Estimation of the Parameters in Exponential Decontamination Models

Richard G. Cornell

February 1, 1967

Technical Report Number 8

NASA Grant Number NGR-10-004-029

Department of Statistics
Florida State University
Tallahassee, Florida

ESTIMATION OF THE PARAMETERS IN EXPONENTIAL DECONTAMINATION MODELS

Richard G. Cornell

1. INTRODUCTION AND PROBABILITY MODELS

This report is concerned with the estimation of parameters in decontamination models based upon the assumption that the probability, p_t, that a single microorganism is alive at time t in a given environment is exponential, that is, is of the form

$$p_t = \mu^t. \tag{1}$$

Although such models are of general interest in microbiology, they are of particular interest in the development of decontamination standards and procedures for the spacecraft sterilization program of the National Aeronautics and Space Administration. There are several equivalent ways in which μ is commonly expressed, namely,

$$\mu = e^{-\beta},\ 0 < \beta < \infty, \tag{2}$$

$$\mu = 10^{-\alpha},\ 0 < \alpha = (\log_{10} e)\beta = (0.43429)\beta < \infty, \tag{3}$$

$$\mu = 10^{-1/D},\ 0 < D = 1/\alpha = 1/(0.43429)\beta < \infty, \tag{4}$$

where D is defined as the t value for which $p_t = 0.10$. Note that $0 < \mu < 1$. Assumption (1) implies that if n_o viable microorganisms of a given type are exposed initially and if the deaths of these microorganisms are independent of each other, then the probability of n organisms being

alive at time t is

$$f(n) = \binom{n_0}{n} (\mu^t)^n (1-\mu^t)^{n_0-n}, \tag{5}$$

where $n_0 > 0$ and $\binom{n_0}{n}$ is the binomial coefficient representing the number of ways of selecting n items from among n_0 items. Since p_t is usually small relative to n_0 for $t > 0$ because of the exponential die-off assumed in equation (1), the binomial probability model given by (5) can instead be expressed using the Poisson probability function

$$g(n) = \frac{e^{-\lambda_t}(\lambda_t)^n}{n!}, \tag{6}$$

where

$$\lambda_t = n_0 p_t = n_0 \mu^t, \tag{7}$$

which is the mean of the probability functions given by both (5) and (6).

For large t, λ_t is very small and the probability of more than one viable organism is especially small, so that instead of dealing with the count n of viable organisms, it is more reasonable to record merely whether or not any microorganisms have survived to time t. In this instance, the probability no microorganisms survive is $\exp(-\lambda_t)$ as given by substituting $n = 0$ in (6). Therefore, the probability at least one microorganism survives is $1-\exp(-\lambda_t)$. In this situation it is common to examine, say, m_0 containers for survival at each value of t selected, where each container contains n_0 microorganisms initially. Then for any given exposure t, the probability that m containers show survival is given by the binomial probability function

$$h(m) = \binom{m_0}{m} (1-e^{-\lambda_t})^m (e^{-\lambda_t})^{m_0-m} \qquad (8)$$

The assumption of exponential die-off given by (1) is commonly made in analyzing decontamination data. Also, it is common practice to analyze counts of organisms with expected counts λ_t as given by (7), for probability models given by (5) and (6), when the expected number of survivors is larger than one and to analyze quantal survivor data, using a model like (8), when λ_t is thought to be less than one. However, experimental data often do not exhibit die-off as described by equation (1). The most common departure is for the rate of die-off, that is, the parameter μ in (1), to change with time. One way to overcome this difficulty is to apply equation (1) and the probability models (5), (6) and (8) derived from it, over only part of the range of t for which experimental data has been obtained. For instance, rapid early die-off is often ignored and the models presented here are then used to describe the data after a time when the rate of die-off is more moderate.

Another approach for dealing with departures from assumption (1) is to assume a different distribution of the underlying tolerances to exposure of the microorganisms being studied. Assumption (1) can be thought of as an assumption that the probability a single microorganism dies before time t is $1-p_t = 1-\exp(-\beta t)$. This is a cumulative tolerance distribution for the single-parameter exponential probability density function given by

$$\frac{\partial}{\partial t}(1-e^{-\beta_t}) = \beta e^{-\beta t} . \qquad (9)$$

Thus assumption (1) is equivalent to an assumption of exponentially distributed tolerances in the population of microorganisms. The well-known probit and logit methods of analysis have been developed and extensively applied to a probability model similar to that given by (5) for the more flexible two-parameter normal and logistic tolerance distributions, respectively. Both of these tolerance distributions may be expressed on either an arithmetic or logarithmic t scale. Another tolerance distribution which is a direct extension of (1) is the two-parameter gamma distribution. Research is certainly needed to determine appropriate tolerance distributions for a variety of organisms and types of exposure. However, assumption (1) holds in many experimental situations, so it is appropriate to concentrate in this report on methods of estimation for models derived from that assumption.

In Sections 2 and 3 of this report, the estimation of μ, and n_0 when applicable, is discussed for the probability models given by equations (5) and (6), respectively. The application of techniques already presented in the statistical literature is emphasized in these sections although some new results are stated in Section 3. Once μ is estimated, estimates for α, β or D may be computed by substituting the estimate of μ into equation (2), (3) or (4). Estimation for the model given by equation (8) is not discussed further because little work has been done on this model when independent observations are taken at each exposure t. This model has been studied when the same organisms are observed at a succession of t values. In such contexts it is usually referred to as the Gompertz curve.

Section 4 gives a summary of research underway or planned for the future on the statistical aspects of microbial assay.

2. ESTIMATION FOR MODEL (5)

If n_0 is assumed to be known, the model given by equation (5) for each of k exposures t_i with corresponding independent observed counts n_i, i=1,2,...,k, is one often encountered in bioassay and is a model for which several statistical procedures have been proposed. Several situations in the biological sciences in which this model arises are described in Technical Report Number 3 by Cornell and Speckman [1966], which was previously sent to NASA. This report also presents, illustrates and compares several estimation procedures and it contains an extensive bibliography. The conclusion of that report is that the method of maximum likelihood is a good technique for any sample size. This method is discussed, for instance, by Peto [1953], Finney [1964] and Cochran [1950] and also called the most probable member technique. It can be used for any spacing of exposures. The simple method of partial totals, developed by Speckman and Cornell [1965], was suggested as an alternative to maximum likelihood for small sample sizes and equally spaced t values. The Fisher [1921] and Spearman methods (see Johnson and Brown [1961]), which are similar and computationally easy, were recommended alternatives to maximum likelihood regardless of the sample size for t values equally spaced on a logarithmic scale. Other discussions of estimation techniques for this model are given by Cornell [1965], Cornfield [1954] and Finney [1964].

If n_0 is not known, one possibility is to use the fact that n_0 is large relative to p(t) so that the Poisson model (6) can be used as an approximation to (5), as stated in the introduction. Estimation for model (6) is discussed in the next section. Another possibility is to retain model (5) and estimate both n_0 and μ. Bowman and David [1962] develop and illustrate an iterative maximum likelihood solution for this purpose. However, this model has not received a great deal of attention in the statistical literature when n_0 is unspecified.

3. ESTIMATION FOR MODEL (6)

Let us now investigate the estimation of n_0 and μ for model (6) for independent observations n_i made at corresponding exposure times, t_i, i=1,2,...,k. The likelihood function for (n_0,μ) is

$$L(n_0,\mu) = \prod_{i=1}^{k} g(n_i) = \exp(-n_0 \sum_{i=1}^{k} \mu^{t_i}) n_0^N$$

$$= \exp(-n_0 \sum_{i=1}^{k} \mu^{t_i}) \; n_0^N \mu^T / \prod_{i=1}^{k} (n_i!), \qquad (10)$$

where $g(n_i)$ is given by (6) and (7) and where

$$N = \sum_{i=1}^{k} n_i; \; T = \sum_{i=1}^{k} t_i n_i. \qquad (11)$$

Then the equations for the maximum likelihood estimators of n_0 and μ, denoted by $\hat{n}_0$ and $\hat{\mu}$, respectively, are

$$\frac{\sum_{i=1}^{k} t_i \hat{\mu}^{t_i}}{\sum_{i=1}^{k} \hat{\mu}^{t_i}} - \frac{T}{N} = 0, \tag{12}$$

$$\hat{n}_0 = N / \sum_{i=1}^{k} \hat{\mu}^{t_i}. \tag{13}$$

If n_0 is known, $\hat{\mu}$ is given by

$$n_0 \sum_{i=1}^{k} t_i \hat{\mu}^{t_i} - T = 0. \tag{14}$$

Equations (12) and (14) for $\hat{\mu}$ require iterative solution and they lead to solutions for $\hat{\mu}$ only if not all of the t_i values are the same. Also, from the form of (10), it can be seen that (N,T) is sufficient for (n_0,μ). Since the maximum likelihood estimation equations (12), (13) and (14) only involve the n_i observations in the fomation of N and T as given by (11), the maximum likelihood estimators are also sufficient and there is no need to consider other possible estimators.

It is common practice to use equally spaced t_i values, either on and arithmetic or logarithmic scale. Williams [1961] considers maximum likelihood estimation and the significance of departures from the model, as well as asymptotic formulas for tail probabilities, for t_i values equally spaced arithmetically, that is, for $t_i = i - 1$, $i = 1,2,\ldots,k$. To avoid iterative calculations, he presents a table of $\hat{\mu}$ values for steps of 0.05 of the argument $V = T/(k-1)N$ for $k = 2,3,4$ and 5, which is

the range of k values most likely to be used in biological experimentation. His table may be used whether or not n_0 is known. Williams also gives the asymptotic covariance matrix for $(\hat{n}_0, \hat{\mu})$, which has an asymptotic bivariate normal distribution when the t_i's are equally spaced.

Tables have been computed, as a part of the research reported here, from which $\hat{\mu}$ can be determined for t_i values which are equally spaced on a logarithmic scale, that is, for

$$t_i = bc^{i-k}, \; i = 1,2,\ldots,k; \; b > 0. \tag{15}$$

These tables will be included in a later report which will be prepared when current research on estimation for model (6) when (15) holds is completed. These tables were computed using equation (12) under the assumption that n_0 is unknown. It is assumed in (15) that the t_i are ordered from smallest to largest, that is, that $c > 1$. Common values of c used in experimentation are 2, 4 and 10. The tables are given in terms of the statistic

$$V = \frac{T}{t_k N} = \frac{T}{bN}, \; 0 \leq V \leq 1, \tag{16}$$

and the estimator

$$\hat{\gamma} = \hat{\mu}^b. \tag{17}$$

Values of $\hat{\gamma}$ are tabled for V ranging by steps of 0.01 from 0 to 1 for all possible combinations of $c = 2$, 4 and 10 and $k = 2,3,\ldots,10$. Once $\hat{\gamma}$ is determined, $\hat{\mu}$ is computed as $\hat{\gamma}^{1/b}$. In the preparation of these tables, instead of specifying V and solving iteratively for $\hat{\gamma}$, $\hat{\gamma}$ was varied from

0 to 1 in steps of 0.01 and the corresponding value of V was calculated. Arranging the tables with V as the argument has been easily done because V is a monotomically increasing function of $\hat{\gamma}$ for the model under consideration:

Examination of the tables relating V and $\hat{\gamma}$ has shown that for most of the range of V from 0 to 1 and for most of the combinations of c and k investigated, $\hat{\gamma}$ can be closely approximated by a linear function of V. This means that much of the extensive tabling described above can be done by citing the two coefficients of such a linear function and a range for V for each combination of c and k. This fitting of straight lines has not been completed and this is the reason for delaying the presentation of these tables. Apart from reducing the amount of tabling required, the determination of simple explicit relationships between V and $\hat{\gamma}$ is important because it would make it possible to compute small sample variances for $\hat{\gamma}$ which in turn could be used to approximate small sample variance for $\hat{\mu}$, $\hat{\alpha}$, $\hat{\beta}$ or $\hat{D}$.

The large sample distribution for large k of the vector $(\hat{n}_0, \hat{\mu})$ of maximum likelihood estimators has been investigated for model (6) when (15) holds. The asymptotic covariance matrix has been shown to exist and be positive definite provided that the constant c in (15) be replaced by a function of k which has the asymptotic properties as $k \to \infty$ of $c_k = c^{1/k}$. That is, when a large number of exposures are used in an experiment, the ratio of successive exposures should be closer to one than when a small number of exposures is used. The mathematical rational behind this statement is complicated even though the conclusion is simple and

reasonable. This rational has been worked out and will be given in the subsequent report which will concentrate on model (6) when the spacing of t values is given by (15).

4. FUTURE RESEARCH

The discussion of decontamination probability models and estimation procedures in earlier sections has pointed out the need for further research on the statistical aspects of microbial assay. The presentation of models in Section 1 includes a discussion of possible alternative models and alludes to the fact that the statistical literature already contains considerable work on such models. Reference is also made in Section 1 to the need for estimation techniques for model (8). Research on new models is presently underway by several investigators associated with the spacecraft sterilization program. The author of this report plans to continue his work on models and estimation reported here with the following goals:

1) To relate research on model building and estimation in microbial assay with work already in the statistical literature.
2) To develop methods to choose between alternative models.
3) To develop estimation techniques for model (8).
4) To complete the work referred to in Section 3.
5) To develop small sample variance formulas for the estimation methods referred to in Section 2.

A sixth related and very important goal is as follows:

6) To develop probability statements concerning the time required to achieve sterility. In other words, to develop a

probability statement concerning the exposure required to extrapolate, say, model (6) to a specified probability level such as 10^{-4}.

Another goal upon which considerable progress has already been made is as follows:

7) To develop Bayesian estimation and experimental design procedures for models (5) and (6). Such methods allow past experimental evidence to be explicitly incorporated into the analysis of current experimental results.

REFERENCES

1. Bowman, K. A. and David, H. A. [1962]. Estimation of the strength of a radioactive source. Technical Report Number 54 for Army Research Office Grant DA-ORD-29 to Virginia Polytechnic Institute, Blacksburg, Virginia.

2. Cochran, W. G. [1950]. Estimation of bacterial densities by means of the "most probably number". Biometrics 6, 105-116.

3. Cornell, R. G. [1965]. Spearman estimation for a simple exponential model. Biometrics 21, 858-864.

4. Cornell, R. G. and Speckman, J. A. [1966]. Estimation for a simple exponential model. Technical Report Number 3 for NASA Grant NGR-10-004-029 to Florida State University, Tallahassee, Florida.

5. Finney, D. J. [1964]. Statistical methods in biological assay (2nd ed.), Section 21.5. New York: Hafner Publishing Company.

6. Fisher, R. A. [1921]. On the mathematical foundations of theoretical statistics. Phil. Trans. Roy. Soc. A 222, 309-368.

7. Johnson, E. A. and Brown, B. W., Jr. [1961]. The Spearman estimator for serial dilution assays. Biometrics 17, 79-88.

8. Peto, S. [1953]. A dose-response equation for the invasion of micro-organisms. Biometrics 9, 320-335.

9. Speckman, J. A. and Cornell, R. G. [1965]. Estimation for a one-parameter exponential model. J. Amer. Stat. Ass. 60, 560-572.

10. Williams, E. J. [1961]. Fitting a geometric progression to frequencies. Biometrics 17, 584-606.