

N73-23829
NASA CR-132001

CASE FILE COPY

PROBABILISTIC MODELS OF PLANETARY CONTAMINATION

Prepared for:

PLANETARY PROGRAMS OFFICE
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546

CONTRACT NASW-2451



STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 • U.S.A.



STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

April 1973

PROBABILISTIC MODELS OF PLANETARY CONTAMINATION

By: J. MICHAEL HARRISON
D. WARNER NORTH

Prepared for:

PLANETARY PROGRAMS OFFICE
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546

CONTRACT NASW-2451

SRI Project MSU-2274

CONTENTS

I	INTRODUCTION	1
II	THE SAGAN-COLEMAN FORMULA	5
III	A CLASSICAL PARALLEL PROBLEM	10
IV	THE EFFECT OF DEPENDENCIES IN THE CONTAMINATION PROBLEM	16
V	ENRICHING THE PROLIFERATION MODEL	31
VI	SUMMARY AND CONCLUSIONS	43
APPENDIX	DERIVATION OF THE ERROR BOUND	46
REFERENCES		48

ILLUSTRATIONS

1	The Hierarchy of Planetary Quarantine Decisions	2
2	Distribution Implying Independent Trials	13
3	Distribution Implying Perfectly Dependent Trials	13
4	Uniform Distribution	13
5	Probability of All Heads Versus Number of Trials	15
6	Basic Logical Structure of the Sagan-Coleman Model	16
7	Basic Logical Structure of the Proposed Model	17
8	Illustrative Landing Model	17
9	Probability Tree Form of the Proposed Model	19
10	Illustrative Release Model	21
11	Illustrative Distribution of Region 1 Survivability	22
12	Discrete Approximation for the Region 1 Survivability Distribution	22
13	Illustrative Distribution of Region 2 Survivability	23
14	Discrete Approximation for the Region 2 Survivability Distribution	23
15	Some Martian Characteristics and Physical Processes Critical to Biological Contamination	32
16	An Illustrative Proliferation Model (for a Single Released Organism) to Generate $P(G B_j)$	34
17	Illustrative Distributions for Extent of Water in Regions 1 and 2	37
18	Illustrative Distribution of Mean Lifetime for Areophiles Released on the Martian Surface	37

I INTRODUCTION

A continuing source of concern in man's exploration of space is the potential for biological contamination of other planets. The possibility exists that a planet such as Mars is capable of supporting terrestrial microorganisms; hence, the landing of unsterilized space vehicles could lead to widespread contamination.

Several reasons for concern of planetary contamination are evident. One cause for alarm is the possibility of greatly impeding subsequent attempts to detect indigenous life. In addition, many experts simply feel that contamination would be morally reprehensible. Another concern suggests that a proliferation of earthly microbes might impede future efforts toward colonization.^{1*} The apparent solution to these concerns is vigorous sterilization of all spacecraft engaged in planetary exploration. Sterilization can be costly, however, in terms of both direct expense and reduced spacecraft reliability.

In response to concern over potential contamination of Mars, a planetary quarantine resolution was adopted in 1966 by the Committee on Space Research (COSPAR) of the International Council of Scientific Unions. This resolution required that the spacefaring nations conduct their unmanned explorations of Mars in such a way that the total probability of contamination during a specified quarantine period not exceed 10^{-3} . In its desire to conform with this standard, the National Aeronautics and Space Administration (NASA) established a planetary quarantine policy, administered by a planetary quarantine officer (PQO) whose responsibilities are

* All references are listed at the end of this report.

broadly depicted in Figure 1. Given tentative plans as to the number and character of future missions to Mars, the quarantine officer establishes a maximum permissible value for the probability of contamination from individual missions. These upper bounds, widely known as mission "allocations," are set in such a way as to ensure program conformance with the COSPAR constraint. The final level of decision-making is then to determine the engineering specifications that will most efficiently meet the mission constraint. Some primary means of reducing contamination probabilities from individual missions are trajectory biasing for fly-bys, altitude restrictions for orbiters, and spacecraft sterilization for landers. The central analytical problem of the planetary quarantine program is to determine the probability of contamination under a variety of different mission specifications. This we call the "mission contamination problem." Its solution is a prerequisite for the determination of which specifications meet a given mission constraint.

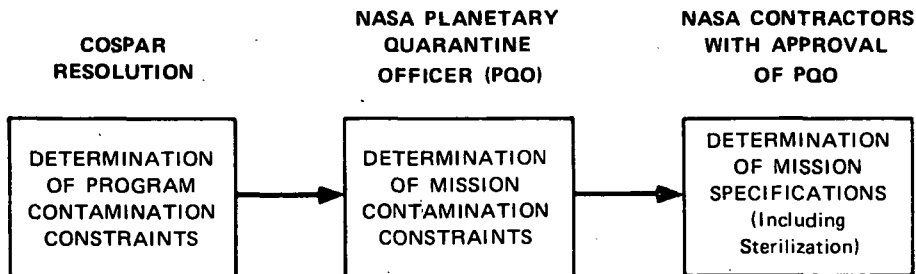


FIGURE 1 THE HIERARCHY OF PLANETARY QUARANTINE DECISIONS

The purpose of this report is to discuss at a broad conceptual level the analytical basis for planetary quarantine standards and procedures. Attention is focused on the final phase of the decision process shown in Figure 1, and discussion is restricted to the case of landing missions. To a large extent, current standards and procedures are based on the probabilistic model of planetary contamination advanced by Sagan and Coleman, whose work provided both the stimulus and the theoretical foundation for the COSPAR resolution.²

Section II begins with a brief account of the Sagan-Coleman model. We note that certain independence assumptions critical to their analysis give an overly simplified characterization of the problem. It should be emphasized, however, that our criticisms concern the structure of the Sagan-Coleman model rather than the numerical values assigned to various input parameters. Considerable debate and discussion^{3,4} of parameter values have taken place but the basic structural assumptions and resulting formulas are widely accepted by COSPAR, NASA, and NASA contractors as a means of determining sterilization requirements for Project Viking and other future unmanned planetary missions.

In Section III, we discuss a classical problem in probability theory that provides--at least in our opinion--a close conceptual parallel to the type of dependence present in the contamination problem. The Sagan-Coleman model fails to account for this dependence.

In Section IV, we indicate how the shortcomings of the Sagan-Coleman analysis can be remedied by the adoption of a slightly richer model structure. By contrasting this formulation of the problem with those now in use, we can identify both the type and the probable magnitude of error introduced by current analytical procedures. Our conclusion is that a relatively minor modification of the basic Sagan-Coleman formula yields a good approximation to the true probability of contamination in the range of interest. The question remains, however, of how to obtain the input parameters called for in this formula.

In Section V, we argue that those parameter values relating to the survivability of terrestrial organisms in various regions of Mars should be based on a model of potential mechanisms for growth and proliferation. An example of such a model will be presented in the section.

A general theme underlying our entire presentation is the firm conviction that any analytical treatment of planetary quarantine requirements

must lie squarely in the domain of subjective probability assessment. Much physical evidence exists relevant to the contamination issue (such as the findings of Mariner 9), and large amounts of data (such as the lethality of various sterilization treatments) are available on various aspects of the problem. Inevitably, however, the quantification of expert scientific judgment will provide the ultimate means for integrating all available information into an assessment of the total risk of contamination.

The model builder's task, therefore, is to structure a problem so as to ease the burden of direct assessment as much as possible without imposing or presuming relationships that may conflict with the expert's judgment. The weaknesses of the Sagan-Coleman model do not arise from any erroneous characterization of physical processes or objects; the question is whether the assumptions of the model are consistent with the current state of scientific information. A separate issue is the ability of human beings--even very sophisticated ones--to directly assess probabilities as small as 10^{-4} . The Sagan-Coleman model requires experts to make just such assessments. A relatively detailed proliferation model of the type discussed in Section V is designed primarily to help overcome this difficulty, although it yields additional benefits in improved communication among experts.

Our conclusions and recommendations will be presented in Section VI.

II THE SAGAN-COLEMAN FORMULA

Let us consider a single mission to Mars. To be specific, we will assume that a mission is intended to land a vehicle at some specified equatorial site on Mars during July 1976. Given all mission specifications, including a proposed sterilization procedure, the mission contamination problem is to determine the probability of

C = the event that Mars will be biological contaminated by organisms aboard this spacecraft

in terms of more fundamental descriptors of the mission. Toward that end, we further define

N = the number of viable organisms released to the Martian environment or into its atmosphere from the spacecraft (a random variable),

$$E(N) = \sum_{k=0}^{\infty} k P(N=k)$$

= the expected (or mean) number of viable organisms released,

G = the event that a single released organism will survive, multiply, and contaminate a significant fraction of the planet.

The solution advanced by Sagan and Coleman for the mission contamination problem was the simple approximation

$$P(C) \approx E(N)P(G) \quad (1)$$

for which no justification was offered. The apparent rationale for this approximation is as follows. With the definition

E_i = the event that the i^{th} released organism does not survive to multiply and cause contamination,

it follows directly that

$$\begin{aligned} 1-P(C) &= P(E_1 \text{ and } \dots \text{ and } E_N) \\ &= \sum_{k=1}^{\infty} P(N=k) P(E_1 \text{ and } \dots \text{ and } E_k | N=k) + P(N=0) \quad . \quad (2) \end{aligned}$$

If, given that k organisms are released, we assume the events $E_1, E_2, \dots, E_k$ to be mutually independent and of equal probability,* then

$$P(E_1 \text{ and } \dots \text{ and } E_k | N=k) = [P(E_1 | N=k)]^k \quad . \quad (3)$$

Moreover, if we assume that the survival of any one organism is independent of the number of organisms released, then

$$P(E_1 | N=k) = P(E_1) = 1-P(G) \quad . \quad (4)$$

Substituting Eqs. (3) and (4) into Eq. (2) then gives us

$$\begin{aligned} 1-P(C) &= \sum_{k=0}^{\infty} P(N=k) [1-P(G)]^k \\ &\approx \sum_{k=1}^{\infty} P(N=k) [1-kP(G)] = 1 - E(N)P(G) \quad , \end{aligned}$$

* The events $E_1, E_2, \dots, E_m$ are called mutually independent if for all combinations $1 \leq i < j < \dots \leq m$ the multiplication rules

$$\begin{aligned} P(E_i \text{ and } E_j) &= P(E_i)P(E_j) \\ P(E_i \text{ and } E_j \text{ and } E_k) &= P(E_i)P(E_j)P(E_k) \\ &\vdots \\ P(E_1 \text{ and } \dots \text{ and } E_m) &= P(E_1)P(E_2) \dots P(E_m) \end{aligned}$$

apply.

the approximation being justified by the fact that $P(G)$ is very small, and only moderate values of N have significant probability.* This relationship is equivalent to Eq. (1).

Let us now consider the independence assumption underlying Eq. (3). If the assumption is accepted, then the following must hold by definition. Having learned that exactly k organisms were released and that none of the first $k-1$ survived, we would not be inclined to alter our initial assessment for the probability that the last will survive and proliferate. We do not think this is a reasonable statement, and the reason lies in a rather fundamental question. Why are we uncertain about the survival of any single organism? To a large extent, many important characteristics of the Martian environment are not yet known with certainty. The events $E_1, \dots, E_k$ are mutually dependent on the actual character of that environment. Learning the fate of the first $k-1$ organisms tells us something about Mars itself and in turn tells us something about the k^{th} organism's chances of surviving and proliferating.

The extreme case of total dependence is illustrated by the following scenario. Suppose that the current state of scientific knowledge admits only two possibilities with respect to the Martian environment: either it is so hostile that no terrestrial organism could possibly survive, or it is so hospitable that the survival and proliferation of any such organism are ensured. In this case, the events $E_1, E_2, \dots$ are totally dependent because knowing the fate of any one organism would automatically tell us the fate of the others. If we assign a probability of 10^{-4} to the hospitable situation and $1-10^{-4}$ to the hostile one, then $P(G) = 10^{-4}$. Note, however, that

$$P(C) = P(N \geq 1) \times P(G) = P(N \geq 1) \times (10^{-4}) \quad (5)$$

* Under these conditions, the second and higher order terms in the expansion of $[1-P(G)]^k$ are negligible.

The value for $P(C)$ given by Eq. (5) will typically be much smaller than that given by Eq. (1). The actual state of current scientific information would put us somewhere between the case of independence, on which Eq. (1) is based, and the case of total dependence, on which Eq. (5) is based.

Our objection to the independence assumption underlying Eq. (4) is more subtle. As our spacecraft arrives on Mars, imagine that a scientific expert is sequestered and denied any information other than unimpeachable evidence that exactly ten viable organisms have been released. He is then asked to assess the probability that the first organism released will survive and proliferate. Would this assessment be any different if he were told that 1000 organisms had been released? Although we are not certain, there is at least one reason why he might give different answers in the two circumstances. Having learned that 1000 organisms were released, the scientist might conclude that some engineering failure had led to a hard landing, and thus released many encapsulated organisms. This in turn increases the likelihood that the vehicle has impacted far from its targeted landing site, which we shall assume is in the equatorial region. A critical uncertainty regarding the life-supporting capability of Mars concerns the existence of liquid water and most scientists feel that even if liquid water is on the planet, it is unlikely to be in the equatorial region. Consequently, knowing that many organisms have been released might alter the expert's assessment as to the accessibility of liquid water and hence to the survivability of an individual organism. In summary, the number of released organisms depends on the mode of landing, which in turn affects survivability. Thus the events $N = k$ and E_1 are not independent; hence, Eq. (4) is invalid. Another way of looking at this dependence is to say that Mars is not in fact characterized by a single $P(G)$ value. The probability that a single, released organism will survive and multiply depends on where the organism is released. Consequently, an expert has difficulty

in making meaningful probability assessments within the framework of a model that does not take explicit account of uncertainty about the actual landing site and impact velocity. This point is pursued further below.

The Sagan-Coleman linear approximation $P(C) \approx E(N)P(G)$ is a cornerstone of current quarantine planning procedures. To illustrate its use, let us suppose that the mission allocation requires $P(C) \leq 10^{-4}$. Then the corresponding constraint on the mean number of viable organisms released is $E(N) \leq 10^{-4}/P(G)$. The probability $P(G)$, which is obviously very small, has been directly assessed by scientific experts, who have internally integrated the many factors that affect an organism's survivability on Mars. In contrast, a great deal of laboratory experimentation and additional modeling has been used in developing assessments for $E(N)$ under various sterilization procedures.⁵ We have argued that the linear approximation is based on independence assumptions that are highly questionable. In the following sections, we discuss the general notion of dependent events in a more familiar setting and then indicate how the Sagan-Coleman formulation can be modified to eliminate its current weaknesses.

III A CLASSICAL PARALLEL PROBLEM

The following problem, which has been discussed in a slightly different form by Howard,⁶ provides an example of a familiar physical process having identical but informationally dependent trials. Its significance to the mission contamination problem will be discussed shortly. Let us suppose that a tack is dropped onto a large flat surface; it has two possible landing positions that are labelled "heads" and "tails" as in the following diagram. You are told only that the tack in the diagram is



drawn to scale, that a human being will drop it from a height of four feet and that the landing surface is very flat. Your first problem is to assess the probability of a head in one toss; your second problem is to assess the probability of ten heads in ten tosses. To respond that you do not know the probabilities, never having watched any tack tossing, is unacceptable. The questions do not concern frequencies or any other type of "physical fact." We ask only a quantification of your judgment and recognize that different people will typically make different assessments. Now suppose that after much scrutiny of the diagram, you assess the probability of a head in one toss to be one-half. Using only the

rules of consistency imposed by probability theory, is it possible to deduce your response to the second question from this? The answer is no. You simply have not told us enough about your judgment (or state of information). Before any calculations can be done (on your behalf), we need to know something about how you feel the individual tosses relate to one another. To fill in this gap, you might say that you view the events,

B_i = the event of a head on the i^{th} toss ($1 \leq i \leq 10$),

as independent, in which case we immediately have

$$P(10 \text{ heads in } 10 \text{ tosses}) = (1/2)^{10} \approx 10^{-3}.$$

But, considering the characterization of independence given earlier, does this assumption accurately reflect your state of information? It seems unlikely, for undoubtedly you would be inclined to alter your initial assessment for the probability of a head in one toss if we told you the results of the first nine tosses.

Having rejected the independence assumption, how can you compactly express the degree of dependence that you perceive to exist among the results of the separate trials? Under very mild assumptions,* it can be shown that the following characterization provides all the required information. Let

ϕ = the fraction of heads that would be observed
in a very long sequence of tosses,

which can be viewed only as a random variable with your current state of information. What we need is your subjective (prior) probability distribution for the random variable ϕ . This is conveniently expressed by the cumulative distribution function

$$F(x) = P(\phi \leq x), \quad 0 \leq x \leq 1.$$

* The assumption is that the trials be exchangeable. For a definition and discussion of exchangeable trials see de Finetti.⁷

The mean (or expected value) of this distribution is given by*

$$E(\phi) = \int_0^1 [1-F(x)] dx ,$$

and consistency demands that it equal 0.5. That is, the axioms of probability theory require that your subjective probability of a head in one trial equal the mean of your subjective distribution for the fraction of heads in a great many trials.

Figures 2 through 4 show three possible distribution functions for the random variable ϕ , each of which is consistent with the earlier assessment that $P(\text{head}) = \text{one-half}$. The first of these distributions corresponds to the case of independent trials, the subject being absolutely certain that the long-run fraction of heads will be 50 percent.[†] Such a distribution might be assessed by an individual who has spent the last few months tossing this same tack onto this same surface. Although he is uncertain as to what will happen in a few trials, his complete knowledge of the basic environment leads him to view the tack as equivalent to an unbiased coin.

The second distribution (Figure 3) corresponds to the case of totally dependent trials. The subject is absolutely certain that the tack will always either come up heads or come up tails, but he is not sure which of these cases pertains. (He might have an acquaintance who has tossed the tack many times, told him it always falls one way, but left him to guess from there.) He has assessed the probability of all heads to be one-half and that of all tails to be one-half. Note that if this subject

* Integration by parts shows this formula equivalent to the usual one in terms of the density function or probability mass function.

[†] Of course, the subject would also view the trials as independent if he were certain that the long-run fraction would be 40 percent or any other specific number.

$$F_1(x) = P(\phi \leq x)$$

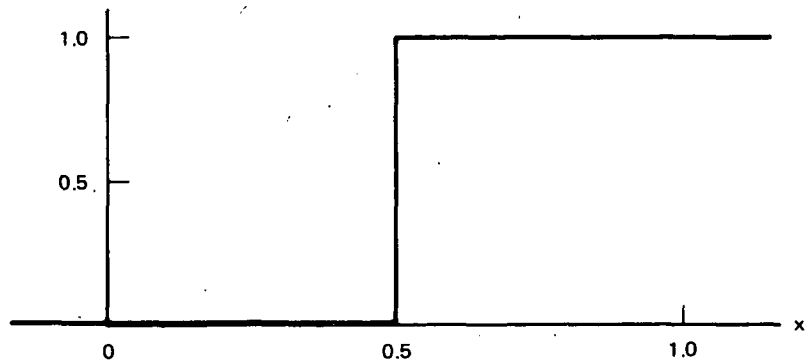


FIGURE 2 DISTRIBUTION IMPLYING INDEPENDENT TRIALS

$$F_2(x) = P(\phi \leq x)$$

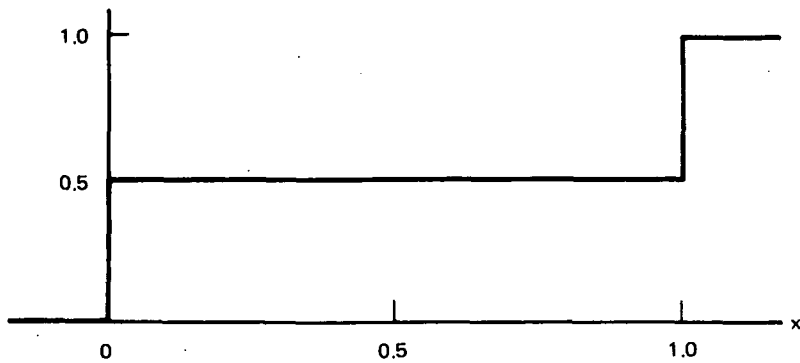


FIGURE 3 DISTRIBUTION IMPLYING PERFECTLY DEPENDENT TRIALS

$$F_3(x) = P(\phi \leq x)$$

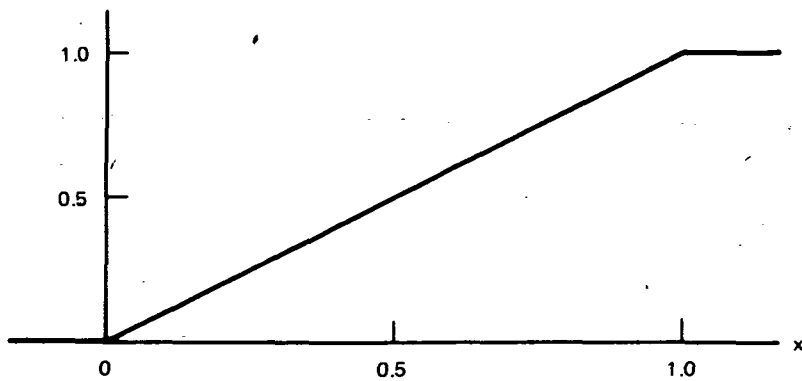


FIGURE 4 UNIFORM DISTRIBUTION

were able to observe one toss, it would resolve all his uncertainty regarding the outcomes of subsequent tosses.

The type of distribution that we would generally expect, intermediate to the preceding extreme cases, is shown in Figure 4. Here the subject reveals great uncertainty as to the experiment's environment and assigns a uniform distribution over the interval of possible values. The mean of his distribution, like that of the others, is $E(\phi) = \text{one-half}$.

Given the probability distribution for ϕ , we can calculate the probability of all heads in n trials using the formula*

$$P(\text{all heads in } n \text{ trials}) = E(\phi^n), \quad n \geq 1.$$

From this we have computed the relationships shown in Figure 5 for each distribution discussed earlier. The subject who views the trials as independent thinks it very unlikely (less than one chance in a thousand) that we could survive the ten trials without observing a tail. In contrast, the subject who views the trials as perfectly dependent continues to assign a probability of one-half to the event of all heads, regardless of how many times the tack is tossed. The corresponding relationship for the third subject lies between these two extremes. In particular, he assesses the probability of ten heads in ten trials to be about 9 percent, one hundred times the probability implied by the first distribution. Thus, we find that the three individuals differ greatly in their assessment of what is likely to occur in repeated trials, although they agree perfectly as to the probability of a head in a single trial. It is the degree of informational dependence among trials that differs from one subject to another, and these differences have significant implications.

* This is an application of de Finetti's theorem.⁷

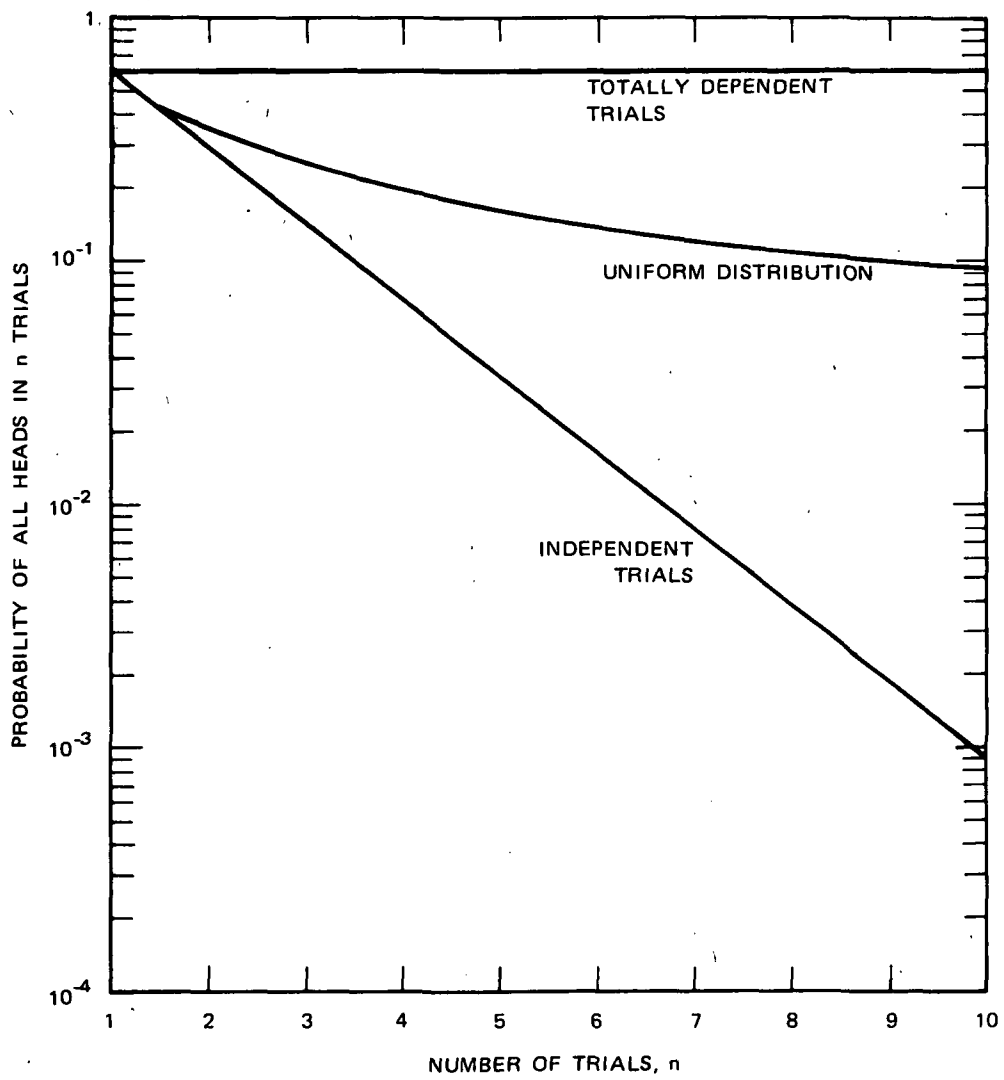


FIGURE 5 PROBABILITY OF ALL HEADS VERSUS NUMBER OF TRIALS

IV THE EFFECT OF DEPENDENCIES IN THE CONTAMINATION PROBLEM

Returning now to the mission contamination problem, we wish to indicate how the conceptual deficiencies of the Sagan-Coleman formulation can be remedied by adopting a somewhat richer model structure. The discussion in this section, however, should not be construed as a recommended procedure for actually analyzing the problem. Drawing heavily on the thumbtack parallel presented in Section III, we shall present an approach to the problem that, although conceptually correct, would be very difficult to implement. It provides, however, a means for discussing both the type and the likely magnitude of error introduced by the Sagan-Coleman approximation.

Figure 6 shows the level of modeling detail proposed by Sagan and Coleman. The basic structure includes a biorelease model, whose output is the mean number of released organisms; and a proliferation model that consists of the linear relationship between contamination probability and number of released organisms. We believe that the Sagan-Coleman model structure must be enriched to at least the extent shown in Figure 7. Here we have added an initial landing model, through which uncertainty as to

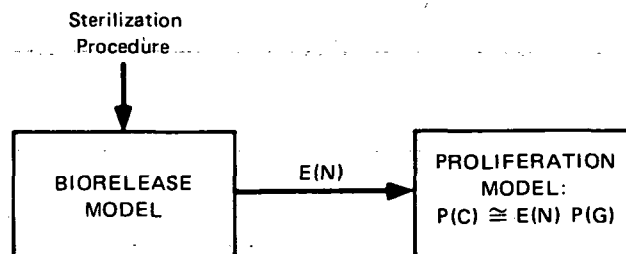


FIGURE 6 BASIC LOGICAL STRUCTURE OF THE SAGAN-COLEMAN MODEL

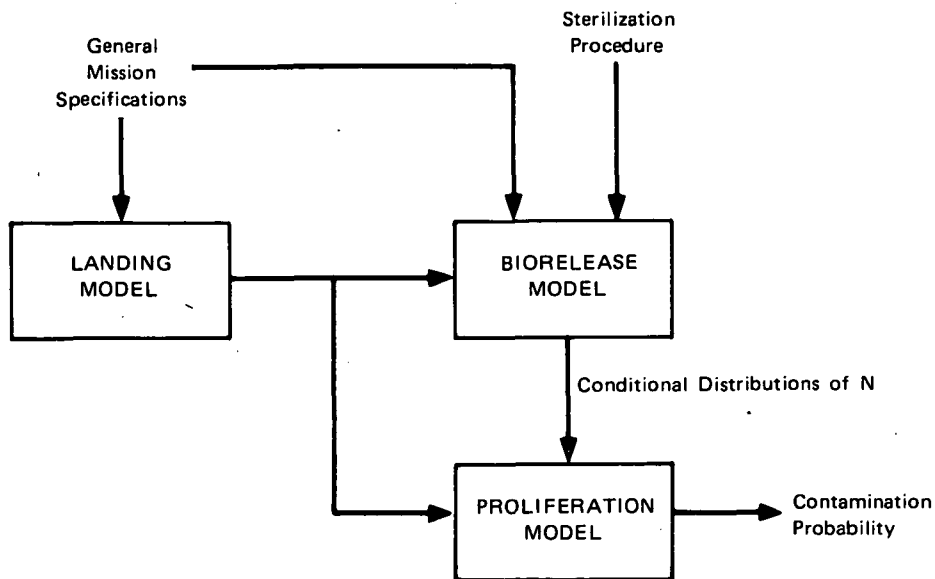


FIGURE 7 BASIC LOGICAL STRUCTURE OF THE PROPOSED MODEL

the technical success of the mission is expressed explicitly. As Figure 8 illustrates, this model component can take the form of a simple probability tree, can identify a relatively small number of possible mission outcomes, and can associate a probability with each. For purposes of illustration, we have distinguished two possibilities with respect to landing site and three with respect to impact velocity. In particular, letting

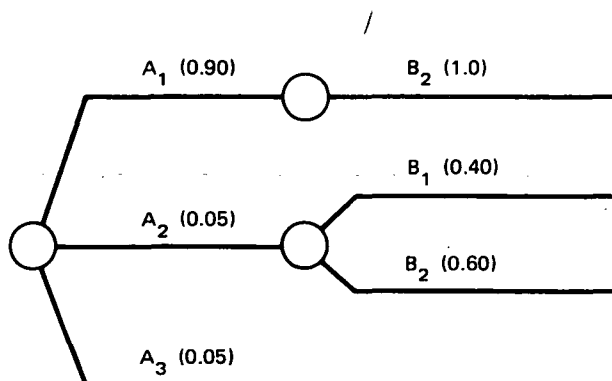


FIGURE 8 ILLUSTRATIVE LANDING MODEL

Region 1 = that portion of Mars between 50° and 70° north latitude,

Region 2 = the remainder of the planet,

we define the events

A_1 = landing is soft,

A_2 = landing is hard,

A_3 = mission fails to impact Mars,

B_1 = landing is in Region 1,

B_2 = landing is in Region 2.

The branches emanating from the first node in Figure 8 give a probability distribution over A_1 , A_2 and A_3 . Branches emanating from subsequent nodes correspond to a conditional distribution for the landing area, given impact velocity. We emphasize that both the level of detail (number of possibilities identified) and the probabilities in Figure 8 are purely illustrative. We have assumed here that the mission is targeted for an equatorial landing site.

To determine an appropriate level of detail for the description of landing areas, the central consideration is the following. Each area identified must be internally undifferentiated (given our current state of information) in its ability to support growth of terrestrial organisms. (We shall indicate in Section V why the particular geographical regions identified above have been chosen for purposes of illustration.) After the planet is partitioned in accordance with this criterion, the following parallel to the thumbtack example can be drawn. Let us associate the fates of individually released organisms with successive tosses of indistinguishable tacks. The outcome "heads" will correspond to the event of survival and proliferation, while "tails" will correspond to the organism's demise. The landing areas can be compared with homogeneous surfaces

of varying roughness; on each surface, we feel that the behavior of a tack will be different. The number of released organisms corresponds to the total number of tosses observed, and contamination corresponds to the occurrence of at least one head. Our uncertainty as to whether or not this will occur can be represented in stages by the probability tree in Figure 9. Here we let

N_1 = the number of organisms released in the event of a soft landing (a random variable),

N_2 = the number of organisms released in the event of a hard landing (a random variable),

f_j = the fraction of a great many organisms released in Region j that would survive and proliferate (a random variable), $1 \leq j \leq 2$.

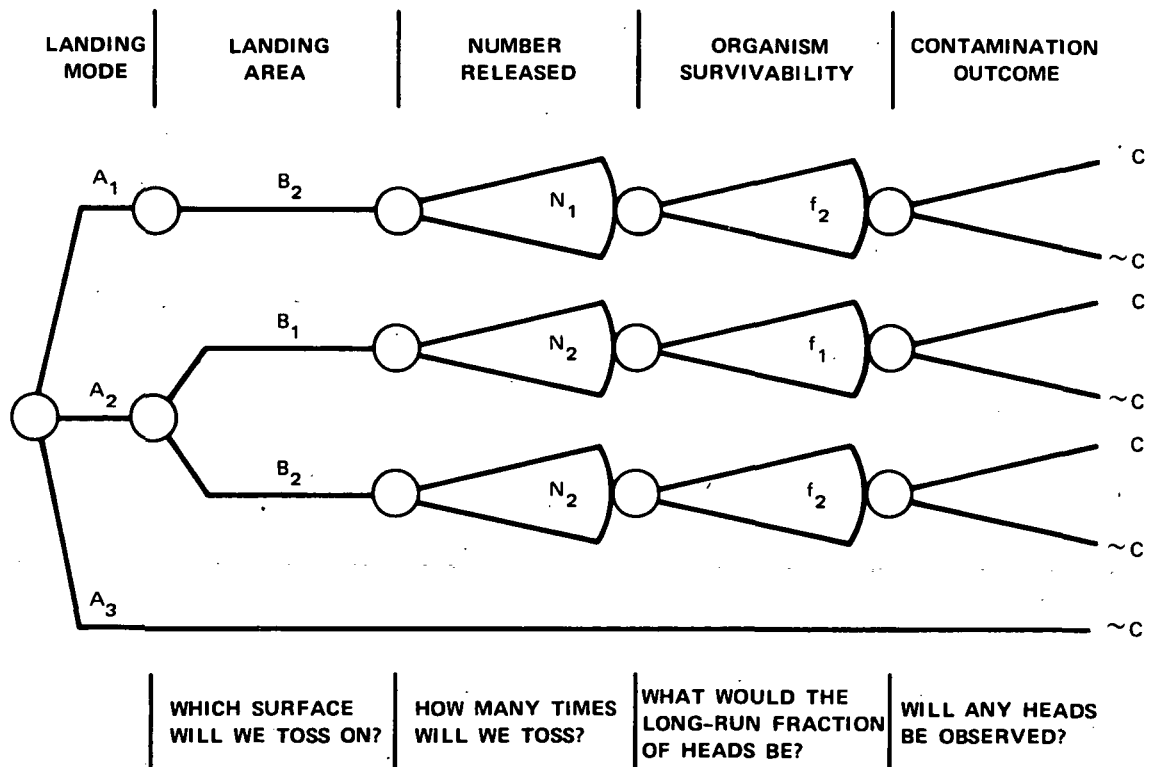


FIGURE 9 PROBABILITY TREE FORM OF THE PROPOSED MODEL

The random variable f_j corresponds to the long-run fraction of heads that would result from a great many tack tosses on surface j . As we have seen earlier, the distribution of f_j is the essential element required to compute the probability of observing a head in repeated trials.

In drawing Figure 9, we have assumed that the distribution for a number of released organisms depends on impact velocity (hard versus soft landing) but not on landing area. Although this is plausible and simplifies calculations, it is by no means critical to our computational procedure. A different distribution could be used, if appropriate, for each combination of impact velocity and landing area. With our assumptions, the biorelease model portrayed schematically in Figure 7 takes the simple form of a probability distribution for N_1 and another for N_2 . Two such distributions are displayed for purposes of illustration in Figure 10. Certain theoretical justifications exist for the assumption that each distribution is Poisson; this simplifies calculations somewhat, but the specific distributional form is unimportant to our general approach.

The last data required for the overall probability tree in Figure 9 are probability distributions for f_1 and f_2 . These two distributions together constitute the proliferation model schematically represented earlier in Figure 7. An illustrative distribution for f_1 is displayed in Figure 11. Because the numbers are extremely small, it is graphically convenient to show a complementary cumulative distribution function, which simply equals one minus the ordinary cumulative distribution function at all points. The dotted lines give a step-function approximation to the continuous curve; this step function corresponds to the discrete distribution for f_1 shown in Figure 12. In all subsequent calculations, we shall use this discrete approximation to the original distribution. Figure 12 shows the first two moments of the discrete distribution.

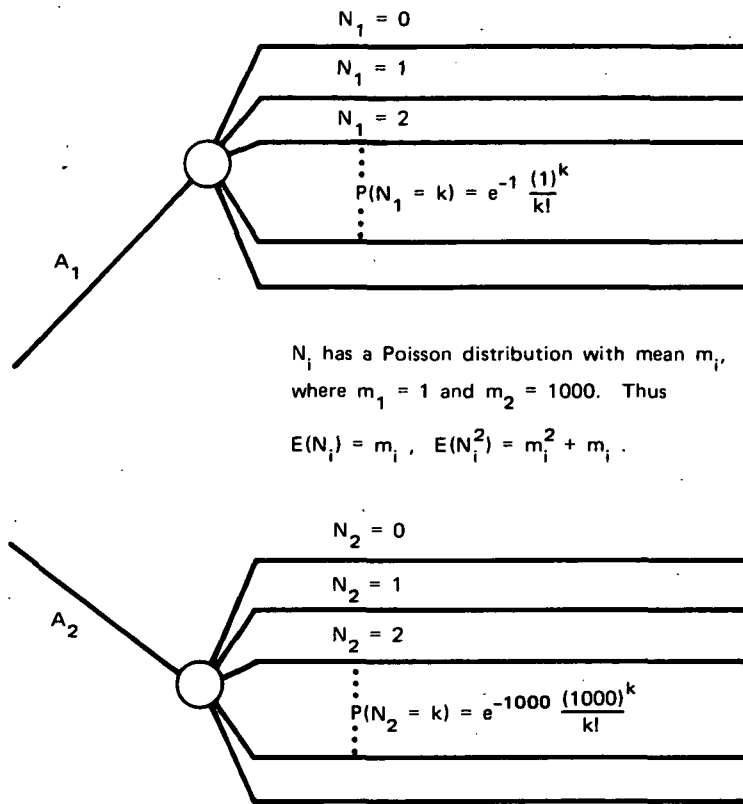


FIGURE 10 ILLUSTRATIVE RELEASE MODEL

Recall from the thumbtack problem that $E(f^k)$ equals the probability of observing k heads in as many tosses. Analogously, $E(f_1^k)$ is the probability that, if k organisms were released in Region 1, all of them would multiply and proliferate. In particular, $E(f_1)$ gives the probability that a single released organism will multiply and proliferate, meaning that it represents a $P(G)$ assessment for Region 1. Figure 12 shows this to be 1.0×10^{-5} with our illustrative data. An illustrative distribution for f_2 is shown in Figure 13, and the corresponding discrete approximation is given in Figure 14. Observe that $E(f_2)$ is 1.9×10^{-6} , so Region 1 is perceived as more hospitable to terrestrial organisms than is Region 2.

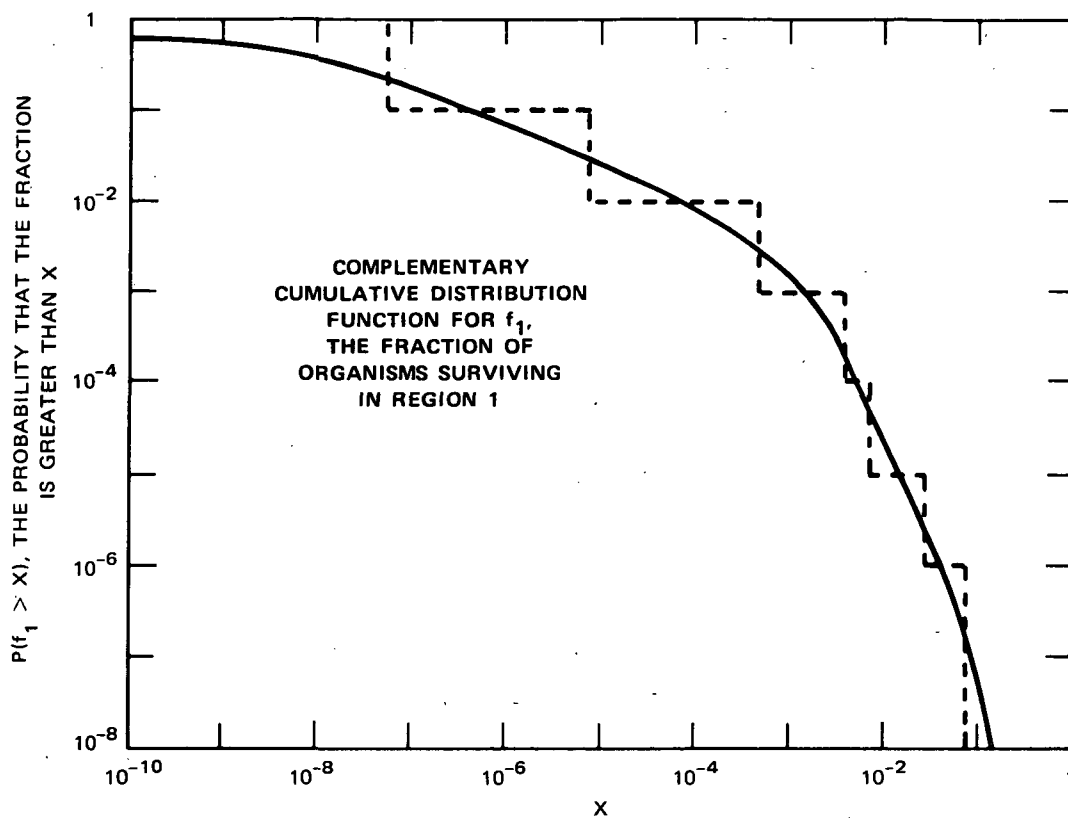


FIGURE 11 ILLUSTRATIVE DISTRIBUTION OF REGION 1 SURVIVABILITY

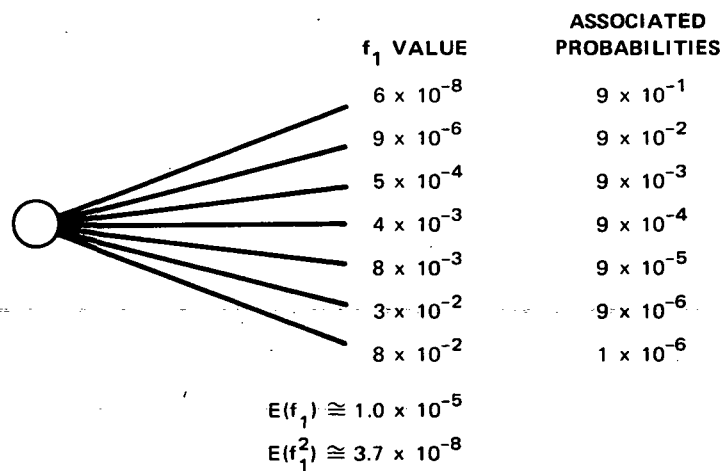


FIGURE 12 DISCRETE APPROXIMATION FOR THE REGION 1 SURVIVABILITY DISTRIBUTION

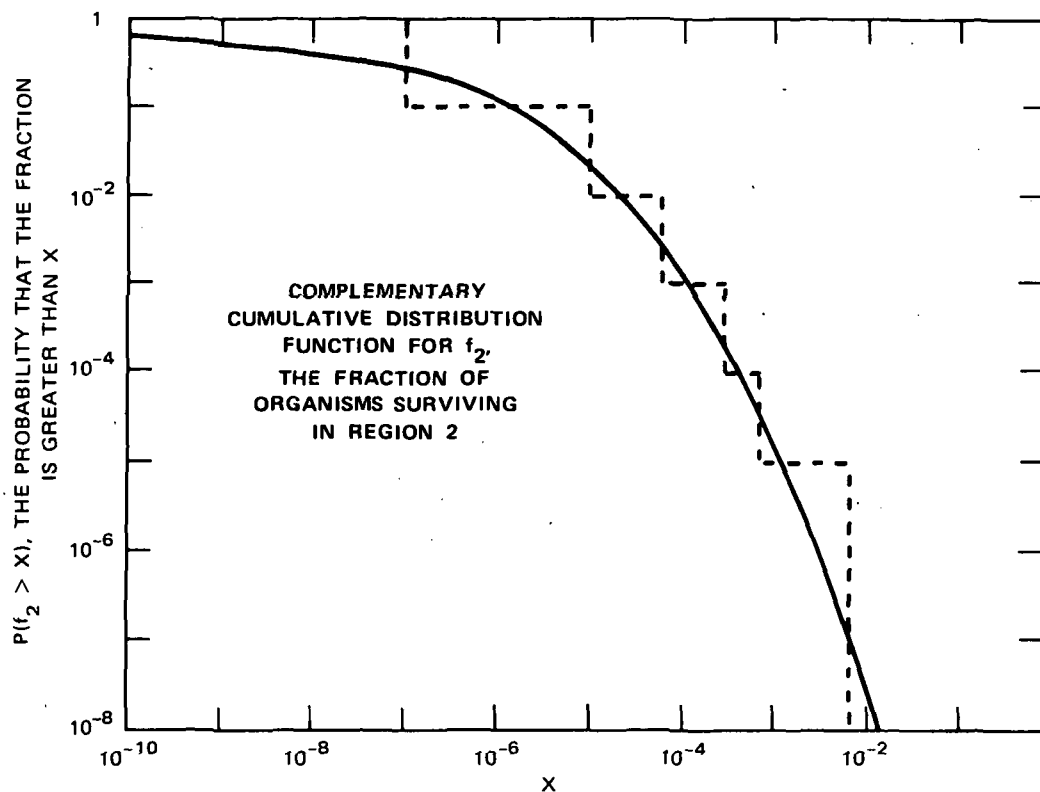


FIGURE 13 ILLUSTRATIVE DISTRIBUTION OF REGION 2 SURVIVABILITY

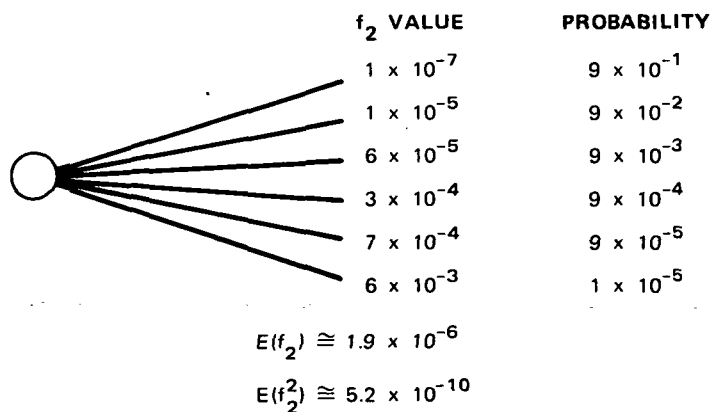


FIGURE 14 DISCRETE APPROXIMATION FOR THE REGION 2 SURVIVABILITY DISTRIBUTION

We now wish to roll back* the probability tree shown in Figure 9 so as to determine $P(C)$, the total probability of contamination. If we knew that k organisms were released in region j and we were told the value for f_j , the corresponding conditional probability of contamination would be

$$P(C|A_i, B_j, N_i = k, f_j) = 1 - (1 - f_j)^k .$$

Removing the conditioning on f_j and N_i then yields

$$P(C|A_i, B_j) = 1 - E\left((1 - f_j)^{N_i}\right) .$$

Here the expectation is taken over both f_j and N_i , and we shall assume throughout that these random variables are conditionally independent given A_i and B_j . Completing the rollback then yields the final formula

$$P(C) = \sum_{i=1}^2 \sum_{j=1}^2 P(A_i B_j) \left[1 - E\left((1 - f_j)^{N_i}\right) \right] . \quad (6)$$

With our assumption that N_i has a Poisson distribution, we note that

$$E\left((1 - f_j)^{N_i}\right) = E\left(\sum_{k=0}^{\infty} e^{-m_i} \frac{m_i^k}{k!} (1 - f_j)^k\right) = E\left(e^{-m_i f_j}\right) .$$

Thus, the master formula, Eq. (6), simplifies to

$$P(C) = \sum_{i=1}^2 \sum_{j=1}^2 P(A_i B_j) \left[1 - E\left(e^{-m_i f_j}\right) \right] , \quad (7)$$

where the expectation is over f_j only. Using the distributions for f_1 and f_2 given in Figures 12 and 14, we find that

*That is, compute total probability by multiplying conditional probabilities in the usual way, working from the final stage of the tree back to the initial stage.

$$1 - E \left(e^{-m_1 f_2} \right) = 1 - E \left(e^{-f_2} \right) = 1.9 \times 10^{-6}$$

$$1 - E \left(e^{-m_2 f_1} \right) = 1 - E \left(e^{-1000 f_1} \right) = 5.4 \times 10^{-3}$$

$$1 - E \left(e^{-m_2 f_2} \right) = 1 - E \left(e^{-1000 f_2} \right) = 1.8 \times 10^{-3}$$

Also, from Figure 8,

$$P(A_1 B_2) = 0.90, \quad P(A_2 B_1) = 0.02, \quad P(A_2 B_2) = 0.03,$$

so formula (7) gives us

$$\begin{aligned} P(C) &= (0.9)(1.9 \times 10^{-6}) + (0.02)(5.4 \times 10^{-3}) + (0.03)(1.8 \times 10^{-3}) \\ &= 1.6 \times 10^{-4} \end{aligned}$$

Thus, with our illustrative data, the total probability of contamination is slightly larger than the current mission allocation for Project Viking.

Again with our illustrative data, the reader may verify that the nominal landing $A_1 B_2$ (which has a 90% probability) makes an insignificant contribution to $P(C)$. The dominant factor in the computation is the large expected number of released organisms upon hard landing, with the relatively higher probabilities of proliferation associated with Region 1 also playing an important role. The contribution of the nominal landing to $P(C)$ would remain insignificant if m_1 , the expected number of organisms released by a soft landing, were increased from 1 to as much as 40 or 50.

Although the model structure described in this section is conceptually sound and accounts for all of the dependencies ignored in the Sagan-Coleman formulation, the resulting analytical procedure appears impractical. The reason is that the distributions for f_j would be difficult to obtain, at least directly, as we shall explain later. It is natural then

to seek convenient approximations, such as those used by Sagan and Coleman. For each landing outcome $A_i B_j$, we might approximate the conditional probability of contamination by

$$P(C|A_i B_j) = 1 - E \left[(1 - f_j)^{N_i} \right] \approx E(N_i) E(f_j)$$

since $E(f_j)$ is simply a region-specific $P(G)$ assessment. Substituting this in Eq. (6) would have us approximate $P(C)$ by

$$Q(C) = \sum_{i=1}^2 \sum_{j=1}^2 P(A_i B_j) E(N_i) E(f_j) \quad (8)$$

The approximation $P(C) \approx Q(C)$ is identical to the Sagan-Coleman formula, except we have added an initial conditioning on landing outcome. This approximation has some very attractive properties. In Appendix I, it will be shown under very mild assumptions that

$$Q(C) \geq P(C) \geq Q(C) - EB, \quad (9)$$

where the error bound EB is given by

$$EB = \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 P(A_i B_j) E(N_i^2) E(f_j^2) \quad (10)$$

That is, the approximation $P(C) \approx Q(C)$ is conservative in that it can only overstate the true probability of contamination, and the degree of overstatement is bounded above by the second-moment expression EB . With the illustrative data presented in Figures 8, 10, 12 and 14, we find that

$$\begin{aligned} Q(C) &= (0.90)(1.0)(1.9 \times 10^{-6}) + (0.02)(1000)(1 \times 10^{-5}) \\ &\quad + (0.03)(1000)(1.9 \times 10^{-6}) = 2.6 \times 10^{-4}, \end{aligned}$$

while

$$\begin{aligned} EB &= \frac{1}{2} \left\{ (0.90)(2)(5.2 \times 10^{-10}) + (0.02)(10^6)(3.7 \times 10^{-8}) \right. \\ &\quad \left. + (0.03)(10^6)(5.2 \times 10^{-10}) \right\} = 1.2 \times 10^{-4}. \end{aligned}$$

Thus, using Eq. (9), we arrive at the upper and lower bounds

$$1.4 \times 10^{-4} \leq P(C) \leq 2.6 \times 10^{-4},$$

which agree with the exact value of $P(C) = 1.6 \times 10^{-4}$ computed earlier. With our illustrative data, the first-moment approximation $Q(C)$ overstates the true value of $P(C)$ by a factor of about 1.6, but the absolute error (1×10^{-4}) is, of course, very small.

We have indicated that the exact computational procedure described earlier would be difficult to implement. To understand the reason for this, one need only look at the scales in Figures 11 and 13. To obtain the distributions of f_1 and f_2 directly from scientific experts, one would have to ask such questions as, "Would you assign a probability as large as 10^{-4} to the event of more than one in a thousand organisms surviving and proliferating in Region 1?" Experience in the encoding of subjective probability distributions indicates that even very sophisticated subjects have difficulty dealing with numbers smaller than about one in a hundred. To encode a distribution for f_1 or f_2 would require that the subject deal with pairs of numbers that both lie below this threshold of comprehensibility.

To compute the quantity $Q(C)$ requires only the means of f_1 and f_2 , which are equivalent to region-specific $P(G)$ assessments rather than to their entire distributions. It is then natural to ask whether the approximation $P(C) \approx Q(C)$ is adequate for purposes of the planetary quarantine program. Although this is difficult to answer definitively, our feeling is that the approximation does suffice for the following reasons.

First, the only possible error that can be introduced is on the conservative side, so any sterilization program that yields a $Q(C)$ value smaller than the mission allocation is guaranteed to meet program standards. Second, computations that we have done indicate that the error

introduced by the approximation is insignificant in the following sense. If the data are such that the true $P(C)$ value lies below 1×10^{-4} (the current mission allocation for Project Viking), then $Q(C)$ approximates $P(C)$ quite closely except when the distributions of N_i and f_j have certain very unlikely characteristics. (These characteristics will be described later.) In contrast, one can construct plausible-looking distributions that yield $P(C)$ values in the range of 10^{-2} to 10^{-3} , under which the approximation is very poor (meaning that $Q(C)$ overstates $P(C)$ by as much as an order of magnitude). Such possibilities give little cause for alarm, however, because using the approximation $P(C) \approx Q(C)$ in such cases, we would correctly reject as inadequate the corresponding sterilization procedure.

In summary, our general feeling is that the approximation $P(C) \approx Q(C)$ is a good one in the range of interest, but one must remain alert for the pathological situations in which it becomes overly conservative. The lower bound in Eq. (9) gives a ready means of identifying, at least qualitatively, what these pathological situations are. Note that the expression in Eq. (10) for the error bound can be rewritten as

$$EB = \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 P(A_i B_j) \left\{ [E(N_i)]^2 + \text{Var}(N_i) \right\} \left\{ [E(f_j)]^2 + \text{Var}(f_j) \right\} .$$

If the variances of each N_i and f_j are set to zero, then EB becomes insignificant in comparison with $Q(C)$ over the range of interest. As these variances become very large, however, the error bound grows correspondingly, and the accuracy of our first-moment approximation is ended. Thus, unfavorably, the distributions of some N_i and f_j are very widely spread, giving rise to large variances. Actual computation indicates that the degree of spread must be extreme to seriously affect the approximation in the range of interest.

As already noted, the approximation $P(C) \simeq Q(C)$ is identical to the Sagan-Coleman formula except that an initial conditioning on landing outcome has been added. Since the initial conditioning adds a certain amount of complexity to the data gathering and subsequent analysis, it is natural to ask if it is really necessary. We strongly believe that it is, and our illustrative data demonstrate the reasons.

With our data, the expected number of organisms released by the mission is

$$\begin{aligned} E(N) &= \sum_{i=1}^2 \sum_{j=1}^2 P(A_i B_j) E(N_i) \\ &= (0.90)(1) + (0.02)(1000) + (0.03)(1000) \simeq 50 . \end{aligned}$$

Similarly, the probability that any one organism released by the mission will survive and proliferate is

$$\begin{aligned} P(G) &= \sum_{i=1}^2 \sum_{j=1}^2 P(A_i B_j) P(G|B_j) = \sum_{i=1}^2 \sum_{j=1}^2 P(A_i B_j) E(f_j) \\ &= (0.90)(1.9 \times 10^{-6}) + (0.02)(1 \times 10^{-5}) + (0.03)(1.9 \times 10^{-6}) \\ &= 1.96 \times 10^{-6} . \end{aligned}$$

Thus, the original Sagan-Coleman formula would give us

$$P(C) \simeq E(N)P(G) = (50)(1.96 \times 10^{-6}) = 9.8 \times 10^{-5} .$$

Our exact calculations showed that $P(C) = 1.6 \times 10^{-4}$, so this approximation actually understates the probability of contamination. Taking a product of expected values rather than the expected value of the product, we are no longer assured that the final approximation is an upper bound for the true value. The reason is that the original Sagan-Coleman formula fails to account for the potential dependence between the number of released organisms and the survivability of any one organism, as was pointed

out in Section II. This dependence is very strong with our illustrative data because the unfavorable events of landing hard (releasing many organisms) and landing in Region 1 (which is judged more hospitable to terrestrial organisms) tend to occur together. By forming a Sagan-Coleman approximation for each landing outcome and then taking an expected value over the landing outcomes, we account for this dependence and ensure that an upper bound for $P(C)$ is ultimately derived.

V ENRICHING THE PROLIFERATION MODEL

In the previous section, we indicated that for most cases of interest in the quarantine analysis for Project Viking, $P(C)$ can be adequately approximated by

$$Q(C) = \sum_i \sum_j P(A_i B_j) E(N_i) P(G|B_j) .$$

This conclusion results in the following outstanding analytical problems.

- (1) The level of detail to represent landing outcomes (that is, the number and type of events A_i and B_j identified in the landing model).
- (2) The manner of determining the parameters $P(A_i B_j)$, $E(N_i)$, and $P(G|B_j)$.

The first of these issues has been discussed earlier. On the matter of parameter values, much analytical effort has been spent in building submodels to determine $E(N_i)$ under various types of sterilization procedures and with various landing mode events A_i . In a similar way, the landing outcome probabilities can be assessed by a submodel that accounts for the laws of orbital mechanics, historical data on component reliability, and so forth. To determine the proliferation probabilities, however, the possibilities for further modeling are not quite so obvious. In the past, the single proliferation probability $P(G)$ called for in the Sagan-Coleman formula has been directly assessed by scientific experts. In making such an assessment, the expert must internally integrate all of the various considerations that contribute to his uncertainty about an organism's survival on Mars. Some of the most important such considerations are schematically represented in Figure 15. The center line of

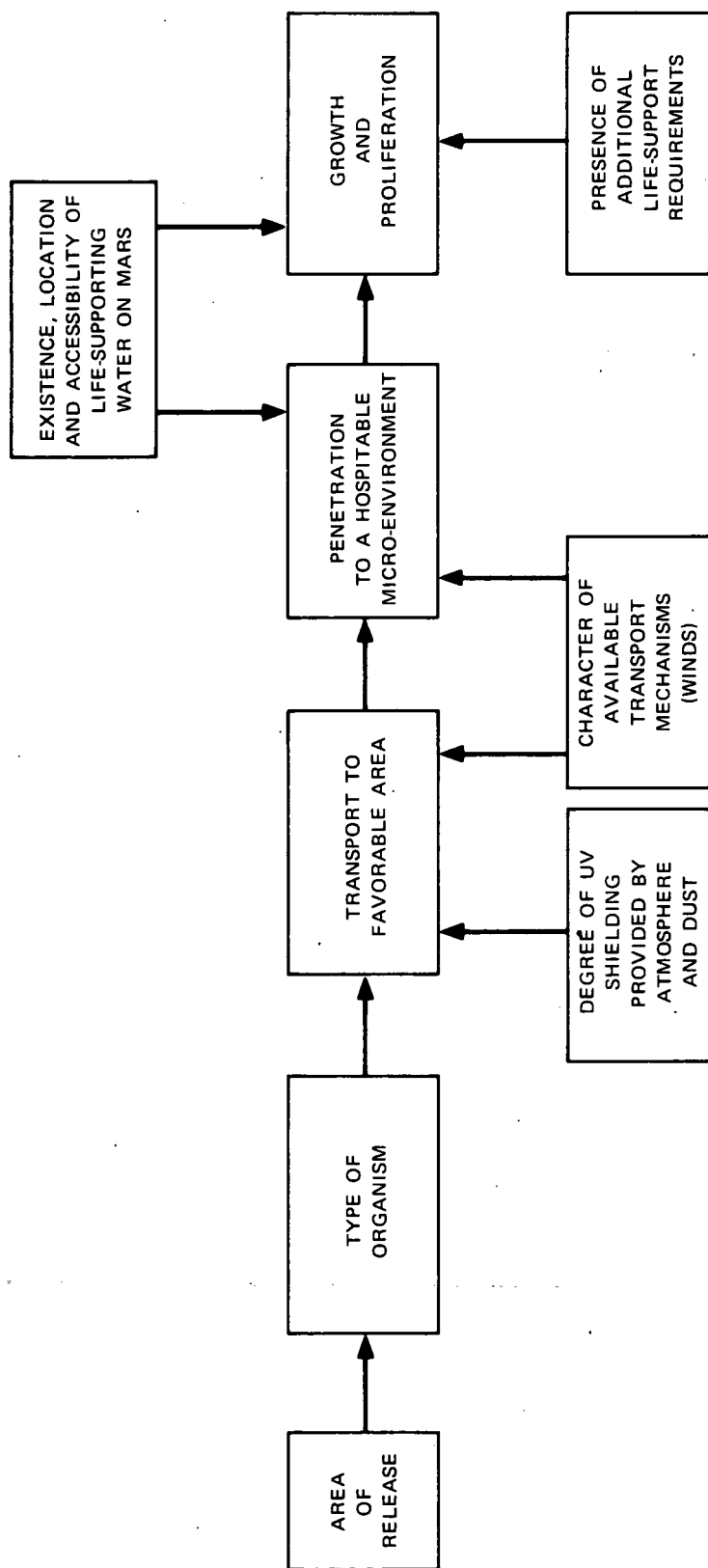


FIGURE 15 SOME MARTIAN CHARACTERISTICS AND PHYSICAL PROCESSES CRITICAL TO BIOLOGICAL CONTAMINATION

the figure contains characteristics of the mission, characteristics of the released organism, and specific physical events that are required for growth and proliferation. The upper and lower lines portray characteristics of the Martian environment that influence the likelihood of those events.

Water is of particular importance among the environmental characteristics. No known terrestrial organisms can proliferate in the absence of H_2O , and most organisms require it in the liquid state. The atmospheric pressure almost everywhere on Mars is thought to be less than the triple-point pressure of water. Thus, proliferation of most species would require transport to some nonequilibrium microenvironment, such as a pressure pocket under a polar cap, where water can exist in the liquid state. If such microenvironments exist at all, they are probably rare and relatively inaccessible.

In contrast, relatively large amounts of water ice are known to exist on the Martian surface. Some species of cryotolerant terrestrial organism are capable of obtaining enough moisture from ice in the temperature range between $-15^{\circ}C$ and $0^{\circ}C$ to sustain growth and proliferation. We shall use the term aerophiles to mean facultative anaerobes that have this capability. As the name indicates, this is an organism particularly well-suited for growth in the very dry and essentially oxygen-free Martian environment.

Another environmental characteristic of importance is the amount of shielding against UV radiation that is afforded by Martian dust. Because the daytime UV flux at the planet's surface is severe enough to render almost any organism nonviable within a few hours, the long-distance transport would require some such shielding mechanism. Another major source of uncertainty relating to transport is the strength and direction of Martian winds at the point and time of release.

The basic goal in building a proliferation model is to reduce its complexity. A structure must be created that can replace a few large questions with a series of smaller, more readily comprehensible ones. Hence, the scientific expert could develop personal assessments for the key parameters $P(G|B_j)$ from more fundamental assessments that reflect his uncertainty about such factors as those depicted in Figure 15.

In this section, we shall present a simple example of such a model. We emphasize, however, that both the numerical data and the model structure are intended solely as illustrations. The model is predicated on the supposition that the total threat of contamination is traceable almost entirely to the class of terrestrial organisms that we have called areophiles. The overall model structure is shown in Figure 16. It takes the form of a probability tree that represents in stages our uncertainty about proliferation of a single, released organism, given a specific landing area event B_j . (We shall continue to speak in terms of the two landing regions defined in Section IV, i.e., Region 1 corresponds to B_1

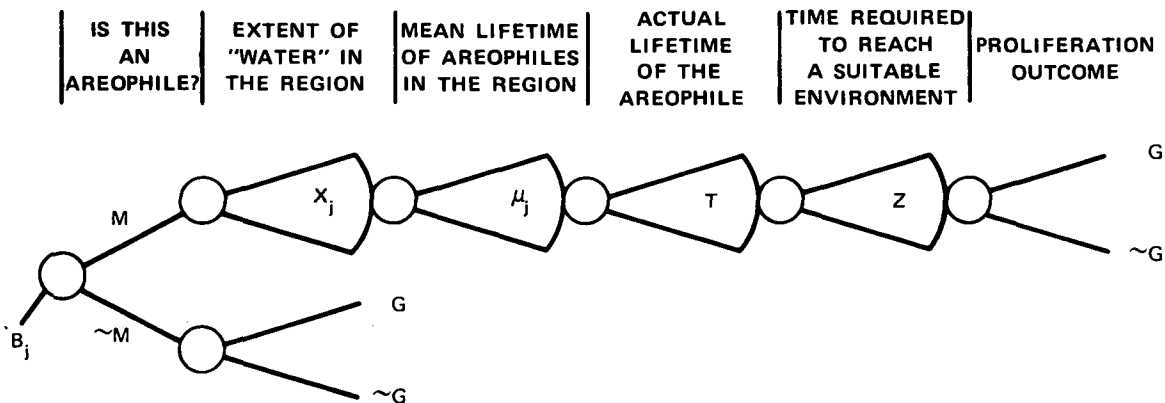


FIGURE 16 AN ILLUSTRATIVE PROLIFERATION MODEL (FOR A SINGLE RELEASED ORGANISM) TO GENERATE $P(G|B_j)$

and Region 2 corresponds to B_2 . The first node of the tree represents the occurrence or nonoccurrence of the event

M = the released organism is an areophile,

and we shall assume the probability assessment

$$P(M) = 0.001,$$

independent of the landing area B_j . Such an assessment would of course reflect a belief that areophiles are rare among those viable organisms aboard the spacecraft. This point will be discussed later. For the case $\sim M$, meaning that the organism is not an areophile, we assess the proliferation probability directly with no further modeling. We assume the assessments

$$P(G|B_1, \sim M) = 1 \times 10^{-7}, \quad P(G|B_2, \sim M) = 1 \times 10^{-8},$$

meaning that proliferation by a nonareophile is considered somewhat more probable in Region 1 but very unlikely in either event.

Given that the organism is areophilic, the next two stages of our probability tree reflect the uncertainty surrounding critical features of the landing area. We define

X_j = the fraction of Region j that is characterized by the intermittent presence of water ice in the temperature range between -15°C and 0°C , $1 \leq j \leq 2$,

μ_j = the mean time of viability (in hours) that would be observed if a great many areophiles were released in Region j , $1 \leq j \leq 2$.

Both of these definitions deserve some comment. The random variable X_j is intended to represent the fraction of Region j that provides the necessary moisture for growth and proliferation of areophiles. Transport to such a site is then required for areophilic growth; however, growth is

not guaranteed. The definition of μ_j is rather subtle. It requires a hypothetical experiment that releases many areophiles into the region; then they are observed until they are rendered nonviable. Because such an experiment has not been performed (and probably never will be), the resulting long-run average lifetime can only be viewed as an uncertain quantity. In this regard, it is similar to the long-run fraction of heads, ϕ , that we used in conceptualizing the thumbtack problem. Our uncertainty as to the value of μ_j results primarily from lack of knowledge about the degree of shielding provided by the Martian dust and the fraction of organisms that would acquire it before becoming nonviable.

Illustrative distributions for X_1 and X_2 are given in Figure 17. The reader may verify that the expected value of X_1 is about one percent, while the expected value of X_2 is about one-tenth of a percent. Thus, Region 1 is considered roughly about ten times more hospitable to areophiles than is the remainder of the planet. The reason for our original differentiation between Region 1 and the remainder of the planet is that the portion of Mars between 50° and 70° north latitude provides optimal conditions for the existence of water ice in the critical temperature range. Temperatures throughout the area are below freezing during the night, but they approach the freezing point in daylight. It is of course simplistic to say that conditions throughout Region 1 are uniform, as the level of detail in our model would imply. In a similar way, to differentiate among other areas of the planet is certainly possible; but for ease of exposition, we shall proceed as if both Region 1 and Region 2 were internally homogeneous (given our current state of information).

An illustrative discrete distribution for μ_1 is given in Figure 18. We shall suppose that μ_2 has the same distribution, and for both $j = 1$ and $j = 2$ it will be assumed that X_j and μ_j are independent random variables.

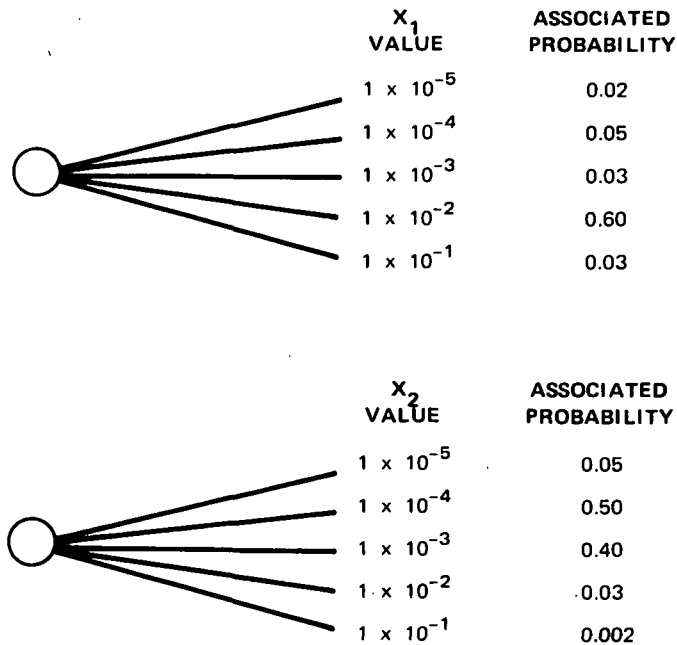
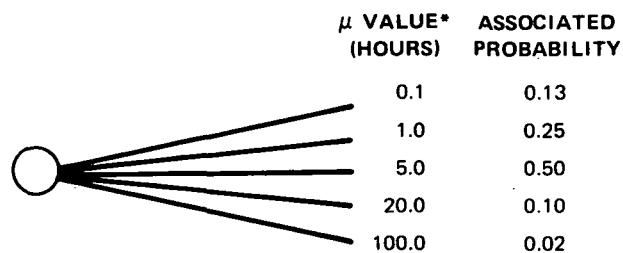


FIGURE 17 ILLUSTRATIVE DISTRIBUTIONS FOR EXTENT OF "WATER" IN REGION 1 AND REGION 2



*Distributions of μ_1 and μ_2 are identical.

FIGURE 18 ILLUSTRATIVE DISTRIBUTION OF MEAN LIFETIME FOR AREOPHILES RELEASED ON THE MARTIAN SURFACE

The final stages of the probability tree shown in Figure 16 reflect uncertainty surrounding the critical quantities.

T = the actual time of viability for the released areophile,

Z = the time required for the released areophile to reach a suitable environment (i.e., a site characterized by the intermittent presence of water ice between -15°C and 0°C).

It will be assumed that the conditional distribution of T is given by

$$P(T \leq t | B_j, M, X_j, \mu_j) = 1 - e^{-t/\mu_j}, \quad t \geq 0.$$

That is, T has a negative exponential distribution that has mean μ_j but does not depend on X_j . (Although there are certain possible theoretical justifications for the use of a negative exponential distribution, we are using this particular distributional form only to simplify our illustrative computations.) In a similar way, we shall assume that A has a conditional distribution of a negative exponential form that depends on X_j but not on μ_j ; the general form is

$$P(Z \leq z | B_j, M, X_j, \mu_j, T) = 1 - e^{-\lambda(X_j)z}, \quad z \geq 0.$$

The distributional parameter $\lambda(X_j)$ will be set so that the probability of the areophile reaching a suitable environment during the first hour of its transport equals X_j ; the fraction of Region j is classified as suitable. That is, we take

$$\lambda(X_j) = -\ln(1 - X_j).$$

The final stage of the probability tree in Figure 16 represents our uncertainty as to whether the areophile will proliferate, given all information contained in previous stages of the tree. If $T < Z$, meaning that the organism is rendered nonviable before reaching a suitable environment, then of course the probability of proliferation is zero. If $T \geq Z$, we are still not certain of proliferation, because such questions

remain as the availability of on-site UV shielding, the presence of pre-formed organic materials necessary for anaerobic growth, and so forth. To account for this remaining uncertainty, we assume the assessment

$$P(G|M, T \geq Z) = 0.10,$$

this probability being independent of the area of release.

We now have all the data necessary to roll back the probability tree in Figure 16 to obtain an overall assessment for $P(g|B_j)$. With our distributional assumptions, it is easy to show that

$$P(T \geq Z|B_j, M, X_j, \mu_j) = \lambda_j(X_j)\mu_j / [1 + \lambda_j(X_j)\mu_j],$$

so that

$$P(G|B_j, M) = (0.10) E \left(\lambda_j(X_j)\mu_j / [1 + \lambda_j(X_j)\mu_j] \right).$$

Here, the expectation is taken over both X_j and μ_j , which we have assumed independent. Using this formula and the distributions given earlier for X_j and μ_j , the reader may verify that $P(G|B_j, M) = 4.25 \times 10^{-4}$,
 $P(G|B_2, M) = 5.12 \times 10^{-4}$.

The rollback is then completed by

$$P(G|B_j) = P(M|B_j)P(G|B_j, M) + P(\sim M|B_j)P(G|B_j, \sim M),$$

which yields

$$P(G|B_j) = (0.001)(4.25 \times 10^{-3}) + (0.999)(1 \times 10^{-7}) = 4.35 \times 10^{-6},$$

$$P(G|B_2) = (0.001)(5.12 \times 10^{-4}) + (0.999)(1 \times 10^{-8}) = 5.22 \times 10^{-7}.$$

At this point, the objective of the proliferation model has been met. The proliferation probabilities generated by the model can be integrated with other data by means of the modified Sagan-Coleman approximation discussed in Section IV and repeated at the beginning of this section. The use of the landing and release data given in Section IV yields a total

contamination probability of

$$\begin{aligned}
 P(C) \approx Q(C) &= \sum_{i=1}^2 \sum_{j=1}^2 P(A_i | B_j) E(N_i) P(G | B_j) \\
 &= (0.90)(1)(5.22 \times 10^{-7}) + (0.02)(1000)(4.35 \times 10^{-6}) \\
 &\quad + (0.03)(1000)(5.22 \times 10^{-7}) = 1.0 \times 10^{-4}
 \end{aligned}$$

Thus, with our illustrative data and model structure, the total contamination probability is approximately equal to the current mission allocation for Project Viking. In this regard, it is relevant to note that some of our illustrative data may be quite optimistic from the standpoint of potential contamination. For example, the mean of our distribution for X_1 is about one percent, but we have encountered some expert belief that a more reasonable figure would be ten or even twenty percent. Also, we have used 0.10 as the probability that an areophile, having reached a suitable environment, will ultimately proliferate. Some expert opinion would have this as high as 0.50. Perhaps most important is the probability of 0.001 that we have associated with the event of a released organism being areophilic. This might be as high as 0.01 or even 0.10 in the opinion of some experts. The question here is what percentage of those organisms that are sufficiently heat resistant to survive terminal sterilization would also fall into our category of areophiles. Much uncertainty appears to cloud the answer, but laboratory experimentation could of course shed much light on the issue.

We have already indicated that a richer model structure reduces complexity and replaces a few large questions with a sequence of smaller, more readily comprehensible ones. Another advantage of additional modeling relates to the well-documented difficulty of directly assessing very small probabilities. If a person says he assesses the probability of event E_1 to be 10^{-4} and that of event E_2 to be 10^{-6} , then you can be quite sure that he considers both events unlikely, with E_1 more likely than E_2 .

Experience indicates, however, that caution should be taken in attaching any absolute significance to the numerical assessments. Would he, for example, rather bet on E_2 occurring or on being dealt a royal flush in a game of 5-card stud poker? Since calculation will show that the latter event has a probability of about 1.5×10^{-6} , the person's assessment would lead to the conclusion that he prefers to bet on the royal flush. Yet one can hardly be confident that such a preference would emerge if the question were asked directly. The problem is that, when asked to assess probabilities smaller than (say) 1/100, we all have difficulty conjuring up familiar reference events that we perceive to be of comparable likelihood. In many applications, a probability of 10^{-2} or 10^{-3} can in fact be used as a working definition of impossibility. One might argue that scientists are unusually comfortable working with numbers as small as 10^{-3} , but we are not convinced that they are accustomed to explicitly dealing with subjective probabilities of this magnitude. One aid that the analyst can provide is a set of familiar reference events--such as the royal flush example if the expert happens to be a poker player--against which relative likelihood can be compared. Greater assistance is usually provided by enriching the model structure and by recognizing that most rare events can be decomposed into a sequence of requisite component events. By modeling that sequence and encoding the conditional probability of each given the occurrence of its predecessors, we enable the expert to assess only probabilities of a readily comprehensible magnitude.

Up to now we have addressed only the problem of how a single scientific expert can be aided in developing a quantitative assessment of planetary contamination risks that is logically consistent with his information and judgment. There is no particular reason to think that two different experts will arrive at the same assessments. In developing a planetary quarantine policy, however, NASA is clearly concerned with

the matter of consensus. One would hope that disagreement among experts is traceable to differences in the information available to them and hence is resolvable by exchanging information. Such intercourse can be greatly facilitated if assessments are built up from a model structure fine enough to demonstrate which specific aspects of the problem (such as existence of liquid water in the equatorial region) are judged differently. Finally, it is apparent that the relative expertise of most scientists is not uniform over all considerations relevant to Martian contamination. This has already been implicitly recognized in the use made of the Sagan-Coleman model. The community of experts asked to assess the value of m under various sterilization procedures is almost totally separate from the community considered most expert in assessing P_g . Therefore, modeling more of the uncertain Martian characteristics that contribute to our uncertainty about contamination would allow further differentiation of expertise.

VI SUMMARY AND CONCLUSIONS

In the initial sections of this report, we have discussed the likely fundamental inadequacies in the model of planetary contamination advanced by Sagan and Coleman. Subsequent investigation shows, however, that a relatively minor modification of the basic Sagan-Coleman formula yields approximations that are generally adequate with data in the range of interest. This approximation formula differs from the original Sagan-Coleman version only through an initial conditioning on landing outcome. It always yields an upper (conservative) bound for the total probability of contamination. We have shown that this appealing feature is lost if the conditioning on landing outcome is deleted.

Because the initial conditioning requires that we differentiate between various geographical areas of the planet, it yields additional benefits. If several missions are sent to Mars and targeted for different landing sites, the individual mission contamination probabilities may not be the same. With the analytical structure that we have proposed, such differences are automatically accounted for by appropriately adjusting the landing model probabilities. The reader may verify that if the illustrative landing data in Figure 8 are changed to indicate a target site in Region 1, the total probability of contamination will increase by almost an order of magnitude.

In Section V, we have discussed the question of how parameter values called for in the approximation formula should be assessed. We have recommended that the region-specific probabilities of growth and proliferation be built up from a supporting model structure. There are three basic reasons for this recommendation. First, such a model helps the scientific

expert to systematically organize his information and judgment about all the diverse factors that are relevant to the process of proliferation on Mars. Second, by building up an assessment in stages, we reduce the extent to which experts must deal with probabilities of incomprehensively small magnitude. Finally, a supporting model structure provides a means for better communication when disagreements arise among experts.

The uncertainty surrounding the possible contamination of Mars by terrestrial microbes can only be addressed through the subjective judgment of experts. This judgment depends on very limited available knowledge, particularly on the Martian environment. A quantitative summary of this judgment can be made in the form of a subjective possibility assessment. Assessing a probability of contamination is a difficult task, but the need for such an assessment was clearly established by 1966 as part of the NASA planetary quarantine policy.

The contribution of this report has been to illustrate how structuring the available information can yield better overall assessment of the probability of contamination. Explicitly including uncertainty on landing mode and landing site appears to be the most important element in carrying out an improved assessment. We have shown in Section V how further refinement can be obtained by modeling the process of transporting certain organisms in a viable state to a location on Mars that would permit them to proliferate. Such a model would allow expert judgment to be assembled and critically examined in a more careful and comprehensive manner.

The great advantage of describing subjective uncertainty through formal probability assessments lies in the role of such assessments within the broader conceptual framework of decision analysis.⁸ The central idea of this analytical discipline for the treatment of decisions under uncertainty is to separate the issues of value and likelihood. On the one hand, we assess (appropriately conditional) probabilities for the uncertain

events that impinge on our decision; on the other hand, we associate value assessments with each of the outcomes ultimately possible as a consequence of the actions taken. The probabilities and value assessments can then be integrated in a systematic and rational way. An application of this procedure to space project planning is given by Matheson and Roths,⁹ and the same conceptual framework underlies the discussion of Levinthal, Lederberg, and Sagan.¹ If uncertainty is measured in any way other than through probability assessments, then we no longer know how to integrate such measures with value assessments to reveal the preferred course of action. We believe that the language of subjective probability, as part of the broader conceptual framework of decision analysis, can be of great value in solving the resource allocation problems that arise at all levels of scientific investigation.

Appendix

DERIVATION OF THE ERROR BOUND

Recall that in Section II we defined E_n to be the event that the n^{th} released organism does not survive to multiple and cause contamination. It is then immediate that

$$P(C|A_i, B_j, N_i = k) = P\left(\bigcup_{n=1}^k (\sim E_n) | A_i, B_j, N_i = k\right).$$

That is, contamination is equivalent to one or more of the events $E_1, \dots, E_k$ not occurring. The probability of this union can be bounded both above and below by the use of the following relationships. For an arbitrary collection of events $Z_1, Z_2, \dots, Z_k$, it is well known that

$$P\left(\bigcup_{n=1}^k Z_n\right) \leq \sum_{n=1}^k P(Z_n),$$

while

$$P\left(\bigcup_{n=1}^k Z_n\right) \geq \sum_{n=1}^k P(Z_n) - \sum_{n=2}^k \sum_{m=1}^{n-1} P(Z_n, Z_m).$$

For our problem, this implies

$$P\left(\bigcup_{n=1}^k (\sim E_n) | A_i, B_j, N_i = k\right) \leq \sum_{n=1}^k P(\sim E_n | A_i, B_j, N_i = k),$$

while

$$\begin{aligned} P\left(\bigcup_{n=1}^k (\sim E_n) | A_i, B_j, N_i = k\right) &\geq \sum_{n=1}^k P(\sim E_n | A_i, B_j, N_i = k) \\ &\quad - \sum_{n=2}^k \sum_{m=1}^{n-1} P(\sim E_n, \sim E_m | A_i, B_j, N_i = k). \end{aligned}$$

We have stated in Section IV that the contamination problem can be treated by analogy to our thumbtack example. In precise mathematical terms, this means that we think it is reasonable to assume that, given a specific landing area, the events $E_1, E_2, \dots$ can be viewed as exchangeable? This, along with our other assumptions, implies that

$$P(\sim E_n | A_i, B_j, N_i = k) = E(f_j), \quad 1 \leq n \leq k,$$

and for any $m \neq n$,

$$P(\sim E_n, \sim E_m | A_i, B_j, N_i = k) = E(f_j^2).$$

Combining all these relationships then gives us

$$\begin{aligned} k E(f_j) &\geq P(C | A_i, B_j, N_i = k) \\ &\geq k E(f_j) - \frac{1}{2} k(k-1) E(f_j^2) \geq k E(f_j) - \frac{1}{2} k^2 E(f_j^2). \end{aligned}$$

Taking the expected value with respect to N_i then yields

$$E(N_i)E(f_j) \geq P(C | A_i, B_j) \geq E(N_i)E(f_j) - \frac{1}{2} E(N_i^2)E(f_j^2).$$

Finally, taking the expected value over all possible landing outcomes gives us

$$Q(C) \geq P(C) \geq Q(C) - EB,$$

where $Q(C)$ and EB are defined as in Section V.

REFERENCES

1. Levinthal, E. C., Lederberg, J., and Sagan, C., "Relationship of Planetary Quarantine to Biological Search Strategy," Life Sciences and Space Research, North-Holland, Amsterdam (1968).
2. Sagan, C. and Coleman, S., "Spacecraft Sterilization Standards and Contamination of Mars," Astronautics and Aeronautics, Vol. 3, p. 22 (1965).
3. Horowitz, N. H., Sharp, R. P., and Davies, R. W., "Planetary Quarantine I; The Problems and the Agreements," Science, Vol. 155, p. 1501 (1967).
4. Murray, B. C., Davies, M. E., and Eckman, P. K., "Planetary Quarantine II: Soviet and U.S. Practices and Policies," Science, Vol. 155, p. 1505 (1967).
5. Schalkowsky, S. and Kline, R. C., Jr., "Analytical Basis for Planetary Quarantine," Environmental Biology and Medicine, Vol. 1, p. 9 (1971).
6. Howard, R. A., "Decision Analysis: Perspectives on Inference, Decision and Experimentation," Proceedings of the IEEE, Vol. 58, p. 632 (May 1970).
7. de Finetti, B., "Foresight: Its Logical Laws, Its Subjective Sources, Studies in Subjective Probability, Kyberg and Smokler (eds), p. 97, (Wiley, New York, 1967).
8. Howard, R. A., "The Foundations of Decision Analysis," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3 (1968).
9. Matheson, J. E. and Roths, W. J., "Decision Analysis of Space Projects: Voyager Mars," presented at the National Symposium, Saturn/Apollo and Beyond, American Astronomical Society (June 11, 1967).