NRL Memorandum Report 4903

The Bayesian Inference Method and Its Application to Reliability Problems

R. LOWELL SMITH

Texas Research Institute, Inc.
5902 W. Bee Caves Road
Austin, Texas 78746

September 15, 1982

NAVAL RESEARCH LABORATORY
Washington, D.C.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Memorandum Report 4903	2. GOVT ACCESSION NO. AD-A121880	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (U) The Bayesian Inference Method and Its Application to Reliability Problems		5. TYPE OF REPORT & PERIOD COVERED Final report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. Lowell Smith*		8. CONTRACT OR GRANT NUMBER(s) Contract N00173-81-W-T201
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas Research Institute, Inc. 5902 W. Bee Caves Road Austin, Texas 78746		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element: 64503N Project Number: S0219-AS NRL Work Unit: (59)-0584
11. CONTROLLING OFFICE NAME AND ADDRESS Underwater Sound Reference Detachment Naval Research Laboratory P.O. Box 8337, Orlando, Florida 32856		12. REPORT DATE September 1, 1982 13. NUMBER OF PAGES 42
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

*Author is an employee of Texas Research Institute, Inc. This report is a final report on work performed under contract N00173-81-W-T201 for the Sonar Transducer Reliability Improvement Program (STRIP) which is sponsored by NAVSEA (63X5).

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayesian Reliability Statistical inference

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Although it is not new, there has been a recent resurgence of interest in the Bayesian approach to dealing with statistical inference problems. Statistical inference, of course, is the activity of characterizing the parameters of mathematical models by utilizing available sampling data. This report discusses as a specific motivation the modeling of reliability problems and for the sake of clarity deals only with inference while avoiding the larger area of decision theory. The classical and Bayesian approaches to evaluating the parameter of the familiar exponential reliability model are (continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

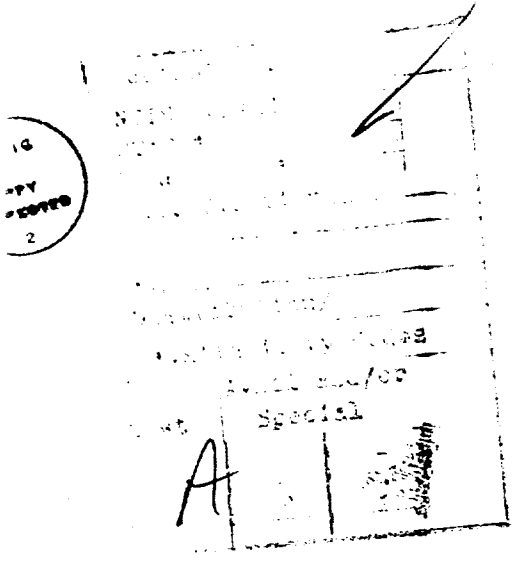
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

compared. Classically, model parameters are unknown constants which can be estimated. From the Bayesian viewpoint model parameters are treated as distributed random variables. As is also true of the classical maximum likelihood method, the determining or informational impact of the sampling data is represented completely by the likelihood function. Operationally, Bayesian inference involves applying Bayes theorem, a celebrated consequence of conditional probability theory. For the sake of completeness, the relevant probability background is developed and Bayes theorem derived. Bayesian inference has the very appealing capacity to incorporate previous information as well as current sampling inputs. Classical results are reproduced in the limiting forms of this involving noninformative prior distributions. Several application examples are discussed illustrating the use of both continuously and discretely distributed data and in one case emphasizing numerical methods.

CONTENTS

TABLES

FIGURES



THE BAYESIAN INFERENCE METHOD AND ITS APPLICATION TO RELIABILITY PROBLEMS

1.0 INTRODUCTION AND SUMMARY

This report represents a very basic introduction to what is referred to as Bayes method in statistics. The author's interest in this subject relates to its application to hardware reliability characterization. Section 2 will clarify this connection in terms of mathematical modeling. Preliminarily one should simply be reminded that the operation or use of nominally identical equipments results ultimately in failures whose times of occurrence may be broadly distributed. Thus there is a strong stochastic or chance aspect to hardware serviceability. Causality is in no way compromised in this. One simply has to recognize that similar hardware items are at least microscopically different and they may see different stresses in service. What one doesn't know about the situation is outcome determining.

Using limited information most efficiently and optimally supporting the decision-making process in the face of uncertainty is the business of statistical inference. Most engineering use of statistics at the present time employs the classical or frequentist approach pioneered by R. A. Fisher. The Bayes method is an appealing alternative based on a somewhat different world view. The two are compared and contrasted in the following pages.

At the heart of both classical and Bayes methods is the concept of likelihood, a measure of the a priori probability that a particular observational outcome will occur given a specification of the statistical model parameters. Likelihood, the idea of summarizing data collectively and completely via sufficient statistics, and confidence statements are all discussed under the reliability mathematical modeling heading.

Bayesian inference processes probability statements via Bayes theorem. Probability in this setting is subjective and conditional in contrast to what is claimed to be the objective classical viewpoint. Bayes theorem itself is derived as a straightforward consequence of one of the axioms of probability theory. As such, as is so often also the case in observational science, it must be evaluated on the basis of its consequences rather than its origins. The important structural properties of Bayesian inference that permit its use and evaluation are developed in this report.

Several examples of the application of Bayesian theory to interesting reliability problems are provided. The first is a textbook kind of pedagogic illustration using the familiar exponential model. Other problems involving different statistical models, continuously or discretely distributed data, and the use of numerical methods help delineate the scope and usefulness of Bayes techniques.

2.0 RELIABILITY MATHEMATICAL MODELING

Mathematical modeling of reliability situations plays essentially the same role as mathematical descriptions in more fundamental settings in the physical sciences. Thus one achieves economy of thought and provides a basis for developing understanding and insights by giving simple mathematical expression to attributes of areas of interest. Here "simple" usually means expressible in terms of relatively few known functions rather than trivial or easy to understand without advanced training. Reliability is a success/failure oriented concern. One asks questions like what is the mean time between failures for a particular piece of hardware or what is the probability of surviving a mission or task of specified duration. Actual failure times depend on hardware construction and use factors, some aspects of which are unknown. As a result, times to failure for similar (nominally identical) equipments used in similar ways differ, i.e., time to failure is a distributed random variable or stochastic quantity. An analytical representation of the distributional aspects of time to failure constitutes what is called a statistical failure model. This takes several interrelated forms in reliability work. Thus, beginning with the time-to-failure probability density function f(t) for completeness

$$f(t) = f(t) \quad . \tag{1}$$

The cumulative (also called the distribution function) of Eq. (1) is the unreliability U(t)

$$U(t) = \int_{-\infty}^{t} f(t)\, dt \quad . \tag{2}$$

Reliability, the probability of successful operation through time t, is

$$R(t) = 1 - U(t) = \int_{t}^{\infty} f(t)\, dt \quad . \tag{3}$$

Finally, hazard rate or the instantaneous rate of failure is

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{\int_{t}^{\infty} f(t)\, dt} \quad . \tag{4}$$

Occasionally we will speak of the quantities exhibited or defined by Eqs. (1) through (4) collectively as the reliability functions associated with a problem of interest. Use of this term "reliability function" in the singular will refer to R(t) alone. Equations (1) through (4) have taken f(t) to be fundamental while the other reliability functions are defined in terms of f(t). Actually there is complete reciprocity among the reliability functions. Any one implies the other three. For example, in terms of the hazard rate

$$\lambda(t) = \lambda(t) \quad , \tag{5a}$$

$$R(t) = \exp\left[-\int_{o}^{t} \lambda(t)\, dt\right] \quad , \tag{5b}$$

$$U(t) = 1 - \exp\left(-\int_0^t \lambda(t)\,dt\right) , \tag{5c}$$

$$f(t) = \lambda(t)\exp\left(-\int_0^t \lambda(t)\,dt\right) . \tag{5d}$$

As a practical matter then, modeling a reliability problem typically translates into characterizing the form of one of the reliability functions and specifying the parameters of this statistical failure model. These two issues, model selection and parameter evaluation, should be treated quite separately. Methods of parameter evaluation don't shed much light on the appropriateness of a model choice. While the discussion in this report is built around some standard important models, motivation for them is not developed here. Statistical failure models are discussed in standard reliability texts [1]. Most of them are familiar distributions that are presented in statistics texts [2] as well. In particular the very commonly used exponential model has been motivated by Epstein [3] and Barlow and Proschan [4]. For our present purpose we suppose an appropriate model has been selected and concentrate on parameter characterization from the classical and especially the Bayesian viewpoints.

2.1 Classical and Bayesian Viewpoints Compared

In reliability work both the classical and Bayesian methods of statistical inference begin from the same point of departure -- a specified statistical model. In each case the problem of interest is to specify the parameter or parameters of the model on the basis of information acquired concerning the operation of the hardware in question. This information is developed in the traditionally most tractable form if a decision is made concerning what constitutes acceptable equipment performance. Then passages through these performance boundaries can be monitored to obtain a set of failure times. It is these failure times that are the observable outcomes of a life testing study. The parameters of the statistical model themselves cannot be directly observed. Rather, inferences concerning the model parameters must be drawn based on the failure times. That is, what must the parameters be to be most consistent with the set of actually observed failure times? We will explore the answers to this question obtained by both classical and Bayesian statisticians.

In the interest of proceeding within a more specific and perhaps familiar framework let us now introduce the exponential failure model, which is characterized by a constant hazard rate. Equations (5) become

$$\lambda(t) = \lambda \qquad (\lambda \geq 0) , \tag{6a}$$

$$R(t) = e^{-\lambda t} \qquad (\lambda \geq 0,\ t \geq 0) , \tag{6b}$$

$$U(t) = 1 - e^{-\lambda t} \qquad (\lambda \geq 0,\ t \geq 0) , \tag{6c}$$

$$f(t) = \lambda e^{-\lambda t} \qquad (\lambda \geq 0,\ t \geq 0) . \tag{6d}$$

In Eqs. (6) the reliability functions depend on time t and the single model parameter λ. Classically λ is understood to be a constant having an unknown

value. Statistical inference is designed to allow one to make as strong a statement as possible about the true value of λ as determined indirectly by observations of failure times. Classically, there are various ways of obtaining these "estimates" of λ. For example, in the case of a complete sample (all items exercised to failure) the mean failure time can be calculated from the data and equated to the expected value of t obtained using Eq. (6d). Thus, if there are n failures labeled t_i, i=1,2,...n,

$$E(t) = \int_0^\infty t\, f(t)\, dt = \frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^{n} t_i \,. \tag{7}$$

This is referred to as the method of matching moments and readily generalizes to yield simultaneous equations for more than one modeling parameter. Other estimation procedures such as probability plotting and regression analysis are also available. However, we shall limit further discussion to the maximum likelihood method. Maximum likelihood estimators have some appealing statistical properties (unbiasedness, minimum variance) and actually incorporate sampling information in the same way as the Bayes approach does (via the likelihood function). This topic will be pursued in Section 2.2.

The Bayesian interpretation of reliability modeling differs from the classical one in a subtle but important way. Again model parameters are taken to be unknown constants; but this terminology has different meanings for classical and Bayesian statisticians. Classically, an unknown constant is a dispersionless scalar quantity of unspecified value. A Bayesian represents the "unknownness" aspect by a probability density function. Mathematically then, a model parameter is treated as a random variable. The Bayesian hastens to emphasize that the parameter is not actually variable in the sense of changing, but that its true value is simply not accessible (in an experiment of finite size). The nomenclature "random quantity" has been introduced to make this distinction.

If we reflect on the matter, thinking of an unknown constant as distributed shouldn't seem too bizarre. Do we not characterize direct (as opposed to indirect or inferential) measurements of stable quantities in exactly this way? Thus, several measurements are taken and processed numerically to yield typically both central tendency and dispersion measures. The quantity in question (length, weight, concentration, etc.) is understood to be constant but unknown within the precision of the measurement technique. Its value is formally represented as distributed.

This can be looked at in another way. Taking a constant to be distributed implies assigning probability to situations that don't occur. Again, there is classical precedent for this. One can shuffle a standard pack of playing cards and inquire with what probability the top most card is the jack of diamonds or some other specified card. Given no further information the answer is 1/52. Distributing probability equally among the alternatives in this way reflects only on our uncertainty of the situation and has nothing to do with any lack of definiteness with respect to how the cards are actually arranged. In thinking about mixing cards, we are dealing with a repeatable process having a denumerable set of possible outcomes. It is possible to realize the frequency limiting behavior that in the long run, on the average the jack of diamonds will turn up on top 100/52 percent of the time. What

the Bayesian does is assert the relevance of assigning probabilities to situations that do not necessarily exhibit a frequency limit.

In subsequent sections of the report we will examine some of the methods of statistical inference in greater detail. To bring this section to a close, let us take note of the major operational differences implied by the two approaches -- classical and Bayesian.

Classically, statistical model parameters are unknown constants. Inference methods yield parameter estimates. These estimators themselves turn out to be distributed (dependent on the unknown true parameter values). Hence a substantial part of classical statistical inference addresses developing the statistical properties (biasedness, efficiency, etc.) of estimators. One implication of this is that confidence statements do not relate in a very satisfactory way to model parameters directly. Another property of classical inference situations is that conclusions often depend on experimental censoring procedures (stopping rules).

Let us contrast the Bayesian situation. Model parameters are random quantities described directly in distributional terms. Inference proceeds by modifying the prior parameter distribution (probability density function) via the sample likelihood to obtain a posterior distribution. Thus in contrast to classical inference the parameter space is directly accessible. There are neither estimators nor complicated estimator statistics. Confidence intervals are developed quite naturally by integrating the posterior density and directly represent valid probability statements on the model parameters. Typically, how a particular experimental outcome happens to be realized is of no consequence -- the stopping rule is said to be noninformative.

Bayes methods provide the capability of integrating previous experience (through the prior) with what is learned from the current round of testing. So far this description makes Bayesian inference sound like a very appealing alternative. It is only fair to temper this somewhat. The key difficulty is choosing an appropriate prior. How does one cast what one knows generally about a hardware item into a distributional description of a modeling parameter? One approach is to ignore this history and construct what is called an ignorance prior. From this point of departure one would like to see conclusions drawn classically and from the Bayesian viewpoint coalesce in reflecting only information developed in the current test. This has been demonstrated under a number of circumstances. However, ignorance priors are typically improper (non-normalizable) and a focus of continuing debate.

2.2 Likelihood

In the next several sections of the report we discuss statistical concepts that are important from both the classical and Bayesian viewpoints. This will be done by developing the classical maximum likelihood approach and then comparing with Bayes method in Section 3.0 and its subsections. The concept of likelihood is quite fundamental in this. At least one author [5] places likelihood at the heart of an approach to statistical inference (method of support) without being either a classicist or Bayesian.

We have talked about likelihood and now need to define it. To do this we need to introduce the concept of conditional probability (Bayesians view

all probabilities as conditional on previous history.). Reliability problems represent an excellent setting in which to discuss conditional probability. Consider a life test that yields time-to-failure data. Then the elements of the discussion are a statistical model not being questioned, a set of statistical hypotheses H being evaluated, and the experimental results or data D. When a model has been specified and a particular hypothesis (such as specification of the model parameters) imagined to be true, probabilities for an exhaustive set of mutually exclusive consequences or outcomes (all possible forms the data might have taken) can be calculated. Since one or another of the potential outcomes must occur with certainty, these probabilities have to sum to unity. The problem of statistical inference involves inverting this philosophy. That is, a particular consequence is available as an experimental fact and one wishes to make an associated statement about the probable validity of one or more hypotheses. In probability language the likelihood L of the hypothesis given the data is defined as

$$L(H|D) \propto P(D|H) \quad , \tag{8}$$

where the notation reads the probability of the data D given the hypothesis H or the probability of D conditioned on H. If a particular hypothesis were known to be true, then $p(D|H)$ would be a true probability (i.e., sum or integrate to unity on D). In the likelihood context Eq. (8) refers to a fixed D and is intended to span a number of candidate hypotheses H (or a range of model parameter values). Viewed in this way Eq. (8) is not a true probability since hypothesis space cannot generally be partitioned in a mutually exclusive, exhaustive manner.

Let us return to consideration of a set of failure times obtained in sampling from an exponential time-to-failure distribution. The data are failure times for failed units and survival times for unfailed units. The hypothesis is that λ is the true value of the model parameter. For the sake of definiteness, let us consider the testing of n nominally identical items until the occurrence of the r^{th} failure. The likelihood is the joint probability given λ that the failure times are the observed t_i, $i = 1,2,\ldots r$, and that $n - r$ units survive to suspension of the test at t_r. Using Eqs. (6d) and (6b), Eq. (8) becomes

$$L \propto \left[\prod_{i=1}^{r} \lambda e^{-\lambda t_i} dt_i\right]\left[e^{-\lambda t_r}\right]^{n-r} \quad . \tag{9}$$

The observation intervals dt_i are present because Eq. (6d) is a probability density function. However, the timing resolution imposed on a life testing experiment is largely irrelevant to the use of the likelihood function as a measure of support provided to different hypotheses by a particular body of data. Thus the quantity $\prod_{i=1}^{r} dt_i$ may be absorbed into the proportionality constant implicit in Eq. (9). Furthermore, since one is ordinarily interested only in relative likelihoods against a particular sampling outcome, Eq. (9) is usually written as the equality (interpreted as dimensionless)

$$L = \left[\prod_{i=1}^{r} \lambda e^{-\lambda t_i}\right]\left[e^{-\lambda t_r}\right]^{n-r} \quad . \tag{10}$$

Equation (10) has been specialized to the exponential statistical failure model for illustrative purposes. However, its structure is similar for any situation where failures are independent. That is, the likelihood is a product of factors representing relative probabilities of observed failures (via the time-to-failure pdf's) and observed successes (via the reliability). If an unfailed unit is withdrawn prior to termination of the test, its proper weighting is $R(t_w)$ where t_w is the time of withdrawal.

2.3 Sufficient Statistics

Equation (10) can be written more compactly via some rearrangement as

$$L = \lambda^r e^{-\lambda T} \quad , \tag{11}$$

where

$$T = \sum_{i=1}^{r} t_i + (n-r)t_r \quad . \tag{12}$$

Equation (12) represents, on a per unit basis, the total exposure (to operating conditions) of the hardware being evaluated. Conventionally then, T is referred to as the total time on test. From Eq. (11) we see that the actual time-to-failure sampling data influence the likelihood function only through r, the number of failures observed, and T, the total test time. These quantities r and T are said to be sufficient for a complete description of the problem at hand. Interestingly the number of items tested n and the individual failure times t_i are not of specific concern beyond their impact on T.

We will see in Section 3.3 that the existence of conjugate distributions is closely related to situations that admit to description in terms of sufficient statistics. In the Bayesian context, at least, it usually doesn't matter how the particular values of r and T are obtained. That is, particular values of r and T may have resulted because:

1. The test plan called for stopping at the r^{th} failure.
2. A time-terminated test was planned and executed.
3. Either of the above plans was altered when one or more units had to be withdrawn during the test for other purposes.

For the Bayesian all of these situations would be characterized by the same r and T and exactly the same inferences drawn. The experimental stopping rule is said to be noninformative in such a case. In contrast, classical procedures will typically distinguish the above situations and treat failure-terminated and time-terminated tests differently.

2.4 Maximum Likelihood

R. A. Fisher [6] introduced the idea that estimates of the values of the parameters of a statistical model could be obtained by maximizing the likelihood function given a particular experimental outcome. That is, for what model parameter values are the observed data collectively more probable than

for any other parameter choices? If the likelihood is a function $L(D|\alpha_i)$ of the data D (or corresponding sufficient statistics) and model parameters α_i, the maximum likelihood estimators are obtained by solving the simultaneous equations (one for each α_i)

$$\frac{\partial}{\partial \alpha_i}\left[L(D|\alpha_i)\right] = 0 \quad . \tag{13}$$

In the case of the one-parameter exponential model and using Eq. (11), this reduces to the single statement

$$\frac{d}{d\lambda}\left[\lambda^r e^{-\lambda T}\right] = 0 \quad . \tag{14}$$

Equation (14) can be solved directly to obtain the maximum-likelihood estimator $\hat{\lambda}$. However, it is equivalent and frequently simpler to maximize the logarithm of the likelihood. For the exponential problem we have been considering, this yields

$$\frac{d}{d\lambda}\left[r \ln\lambda - \lambda T\right] = 0 \quad . \tag{15}$$

Solving Eq. (15) the maximum-likelihood estimator of the model parameter λ is

$$\hat{\lambda} = \frac{r}{T} \quad . \tag{16}$$

As is first apparent from the structure of Eq. (11), the maximum-likelihood estimator for this problem depends only on the sufficient statistics r and T. However, if the entire life testing experiment is repeated with another sample drawn from the same parent population, different values of r or T or both will be obtained. It is clear then that the estimator $\hat{\lambda}$ is itself a distributed random variable. The distributional properties of $\hat{\lambda}$ for this problem have been worked out in a pioneering paper by Epstein and Sobel [7]. They found that the quantity $z = 2r\lambda/\hat{\lambda}$ is χ^2 distributed with 2r degrees of freedom. That is

$$g(z) = \left(\frac{1}{2^r(r-1)!}\right) z^{r-1} e^{-z/2} \quad . \tag{17}$$

Using standard variable transformation methods (see Appendix B of Ref. 8 for example) Eq. (17) implies that $\hat{\lambda}$ is distributed as

$$h(\hat{\lambda}) = \frac{1}{(r-1)!\hat{\lambda}}\left(\frac{r\lambda}{\hat{\lambda}}\right)^r e^{-r\lambda/\hat{\lambda}} \quad . \tag{18}$$

Evaluation of Eq. (18) requires that the true modeling parameter λ be known. In engineering practice one is rarely, if ever, so fortunate as to have this information available. The Bayesian approach turns the problem around. One

is not concerned about the distribution of estimators. Rather, a single such result is recognized as an experimental fact and one inquires about the range of true parameter values that are compatible with it. To assist in visualizing this concept Fig. 1 is a plot of the likelihood function [Eq. (11)] for our example problem. The sufficient statistics, r = 10 and T = 88,827 hours, are developed from the simulated time-to-failure data presented as Table 1. The maximum-likelihood estimator, $\hat{\lambda} = 1.13 \times 10^{-4}$ hours, is the abscissa of Fig. 1 for which the corresponding ordinate is maximum as indicated. However, the figure also shows that there is a high probability of the observed data being associated with any other parameter value in the vicinity of the maximum-likelihood estimator.

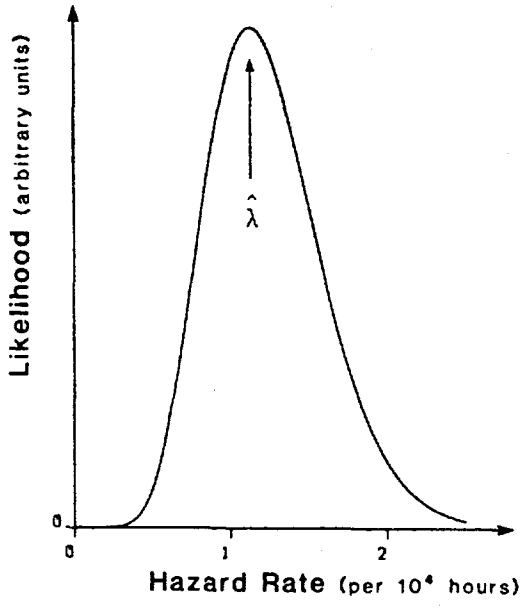


Fig. 1 - Plot of the likelihood function [Eq. (11)] for an exponential population based on r = 10 failures observed in T = 88,827 hours total time on test.

Table 1 - Synthesized ordered failure times representing sampling from an exponential parent population.

TIMES TO FAILURE (Hours)	ANCILLARY INFORMATION
265	n = 20 Samples Placed on Test
934	
1171	Sufficient Statistics:
1350	
2725	r = 10 Failures Observed
3155	
3542	T = 88,827 Hours Total Test Time
4606	
4892	
6017	

2.5 Confidence Statements

In most situations in engineering or science two numbers are used in reporting observational results. These are usually some average or central tendency measure of repeated measurements and a self-consistency or quality descriptor called the uncertainty, standard deviation, probable error, etc. In ordinary observations of directly accessible physical properties (length, weight, voltage, etc.), this represents the characterization of experimental error superimposed on the true values in question. When stochastic variables such as time to failure, number of failures, or total time on test are observed in replicated experiments, the variability is intrinsic rather than associated with some limitation of the measurement tool employed. In either situation, one can ask with what probability yet another (future) observation would fall within a specified range or interval. The interval boundaries in such a description are called confidence limits (upper and lower) and the probability is referred to as the confidence level. A confidence statement is trivially related to the area under (or the cumulative of) the associated probability density function. For example, the probability statement on z [distributed per Eq. (17)] at a confidence level of $1 - \alpha$ is

$$P\left[\chi^2_{(1-\alpha/2),2r} \leq \frac{2r\lambda}{\hat{\lambda}} \leq \chi^2_{\alpha/2,2r}\right] = 1-\alpha \quad , \tag{19}$$

which can be evaluated using tabulated quantiles of the χ^2 distribution. Or, equivalently, two-sided confidence limits on $\hat{\lambda}$ at the $1-\alpha$ confidence level are given by

$$L_2 = \frac{2r\lambda}{\chi^2_{\alpha/2,2r}} \leq \hat{\lambda} \leq \frac{2r\lambda}{\chi^2_{(1-\alpha/2),2r}} = U_2 \quad . \tag{20}$$

Common practice is to invert Eq. (20) to obtain what are claimed to be confidence limits on the model parameter λ, i.e.

$$\left(\frac{\hat{\lambda}}{2r}\right)\chi^2_{(1-\alpha/2),2r} \leq \lambda \leq \left(\frac{\hat{\lambda}}{2r}\right)\chi^2_{\alpha/2,2r} \quad . \tag{21}$$

Mann et. al., in Section 8.1.2 of Ref. 1, point out the inconsistency of this inversion process since classically $\hat{\lambda}$ is distributed and λ itself is not. Equation (21) does have a proper interpretation; namely, that in the long run $100(1 - \alpha)$ percent of the different intervals so constructed will contain the true parameter value λ. From the Bayesian viewpoint this confusion disappears entirely since a distributional model parameter space is always accessible.

3.0 STRUCTURE OF BAYES METHOD

Thus far in the report we have looked at some of the general characteristics of Bayesian inference in conjunction with the development of a preferred classical approach. Now we turn to the exposition of more specific structural properties of the Bayes alternative. Bayes theorem itself is a statement relating conditional probabilities. We open this discussion with a review of the underlying probability ideas.

3.1 Probability

In trying to assimilate in useful ways the results of reliability or life testing experiments, we are dealing with uncertain events. We do not know in advance how much time will be required to induce failures in all items of a test population. Or, if the test time is decided upon initially, the number of failures that will occur is uncertain. Even after this information becomes available, the statistical model parameters introduced as descriptors of the situation remain to some degree uncertain or incompletely specified. Lindley [9] argues that all uncertainties are of the same genre and properly measured on a probability scale. The first three chapters of Ref. 9 contain a very readable discussion of uncertainty and probability including numerous examples from everyday life. We shall have to be content with a more terse presentation here.

Statistics texts typically develop probability ideas using set theory. For our own purposes it will suffice to think in terms of the set of possible outcomes a life test might yield. This we call a sample space. The sample space may be discrete as in the case where one counts the total number of failures occurring in a particular test. Continuous sample spaces are also commonly encountered such as occurs when individual failure times are specified to arbitrarily high precision. In either case it is usually desirable to arrange that the events or particular sampling outcomes be exclusive and exhaustive. Here exclusive means that one testing result preempts all the other possibilities. Exhaustive refers to the completeness of the sampling space description, i.e., no potential outcome has been overlooked in specifying the range of alternatives. Consider an example. Suppose we set up to run a life test of duration τ on n similar equipments. The outcome is that some number r of failures will occur. This result is exclusive; that is, if r equals 3, it cannot also be some other number. Furthermore, saying that r must fall in the range of integers $0,1,\ldots,n$ exhausts all the possibilities for the problem. Thus, the set $(0,1,\ldots,n)$ is exclusive and exhaustive for r.

Let us turn now to making some probability statements with respect to sample spaces. The word "probability" is used as shorthand for the idea of probability of occurrence of some event or specified element of the sample space. In mathematical notation we write p(E) for the probability of occurrence of the event E. In addition to stating what event on which attention is focused, we need to describe the experiment (stresses imposed, duration, etc.) and the criteria for determining what outcome has occurred (failure definition). Lindley [9] calls this ancillary information the history H of the problem. He asserts that every probability statement depends on expression or at least implicit understanding of the relevant history. This can be made explicit in the notation by writing $p(E|H)$ which is read the probability of E given (or conditioned on) H. In what follows we share Lindley's

[9] interpretation and his lead in using the simpler notation omitting H for most purposes.

An event that is certain to occur is conventionally assigned a probability of 1. A probability of zero is taken to describe an event that cannot possibly happen. An event A that is possible but less than certain has a probability between these two extremes. Expressed as an inequality, this statement represents the first law (or convexity rule) of probability

$$0 \leq p(A) \leq 1 \quad . \tag{22}$$

The second law of probability tells us under what circumstances probabilities may be added. If A and B are two exclusive uncertain events, the probability of one or the other occurring is

$$p(A \text{ or } B) = p(A) + p(B) \quad . \tag{23}$$

Equation (23) is called the addition rule of probability and is readily generalized to more than two events. It is important to remember that it refers to exclusive events. For example, in casting a standard six-faceted die the probability of an even number showing in a single throw is p(2) + p(4) + p(6). Under certain circumstances probabilities may also be multiplied. Thus the probability that two uncertain events A and B will both occur is

$$p(A \text{ and } B) = p(A)p(B|A) \quad . \tag{24}$$

This is the third or multiplication law of probability. Clearly if A and B are exclusive, p(B|A) = 0 and Eq. (24) yields the expected result that the probability of the simultaneous occurrence of mutually exclusive events is zero. On the other hand, suppose a package contains 10 metal parts and 16 plastic parts. Let half the metal components be painted black while one-fourth of the plastic items are also black. These aspects -- type and color -- are not exclusive. Thus we can ask the probability of selecting at random from the carton a black metal part in a single trial. Applying Eq. (24) we find p(metal and black) = p(metal)p(black if metal) = (10/26)(1/2) = 5/26. Notice this argument can be reversed yielding p(black and metal) = p(black)x p(metal if black) = (9/26)(5/9) = 5/26. Consider another example using the same package of plastic and metal parts. Suppose we ask the probability of drawing two metal components in a row in two trials. Equation (24) applies to this situation also since the outcome of the first trial affects the odds or chances that apply to the second trial by altering the population being selected from. Thus p(2 metal) = p(metal)p(metal if metal) = (10/26)(9/25) = 9/65.

Equation (24) can also be extended to include any number of events. For the case of three events A, B, and C,

$$P(A \text{ and } B \text{ and } C) = p(A)p(B|A)p(C|AB) \quad , \tag{25}$$

where AB written together in the argument of p(C|AB) means that the probability of C is conditional on both A and B. Equation (25) and its generalization to larger numbers of events applies to the situation where the results of previous trials affect the odds applicable to the next and following trials. Our example in the previous paragraph of a container of mixed parts illustrates

this. However, if the item withdrawn in the first trial were replaced before the second item was selected, the trials would be independent. That is, the odds applying to all trials would be the same because of the restoration of the test population to its original condition prior to each trial. In the case of independent sampling (B independent of A, C independent of B and A etc.), $p(B|A) = p(B)$ and $p(C|AB) = p(C)$ so that Eq. (25) takes the simpler form

$$p(A \text{ and } B \text{ and } C) = p(A)p(B)p(C) \quad . \tag{26}$$

Equation (26) (and its generalization to more events) is a very important result which applies to tossing coins, casting dice, and the observation of independent failure times in reliability and life testing situations.

Equations (23) and (24) respectively deal with the addition and multiplication of probabilities. The reader is reminded to focus attention also on the conditional aspects of probability statements. Thus Eq. (23) applies to exclusive events while the implications of Eq. (24) are more interesting for events which are not exclusive. We close this section with a fourth probability law, a statement in which the operations of addition and multiplication occur together. If A and B are two <u>exclusive</u> and <u>exhaustive</u> events and E is any other uncertain event, then

$$p(E) = p(A)p(E|A) + p(B)p(E|B) \quad . \tag{27}$$

Equation (27) readily generalizes to any number of exclusive and exhaustive events. It is an example of decomposing a quantity of interest in terms of a complete set of basis functions. Analogous procedures include geometric projection in Cartesian vector calculus or expansion in terms of a complete set of orthonormal basis states in the abstract vector calculus of quantum mechanics. In the Bayesian statistics context applying Eq. (27) is often colorfully referred to as "extending the conversation." Actually Eq. (27) can be derived from Eqs. (23) and (24) as is shown in Appendix A. It is therefore an example of a probability theorem. For the axiomatic basis or externally accepted structure of the probability language only Eqs. (22) through (24) are needed. This concise grammar is the key to speaking and understanding the rich calculus of probabilities.

3.2 Bayes Theorem

Bayes theorem was established over 200 years ago [11] as the central probability statement on which the Bayesian inference method is built. Given the background of the previous section, this famous result can be developed with remarkable ease. From Eq. (24) the probability that two uncertain events A and B will both occur is $p(A \text{ and } B) = p(A)p(B|A)$. The order of labeling the events is immaterial so that

$$p(A)p(B|A) = p(B)p(A|B) \quad , \tag{28}$$

a result we have already seen illustrated by an example in the previous section of the report. A trivial rearrangement yields

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \quad , \tag{29}$$

which is Bayes theorem. For our statistical inference or hypothesis testing purposes in reliability or life tesing situations, the event A is the body of data D and B represents some hypothesis H. Equation (29) becomes

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)} \quad , \tag{30}$$

where the factor $p(D|H)$ is recognized as the kernel of the likelihood defined by Eq. (8). It may happen that there are several hypotheses that one wishes to test for compatibility with the data D. If these can be sorted out into an exclusive and exhaustive set having k elements H_i, the denominator of Eq. (30) can be expanded via a generalized Eq. (27) to yield the set of k results

$$p(H_i|D) = \frac{p(D|H_i)p(H_i)}{\sum_{i=1}^{k} p(H_i)p(D|H_i)} \quad . \tag{31}$$

Equation (31) is a commonly encountered form of Bayes theorem for a discrete decomposition of hypothesis space [1, 10]. The analog to Eq. (31) where we have selected a particular statistical model and are dealing with a continuous range of possible parameter values λ is

$$p(\lambda|D) = \frac{p(D|\lambda)p(\lambda)}{\int p(\lambda)p(D|\lambda)d\lambda} \quad . \tag{32}$$

The extension to models having more than one parameter is straightforward. Notice that the appearance of the factor $p(D|H_i)$ or $p(D|\lambda)$ in both the numerator and denominator of Eqs. (31) and (32) allows the corresponding likelihood [Eq. (8)] to be unambiguously substituted without regard to actually evaluating the missing proportionality constant. Thus making the parametric dependence of L explicit, Eq. (32) becomes

$$p(\lambda|D) = \frac{L(\lambda|D)p(\lambda)}{\int p(\lambda)L(\lambda|D)d\lambda} \quad . \tag{33}$$

A similar argument applies to the prior distributions in Eqs. (31) through (33). Thus the probability mass functions $p(H_i)$ for the discrete case and the probability density $p(\lambda)$ for the continuous case need be specified only within an arbitrary multiplicative constant for the purpose of implementing Bayes theorem in the forms displayed. Said another way, the Bayesian posterior $p(H_i|D)$ or $p(\lambda|D)$ will turn out to be normalized (sum or integrate to unity) whether or not the corresponding prior distributions exhibit this property. Normalization is required if a distribution function is to have a proper probability interpretation. Some functions used as Bayesian priors possess infinite norms and are referred to as improper. They are the subject of some interpretational controversy but even these functions cause no trouble in implementing Eqs. (31) through (33).

Bayes theorem is sometimes written for the continuous distribution case, again making explicit reference to the history H of the problem, as

$$p(\lambda|DH) \propto L(\lambda|DH)p(\lambda|H) \quad , \tag{34}$$

The constant of proportionality is established by demanding that the left side of Eq. (34) integrate to one [note the equivalence of this to Eq. (33).]. Equation (34) is a form very suitable for discussing the philosophy of Bayes method. Thus, what one knows about the situation (the history H) motivates the selection of a particular statistical model. This model choice augmented by the current observational results (the data D) fixes the likelihood function $L(\lambda|DH)$. The likelihood modifies or shapes the prior distribution $p(\lambda|H)$ to give within a multiplicative constant factor the posterior distribution $p(\lambda|DH)$. What the prior is, is one's best assessment given previous experience H of the statistically weighted probable range of values expected to include the true value of the model parameter. The posterior is the prior as modified to reflect the impact of the new information D; i.e., the best description given now both H and D. This can be an iterative procedure. Thus the new history embraces both the old history and the current data, the current posterior becomes the new prior, and a new experiment may be conducted. Operationally, this is straightforward enough. The part that is disquieting for some and the area where real creative input is required is selection or specification of an appropriate prior distribution. The prior is often described as subjective or as the observer's personal probability, and is intended to represent true belief in the hypothesis or probable range of the model parameter. This kind of language causes some people to reject the Bayesian approach entirely because they feel that a technique to be used for scientific or engineering purposes must be objective or independent of who implements it. The reader is urged not to be too concerned about this objection. Science, while seeking objectivity, does have its subjective aspects. There seems to be a considerable need for dialogue before agreement on basic issues can be reached. Closer to the problem at hand, the choice of an appropriate statistical model to represent a reliability or life-testing situation is itself a very subjective matter. A more fundamental aspect of this issue is raised by de Finetti [12] who asserts that no probability enjoys an existence independent of the perception of an observer. That is, probability is intrinsically subjective by nature. De Finetti's thesis is no less than revolutionary. Nevertheless, it has already attracted many adherents and, of course, neatly disposes of many of the objections to the Bayesian paradigm or world view. This is so because Bayes priors and posteriors are nothing more than probabilites. Further discussion of the differences between the classical and Bayesian views of probability is presented in Appendix B.

3.3 Conjugate Distributions

Analytical life in mathematics, in statistics, and in the physical and biological sciences is full of compromises. Thus, it is commonplace and usually desirable to give up a bit of rigor in an argument or description in favor of tractable mathematics. Such is the case in Bayesian inference. One can avoid tedious numerical procedures (although these are not so unpalatable in the computer era) by discovering and making use of what are called conjugate distributions. This terminology refers to the situation where the Bayesian prior and posterior distributions belong to the same family of functions. The presentation of examples of conjugacy is deferred to Section 4.0 and its subsections where application examples are discussed. A number of conjugate distributions useful in connection with reliability and life-testing problems are cataloged

in Chapter 3 of Ref. 13. Conjugate distributions are associated with and implied by the likelihod function appropriate to a given problem. Thus one looks for structural features such as factors common to the likelihood and a tentative prior so that their product is a similar mathematical entity. Situations for which conjugate distributions exist are also referred to as closed under sampling. As has been mentioned, this kind of closure is more a convenience than a fundamental concern. Many conjugate families are rich enough in the properties of their members that any of a wide range of prior beliefs can be quite adequately represented for the purposes of Bayesian inference.

3.4 Robustness

Shortly we shall be looking at examples of the use of the conjugate or convenience priors discussed in the preceding section. Their use forces the posterior distribution to be more strongly peaked or localized than the corresponding prior. This, of course, is the proper result of incorporating the new data via the likelihood provided that a reasonable statistical model has been advanced. However, there is nothing in the use of conjugate distributions alone to call attention to a poor choice of model or prevent the Bayesian statistician from being happily deceived by his own analysis in such a case. This difficulty is normally avoided by careful selection of a suitable statistical model for the problem. It is also possible to work with priors that are more forgiving. These functions can be shaped by the current data to become either more peaked or more diffuse and are said to be robust. The former outcomes (more localized posterior) lends support to the choice of model. The reverse is true if the posterior is less localized than the prior suggesting that a more appropriate statistical model be looked for. Robustness is discussed further in the papers by Dempster, Huber, and Rubin in Ref. 14.

3.5 Classical Limiting Behavior

One of the major practical advantages of the Bayesian inference method is that it allows previous and new or current experience to be combined in a natural way. Serious Bayesian protagonists also advance more fundamental arguments that the Bayes approach overcomes certain logical inconsistencies of classical methods. We leave this sort of proselytizing to others since this report is concerned more with the mechanics than the justification of Bayesian inference. The point to be made in this section is that if one chooses to ignore previous history and focus attention only on the results of a current set of observations, then the Bayesian and classical methods can be compared in their processing of the same body of information. Bartholomew [15] has reviewed some of the literature comparing Bayesian and classical inference and addresses some remaining open questions. Under some fairly general circumstances a prior distribution can be chosen such that Bayesian inference has the frequency or confidence property and is said to agree with the classical approach. Such a prior is called a noninformative or ignorance prior since it is intended to represent the absence of previous experience. As an aid to constructing ignorance priors, Jeffreys [16] has formulated an invariance principle dealing with the idea that ignorance about a model parameter implies ignorance about any function of that parameter. Thus admissible ignorance priors must lead to transformed distributions that also appropriately convey ignorance. Bartholomew [15] considers situations where application of these ideas alone does not bring classical and Bayesian

inference into complete agreement. He argues that stopping criteria are informative classically and must be mirrored by adjustments of the Bayesian prior. This is true because different stopping conditions ordinarily leave the Bayesian likelihood unchanged.

The discussion of this section is far from exhaustive. Its purpose, however, is to suggest that evidence is accumulating that Bayesian inference includes classical inference as a special case. The argument involves ignorance priors which often exhibit pathological mathematical properties. Appearance of these infinities does little to help convince frequentists that they should become Bayesians. The situation is largely artificial, however, since it is difficult to envision designing an actual evaluation experiment wherein one knows nothing at all about the hardware involved. Thus prior distributions should normally be informative, noncontroversial, regular functions.

4.0 APPLICATION EXAMPLES

We now turn to demonstrating the use of Bayes method via some examples representative of reliability and life-testing situations. Two statistical failure models are considered. These are the exponential and the Gamma-distributed models. The exponential time-to-failure distribution is known to be appropriate for complex equipment as well as for components such as semiconductors, for which obsolescence usually preempts wearout as a concern. In contrast, the more general gamma distribution can be peaked and localized as is descriptive of many situations where a systematic loss of integrity termed wearout leads to failures in service which are clustered in time. In the case of the exponential model we consider only a failure-terminated test.

Raiffa and Schlaifer, in Chapter 10 of Ref. 13, and Locks, in Chapter 7 of Ref. 17, consider also sampling from the exponential distribution involving two types of time termination (predetermining total time on test or not). All of these situations have the same Bayesian description as one can see most easily because they have the same likelihood kernel [Eq. (11)]. In a final example we interpret data directly from a success/failure point of view and regard reliability as the random variable associated with a Bernoulli process. All the examples follow the general procedural scheme presented in Section 4.1.

4.1 General Procedural Format

In this section of the report we present a concise summary of the steps involved in obtaining a Bayes solution to characterizing the parameters of a statistical model given previous experience and current data. The notation is tailored specifically to the case of a continuously distributed single model parameter. For more than one parameter, make the replacement $\lambda \rightarrow \alpha,\beta,\ldots$ as appropriate. If the model parameter is discretely distributed, substitute mass functions for density functions and replace integrations by sums.

Bayes method consists of:

1. Select a statistical model and obtain an expression for the distribution of the stochastic variable x.

2. Specify a prior (marginal) distribution $\pi(\lambda)$ on the modeling parameter λ.

3. Using the result of step 1, express the conditional probability of the experimental outcome (a set of observations x_i) with respect to a given value of the modeling parameter. This is the likelihood

$$L(\lambda|x_i) \propto p(x_i|\lambda) \quad .$$

4. Multiply the results of steps 2 and 3 to obtain the joint probability of the experimental outcome and the parameter λ:

$$g(x_i,\lambda) = \pi(\lambda)p(x_i|\lambda) \quad .$$

5. Integrate overall parameter space to determine the marginal distribution $\pi(x_i)$ of the experimental outcome

$$\pi(x_i) = \int_{\text{all } \lambda} g(x_i,\lambda)\, d\lambda \quad .$$

6. The quotient of steps 4 and 5 is the Bayesian posterior which is the desired parameter distribution conditioned on the observed data

$$\pi(\lambda|x_i) = \frac{g(x_i,\lambda)}{\pi(x_i)} \quad .$$

Steps 1 and 2 are the essential subjective inputs to Bayesian inference. The rest is straightforward mathematical manipulation which may or may not involve the convenient data summary functions called sufficient statistics. Step 3 may involve integration to find the cumulative time-to-failure distribution to represent survival to time t_c in treating censored data. The integration of step 5 often involves familiar conjugate distributions but may also be carried out numerically to reflect virtually any form of prior belief (presented as a sketched pdf for example).

4.2 Exponential Time-to-Failure Distribution

In this section we treat as an application example the familiar exponential reliability model already introduced in Section 2.1. The exponential model is important because of its simplicity, wide applicability and use, and unique status as the basis for the military handbook reliability prediction methods for electronics components [18]. For comparison purposes we shall consider from the Bayesian viewpoint the same problem Epstein and Sobel treated classically in their celebrated 1953 paper [7]. Thus consider r failures among n items in total time on test T [Eq. (12)].

Proceeding as described in the previous section: The time-to-failure distribution is $f(t) = \lambda \exp(-\lambda t)$ [Eq. (6d)]. As a prior distribution on the parameter we select the improper ignorance prior $\pi(\lambda) = 1/\lambda$ introduced by Jeffreys [16] and used by others [19]. The likelihood in terms of the sufficient statistics r and T for this problem has already been developed as Eq. (11). Thus in terms of the notation of Section 4.1,

$$p(r,T|\lambda) \propto \lambda^r e^{-\lambda T} \quad . \tag{35}$$

Combining Eq. (35) with the prior $\pi(\lambda) = 1/\lambda$ yields the joint distribution of the parameter and the data

$$g(r,T,\lambda) = \lambda^{r-1} e^{-\lambda T} \quad . \tag{36}$$

Integrating over all λ, one obtains the distribution of the data r and T

$$\pi(r,T) = (r-1)!\, T^{-r} \quad . \tag{37}$$

Dividing Eq. (36) by Eq. (37) yields finally the Bayesian posterior

$$\pi(\lambda|r,T) = \frac{1}{(r-1)!}\, T^r\, \lambda^{r-1}\, e^{-\lambda T} \quad . \tag{38}$$

Equation (38) is a member (having parameters r and T) of the gamma family of distributions and is plotted together with the improper prior $1/\lambda$ in Fig. 2 for the case r = 10, T = 88,827 hours. The likelihood function $p(r,T|\lambda)$ for this problem has already been displayed as Fig. 1.

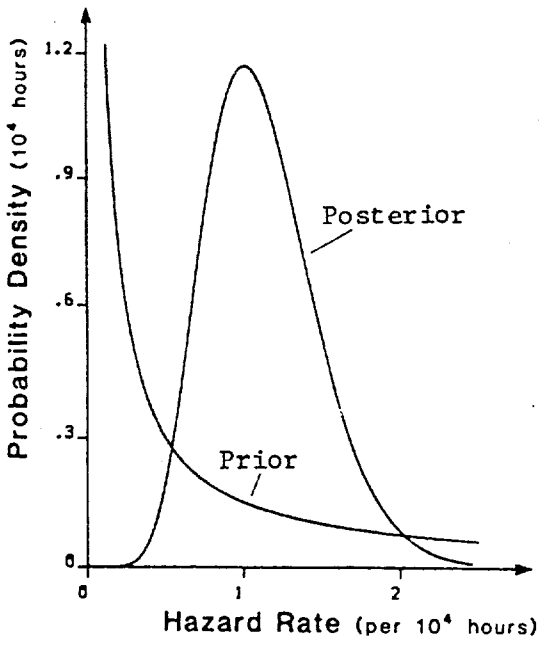


Fig. 2 - Plot of the Bayesian ignorance prior $\pi(\lambda)$ and posterior $\pi(\lambda|x_i)$ based on observing r = 10 exponentially distributed failures in T = 88,827 hours total time on test.

The gamma distribution is the conjugate family for sampling against an exponential time-to-failure density. Thus repetition of the analysis of the preceding paragraph carries the prior $\pi(\lambda) = \Gamma(r',T')$ into the posterior $\pi(\lambda|r,T) = \Gamma(r'+r,T'+T)$. We chose to treat above the special case r' = T' = 0. Let us consider a final vignette before closing this section. Lindley [20] has stated that an ignorance prior ought to be appropriately diffuse, but that otherwise its detailed shape is not very important provided data are plentiful. This is justification for the mathematically convenient and common practice of employing the uniform distribution to represent prior ignorance on a parameter. The gamma family includes the uniform distribution $[\Gamma(r'=1,T'=0)]$ as a special case. This choice of prior leads to the posterior distribution $\pi(\lambda|r,T) = \Gamma(r+1,T)$. Comparing with Eq. (38) the implication is that uniform $\pi(\lambda)$ is more informative than $\pi(\lambda) = 1/\lambda$. This is most easily recognized by comparing the coefficients of variation of $\pi(\lambda|r,T)$ for the two cases. The mean, variance, and coefficient of variation of Eq. (38) are

$$E(\lambda) = r/T \quad , \tag{39a}$$

$$\mathrm{Var}(\lambda) = E(\lambda^2) - [E(\lambda)]^2 = r/T^2 \quad , \tag{39b}$$

and

$$\mathrm{COV}(\lambda) = [\mathrm{Var}(\lambda)]^{1/2}/E(\lambda) = 1/\sqrt{r} \quad . \tag{39c}$$

From Eq. (39c) as r increases the relative sharpness of the posterior distribution also increases.

4.3 Gamma Time-to-Failure Distribution

As the second application example we consider a set of failure times taken to be identically gamma distributed as

$$f(t) = \Gamma(t|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\, t^{\alpha-1}\, e^{-\beta t} \quad . \tag{40}$$

This is a slightly more general form of the gamma distribution than Eq. (38) where the parameter r was integer (here α is not so restricted). Properties of the gamma distribution are discussed in Chapter 4 of Ref. 2. Equation (40) has been discussed as a statistical failure model by Gupta and Groll [21] and found to represent the fatigue life of materials under repetitive loading by Birnbaum and Saunders [22]. The gamma distribution includes the exponential distribution as a special case ($\alpha = 1$). In addition, also much like the celebrated Weibull model [23], the gamma distribution has sufficient flexibility to characterize infant mortality (via $0 < \alpha < 1$) and wearout ($\alpha > 1$). Wearout failures tend to be clustered in time and are often taken to be normally distributed. (Bayesian inference for the independent normal process is discussed in detail in Chapter 11 of Ref. 13.) However, this author feels the gamma distribution is more representative for two reasons -- it can be skewed to the right matching a variety of wearout and fatigue data and its natural range ($0 \leq t \leq \infty$) corresponds exactly to the range of lifetime data. For our purposes the gamma statistical failure model is also more appealing than the Weibull. This is because sufficient statistics exist for the former but not the latter. I have not looked into whether conjugate Bayesian parameter distributions exist for the gamma model (they do for the Weibull). However, sufficient statistics contribute more strongly to the computational simplification of a Bayesian inference problem than does the availability of a conjugate description. We shall proceed with this application example using numerical methods.

For illustrative purposes take the experimental outcome to be the complete (uncensored) set of n = 20 mockup failure data points generated by Monte Carlo simulation and presented as Table 2. As before, the steps involved in obtaining the Bayes posterior are given in Section 4.1. The gamma statistical model has been selected and the time-to-failure probability density function displayed as Eq. (40). To show how previous experience can be built into the Bayesian description, imagine that the prior distributions on the shape parameter α and the scale parameter β have been obtained using regression analysis methods to analyze the outcome of an earlier life test on the same kind of hardware.

Table 2. Uncensored synthesized time-to-failure data representing sampling from a gamma distributed parent population.

TIMES TO FAILURE (Hours)		ANCILLARY INFORMATION
1337	2043	n = 20 Samples Placed on Test
1650	2155	
1657	2190	r = 20 Failures Observed
1738	2323	
1754	2340	Sufficient Statistics:
1798	2376	
1943	2459	r = n = 20
1999	2513	
2010	2561	T_σ = 41,842 hours
2031	2965	
		$T_\pi = 1.857 \times 10^{66}$ hours^{20}

(Regression analysis as it relates to reliability problems, and particularly the use of median ranks, is discussed in Ref. 8.) Standard regression analysis produces as an output that the statistical model parameters are normally distributed with specified mean and standard deviation. Thus, if we identify as the standard form of the normal distribution

$$N_\lambda(\mu_\lambda,\sigma_\lambda) = \frac{1}{\sqrt{2\pi}\,\sigma_\lambda} \exp\left[-\frac{1}{2}\left(\frac{\lambda-\mu_\lambda}{\sigma_\lambda}\right)^2\right] \quad , \tag{41}$$

the prior densities on the model parameters α and β can be written

$$\pi(\alpha) = N_\alpha(\mu_\alpha,\sigma_\alpha) \tag{42a}$$

and

$$\pi(\beta) = N_\beta(\mu_\beta,\sigma_\beta) \quad , \tag{42b}$$

where from previous observation and classical inference

$$\mu_\alpha = 25.2 \quad , \tag{43a}$$

$$\sigma_\alpha = 2.9 \quad , \tag{43b}$$

$$\mu_\beta = 0.0129 \ \text{hours}^{-1} \quad , \tag{43c}$$

and

$$\sigma_\beta = 0.0015 \ \text{hours}^{-1} \quad . \tag{43d}$$

From Eq. (40) and the discussion of Section 2.2 the likelihood function or joint probability of the data given specified model parameters is

$$p(x_i|\alpha,\beta) \propto \prod_{i=1}^{n} f(t_i) = \beta^{n\alpha}\left[\Gamma(\alpha)\right]^{-n} T_\pi^{\alpha-1}\, e^{-\beta T_\sigma} \quad . \tag{44}$$

Equation (44) is expressed in terms of the sufficient statistics

$$n = 20 \quad , \tag{45a}$$

$$T_\sigma = \sum_{i=1}^{n} t_i = 41{,}842 \quad \text{hours} \quad , \tag{45b}$$

and

$$T_\pi = \prod_{i=1}^{n} t_i = 1.857 \times 10^{66} \quad (\text{hours})^{20} \quad , \tag{45c}$$

and the notation (x_i) of Section 4.1 to indicate their simultaneous specification. Since Eqs. (42) are independent, the joint probability of the data and the model parameters is

$$g(x_i,\alpha,\beta) = \pi(\alpha)\pi(\beta)p(x_i|\alpha,\beta) \quad . \tag{46}$$

Using Eqs. (42) through (44) and numerically integrating Eq. (46) over all parameter space for the particular sampling outcomes x_i displayed in Table 2 and summarized by Eqs. (45) yields

$$\pi(x_i) = \int_\beta \int_\alpha g(x_i,\alpha,\beta)\, d\alpha\, d\beta = 1.476 \times 10^{10} \quad . \tag{47}$$

Equation (47) represents the a priori probability of realizing the experimental outcome actually subsequently observed and is the proper normalization or weighting factor required for the Bayesian posterior to have a true probability interpretation. The posterior itself $\pi(\alpha,\beta|x_i)$ is obtained by dividing Eq. (46) by Eq. (47) and for the two-parameter gamma model is still a joint distribution function. To obtain a marginal posterior distribution on each parameter separately requires integration over the full range of the other parameter. Thus finally

$$\pi(\alpha|x_i) = \int_0^\infty \pi(\alpha,\beta|x_i)\, d\beta \tag{48a}$$

and

$$\pi(\beta|x_i) = \int_0^\infty \pi(\alpha,\beta|x_i)\, d\alpha \quad . \tag{48b}$$

These integrations have been performed numerically. The results, together with the corresponding priors [Eqs. (42)] are shown in Figs. 3 and 4. For comparison with Eqs. (43) the means and variances of Eqs. (48) are

$$E(\alpha|x_i) = 26.3 \quad , \tag{49a}$$

$$\mathrm{Var}(\alpha|x_i) = 4.45 \quad , \tag{49b}$$

$$E(\beta|x_i) = 1.26 \times 10^{-2} \text{ hours}^{-1} , \tag{49c}$$

and

$$Var(\beta|x_i) = 1.06 \times 10^{-6} \text{ hours}^{-2} . \tag{49d}$$

In comparing Eqs. (43) and (49) recall that the normal standard deviation is the square root of the variance.

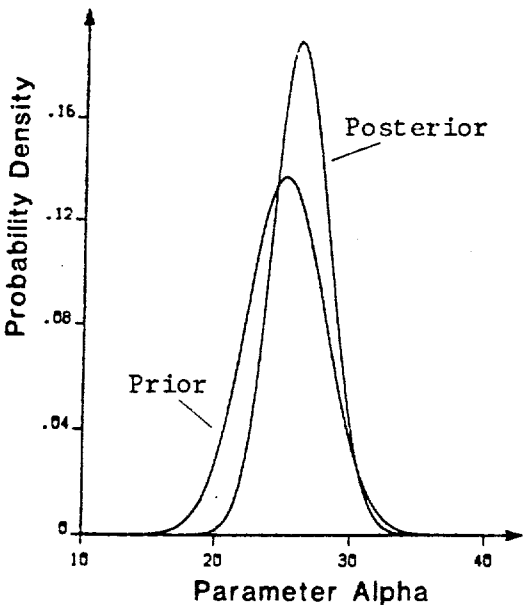


Fig. 3 - Comparison of the Bayesian prior (Eq. (42a)] and posterior (Eq. (48a)] of the gamma failure model parameter α (using the data of Table 2.)

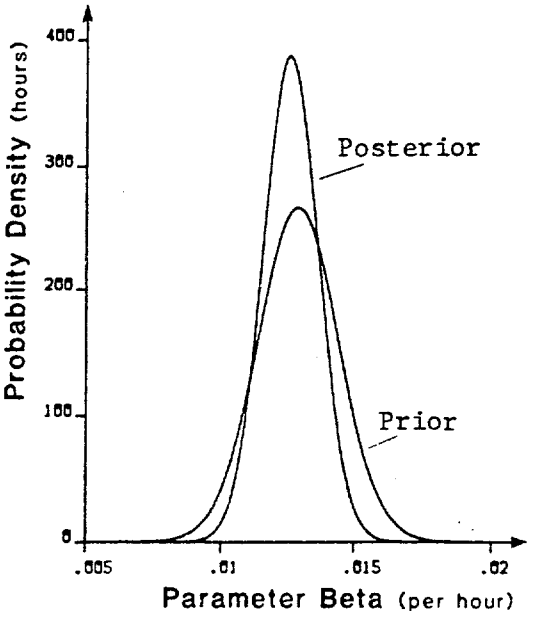


Fig. 4 - Comparison of the Bayesian prior (Eq. (42b)] and posterior (Eq. (48b)] of the gamma failure model parameter β (using the data of Table 2.)

This second application example is a case where very little use has been made of the pedagogic conveniences usually invoked in presentations of Bayesian theory. It is hoped that this will help convince the reader that Bayesian inference can be shaped to address his awkward real-life problems rather than being limited in scope. As Bayes methods gain further acceptance no doubt the necessary computer codes will become the readily available commodity that more conventional statistics packages already are.

4.4 Bernoulli Process -- Reliability Measurement

In Section 4.2 we interpreted time-to-failure data to obtain a Bayesian posterior distribution on the exponential model hazard rate λ [Eq. (38)]. This result may be combined with Eq. (6b) via random variable transformation methods to obtain equivalently that reliability itself is distributed as

$$h(R) = \frac{1}{\Gamma(r)} \left(\frac{T}{t}\right)^r \left(-\ln R\right)^{r-1} R^{(T/t - 1)} , \tag{50}$$

a result that is referred to as the negative-log gamma distribution. In practice it may be that a population of equipments is available for inspection only on a limited opportunity basis. In such a case complete time-to-failure data are not available. However, point observations of reliability can still be made by interpreting the specification of the operational health of the hardware as a set of Bernoulli trials (binomial sampling). That is, at some time t' a total of n equipments are examined with the result that r of them are seen to be failed and n-r unfailed. For an individual unit the probability of successful operation is the reliability R [Eq. (3)] and the probability of failure is 1-R. When appropriate statistical weight is given to the number of ways a particular outcome can be realized (combinations of r from n), one obtains the binomial distribution

$$p(r|n,R) = \frac{n!}{r!\,(n-r)!}\; R^{(n-r)}\; (1-R)^{r} \quad , \tag{51}$$

as the appropriate statistical model for this problem. The number of failures r is a discrete random variable ranging from 0 to n and R [more specifically R(t') here] is a parameter of the model. We will characterize R via Bayesian inference; n is a fixed model parameter of no particular further concern. Equation (51) is already a joint distribution embracing the information that r failure events have occurred. The kernel or R dependence of the likelihood is thus

$$p(r|n,R) \propto R^{(n-r)}\; (1-R)^{r} \quad . \tag{52}$$

Use of Eq. (52) with its conjugate family (beta distribution) is discussed in standard sources [1, 13]. We prefer here to deal with the ignorance prior

$$\pi(R) = \left[-R \ln R\right]^{-1} \quad . \tag{53}$$

Equation (53) is the transformed analog of the λ^{-1} prior of Section 4.2 and may also be seen to result from a noninformative experiment [Eq. (50) specialized to the case r = T = 0]. Combining Eqs. (52) and (53) and using the fact that p(n) = 1, the joint distribution is

$$g(n,r,R) = R^{n-r-1}\; (1-R)^{r}\; \left[-\ln R\right]^{-1} \quad . \tag{54}$$

As before, integration over the entire admissible parameter range (0 to 1 on R) yields the probability of the particular experimental outcome. For illustrative purposes let us again specialize to the case represented by the data of Table 1. Thus taking n = 20 and r = 10 and numerically integrating Eq. (54), we obtain

$$\pi(20,10) = \int_0^1 g(20,10,R)\, dR = 7.776 \times 10^{-7} \quad . \tag{55}$$

And finally combining Eqs. (54) and (55) the Bayesian posterior for this case is

$$\pi(R|n=20,r=10) = (1.286 \times 10^{6})\, R^{9}\; (1-R)^{10}\; \left[-\ln R\right]^{-1} \quad . \tag{56}$$

Equation (56) is plotted as the solid curve in Fig. 5. For comparison purposes we can look at the equivalent result [Eq. (50)] obtained by gamma sampling as described in Section 4.2. That is, the data of Table 1 can be interpreted as a binomial sample (at the time t_r of the r^{th} failure) from a Bernoulli process as we have done in this section. Or the same failure data can be viewed as having been obtained by fixing r in advance and allowing T to be the experimental random variable (gamma sampling from a Poisson process) per Section 4.2. Equation (50) is the description of the latter case expressed in reliability terms rather than as a statement about the hazard rate λ. To make the desired comparison with Eq. (56), Eq. (50) must be specialized to the time $t = t_r$ of binomial sampling. In addition, taking r = 10 and from the Table 1 data t_r = 6017 hours and T = 88,827 hours, Eq. (50) becomes

$$h(R) = (1.355 \times 10^6)(-\ln R)^9 R^{13.763} \quad . \tag{57}$$

Equation (57) is shown as the dashed curve plotted in Fig. 5. As is apparent the two posterior distributions of $R(t_r)$ are quite similar.

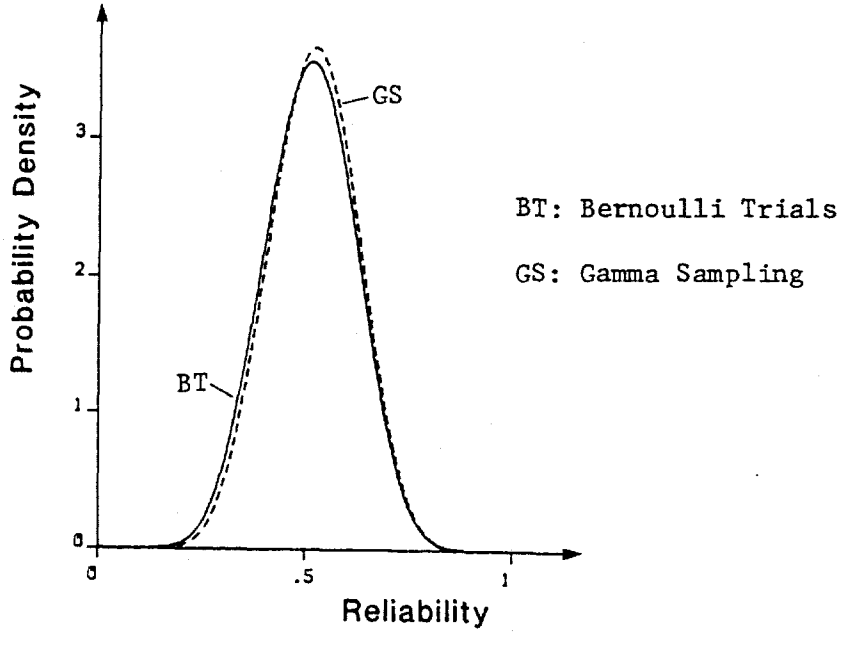


Fig. 5 - Plot of the Bayesian posteriors [Eqs. (56) and (57)] on reliability based on interpreting the Table 1 data as a sequence of Bernoulli trials at time t_r or as a gamma sampling outcome.

5.0 CONCLUSIONS

Bayes theory is very appealing for use in treating reliability problems. This is true in part because reliability issues are quite typically expressed in terms of statistical failure models -- the natural Bayes point of departure. The ability of Bayesian methods to make constructive use of previous experience is also of great benefit for statistical inference situations generally. This is especially true in the reliability and life-testing areas because so often new, improved products are introduced to replace similar equipments for which attributes data are already available. Another major advantage of Bayes methods is that they make model parameter space directly accessible via the prior and posterior distributions. Confidence statements are developed via direct integration of these functions. The whole classical preoccupation with the development of the statistical properties of estimators is avoided entirely.

This report, while necessarily limited in scope, has been structured to touch on the philosophical basis of Bayes theory, to compare classical inference, to develop the operational structure of the method, and to address relevant applications. Even though the focus of this has been the narrow one of completing the specification of statistical or mathematical model parameters using available data, to do justice to the task requires a more heroic effort than this document represents. The reader new to Bayesian inference may find that this report best serves as a study outline helping to place the field and some of its possibilities in perspective.

Many frequentists reject the notion that probability is subjective and object to the admissibility of unnormalized probability density functions as Bayesian priors. This author hopes to address the latter point elsewhere [24]. As to the former -- decide for yourself (some discussion appears in Appendix B). A Bayes solution is often referred to as "learning from experience". Thus one modifies his previous understanding of a situation by assimilating new information to obtain a revised impression. This, of course, goes on every day without mathematical formalization. By basic human nature we are all Bayesians.

Hopefully the discussion of application examples in Section 4 has helped display the versatility of Bayes methods. Priors need not be conjugate, numerical integration can be used as needed, even a digitized graph or sketch is a perfectly acceptable format for introducing prior information.

6.0 ACKNOWLEDGMENT

This work was supported by the Underwater Sound Reference Detachment of the Naval Research Laboratory under Contract N00173-81-W-T201.

REFERENCES

1. N. R. Mann, R. E. Schafer, and N. D. Singpurwalla, *Methods for Statistical Analysis of Reliability and Life Data*, Wiley, New York, 1974.

2. M. H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.

3. B. Epstein, "The Exponential Distribution and Its Role in Life Testing," Industrial Quality Control, 15, 2 (1953).

4. R. E. Barlow and F. Proschan, *Mathematical Theory of Reliability*, Wiley, New York, 1965.

5. A. W. F. Edwards, *Likelihood*, Cambridge University Press, England, 1972.

6. R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society A*, 222, 309 (1922). Reprinted in *Contributions to Mathematical Statistics*, Wiley, New York, 1950.

7. B. Epstein and M. Sobel, "Life Testing," *Journal of the American Statistical Association*, 48, 486 (1953).

8. R. L. Smith, "Reliability and Service Life Concepts for Sonar Transducer Applications," NRL Memorandum Report 4558, Sep 1981.

9. D. V. Lindley, *Making Decisions*, Wiley, London, 1971.

10. W. J. MacFarland, "Bayes Equation, Reliability, and Multiple Hypothesis Testing," *IEEE Transactions on Reliability*, R-21, 136 (1972).

11. T. Bayes, "An Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society of London*, 53, 370 (1763). A modernized version appears in *Biometrika* 45, 293 (1958).

12. B. de Finetti, *Theory of Probability*, Vol. 2 (English translation), Wiley, London, 1975.

13. H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, Harvard University Press, Cambridge, Mass., 1961.

14. S. S. Gupta and D. S. Moore, eds., *Statistical Decision Theory and Related Topics II*, Academic Press, New York, 1977.

15. D. J. Bartholomew, "A Comparison of Some Bayesian and Frequentist Inferences," *Biometrika*, 52, 19 (1965).

16. H. Jeffreys, *Theory of Probability*, Oxford University Press, London, 1961.

17. M. O. Locks, *Reliability, Maintainability, and Availability Assessment*, Hayden, Rochelle Park, New Jersey, 1973.

18. MIL-HDBK-217C; Military Standardization Handbook -- Reliability Prediction of Electronic Equipment, Department of Defense, April 1979.

19. A. H. El Mawaziny and R. J. Buehler, "Confidence Limits for the Reliability of Series Systems," *Journal of the American Statistical Association*, 62, 1452 (1967).

20. D. V. Lindley, *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part II, Inference*, Cambridge University Press, Cambridge, England, 1965.

21. S. S. Gupta and P. A. Groll, "Gamma Distribution in Acceptance Sampling Based on Life Tests," *Journal of the American Statistical Association*, 56, 942 (1961).

22. Z. W. Birnbaum and S. C. Saunders, "A Statistical Model for Life-Length of Materials," *Journal of the American Statistical Association*, 53, 151 (1958).

23. W. Weibull, "A Statistical Distribution Function of Wide Applicability," *Journal of Applied Mechanics*, 18, 293 (1951).

24. R. L. Smith (publication pending).

APPENDIX A

A Probability Closure Theorem

We should like to show that Eq. (27) of the body of the text is not an independent assertion but is implied by Eqs. (23) and (24). Consider an exclusive and exhaustive set of uncertain events A, B, ... and some other uncertain event E. Then the probability of E is

$$p(E) = p(E) \times 1 \quad . \tag{A1}$$

But since A, B, ... are exclusive and exhaustive, Eq. (23) becomes

$$p(A \text{ or } B \text{ or } \ldots) = p(A) + p(B) + \ldots = 1 \quad . \tag{A2}$$

Equation (A2) is correct whether event E also occurs or not. In particular, if E occurs,

$$p(A|E) + p(B|E) + \ldots = 1 \quad . \tag{A3}$$

Substituting Eq. (A3) in Eq. (A1) yields

$$p(E) = p(E)p(A|E) + p(E)p(B|E) + \ldots \quad , \tag{A4}$$

which, from the interchange symmetry of Eq. (24) [i.e., $p(E \text{ and } A) = p(A \text{ and } E)$] becomes

$$p(E) = p(A)p(E|A) + p(B)p(E|B) + \ldots \quad , \tag{A5}$$

which is the desired result.

APPENDIX B

Subjective Versus Objective Probability

At the heart of the differences between frequentists and Bayesians is the interpretation of probability. Classical statisticians view probability as a substantive attribute of an object under study, a state function, objective in the sense of measurable. Perhaps we should speak of the system under study rather than simply the object. For example, in coin tossing, the probability of obtaining heads does not depend on the design of the coin alone, but also on establishing some statistically reproducible flipping procedure. Similarly, the times to failure in a population life test depend on the conditions of use as well as the design of the hardware. The occurrence of heads or the particular life-test failure times also depends in detail on factors that remain unknown to us (flaws or asymmetries for example). Thus, the best that can be managed by way of probability measurement is to replicate the observation, or experiment, or chance setup (as it is sometimes referred to) hoping to symmetrize in the long run the impact of the unknown factors. This is an effort to eliminate systematic bias by homogenizing the representation of the phase space (to borrow a term from statistical mechanics) of these quantities. The long run frequency of occurrence of an event obtained in this way is taken to be the measure of its probability for a single trial. This may seem to be entirely objective and not dependent on who conducts the test. But there are also subjective inputs or judgments to be made. For example, in carrying out 10,000 coin tosses to get a pretty good idea of the long run frequency of turning up heads, are different coins interchangeable? If a single coin is used, might it sustain damage that would progressively alter the property one is trying to measure? Or in the life testing example, if several equipments are tested, are they really alike or is the survival property itself distributed within the population? Another problem in practice is that in most situations of interest one lacks the wherewithal to carry out an experiment heroic enough to yield a statistically well-defined long run frequency.

In the preceding paragraph we have suggested that efforts to objectively measure probability may not actually be successful. One can question whether an entity can have an objective existence if it is unmeasurable. This doesn't trouble Bayesians, for whom probabilities are subjective. Now let's turn the argument around. Suppose we are dealing with a situation that cannot be replicated and therefore is not describable in terms of a long run frequency. Does it make sense to introduce probability into its description? To be specific, suppose we try to assign a probability to the event that a designated individual will receive the Nobel prize in physics next year. Or, we might like to weigh the relative chances of half a dozen potential candidates. Selecting such a list to begin with would elicit very different responses from people with different backgrounds. A non-physicist may be hard put to name individuals with much hope at all of receiving the award. On the other hand, an experienced leader in the physics community, particularly someone close to one of the successful, aggressively pursued subfields of the day, could probably generate a very respectable candidate list. Still, dozens of similar groups of worthy individuals might be identified by others. Individuals named on one of these lists probably have much better chances for the prize than members of the population at large. The outlook for persons named on many lists might be brighter than that of individuals not so recognized. There are even repetitive

aspects of the situation to assuage the classical statistician. Thus, one knows historically how often the prize has gone to a woman; that spectroscopists, solid state, and high energy physicists are more favored than acousticians; and that one's great work is more frequently but not always done early in life. Nevertheless, placing betting odds on Nobel candidates is largely a process of processing information subjectively. A classicist might claim that this is pointless; the Bayesian will argue that progress can be made in no other way. The reader is invited to ponder the issue, check the literature, and sharpen his own interpretation. Is probability objective or subjective?