}000$, $n=9{,}000$. The sampling variance of r is $\frac{100}{81}$, its standard deviation $\frac{10}{9}$, or in decimal logarithms 0·4825. Since $\log\frac{p}{q}=1\cdot9085$ it will be seen that the standard sampling error is ·256 of the difference of logarithms, as compared with ·125 for the difference of mutation frequencies, that is to say a good deal less efficient.

This leads us to the first question, that of a test of significance. As is well known, there is no uniformly best test of significance when some of the numbers in the sample are small. I shall only consider the χ^2 test, and this only when the number of mutants is so large that we can assume that it takes its expected value without too much error. We see from Table 1 that

$$\chi^2 = \frac{(m+n)(an-bm)^2}{mn(m+n-a-b)\ (a+b)}.$$

If a or b is sufficiently large, we may replace a and b by their expectations mp and nq. Then, approximately,

$$\chi^2 = \frac{mn(m+n)(p-q)^2}{(mp+nq)[m(1-p)+n(1-q)]}.$$

The difference due to substituting $m+1$ for m and $n-1$ for n is,

$$\frac{(m+n)^2[-m(m+1)p(1-p)+n(n-1)q(1-q)]\ (p-q)^2}{(mp+nq)(mp+nq+p-q)[m(1-p)+n(1-q)]\ \ [m(1-p)+n(1-q)-p+q]}.$$

We see at once that χ^2 is smallest when this is zero, so

$$\frac{m}{n} = \left[\frac{q(1-q)}{p(1-p)}\right]^{\frac{1}{2}} \text{ nearly.} \tag{6}$$

This is the equivalent of (4) when p and q are small. That is to say the numbers in the series should be inversely as the square roots of the expected mutation frequencies.

$$\text{If } m = \frac{(q-q^2)^{\frac{1}{2}}s}{(p-p^2)^{\frac{1}{2}}+(q-q^2)^{\frac{1}{2}}}, \qquad n = \frac{(p-p^2)^{\frac{1}{2}}s}{(p-p^2)^{\frac{1}{2}}+(q-q^2)^{\frac{1}{2}}}$$

$$\text{Then } \mathcal{E}(\chi^2) = \frac{(p-q)^2 s}{\left[p^{\frac{1}{2}}(1-q)^{\frac{1}{2}} + (1-p)^{\frac{1}{2}}q^{\frac{1}{2}}\right]^2} \qquad \text{roughly,}$$

$$= [p^{\frac{1}{2}}(1-q)^{\frac{1}{2}} - (1-p)^{\frac{1}{2}}q^{\frac{1}{2}}]^2 s.$$

Whereas if we took Auerbach's values $m = \frac{qs}{p+q}$, $n = \frac{ps}{p+q}$,

$$\text{we have } \mathcal{E}(\chi^2) = \frac{(p-q)^2 s}{2(p+q-2pq)}.$$

If p and q are both small, these values approximate to $\frac{(p-q)^2 s}{p+2\sqrt{pq}+q}$, and $\frac{(p-q)^2 s}{2(p+q)}$. The former is always larger, and may be almost twice as large. Thus Auerbach's values are not as efficient as possible.

If, however, the total number of mutants counted is less than 20 or so, the expected value of χ^2 is appreciably greater than the above, and the mathematical theory becomes not only complicated but controversial. It would be of interest to develop it, but not of much biological interest, since, as Auerbach points out on her p.16, experiments and controls should be replicated, and it is not possible to say that replications agree if the total number of mutants is less than 20.

It is obvious that these results can be extended to cover cases where three or more mutation frequencies are being compared. Thus if m is the number of gametes to be tested, and p the previously estimated mutation frequency:

$m \propto [p(1-p)]^{-\frac{1}{2}}$ for the χ^2 test,

$m \propto [p(1-p)]^{\frac{1}{2}}$ for the estimation of a difference,

$m \propto [p^{-1}(1-p)]^{\frac{1}{2}}$ for the estimation of a ratio.

It wish to thank Dr. Auerbach for her clear statement as to an elementary problem. If, as I think, her solution was incorrect, it was better to give an incorrect solution than a vague one. I thank Mr. S. D. Jayakar for helpful criticism.

Summary

In order to get the best assurance that two mutation frequencies differ, or the best estimate of their ratio, the number of gametes tested in a series of experiments should be inversely as the square root of the expected mutation frequency. In order to get the best estimate of the difference of two mutation rates it should be directly as the square root of this frequency.

REFERENCES

Auerbach, C. (1962). *Mutation. An Introduction to Research on Mutagenesis. Part I. Methods.* Oliver and Boyd, Edinburgh.