

MEASURING RESISTANCE TO VIRUS INFECTION

P. A. P. MORAN

Australian National University, Canberra, A.C.T.

Dilution series of viruses often do not fit the exponential curve which they should fit if the usual assumptions of the theory of dilution series were true. Suppose the density of the virus suspension is λ , the average number of virus particles per unit volume of the inoculum (we suppose that unit volume is the amount actually used in inoculating the egg or other test material). Let C be the proportion of infective virus so that $C\lambda$ is the effective density of virus particles, and X is a factor depending on the dilution used such that $C\lambda X$ is the expected number of infective particles in the inoculum. Then the probability of there being no infective virus in the inoculum is

$$e^{-C\lambda X}.$$

Let p be the probability of a particle of infective virus infecting the egg if the particle is present. If p does not vary from egg to egg, the probability of the egg remaining uninfected is therefore

$$e^{-pC\lambda X},$$

and the expected values in the dilution series will follow an exponential curve when plotted against X . In many cases, however, p varies from egg to egg with a probability distribution $f(p)$. Then the probability of the egg remaining uninfected is

$$P(X) = \int_0^1 f(p)e^{-pC\lambda X} dp. \quad (1)$$

The curve $P(X)$ is flatter than an exponential curve, and this will be true unless, and only unless, $f(p)$ is a distribution concentrated at a single point.* A test for this has been given in two previous notes (Moran, 1954*a*, *b*).

What we now want to do is to estimate the mean value of p in the distribution $f(p)$ or some quantity closely related to this mean. But if we do not know $C\lambda$, the density of infective virus, even an exact knowledge of the curve $P(X)$ does not provide any such information about the distribution of p but only about the distribution of $C\lambda p$. In other words, $C\lambda$ has become confounded with p and the mean of p is 'unidentifiable'.

To make any further progress we have to know $C\lambda$, and this can only be found by some other method. We might know what proportion of virus particles are infective and have some physical method for finding the absolute number of particles. Alternatively, the eggs might be treated in such a way that they are completely susceptible to infection by the particles so that if a single particle is present, the egg will become infected. We can then carry out a dilution assay in the ordinary way and estimate the density of complete virus particles.

* It should be emphasized that it is not the host resistance, but the variation in the host resistance which flattens the curve.

We now assume that we know $C\lambda$ with some accuracy. Given a batch of eggs which have not been treated to make them susceptible we suppose as before that $f(p)$ is the distribution, amongst eggs, of the probability, p , that a single particle, if present, will infect the egg. The mean value of p is

$$\bar{p} = \int_0^1 pf(p)dp, \quad (2)$$

and this is what we would like to estimate. Unfortunately it does not seem possible to construct an exact method of estimating p because $f(p)$ is unknown.

Consider, however, for what purposes such an estimator is required. The primary purpose is to be able to compare different methods of treating eggs so that it is possible to say that one treatment makes eggs more susceptible than another. We therefore try to estimate a property of the distribution $f(p)$ which, although not the mean, provides some sort of measure of where $f(p)$ is concentrated in the interval $(0, 1)$. Suppose, for example, that we make up a suspension such that $C\lambda = 1$. If we inject unit volume of this suspension into an egg, the expected proportion of uninfected eggs is

$$\int_0^1 e^{-p}f(p)dp. \quad (3)$$

If r is the number of uninfected eggs out of n , it follows that $1 - (r/n)$ is an unbiased estimator of

$$A = \int_0^1 (1 - e^{-p})f(p) dp. \quad (4)$$

$1 - e^{-p}$ is a function of p which increases steadily from 0 to 0.6321. Although A can therefore deviate from the mean of p by as much as 0.3679 we can expect that it will provide a useful measure for the comparison of two different treatments.

When $f(p)$ is concentrated at the lower values of p , the difference between A and \bar{p} will be small. For example, if p cannot rise above 0.2, the maximum difference between p and $1 - e^{-p}$ is 0.0187 and the difference between A and \bar{p} cannot exceed 0.0187. In practice, the difference will be smaller if p is spread out on the interval $(0, 0.2)$. On the other hand, when $f(p)$ is concentrated near 1 the difference between the means of p 's for two different treatments will be reflected in a somewhat smaller difference in the corresponding A 's.

A numerical example will illustrate the method. Suppose a suspension is made up for which $C\lambda = 1$ and 100 eggs are inoculated. A must lie in the interval $(0, 0.6321)$. Suppose A is 0.4. Then $1 - (r/n)$ will be distributed as a binomial proportion with this probability, thus:

$$\mathcal{E} \left(1 - \frac{r}{n} \right) = 0.4.$$

$$\text{s.E.} \left(1 - \frac{r}{n} \right) = \sqrt{\{(0.4)(0.6)/100\}} = 0.049.$$

This shows that even with 100 eggs, the standard error is fairly large. If we wish to make comparisons between groups of eggs for which the values of \bar{p} (or A) differ

slightly, large numbers of eggs are required. This fact is intrinsic to the problem, and no development of mathematical theory can eliminate it.

Suppose now that we have used the above method to compare two treatments, and we have found that the two A 's are very small and possibly zero. Then, although the standard errors of estimate of the A 's will be small considered absolutely, they will be very large when compared with the difference between the A 's. Although the experiment may then provide fairly accurate estimators of the two \bar{p} 's (since the A 's will be very near to the \bar{p} 's) it will be of no value in comparing the relative effects of the treatments, i.e. the error may be small absolutely but large compared with the \bar{p} 's.

In this case let us use a suspension M times as concentrated as the above one. Then $1 - r/n$ will be an unbiased estimator of

$$\begin{aligned}
 A_M &= \int_0^1 \{1 - e^{-pM}\} f(p) dp \\
 &= \int_0^{M^{-1}} \{1 - e^{-pM}\} f(p) dp + E,
 \end{aligned}
 \tag{5}$$

where

$$E \leq \int_{M^{-1}}^1 f(p) dp.$$

If we can be sure that E is small enough to be neglected, A_M will be a substitute for $M\bar{p}$ which will be close to the latter if $f(p)$ is concentrated at the bottom end of the interval $(0, M^{-1})$. In fact, neglecting E ,

$$0.6321 \leq \frac{A_M}{M\bar{p}} \leq 1.$$

Thus we must choose M sufficiently small for E to be small and for the distribution $f(p)$ to be well covered by the interval $(0, M^{-1})$. On the other hand, the larger M is chosen, the lower will be the relative sampling variance.

It is not possible to choose M with any great confidence about the smallness of E . It is true that $(n - r)/n$ can be used to provide an upper confidence limit for \bar{p} and then, given M , an upper bound for E by using an inequality of Tchebycheff type. This does not work well in practice because the resulting lower bound for M is too small.

An artificially constructed example will help to make the process clear. Suppose that p is, in fact, uniformly distributed over the interval 0.00–0.02 so that its mean is 0.01. Then with unit dilution the probability of an egg being infected is

$$1 - P = 1 - 50 \int_0^{0.02} e^{-x} dx = 0.009950.$$

Suppose that in our first attempt to estimate \bar{p} we use a unit dilution on 50 eggs and that one is infected. Then our estimate of \bar{p} is 0.02 and an upper 5% confidence limit for \bar{p} is 0.11 (Snedecor, 1946; Clopper & Pearson, 1934). We may consider it unlikely that p then has any non-negligible probability of being greater than 0.25.

We therefore put $M = 4$ and try again. $1 - (r/n)$ will then be an unbiased estimator of

$$\begin{aligned} A_4 &= \int_0^{\frac{1}{4}} \{1 - e^{-4p}\} f(p) dp + E \\ &= \frac{5.0}{4} \int_0^{0.08} (1 - e^{-x}) dx \\ &= 12.5 (0.003116) = 0.03895 \end{aligned}$$

and we take

$$\frac{1}{4} (n - r) n^{-1}$$

as an estimator of \bar{p} . The probability of an egg being infected is 0.03895. This time, out of 50 eggs, 3 (say) are observed to be infected. Then we estimate \bar{p} as 0.015 and the 5% confidence interval for $4\bar{p}$ is (0.02, 0.34). It now seems unlikely that \bar{p} is greater than $\frac{1}{4} (0.34) = 0.085$ and we might choose a somewhat higher value of M .

There is another way of approaching the problem of finding \bar{p} . The quantity we want to find is \bar{p} , given by (2), whilst we can find easily $P(X)$, given by (1), for any given value of X . This suggests that we approximate to the function p by a series of exponentials. Consider, for example,

$$\begin{aligned} L_1(p) &= 2.6431 - 1.5047 e^{-1p} - 2.6682 e^{-p} \\ &\quad + 1.8872 e^{-2p} - 0.3587 e^{-4p}. \end{aligned}$$

In the range (0, 1) this never diverges from p by more than 0.0025. It follows that the difference between \bar{p} and

$$l = \int_0^1 L_1(p) f(p) dp$$

never exceeds 0.0025. l can be easily estimated. Choose dilutions such that $C\lambda X = \frac{1}{2}, 1, 2$ and 4 and let $n_1, n_2, n_3,$ and n_4 be the number of eggs tested at these dilutions, and r_1, r_2, r_3 and r_4 be the number uninfected. Then

$$\hat{l}_1 = 2.6431 - 1.5047 (r_1/n_1) - 2.6682 (r_2/n_2) + 1.8872 (r_3/n_3) - 0.3587 (r_4/n_4)$$

is an unbiased estimator of l .

The variance of this estimator depends not only on \bar{p} but also on other characteristics of the distribution $f(p)$. However, it is possible to get, from the experiment itself, an unbiased estimator of this variance. Suppose X is a binomial variate with parameters n, p . Then $\mathcal{E}(X/n) = p, \mathcal{E}(X(1 - X/n)/n) = (n - 1) n^{-1} p (1 - p)$. It follows that an unbiased estimator of the variance of r_i/n_i is

$$\frac{r_i}{n_i (n_i - 1)} \left(1 - \frac{r_i}{n_i} \right),$$

and so an unbiased estimator of the variance of \hat{l}_1 is

$$\hat{V} = 2.265 \frac{r_1 (n_1 - r_1)}{n_1^2 (n_1 - 1)} + 7.118 \frac{r_2 (n_2 - r_2)}{n_2^2 (n_2 - 1)} + 3.561 \frac{r_3 (n_3 - r_3)}{n_3^2 (n_3 - 1)} + 0.129 \frac{r_4 (n_4 - r_4)}{n_4^2 (n_4 - 1)}.$$

It should be noticed, however, that the distribution of \hat{V} is not independent of that of \hat{l}_1 .

The exact variance of \hat{l}_1 has been calculated for a number of theoretical distributions $f(p)$ and is considerably higher, for a given total number of eggs, than that given by the previous method. By using only two exponentials instead of four we can reduce this variance by a factor which is about three or four. Thus

$$L_2 = 1.2549 - 2.4243e^{-2p} + 1.1975e^{-4p}$$

diverges from p , in the range $(0, 1)$, by a maximum error of about 0.05, and the variance of the corresponding \hat{l}_2 is smaller than that of \hat{l}_1 although still larger than previously.

The above methods, using two or more dilutions, have the sole advantage of being approximately linear over the whole range $(0, 1)$. With a given total number of eggs they give substantially less statistical accuracy. It seems therefore that the best thing to do is to use the first method outlined above. Even so it will be found that to compare the host resistance of two lots of eggs which do not differ by a considerable factor, quite a large number of eggs are necessary. This is due to the nature of the problem and cannot be removed by mathematical dexterity.

That this is so may be seen by comparing the situation with an ideal experiment in which individual virus particles could be used by themselves. In such an experiment we would place one particle in each egg. The proportion of eggs in which the virus developed would give a direct estimator of \bar{p} , with binomial variance which will not be very much smaller than the variance involved in the first method above (where we are, in fact, estimating something slightly different from \bar{p}).

REFERENCES

- CLOPPER, C. J. & PEARSON, E. S. (1934). *Biometrika*, **26**, 404.
 MORAN, P. A. P. (1954a). *J. Hyg., Camb.*, **52**, 189.
 MORAN, P. A. P. (1954b). *J. Hyg., Camb.*, **52**, 444.
 SNEDECOR, G. W. (1946). *Statistical Methods*. Iowa State College Press.

(MS. received for publication 25. XI. 54)