THE DILUTION ASSAY OF VIRUSES

By P. A. P. MORAN

Department of Statistics, Australian National University, Canberra, A.C.T.

Suppose that λ is the average density of virus particles per unit volume. If x is a dilution of this and unit volume is applied to an egg (or plate in other problems) the probability that the egg remains sterile is

$$P = e^{-\lambda x}, \tag{1}$$

provided that if a particle is present, it will infect the egg. To make a dilution assay we choose dilutions $x_1, \ldots, x_m$ (m levels) and apply these to $n_1, \ldots, n_m$ eggs. If these result in $r_1, \ldots, r_m$ sterile eggs we can estimate λ by maximum likelihood. The theory has been given by Barkworth & Irwin (1938), and full references to work on this problem will be found in Finney (1952). If we plot the quantities $r_1/n_1, \ldots, r_m/n_m$ against $x_1, \ldots, x_m$ we get a set of point whose fit to the curve (1) can be tested by a χ^2 test. In a number of situations, however, it is found that (1) does not give a good fit. The estimation of λ is then completely invalid. In the present paper we consider why this happens, what types of curve may be fitted to the data and what they imply, and we also give a simple rapid test for such data fitting an exponential curve.

Now suppose that an individual virus particle has a probability $p (o \leqslant p \leqslant 1)$ of infecting an egg. Since the particles are selected at random, we can suppose this probability p to vary from particle to particle, p_0 being the average value. We must not, however, allow it to vary from egg to egg. Then it is easy to see that the probability of an egg remaining sterile is

$$e^{-\lambda p_0 x}, \tag{2}$$

and the whole theory goes through as before except that we are now estimating λp_0 instead of λ. λp_0 is the effective density of the virus particles when applied to eggs and by dilution assay it is not possible to estimate λ and p_0 separately.*

However, if p varies from egg to egg so that the eggs differ amongst themselves in their liability to be infected, the situation is quite different. Suppose that p has a probability distribution whose density function is $f(p) (0 \leqslant p \leqslant 1)$. Then the probability that the egg is sterile is

$$P = \int_0^1 e^{-\lambda p x} f(p)\, dp. \tag{3}$$

It will be noted that this is the moment generating function of $f(p)$. We have now to choose a plausible form for $f(p)$. Since p must lie between 0 and 1 it is natural to choose a distribution of β-type and thus suppose

$$f(p) = \frac{\Gamma(l+m)}{\Gamma(l)\Gamma(m)} \quad p^{l-1}(1-p)^{m-1} dp,$$

* λ and p_0 are in fact 'unidentifiable' in the statistical sense. It is possible to make assumptions about $f(p)$ which will make them 'identifiable' in theory, but not in practice.

https://doi.org/10.1017/S002217240002739X Published online by Cambridge University Press

where l and m are not necessarily integral. Then

$$P=\frac{\Gamma(l+m)}{\Gamma(l)\Gamma(m)}\int_0^1 e^{-\lambda px}p^{l-1}(1-p)^{m-1}dp. \tag{4}$$

However, the fitting of such a function to empirical data is not practicable. Not only is (4) not a tabulated function but it involves three parameters. I have been unable to find a suitable probability distribution on the range (0, 1) which involves only two parameters and for which the moment generating function (3) has a simple form.

Some approximate approach is therefore necessary. Suppose p has a mean value p_0. To allow for variation about this mean let p have a distribution of gamma type so that

$$f(p)=e^{-p/2a}\left(\frac{p}{2a}\right)^{l-1}\Big/\Gamma(l). \tag{5}$$

The mean value of this distribution is al which we equate to p_0. The objection to using this distribution is that it supposes that p can take all values from zero to infinity. This will probably not matter much if p_0 is near zero, but if p_0 is near unity and the variation about p_0 is not small compared with $1-p_0$ we cannot expect a good fit.

Inserting (5) in (3) we get

$$\begin{aligned}P&=\frac{1}{a\Gamma(l)}\int_0^\infty e^{-\lambda px}e^{-p/a}\left(\frac{p}{a}\right)^{l-1}dp\\&=(1+\lambda ax)^{-l}.\end{aligned} \tag{6}$$

This is the zero term in a negative binomial distribution. The mean value of p is $p_0=al$ and the variance of p about this mean value is a^2l.

Now consider the fitting of this formula to an observed dilution series. Suppose that n_i eggs are tested at each of m dilutions $x_1, \ldots, x_m$ and that r_i are found to be sterile at dilution x_i. We then find for the logarithm of the likelihood

$$L=\Sigma r_i \log_e P_i+\Sigma(n_i-r_i)\log_e(1-P_i),$$

where the logarithms are to the base e. Then putting $\lambda a=k$

$$P_i=(1+kx_i)^{-l},$$

and

$$\frac{\partial L}{\partial k}=-\Sigma\left\{\frac{r_i-n_iP_i}{P_i(1-P_i)}\right\}\frac{lx_iP_i}{1+kx_i},$$

$$\frac{\partial L}{\partial l}=-\Sigma\log_e(1+kx_i)\left\{\frac{r_i-n_iP_i}{1-P_i}\right\},$$

$$\frac{\partial^2 L}{\partial k^2}=\Sigma\frac{lx_i^2}{(1+kx_i)^2}\left\{\left(\frac{r_i-n_iP_i}{1-P_i}\right)+lP_i\left(\frac{r_i-n_i}{(1-P_i)^2}\right)\right\},$$

$$\frac{\partial^2 L}{\partial k\partial l}=\frac{1}{l}\frac{\partial L}{\partial k}+\Sigma\frac{lx_i(r_i-n_i)\,P_i\log_e(1+kx_i)}{(1+kx_i)\,(1-P_i)^2}$$

$$\frac{\partial^2 L}{\partial l^2}=\Sigma(\log_e(1+kx_i))^2\frac{P_i(r_i-n_i)}{(1-P_i)^2}.$$

It is convenient to set the calculations of these quantities out on a systematic

https://doi.org/10.1017/S002217240002739X Published online by Cambridge University Press

form. As an example consider some data of Parker (1940, table 1). These are as follows:

Dilution	No. inoculated	No. sterile
1	38	0
2^{-1}	39	5
2^{-2}	40	8
2^{-3}	40	15
2^{-4}	40	21
2^{-5}	40	30
2^{-6}	40	32
2^{-7}	40	35
2^{-8}	40	36

An attempt was first made to fit an exponential curve to these data. After several cycles λ was estimated to be 7·34 with a standard error of 0·71. Calculating the expected number fertile and sterile at each dilution, and grouping together the results with expectations less than 5 or greater than 35, we find $\chi^2 = 19{\cdot}02$ with 4 degrees of freedom. Clearly the exponential does not fit.

To fit the zero term of the negative binomial data guesses for k and l were found by putting the curve (6) through the points where $x = 0{\cdot}25$ and 0·0625. This gave $k = 11{\cdot}53$ and $l = 1{\cdot}1865$. On calculating the expected values with these parameters a χ^2 of 2·13 was found which would have had 3 degrees of freedom ascribed to it if the method of fitting had been most efficient. After two cycles had been calculated estimates $\hat{k} = 10{\cdot}915$ and $\hat{l} = 1{\cdot}2650$ were found. On calculating the expected numbers for these parameters, and grouping in the usual manner, χ^2 was found to be 2·25 with 3 degrees of freedom. This χ^2 was larger than for the initial guess, illustrating the fact that maximum likelihood and minimum χ^2 estimation do not give exactly the same results.

The fit being good, the variance-covariance matrix was calculated as the inverse of the matrix

$$\begin{pmatrix} \frac{\partial^2 L}{\partial k^2} & \frac{\partial^2 L}{\partial k \partial l} \\ \frac{\partial^2 L}{\partial l \partial k} & \frac{\partial^2 L}{\partial l^2} \end{pmatrix}.$$

From this we find the standard error of k to be 4·69 and of l to be 0·365. The correlation between the two estimators is $-0{\cdot}954$. The estimate of the 'degrees of freedom' $(2l)$ in the distribution of $2p$ is $2(1{\cdot}265) = 2{\cdot}53$ with a standard error of 0·73. This would seem to indicate that p varies considerably about this mean.

Formula (6) was also fitted to data of von Magnus (1951). This consisted of five series of dilutions of the same initial suspension of virus particles. A χ^2 of 25·60 was found with 10 degrees of freedom. This is beyond the 1 % point, indicating that formula (6) does not give a good fit.

We may draw the following conclusions from the above discussion:

(1) If the data do not fit an exponential curve the estimation of λ is not possible.

(2) If the distribution of p is approximately fitted by a γ-type distribution (which will probably be true if the mean value of p is near zero or if the standard deviation is small compared with $1 - p_0$) then formula (6) will probably give a good

https://doi.org/10.1017/S002217240002739X Published online by Cambridge University Press

fit and the value of l will give an estimate of the extent of the variation of p about p_0. It is only possible in this case to estimate λp_0 and not λ.

(3) A rapid test whether a given dilution series is likely to fit an exponential curve or be more spread out would be desirable and such a test is considered below:

A test for exponentiality in dilution series

Consider a twofold dilution series so that the mean densities in unit volume of the inoculum are $\lambda 2^m (m = \ldots, -1, 0, 1, \ldots)$, and such that $e^{-\lambda 2^m}$ covers a range from near unity to near zero. Write $p_m = e^{-\lambda 2^m}$ and then

$$P_{m+1} = P_m^2. \tag{7}$$

Suppose that n eggs (or plates) are tested at each dilution and the observed number sterile is f_m. As a test criterion calculate

$$T = \Sigma f_m (n - f_m). \tag{8}$$

If the series does not fit an exponential formula but is more spread out we will expect T to be inflated and thus T will provide as a useful test whether this is so. Assuming (7) we easily prove that

$$E\{f_m(n - f_m)\} = n(n-1)p_m(1 - p_m).$$

and

$$E(T) = n(n-1).$$

T is the sum of a number of bounded independent variates, and its distribution should not be far from normality. Consider its variance. We have

$$\text{var}(T) = \Sigma E\{f_m^2 (n - f_m)^2\} - \Sigma\{E f_m (n - f_m)\}^2$$

and after some reduction this is equal to

$$n(n-1)\Sigma\{(n-1)p_m - (5n-7)p_m^2 + (8n-12)p_m^3 - (4n-6)p_m^4\}.$$

Summing over all values of m and remembering that $p_{m+1} = p_m^2$ we get

$$\text{var}\ (T) = n(n-1)\{5 - 3n + (8n-12)(\Sigma(p_m^3 - p_m^4))\}.$$

The sum $\Sigma(p_m^3 - p_m^4)$ varies with p_0 but belongs to an interesting class of series which are almost independent of a parameter in them. In fact

$$\Sigma(p_m^3 - p_m^4) = 0{\cdot}4150375,$$

with an error never greater than 0·000004, i.e. one part in 10^5. Taking the value as 0·4150375 we get

$$\text{var}(T) = n(n-1)\{0{\cdot}3203n + 0{\cdot}0195\},$$

and the standard error of T is the square root of this. The power of this test, for alternatives of the type considered, may well be as great or greater than that of the χ^2 test since the latter is an 'overall' test of divergence from expectation.

Taking Parker's data as an example, we cannot apply the test strictly since the number of eggs is not the same at each dilution. If, however, we altered the number tested at dilution 1 and 2^{-1} to 40 and kept the number sterile constant, and also supposed that at dilutions beyond 2^{-8} the number sterile would have been 40, we get

$$T = 2080, \quad E(T) = 1560, \quad \text{S.E.}(T) = 141{\cdot}5;$$

$$\frac{T - E(T)}{\text{S.E.}(T)} = 3{\cdot}67;$$

so that a significant deviation is observed.

https://doi.org/10.1017/S002217240002739X Published online by Cambridge University Press

The above test can only be applied when the series is long enough for p_m to be nearly unity and zero at the two ends. A similar test can be set up for fourfold and tenfold dilutions, but is less satisfactory because $E(T)$ then varies somewhat with p_0.

I am indebted to Dr S. Fazekas de St Groth for proposing this problem and for some helpful discussions.

REFERENCES

BARKWORTH, H. & IRWIN, J. O. (1938). *J. Hyg., Camb.*, **38**, 446.
FINNEY, D. J. (1952). *Statistical Method in Biological Assay*. London: Charles Griffin and Co.
PARKER, R. F. (1940). *J. exp. Med.* **71**, 439.
VON MAGNUS, P. (1951). *Acta path. microbiol. scand.* **28**, 1951.

(*M.S. received for publication* 19. XII. 53)