PROGRESS REPORT: NASA GRANT NGR-10-004-029 to
FLORIDA STATE UNIVERSITY

PROJECT TITLE: Biostatistics of Space Exploration:
Microbiology and Sterilization

PRINCIPAL INVESTIGATOR: Richard G. Cornell

August 24, 1966

Department of Statistics
Florida State University
Tallahassee, Florida

PROGRESS REPORT: NASA GRANT NGR-10-004-029 to FLORIDA STATE UNIVERSITY

PROJECT TITLE: Biostatistics of Space Exploration: Microbiology and Sterilization

PRINCIPAL INVESTIGATOR: Richard G. Cornell

August 24, 1966

Introduction

The purpose of this grant is to provide continuing research support on statistical problems arising in investigations on the decontamination of spacecraft and related studies of microbial life. The work under this grant is of two different types, namely, statistical consultation to Mr. Lawrence B. Hall, Planetary Quarantine Officer, to the Spacecraft Sterilization Advisory Commmttee and to NASA contractors; and basic research on contamination probability models and on microbial assay.

With regard to consultation on research problems, the principal investigator has attended two meetings of the Spacecraft Sterilization Committee during the last six months. He also spent a week during March, 1966 consulting with the personnel of the Sterilization Group of the Jet Propulsion Laboratory and he has evaluated reports on decontamination probability models for Mr. Lawrence B. Hall. He recently agreed to prepare for Mr. Hall a standard nomenclature of mathematical terms to be used in the computation of probabilities of planetary contamination. This nomenclature is given in the enclosed Technical Report Number 5.

The remainder of this report will be concerned with a description of progress made on research of interest to NASA. In each instance a brief description of the research project and its relevance to the space quarantine program will be given. The current status of each project will be indicated.

In this and subsequent progress reports, a detailed description of research will be given only for projects which are complete enough for the issuance of technical reports and usually subsequent submission for publication in professional journals. For such projects, technical reports or reprints will be enclosed if not sent earlier, or will be cited if copies have already been sent to NASA. For other projects target dates for the completion of technical reports will be given.

Spacecraft Sterilization Probability Models

One way to represent the information available on the probability of contaminating Mars is in the form of a probability distribution function which would relate a probability based on current knowledge with any specified interval on a contamination probability scale. For instance, if one were to ask: "Based on our current knowledge, what is the likelihood that the probability of contaminating Mars (on a given probe or during a given period of time) is less than 10^{-4}?", then the answer would require that the area under the probability distribution function to the left of 10^{-4} be determined. It should be emphasized that under specified conditions, the true probability of contaminating Mars is constant. The distribution function proposed would not involve the true probability of contaminating Mars as a random variable, but would instead be a distribution on our knowledge of the probability of contaminating Mars.

The distribution function being devised would have to be defined over the domain from zero to one, the possible range of a contamination probability, and would have to be flexible enough to give different patterns as our information is increased and refined. The parameters of the distribution should have meaning in themselves. The distribution should be of simple functional form, or at least be extensively tabled. It should be sensitive and adaptable to new information, that is, it should be such that methods can be developed to modify it in accord with new research findings and new strategies for Martian exploration.

Work has been begun here to develop a probability distribution function on our knowledge on the probability of contaminating Mars. Our approach has been to develop appropriate distributions for spacecraft parts or components which are of a type which can be subjected to microbial assay. Later methods will be developed to combine such distributions determined for different components in order to obtain a distribution on our information on the probability that the spacecraft as a whole is contaminated. Then ways will be sought to modify this spacecraft probability distribution in accord with knowledge of missions to be undertaken.

It is common practice and reasonable, based on the assumption that the number of contaminating particles on a componnet of a specified kind will follow a Poisson distribution, to assume that the probability that the component is contaminated is given by $p = (1-e^{-\lambda})$ where λ is the mean number of contaminants per component, or, more generally, by $p = (1-e^{-\lambda d})$, where d is a dosage in area, time or dilution factor, and λ is a mean per dosage unit. Putting a probability distribution on our knowledge of p over the domain from zero to one is equivalent to determining a probability distribution on our knowledge of λ over the domain from zero to infinity. A flexible, unimodal distribution which can be used to describe information on λ is the gamma-distribution with density function

$$g(\lambda) = \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} e^{-\lambda/\beta}\lambda^{\alpha}, \; 0 < \lambda < \infty, \; \alpha > -1, \; \beta > 0, \tag{1}$$

where here λ is regarded as a random variable even though it is our information about λ which is actually variable. Initially, the parameters α and β would be chosen to represent current evidence concerning the contamination load on a component, or a unit area of a component. In case no such information is available, the initial values taken for α and β would be such that $g(\lambda)$ would be very diffuse. Next suppose that n components of the type under consideration are examined and x of them are found to be free of contamination. The probability that this occurs given λ is

$$\ell(x|\lambda) = \binom{n}{x} e^{-\lambda dx}(1-e^{-\lambda d})^{n-x} \tag{2}$$

using the binomial distribution. This probability function can be combined with $g(\lambda)$ as given by (1) using Bayes' formula to give a new probability distribution on our knowledge of λ which would replace (1) and would be responsive to the experimental evidence. If the new density function for λ should be of the same form as (1), that is, be a gamma density function, but possibly with different parameter values, then this approach would allow us to easily incorporate new evidence into our probability distribution on our information on λ.

Mr. Andres Petrasovits has worked on this problem as a general method of analysis for serial dilution experiments where in addition to modifying the density function (1) at each stage of experimentation, he also estimates λ by $\hat{\lambda}$. He also considers the possibility of an experiment containing several dilution (dosage) levels, d_i, $i = 1,\ldots,k$. His work is formulated in Technical Report Number 6, which is a preliminary report.

His estimate of $\hat{\lambda}$ is given by formula (2.10) or (4.4) in that report, and is too complicated for practical use, although he shows in the last part of the report that it is a ratio of rational combinations of polygamma functions which are tabled. Also, the modified density function he obtains for our knowledge of λ is not of the same form as (1), but is much more complicated. He calls this density function the posterior density of λ and it is given in formula (2.9).

Since he wrote the essay which constitutes Technical Report Number 6, Mr. Petrasovits has succeeded in approximating $\hat{\lambda}$ by a simple function of n and x and the parameters of (1) and he also has been able to closely approximate the posterior density of λ by a gamma density of the form of (1) but with parameters modified by the experimental evidence. His derivation and evaluation of his approximate method is mathematically involved, but the results are simple and can be incorporated into the component level of the general contamination probability level we are seeking. His evaluation of his results is not complete, but numerical comparisons with the exact but cumbersome formulas given in Technical Report Number 6 indicate that his approximate approach is highly efficient in terms of a comparison of relative risks, in fact, usually more than 98% efficient for most parameter values investigated. It is anticipated that these results will be presented to NASA in the form of a technical report by January 1, 1967. Mr. Petrasovits is being employed under the NASA grant this summer but will write a report on the recent results this fall after he returns to his job with the Department of Agriculture of Canada.

In a manner similar to that employed by Mr. Petrasovits, it is expected that count data on contamination will be incorporated into our probability model. Actually, it is thought that this is an easier problem than that encountered with the binomial type data already considered. After this is done, then consideration will be given to combining probability distributions of the information available on different components and to incorporating non-experimental evidence into our model.

In addition to the work done this summer by Mr. Petrasovits and the principal investigator, Dr. Duane Meeter has been carrying out research and taking part in seminars on astronautics and exobiology at the Ames Research Center during the last two months while being supported by this grant. It is thought that his experience there will add fresh insight into our work under this grant on the development of contamination probability models. The research he has started this summer will be reported on later. However, enclosed is Technical Report Number 7 which was prepared by Dr. Meeter before he left for Ames. Although this report as it stands is a highly theoretical investigation of nonlinear estimation, it is thought that further consideration of its implications will be helpful in more applied nonlinear estimation problems, including bioassay.

In the February, 1966 Progress Report, reference was made to the dissertation research of Mr. William E. Lever on a decision theory approach to testing hypotheses with likelihood ratio tests which will be applicable to making decisions with regard to the achievement of spacecraft sterilization probability requirements. Mr. Lever is presently writing his dissertation and his research is virtually completed. A technical report on this project will be prepared this fall. The only support being given Mr. Lever by NASA consists of paying for part of the computer expenses encountered in his research.

Statistical Procedures for Microbial Assays and Related Nonlinear Statistical Estimation Techniques

The work of the principal investigator with Mr. Petrasovits, which has already been described, is really a bioassay estimation technique, but has been described in the previous section because of its close relationship to our contamination probability model. Similarly, although Dr. Meeter is concentrating on model building considerations this summer, his work in Technical Report Number 7 is on nonlinear estimation. In this section consideration will be given to research on nonlinear statistical estimation which is directly applicable to microbial assays and which is being carried out without reference to more general contamination probability models.

The February 1966 Progress Report mentioned work of the principal investigator with John J. Beauchamp on simultaneous estimation for several nonlinear regression equations. A technical report on this work was included with that Progress Report and later reprints of the published manuscript (Beauchamp, J. J. and Cornell, R. G., "Simultaneous Nonlinear Estimation," Technometrics 8 (1966), 319-326) were sent to NASA. That paper dealt with generalized least squares estimation. Mr. Beauchamp and I are now working with two other estimation schemes for simultaneous nonlinear estimation which are particularly applicable to the description of the movement of a tracer, say microbial contamination, through a compartmental system. Research on this project is almost completed and Mr. Beauchamp is writing up the results in dissertation form. It is anticipated that two additional technical reports will be prepared from this work during the next six months. The only NASA support being given Mr. Beauchamp is the payment of some of the cost of using a computer in his research.

Another project which was mentioned in the previous Progress Report and which will only require support from NASA for computing costs, is that being carried out jointly with Mr. Donald C. Martin. This project involves evaluating the small sample variance and mean square errors of various estimators of the instantaneous die-off rate λ in microbial assays for the binomial model given in equation (2), but without specifying a distribution on λ as we previously did in (1). This work has been started successfully but has been delayed because Mr. Martin, who is not being paid from the grant, has been completing a Ph.D. dissertation on another topic. Work will be resumed on this project in the fall and it is anticipated that a Technical Report will be written early in 1967 summarizing the results. A graduate student assistant, Mr. Richard Gero, has been employed, starting on September 1, to work one-quarter time on NASA research projects. He is an experienced computer programmer and he will initially spend most of his time on this project.

Dr. Myles Hollander is currently developing nonparametric, distribution-free, procedures for testing parallelism of regression lines, a problem which often occurs in bioassay. Dr. Hollander has developed easily applied procedures under very general assumptions for comparing two regression lines when the two samples, from which the two lines are calculated, are of equal size. He has not completed the evaluation of these procedures although work thus far indicates that he has developed procedures which are highly efficient relative to the corresponding normal-theory t-test. He anticipates completing this work and perhaps extending it to unequal sample sizes during the coming year. A Technical Report will be prepared when this work is completed. Dr. Hollander has received a major portion of his support during the past year from funds from the NASA grant, but the principle investigator, who has been more actively engaged in the work under this grant, has only been supported for four months by this grant.

Work by the principal investigator, Dr. Richard G. Cornell, is underway on the estimation of parameters in a bioassay model with logarithmic die-off and with microbial counts observed for each plate studied. Specifically, attention is focused on the estimation of the parameters α and β in the model

$$E(Y_i) = \alpha e^{-\beta t_i}, \ i = 0,1,\ldots,k-1; \ \alpha,\beta > 0,$$

where E denotes "expected values of" and where it is assumed that the Y_i are independently distributed Poisson variables. For equally spaced t_i, Williams (Williams, E. J. "Fitting a geometric progression to frequencies," Biometrics 17 (1961), 584-606), considers the problem of estimation and the significance of departures from the model. The purpose of the current work is to extend this work to the situation where the t_i are equally spaced on a logarithmic scale. For this model, the maximum likelihood estimate is a function of sufficient statistics, so only maximum likelihood estimation has been studied. Equations for the maximum likelihood estimates have been obtained, but they can only be solved iteratively. Tables of maximum likelihood estimates have been computed for several spacings of dosages on the logarithmic scale. The large sample properties of the estimators have been studied. The small sample distributions of the estimators will be studied next through the use of empirical polynomial regression equations for the estimates. Goodness-of-fit tests of the underlying model will also be investigated. It is expected that this work will be completed this fall with the help of Mr. Gero in obtaining the empirical equations for the maximum likelihood estimates from the tables already prepared. A technical report on this work should be completed before the end of this year.