NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Memorandum 33-614

Evaluation of Errors in Prior Mean and Variance in the Estimation of Integrated Circuit Failure Rates Using Bayesian Methods

B. C. Fletcher

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA
June 1, 1972

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Memorandum 33-614

Evaluation of Errors in Prior Mean and Variance in the Estimation of Integrated Circuit Failure Rates Using Bayesian Methods

B. C. Fletcher

JET PROPULSION LABORATORY

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA

June 1, 1972

PREFACE

The work described in this report was performed by the Quality Assurance and Reliability Division of the Jet Propulsion Laboratory.

CONTENTS

ABSTRACT

Under the constraint of limited testing time, many attempts
have been made to incorporate prior knowledge and exper-
ience into a quantitative assessment of reliability. This
type of technique is known as Bayesian statistics. Since
the length of time available for testing integrated circuits
is frequently very limited, an analysis of Bayesian methods
when applied to the integrated circuit testing problem was
conducted.

The critical point of any Bayesian analysis concerns the
choice and quantification of the prior information. This
report is a study of the effects of prior data on a Bayesian
analysis. Comparisons of the Maximum Likelihood estimator,
the Bayesian estimator and the known failure rate are pre-
sented. The results of the many simulated trials are then
analyzed to show the region of criticality for prior information
being supplied to the Bayesian estimator. In particular,
effects of prior mean and variance are determined as a
function of the amount of test data available.

1. Introduction

The lack of integrated circuit reliability data is a chronic problem for those concerned with the problem of reliability estimation. The data limitations are often manifested in terms not only of quantity but also quality. For example, good quality data would not only include all the conditions under which the data was obtained but would also include the assumptions and restrictions to be used with such data. This needs to be considered since quite often data from various sources are not compatible with the application the analyst is trying to treat. Data restrictions such as these can severely limit the amount of useful data to be applied to reliability estimation.

Faced with situations in which data is not representative or is inadequate, it is not uncommon for the reliability analyst to devise a rationale which may, in part, be based on "engineering judgment". Unfortunately, this rationale often tends to be arbitrary and leads to inconsistency with respect to device types and applications. Therefore, a defined rationale was sought which would allow the incorporation of both judgment and actual data in a consistent manner. Specifically, judgment must be quantified and combined with data.

There are several rationales for estimating the probability of success for a given type of integrated device. The usual basis for derivation of such models is the failure-rate-estimate. For these reasons a rationale is sought which can estimate integrated circuit failure rates and still be compatible with limited data.

The basic approach taken was to develop a technique which would assess integrated circuit parameters, such as failure-rate and life time.

This technique of assessing integrated circuit parameters was achieved by combining limited statistical data and engineering judgment. A rationale was sought which would combine both data and judgment so that an analytical estimation of the failure rate for integrated circuit devices could be achieved.

This report emphasizes the selection of appropriate statistical distributions for incorporation with Bayesian statistical theory. The resulting Bayesian application with various data inputs was then compared with classical estimators of failure rate. +

Since Bayesian Statistics provides a convenient way of incorporating prior knowledge or judgment regarding probabilistic events, the treatment of the previously stated reliability estimation problem was approached through an application of Bayesian statistics. Bayes' Theorem, or Bayes' Rule, can be expressed by the formulation given in Equation 1.

$$P(C_i|B) = \frac{P(B|C_i)\,P(C_i)}{\sum_{j=1}^{m}[P(B|C_j)\,P(C_j)]} \tag{1}$$

The C_1, C_2, C_3, . . . , C_n in Equation 1 represent n mutually exclusive events. The respective probabilities of these events are denoted by $P(C_i)$ $i = 1, 2, . . ., n$.

The event B is an event for which one knows the conditional probabilities $P(B|C_i)$. This is read: the probability of event B given event C_i. Once these probabilities have been defined, the probability of obtaining a particular event C_i given event B is given by Equation 1. In the context of this reliability estimation problem, Equation 1 must be redefined for the continuous case with regard to a specific notation.

+defined in reference (9), pages 102-4

$$f(\lambda \mid F) = \frac{f(F \mid \lambda) \, f(\lambda)}{\int_\lambda f(F \mid \lambda') \, f(\lambda') \, d\lambda'} \tag{2}$$

Where:

λ =   the values assumed by the propulation failure rate.

F =   the event called failure.

$f(\lambda)$ =   a probability density function for λ.

$f(F \mid \lambda)$ =   a function giving the probability of failure given a particular λ.

$f(\lambda \mid F)$ =   a probability density function of λ given the failure event F.

The controversy about Bayes' Theorem concerns the uncertainty in knowing the probabilities $P(C_i)$ i = 1,2, . . ., n and $f(\lambda)$ in the respective equations 1 and 2. This problem results from a lack of appropriate data. Thus, engineering judgment must be applied in the initial phases of the applied Bayesian analysis. The effects of this judgment on $P(C_i)$ and $f(\lambda)$ are lessened by an allowance for incorporation of the appropriate data. In this case Bayes' Theorem is applied in an iterative fashion which allows the incorporation of data as it becomes available. A better probability estimate is obtained as increasing amounts of data are incorporated.

The initial data will be referred to as prior information. It will be defined by a prior formulation or function. Subsequent data will result in a posterior formulation. In the case of Equation 2 an iterative technique describing this situation is shown below.

Prior information is quantified by $f_o(\lambda \mid \overline{F}_o, t_o)$ given the initial information F_o and time t_o; Substitution of this notation into Equation 2 yields:

$$f_1(\lambda \mid \overline{F}_1) = \frac{f(\overline{F}_1 \mid \lambda) \, f_o(\lambda)}{\int_\lambda f(\overline{F}_1 \mid \lambda') \, f_o(\lambda') \, d\lambda'} \tag{3}$$

Where $f(\lambda)$ is the more convenient notation for $f_o(\lambda \mid \overline{F}_o, t_o)$

$f_o(\lambda \mid \overline{F}_1)$ is the posterior function which replaces

$f_o(\lambda)$ and reflects the failure event new F_1 (t_1 understood)

With more data, $f_1(\lambda \mid \overline{F}_1)$ may be used to derive another function which incorporates this data. Again Equation 2 is used.

Therefore: $$f_2(\lambda \mid \overline{F}_2) = \frac{f(\overline{F}_2 \mid \lambda) \, f_1(\lambda)}{\int_\lambda f(\overline{F}_2 \mid \lambda') \, f_1(\lambda') \, d\lambda'} \tag{4}$$

$$f_1(\lambda) = f_1(\lambda \mid \overline{F}_1, t_1)$$

Depending upon the quantity of data, this type of iteration can be carried out to an arbitrary number of steps. The nth iteration will take the following form.

$$f_m(\lambda \mid \overline{F}_m) = \frac{f(\overline{F}_m \mid \lambda) \, f_{m-1}(\lambda)}{\int_\lambda f(\overline{F}_m \mid \lambda') \, f_{m-1}(\lambda') \, d\lambda'} \tag{5}$$

Thus, an incorporation of new data is used to redefine the functional description of values associated with a population of particular device types.

An essential constituent of this Bayesian application is the determination of the functional relations which can be used to satisfy the iteration. This determination must start with formulations of the prior information. Toward this end there will be an effort below to identify the usefulness of given distributions for best describing this information; to determine the criticality of distribution shape parameters; and to find the influence of varying amounts of data.

The final factor in the approach of this Bayesian application concerns the analysis at iteration termination. For this paper, iteration termination will be defined as the point where data will cease to be incorporated into the Bayesian formulations. This termination point will be determined through a comparison of Bayesian failure-rate-estimates with classical failure-rate-estimates for equivalent data.

2. DISTRIBUTION FUNCTION STUDY

Before the details of the Bayesian investigation can be pursued, the functional relations presented in Equation 2 must be defined.

The probability of failure given a particular failure-rate was denoted by $f(F|\lambda)$. It will be assumed that the exponential distribution for times-to-failure holds. Because of the relation between the exponential and Poisson distributions, $f(F|\lambda)$ can be expressed in the following Poisson form:

$$f(F|\lambda) = \frac{e^{-\lambda t}(\lambda t)^F}{F!} \tag{6}$$

where F = the number of failures.

The interpretation of this formula is that, for failures occurring at a given rate λ, the probability of having exactly F failures in time t is given by $f(F|\lambda)$ (i.e., Equation 6).

The lambda (λ) given in equation (2) can be thought of as the constant rate of failure for an exponential distribution of times-to-failure. Furthermore, the density function $f(\lambda)$ represents the prior estimate of the failure rate for a given integrated circuit device type. This density function can be thought of as describing the prior probability that λ is one of the values between zero and infinity. This probability is a theoretical description of the distribution for the entire population (not sample) of integrated circuits of a particular type.

The selection of the formulation for $f(\lambda)$ in Equation 2 will have to be subject to some fundamental selection criteria. This is because all

distributions selected for $f(\lambda)$ may not be amenable to the iteration process already discussed. As was stated before, the iteration process lessens the initial uncertainty by building the true shape for the function $f(\lambda)$ as data becomes available.

Because of this uncertainty, the function selected must be viable — this means that incorporation of data through iteration is not only perceptible but meaningful. This criterion can be thought of as a sensitivity of shape variability due to the iteration process. The second criterion sought regards the result of Equation 2, i.e., $f(\lambda|F)$. For convenience in the iteration process, the result of Equation 2 must be amenable to reuse as $f(\lambda)$ in the next successive stage of computation. A closed function form is sought for ease of computation.

The first functions considered were of standard form and can be found in basic statistical texts (Ref. 5). The formulations and conditions for these functions are stated in the following section.

3. DISTRIBUTION SELECTION

If one considers the criteria previously discussed and the functional conditions, a likely candidate for this Bayesian application can be selected. Table I provides some remarks necessary for this determination.

TABLE I

NAME	
Uniform	Function does not yield closed form when iterated - computation difficult. Not defined on the entire $(0, \infty)$ domain
Exponential	Iteration yields a gamma form
Gamma	Iteration yields a gamma form - see Section 4.
Normal	Computationally difficult when iterating
Beta	Only defined for the domain $(0,1)$
Cauchy	Moments do not exist therefore no expected value can be calculated.
F-Distribution	Only defined for the domain $(0,1)$
Chi-Square	Special case of Gamma
Inverted Gamma	Does not describe the correct random variable describes the estimate given the true.
Converse Inverted Gamma	Parameter definitions difficult to make resulting in computational difficulty.

PROBABILITY DENSITY FUNCTIONS

Equation	Name	Formulations	Conditions
(7)	Uniform	$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} \\ 0 \end{cases}$	$\alpha < x < \beta$ Elsewhere
(8)	Exponential	$f(x) = \begin{cases} 1/\theta \; e^{-(x/\beta)} \\ 0 \end{cases}$	$x > 0 \quad \theta > 0$ Elsewhere
(9)	Gamma	$f(x) = \begin{cases} K x^{\alpha - 1} \; e^{-(x/\beta)} \\ 0 \end{cases}$	$x > 0 \quad \alpha > 0, \; \beta > 0$ $x \leq 0$ K such that $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1$ $K = \dfrac{1}{\beta^{\alpha} \, \Gamma(\alpha)}$
(10)	Normal	$f(x) = K \; e^{-1/2 \left(\frac{x - \alpha}{\beta} \right)^2}$	$-\infty < x < +\infty, \; \beta > 0$
(11)	Beta	$f(x) = \begin{cases} K \, x^{\alpha - 1} \, (1-x)^{\beta - 1} \\ 0 \end{cases}$	$0 < x < 1$ Elsewhere
(12)	Cauchy	$f(x) = 1/\pi \; \dfrac{a}{a^2 + x^2}$	$-\infty < x < +\infty$
(13)	F-Distribution	$f(x) = \begin{cases} K \, x^{(\sqrt{\gamma_1}/2) - 1} \, (1-x)^{(\sqrt{\gamma_2}/2) - 1} \\ 0 \end{cases}$	$0 < x < 1$ Elsewhere
(14)	Chi-Square	$f(x) = \begin{cases} \dfrac{1}{2^{\gamma/2} \, \Gamma(\gamma/2)} \; x^{(\gamma - 2)/2} \; e^{-(x/2)} \\ 0 \end{cases}$	$x > 0$

There are two further functions which can be derived for this Bayesian application. It was derived from basic statistical and reliability theory. The results of this derivation are stated below.

Function	Name	Formulation	Conditions
(15)	Inverted Gamma	$f(x) = \begin{cases} \dfrac{(\mu\lambda)^{\lambda}}{\Gamma(\lambda)}\left(\dfrac{1}{x}\right)^{\lambda+1} e^{-\left(\mu\lambda/x\right)} & x > o,\ \mu\lambda > o,\ \lambda > o \\ \\ 0 & \text{Elsewhere} \end{cases}$	

	Name	Formulation	Conditions
(16)	Converse Inverted Gamma	$f(\mu\mid x) = \begin{cases} \dfrac{(\phi + \lambda/x)^{\lambda+\alpha}}{\Gamma(\alpha+\lambda)}\ \mu^{(\lambda+\alpha-1)} e^{-(\phi+\lambda/x)\mu} & \mu > o \\ & \lambda > o \\ & \phi > o \\ & \alpha > o \\ 0 & \text{Elsewhere} \end{cases}$	

Equation 15 is derived via the standard change of variable technique. The basic function treated was a density function for the random variable associated with times-to-failure for a given number of failures (μ) given their rate of occurrence (λ). Equation 15 represents the distribution of the random variable called a failure-rate-estimator.

Equation 16 is a description of the random variable called true-failure-rate given the failure-rate-estimator information.

The derivations of formulations 15 and 16 are contained in Appendix A. It will also be noted that, after simplification, these derivations yield the same result as the formulations presented in Section 4. The formulations of Section 4 are, however, more easily derived and understood.

The gamma compliance to the defined selection criterion can be demonstrated

by the formulations below. These formulations are used in Section 4 to

illustrate the Bayesian iteration technique and parameter determination for

the general gamma formulation. Complete derivations are deferred to Appendix B.

4. PROPOSED BAYESIAN ITERATION TECHNIQUE

Substituting Equations 6 and 9 into Equation 2 yields the following formulation:

$$
f(\lambda \mid F) = \frac{\left(K/F! \right) \, e^{-\lambda t} \, (\lambda t)^{F} \, \lambda^{\alpha-1} \, e^{-(\lambda/\beta)}}{\displaystyle\int_{\lambda=0}^{\infty} (K/F!) \, e^{-\lambda' t} (\lambda' t)^{F} \, \lambda'^{\alpha-1} \, e^{-(\lambda'/\beta)} \, d\lambda'}
\tag{17}
$$

where $K = \dfrac{1}{\beta^{\alpha} \, \Gamma(\alpha)}$

F = the number of failures in time t

The derivation for K is given in Ref. (5).

$\Gamma(\alpha)$ is the gamma function

Using the notation outlined for iteration in Section 1, prior information is factored into the following equation:

$$f_0 \left(\lambda \mid F_0 \right) = K_0 \; \lambda^{\alpha - 1} \; e^{-\lambda / \beta}$$

(18)

$$\text{where } K_0 = \frac{1}{\beta^{\alpha} \Gamma(\alpha)}$$

K_0 is determined by the initial constants α and β.

With the above prior information and iteration algorithm, the following sequence may be derived. The successive steps are presented below, however the derivations appear in Appendix B. It is important to note that F_i ($i = 1, 2, 3, \ldots, n$) is being used as the number of failures occurring in time t_i. Since the times-to-failure (t_i) are being generated one at a time, F_i is always unity (i.e., $F_i = 1$, $i = 1, 2, \ldots, n$).

$$f_1 \left(\lambda \mid F_i \right) = K_1 \; \lambda^{(\alpha + F_1) - 1} \; e^{-\lambda(1/\beta + t_1)}$$

(19)

$$\text{where } K_1 = \left[(1/\beta + t_1)^{\alpha + F_i} \right] \Big/ \Gamma(\alpha + F_1)$$

$$f_2 \left(\lambda \mid F_2 \right) = K_2 \; \lambda^{(\alpha + F_1 + F_2) - 1} \; e^{-\lambda(1/\beta + t_1 + t_2)}$$

(20)

$$\text{where } K_2 = (1/\beta + t_1 + t_2)^{\alpha + F_1 + F_2} \Big/ \Gamma(\alpha + F_1 + F_2)$$

$$\vdots$$

$$f_m \left(\lambda \mid F_m \right) = K_m \; \lambda^{\left(\alpha + \sum_{i=1}^{m} F_i\right) - 1} \; e^{-\lambda\left(1/\beta + \sum_{i=1}^{m} t_i\right)}$$

(21)

$$\text{where } K_m = \left(1/\beta + \sum_{i=1}^{m} t_i\right)^{\alpha + \sum_{i=1}^{m} F_i} \Big/ \Gamma\left(\alpha + \sum_{i=1}^{m} F_i\right)$$

It should be noted that iteration of Equation 18 through 20 produces successive gamma formulations.

4.1 INFORMATION DERIVED FROM BAYESIAN TECHNIQUE

The information derived from the Bayesian technique will come from the probability density yielded by Equation 2. The standard methods of treating the problem of finding an average value for a random variable given its density function is via the theory of mathematical expectation (Ref. 5). The expection of λ is defined by the following formula:

$$E(\lambda) \equiv \int_{\lambda} \lambda' f(\lambda') \, d\lambda' \qquad\qquad 0 \le \lambda \le \infty \qquad\qquad (22)$$

where λ = the random variable representing failure rate for a device population.

$f(\lambda)$ = the probability density function for λ.

Equation 22 represents the mean for a particular density function $f(\lambda)$. A function which is related to (22) is the formulation for the variance and deviation for the density function $f(\lambda)$. These are respectively defined with the following notation.

$$E\left[(\lambda - E(\lambda))^2\right] = \text{VAR}(\lambda) = \int_{\lambda} (\lambda' - E(\lambda'))^2 \, f(\lambda') \, d\lambda' \qquad (23)$$

The standard deviation for the density function $f(\lambda)$ is given by the square root of Equation 23.

As regards this particular application, the mean and variance for $f(\lambda)$ are derived in Appendix C. For convenience, the respective formulations are stated on the following page.

$$E_n(\lambda) = \frac{\alpha + \sum_{i=1}^{n} F_i}{1/\beta + \sum_{i=1}^{n} t_i} \tag{24}$$

$$E_n\left[(\lambda - E(\lambda))^2\right] = \frac{\alpha + \sum_{i=1}^{n} F_i}{\left[1/\beta + \sum_{i=1}^{n} t_i\right]^2} \tag{25}$$

Where n = the number of Bayesian iterations

$$F_i = 1 \quad (\text{i.e.,} \quad \sum F_i = m)$$

The above equations are read as the mean and variance after n iterations

4.2 INCORPORATION OF PRIOR INFORMATION

As was stated previously prior information can be derived either from experience or from "engineering judgment". A problem with both experience and judgment is the quantification of these sources. It should be noted from Equation 18 that the functional relation for $f_o\left(\lambda \mid F_o\right)$ is useable only if the gamma constants α and β are determined.

Consider the case where prior data can be utilized in determining the constants α, β. Suppose that data is available which indicates that the mean and variance of failure-rate-estimators for a particular device are M and V respectively, then the parameters α, β of the given distribution may be computed via a system of simultaneous equations. The following equations are for the mean and variance of the gamma distribution of Equation 18.

$$E(\lambda) = \alpha\beta \tag{26}$$

$$VAR(\lambda) = \alpha\beta^2 \tag{27}$$

When Equations 26 and 27 are set equal to M and V respectively,

$$\alpha = \frac{M^2}{V} \tag{28}$$

$$\beta = \frac{V}{M} \tag{29}$$

If engineering judgment is the only method available for assessing prior information, quantification of α, β is similar to the above. For this case, the values for M and V may be judiciously chosen. Insertion into Equations 28 and 29 will yield the appropriate distribution constants. The consequences of mean and variance choices will be deferred to Section 5.

Once the constants α and β have been chosen, one may proceed with the iteration algorithm already defined. Any type of data can be used for this iteration providing failures and times can be specified. There is only one caution with regard to the data used. Data entered into each successive iteration must be derived from similar device lots and application.

4.3 CONVERGENCE PROPERTIES

Evaluation of this Bayesian technique will be conducted by comparing it with the classical maximum-likelihood-estimator for failure rate in Equation 9. It can be shown that the two estimators, Bayesian (Equation 24) and Classical converge as the amount of applied data increases. This comparison may seem to be a rather trivial aspect; however, it is important. The convergence shows that the prior bias supplied to the Bayesian is not only removed but converges to the true rate with the incorporation of sufficienct data.[+]

There is a further convergence property which is of interest. It can be shown that, as the prior variance of the Bayesian becomes large, the Bayesian and classical estimators converge. This convergence is independent of the amount of data incorporated. To see this consider the following argument.

Prior Mean $= M_p = \alpha\beta$

Prior Variance $= V_p = \alpha\beta^2$ $\qquad\qquad$ (30)

From these equations,

$$\beta = \frac{V_p}{M_p}$$

This implies that $1/\beta$ is inversely proportional to V_p for a given M_p. Also from Equations 30,

$$\alpha = \frac{M_p^{\,2}}{V_p}$$

This implies that α is inversely proportional to V_p for a given M_p.

[+] This formula is true provided that the prior density is **not** zero at the true value. While this is not a problem with the gamma distribution, it warrants special attention if a discrete prior is used (see Appendix D).

Consider the Bayesian expected value (24)

$$E(\lambda) = \frac{\alpha + \Sigma F}{1/\beta + \Sigma t}$$

It has been shown that as V_p increases with fixed M_p, α and $1/\beta$ decrease (the limit for both being zero). Therefore $E(\lambda)$ approaches $\Sigma F / \Sigma t$ in the limit. This ratio is the classical maximum-likelihood-estimator.

5.   SIMULATION TECHNIQUE

Because of the difficulties and unreliability in actual failure data
collection, failure simulation was   used   as.   a workable solution.
The method of this computer simulation was to use a random-number generator
to generate exponentially-distributed times-to-failure. This method can be
thought of as a way of collecting "perfect" data.

Failure-time simulation is accomplished when values of $F(t)$ are supplied
for a given rate.

The simulation generates failures and failure-times based on a failure
rate $\lambda *$. The results, allow computation of a failure-rate-estimator λ ,
which approximates the true rate, $\lambda *$, and can be compared to it. Simulation
also provides an ideal way in which both Bayesian and classical estimations
of failure-rate can be compared while data are incorporated.

This comparison is done for the purpose of checking Bayesian and classical
changes with respect to data and convergence properties. The simulated data
also allows for the performance (with respect to time and failures) of the
Bayesian and classical estimators to be studied given any amount to simulated
data. Performance will be defined as an estimator's nearness to the true rate
($\lambda *$). As an illustration, suppose that a given data base is supplied and that
estimate $\mathcal{E}_1$, and estimate $\mathcal{E}_2$ are calculated. The estimate which is nearest
to λ^* after the incorporation of the same data is said to perform better.

Given a specific prior mean (M_p) and prior variance (V_p), it has been
shown above (Section 4.2) how simulated failures and their corresponding times
are incorporated into the Bayesian formulation (See also Appendix B below).
The respective estimators can now be compared to each other and also to λ^*.
This can be done data point by data point. An example of this can be seen in
Figure 1 which represents the incorporation of 50 simulated data points.

Since this simulated data is generated in a pseudorandom fashion, one simulation run may yield a different comparison chart than another. Variation in the random-number-starts causes this situation even though prior information (M_p, V_p) remain constant. An example of this difference is shown in comparison of Figure 1 to Figure 2.

To average out the effects of these random-number starts on the data to be analyzed, comparison information (as Figures 1 and 2) will be generated a given number of times. In the examples which follow, 500 trials were computed then compared for performance. The results were then normalized to find the percent of time for which the respective estimators performed better.

Given M_p and V_p, suppose that the Bayesian and classical estimators are to be compared with respect to λ^*. This comparison will be done after the incorporation of the 50th data point for each of the 500 trials. Thus, a measure of the averaged performance of the respective estimators can be obtained.

This measure may be defined in the following manner.

Bayesian: Measure$_i$ $= \left| E_{50}(\lambda) - \lambda^* \right|$ i=1,2,3,...500 trials

Classical: Measure$_j$ $= \left| EST_{50}(\lambda) - \lambda^* \right|$ j=1,2,3,...500 trials

where: $EST_{50}(\lambda)$ is the classical failure rate estimator after fifty data points are incorporated.

As each of these 500 measures is computed, they can be compared and counted. A simple counting routine will keep track of the percentage of the 500 trials for which each of the measures is best. Since the number of trials is large, this percentage statement is, essentially, a probability-of-occurrence.

The previous discussion was conducted under the assumption that M_p and V_p remain fixed. If M_p were allowed to vary in such a way as to bracket λ^*, then the portion of time for which measure i is less then j can be plotted vs M_p. A Bayesian performance distribution is thus achieved. This distribution would, in effect, show how Bayesian performance is affected by errors in the selection of the prior mean with respect to λ^*.

In order to have V_p and M_p follow the same relation while M_p is changed, a simple formulation was derived. A relation was chosen between V_p and M_p such that V_p would have a fixed relation to M_p while M_p is varied with respect to λ^*.

$$\text{Suppose} \qquad K = \frac{M_p}{D_p} \qquad (32)$$

Where D_p = prior standard deviation +

          K = arbitrary constant

+ The spread or dispersion of a given distribution (in this case, of gamma gamma form) is specified by the variance of that distribution. The positive square root   ($+\sqrt{V_p}$) is defined as the standard deviation.

The standard deviation was used in this ratio since it is usually closer in magnitude to the mean than the variance and therefore is more easily visualized.

Solving for D_p yields:

$$D_p = \frac{M_p}{K}$$

By definition, the corresponding variance (V_p) is:

$$V_p = \left(\frac{M_p}{K}\right)^2$$

Thus, for a given K, the fixed relation between the prior mean and prior variance is established.

Figures 1 and 2 represented a large simulated data base of fifty points. Since an objective of this report is to treat the situation where only limited data is available, the previous procedure will have to be restricted. To represent this situation, simulated data was generated and the measures i, j were applied at the tenth data point. This was an arbitrary selection since other points could have been chosen.

Simulation was used to determine how the Bayesian performance distribution was affected by various K factors. The true rate (λ^*) was fixed so that the effects on Bayesian performance with respective wide and narrow variances could be observed.

Figures 3-11 demonstrate the effect of a wide variance incremented toward a narrow variance. The constant K of Equation 32 increases with each figure. These figures are computed with $\lambda^* = .05 \times 10^{-6}$ failures/hour. For the purpose of this analysis λ^* was kept at the rate specified. The same effects could be

observed for other λ^* values. However, the range of K factors
may change somewhat in order to produce the same demonstrated effects.

Each figure represents Bayesian performance over a range of prior
mean (M_p) values. Only the K factor was changed in each figure. Each point
depicted in the figures was computed from 500 trials.

Analysis of Figures 3-11 will be deferred to Section 6.2.

6. RESULTS

6.1 Analytical Results

The analytical results which were derived from this study may be summarized as follows:

1. Computationally, incorporation of data is independent of unit size. There is no difference between <u>one</u> computation containing n failures and <u>n</u> computations considering n successive failure times. However, trend indications and evaluations are facilitated by making computations at the smallest unit available.

2. In the limit, as the amount of data becomes large, Bayesian and classical estimators converge. (Section 4.3)

3. For a given amount of data, a great increase in the Bayesian prior variance implies that the Bayesian and classical estimators give the same results. (Section 4.3)

4. The prior parameters α, β of the assumed gamma density can be computed form easily understood statistical mean (M_p) and variance (V_p) statements (Section 4.2)

6.2 Analysis of Observations Generated from Simulated data.

Until now it has been understood that the Bayesian technique is predicated upon the selection of a prior mean and variance (M_p, V_p) for the gamma density. These values are arrived at using some prior belief and/or data. However, the qualitative relation between M_p and V_p must be understood before the Bayesian technique can be discussed further. What does this relation between M_p and V_p mean?

In the simplest terms the value of V_p expresses the confidence one has

in the value selected for M_p. This is because the variance, by definition,

is a measure of the spread or dispersion of the selected prior gamma density.

A narrow variance implies high confidence in M_p (ie., you believe λ^* is at,

or near to, M_p). A wide variance implies that one is less sure about the

position of λ^* with respect to M_p.$^+$

Related to this variance/confidence relation is the impact of V_p upon

successive Bayesian estimators (ie, incorporation of data). It has been

shown in Section 4.3 that as V_p increases, $E(\lambda)$ approaches the classical

maximum-likelihood-estimator. One would expect to see fluctuation in the

initial classical estimator, because of the random inputs. One would

likewise see a fluctuating $E(\lambda)$ for a sufficiently wide V_p. This relation

can be demonstrated by Figure 12. (Note that when only one plot character

appears, ordinate equality is indicated.)

Empirical evidence shows that as V_p becomes narrow, indicating confidence

in M_p, the Bayesian estimation sequence has much more inertia with respect to

data inputs. This means that the sequence is much more stable. As a consequence,

large quantities of data are required to achieve good estimation of λ^* should

M_p be largely in error from λ^*.

However, if $M_p \approx \lambda^*$, the stability aspect is an obvious asset to the

Bayesian estimation technique. These observations are demonstrated in

Figures 13 and 14 respectively.

+ It must be understood that λ^* is known only for simulation purposes and
  analysis .· Actual testing does not allow this to be the case.

Since in reality one can not be sure where M_p is with respect to λ^*,
it is obvious that M_p, with respect to λ^* and selected V_p values must be
investigated. The Bayesian performance-distribution is a means of invest-
igating these features. Bayesian performance-distributions were defined
and discussed in Section 5. The results of these empirical studies is
discussed below.

The major questions to be answered by this analysis will be concentrated in
two areas.

1. What is the effect on the Bayesian performance distribution when
   V_p changes?

2. What is the effect on the Bayesian performance distribution when
   M_p is in error with respect to λ^* ?

In addition to the relation between M_p and V_p previously discussed, V_p
also has definite effects upon the Bayesian performance-distributions. In
order to examine this effect, three items were fixed. First, the simulated
true failure-rate remained fixed at .05 failures/10^6 hours. The range of
prior mean values was fixed at $[.012, .09]$ failures/10^6 hours. The relation
defined between M_p and V_p also remained constant (Equation 32). For each of
the successive figures (3,11), the prior variance was allowed to become
narrower by increasing the constant K or Equation 32.

The following values will be used in the discussion below.

Lower distribution tail $\equiv$ < .04 failures/10^6 hours

Central distribution region $\equiv [.04, .062]$ failures/10^6 hours

Upper distribution tail $\equiv$ > .062 failures/10^6 hours

It is apparent (Figure 3) that, for a sufficiently wide V_p (k= .0006),
the performance of Bayesian estimation over classical is equivalent to a
coin toss. This holds for the range of M_p given 10 data points for each
simulation.

The remainder of the distribution figures (4-11) are summarized in the following table.

TABLE II

K FACTOR	EFFECT ON BAYESIAN PERFORMANCE DISTRIBUTION
$.0006 \leq K < .003$	. tail probabilities remain relatively constant . central region probabilities increase
$.003 \leq K \leq .3$	. central region and tail probabilities remain relatively constant
$K > .3$	. tail probabilities decline first . central region probabilities decline next . final phase leaves spike which coincides with λ^* .

There is a particularly significant feature of Table II. This feature concerns the "K" factor interval $[.003, .3]$ where the Bayesian performance distribution remains relatively stable for the prior mean (λ^* $\pm$ .04) $\times 10^{-6}$. This stability is not affected by the value of V_p dictated by the K factor. This implies that there is a range of prior values (M_p, V_p) which produce optimum performance of the Bayesian technique. The lack of distribution uniqueness for the above K interval is important since we are primarily concerned with limited prior knowledge and/or data. This stability implies that one does not have to be super-critical about the selection of M_p or V_p.

The following fact should be noted. The true or simulated rate was specified to be .05 failures/10^6 hours. Other studies indicate that the results generated from other simulation rate values would be similar, but not necessarily the same, quantitatively.

Analysis regarding the criticality of choosing particular values
for the prior Bayesian formulation will be treated below.

The experiments used to generate Figures 3-11 were done for the purpose
of graphically representing the probability distribution of Bayesian failure-
rate-estimators. This was done on the domain of prior mean values, M_p =
$\lambda^* \pm .04$ failures/10^6 hours for fixed K.

Given a pre-selected prior mean (M_p), criticality of selecting a prior
value of K from Equation 32 will be determined. If the Bayesian technique is
to be used to the best advantage, K (given M_p) should be selected such that
the highest probability of getting a better Bayesian estimate (ie., closer to
λ^*) relative to the classical is obtained.

It can be observed, from Figures 3-11, that a uniquely optimum K value
relating to the above criterion does not exist. It appears that the highest
probabilities are obtained when K is in the interval $[.003, .3]$ (see Figures
6-8). This means that, for a given M_p, K in the above interval can be selected
without essentially lowering the probability of obtaining a better Bayesian
estimate (ie, closer to λ^*) relative to the classical. From Equation 32,
it can be shown that the corresponding values of prior deviation (D_p) resulting
from the end points of the K interval are given as follows:

$$3.33 \; (M_p) \; \lesssim \; D_p \; \lesssim \; 333.33 \; (M_p)$$

One can therefore expect good Bayesian performance with respect to
classical without severe restrictions placed upon either prior mean or variance.

6.3 SUMMARY

1. Appropriate functions and distributions were selected for the application of the continuous Bayesian technique.

2. Treatment of a discrete form of the Bayesian technique was analyzed and is presented in Appendix D.

3. An easily computed iteration technique was presented. This allows easy incorporation of new data as it becomes available.

4. Simulation results and graphics were presented in hopes of facilitating Bayesian understanding.

5. The criteria are defined under which the Bayesian technique is superior to classical estimation techniques.

6. It was shown that severe restrictions on prior Bayesian information are not necessary.

7.0 REFERENCES

1) Babillis, R. A., Smith, A.M., "Application of Bayesian Statistics in Reliability Measurement", General Electric Technical Information Series, No. 65SD250, April 1965.

7.0 <u>REFERENCES</u> (cont'd)

2) Breen, T. J., Koehler, R. D., "Implementation of Bayesian Techniques at GE - RDS", General Electric Technical Information Series on Bayes Theorem and Its Applications to Reliability Evaluation, No. 67SD213, April 1967.

3) Canavos, George C. "A Bayesian Approach to Parameter and Reliability Estimation in the Poisson Distribution", <u>IEEE Transactions on Reliability</u> V.R-21, No. 1, February 1972, pp. 52-55.

4) Easterling, Robert G., Weeks, David L., "An Accuracy Criterion for Bayesian Tolerance Intervals", <u>Royal Statistical Society Journal,</u> No. 2, pp. 236-240, 1970

5) Fruend, J. E., <u>Mathematical Statistics,</u> Prentice-Hall, Inc., 1965.

6) Hoel, Paul G., <u>Introduction to Mathematical Statistics</u>, John Wiley and Sons, Inc. 1966.

7) Parzen, Emanuel, <u>Modern Probability Theory and Its Applications,</u> John Wiley and Sons, Inc. 1962.

8) SimKins, D. J., "Weibull, Bayes and Entropy", Proceedings, 1972 Annual Reliability and Maintainability Symposium, January 1972, San Francisco, pp. 190-199

9) Van Alven, William H., Ed. ARINC Research Corporation, <u>Reliability Engineering,</u> Prentice-Hall, Inc., 1965.
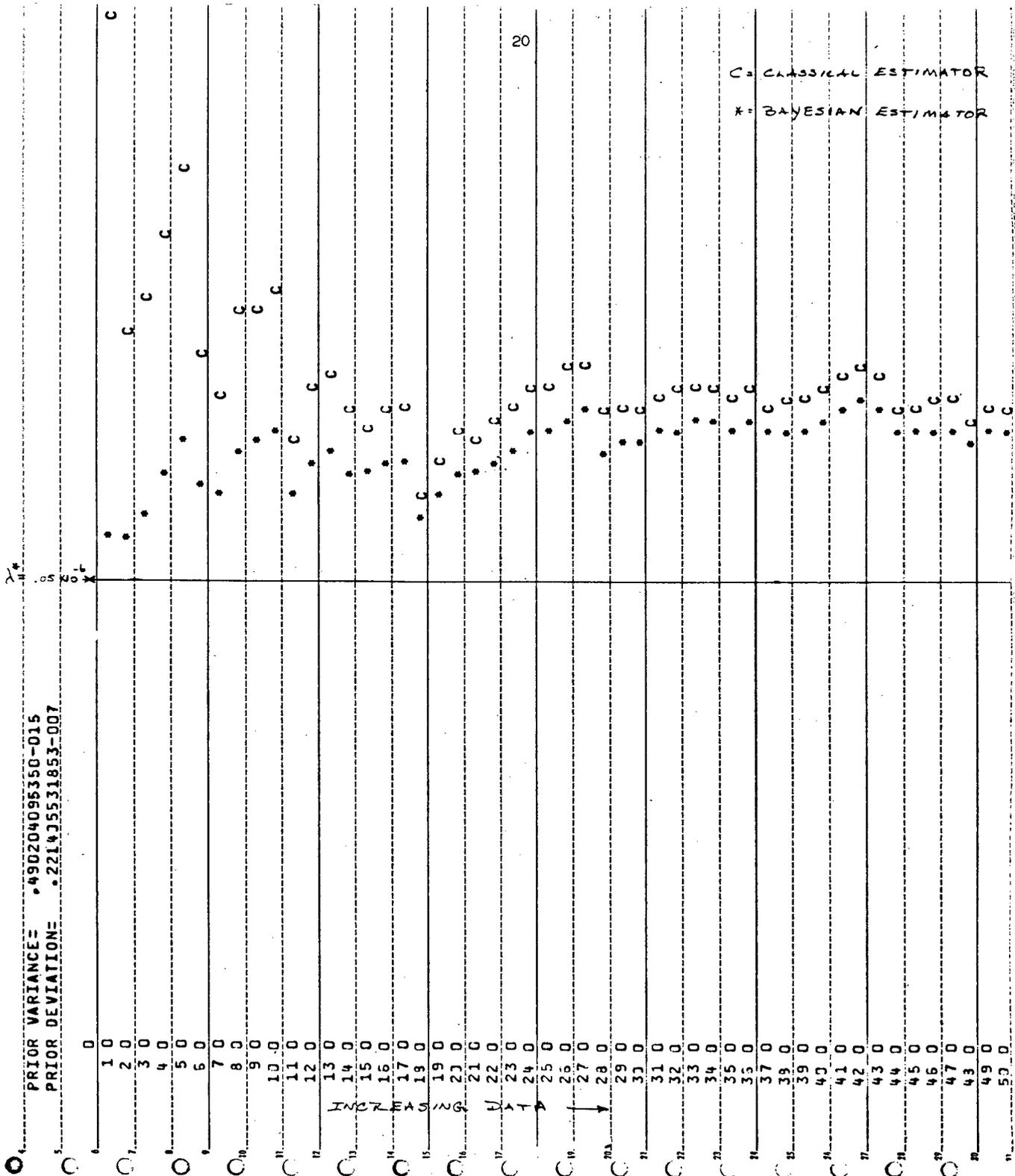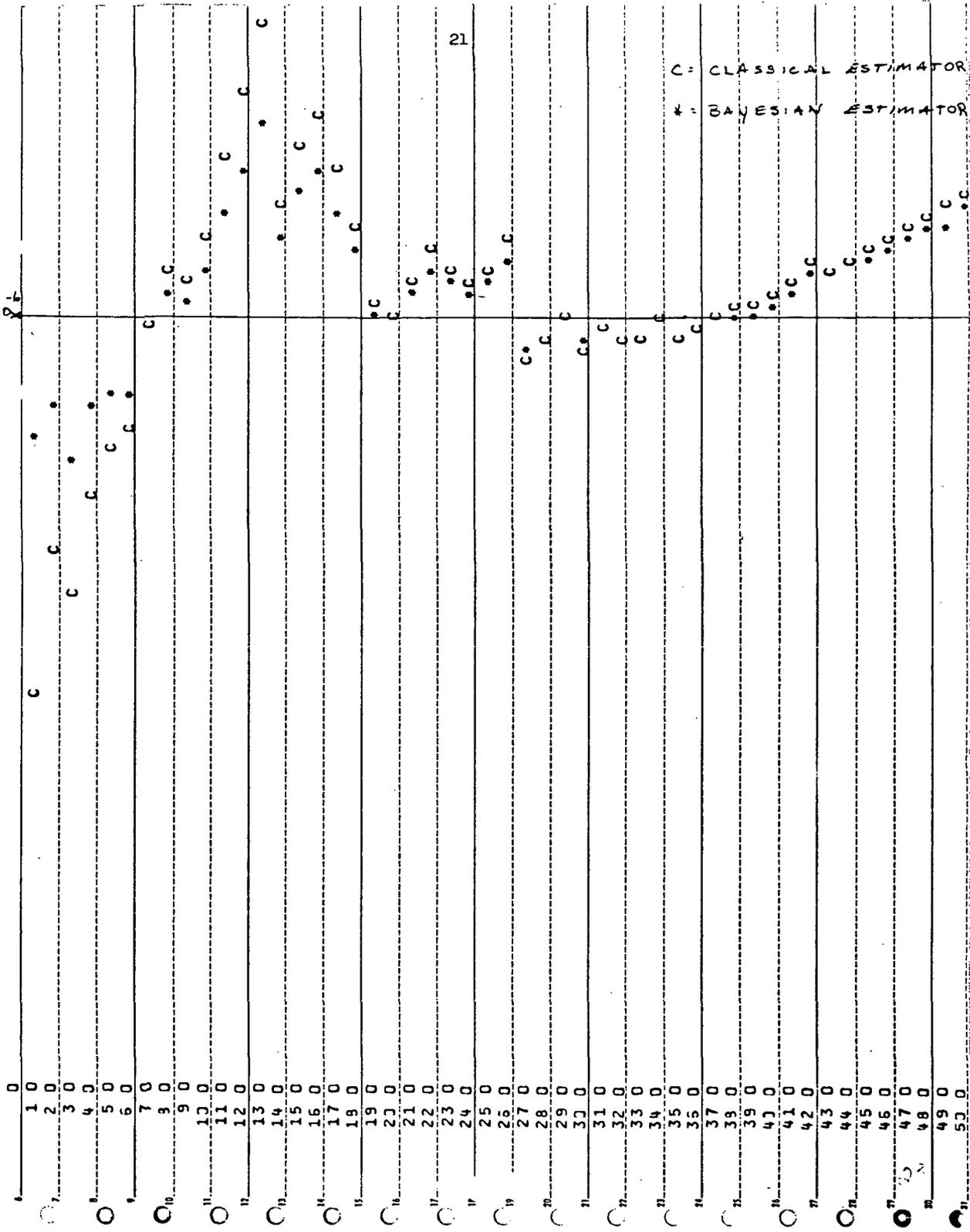
Fig. 1. True failure rate, classical estimator, and Bayesian estimator with respect to number of simulated trials: $\lambda^* = 0.05 \times 10^{-6}$, $V_p = 0.49 \times 10^{-15}$, $D_p = 0.22 \times 10^{-7}$ failures/hour

Fig. 2. True failure rate, classical estimator, and Bayesian estimator with respect to number of simulated trials: $\lambda^* = 0.05 \times 10^{-6}$, $V_p = 0.49 \times 10^{-15}$, $D_p = 0.22 \times 10^{-7}$ failures/hour

Fig. 3. Probability of Bayesian estimator being better than classical estimator, K = 0.006

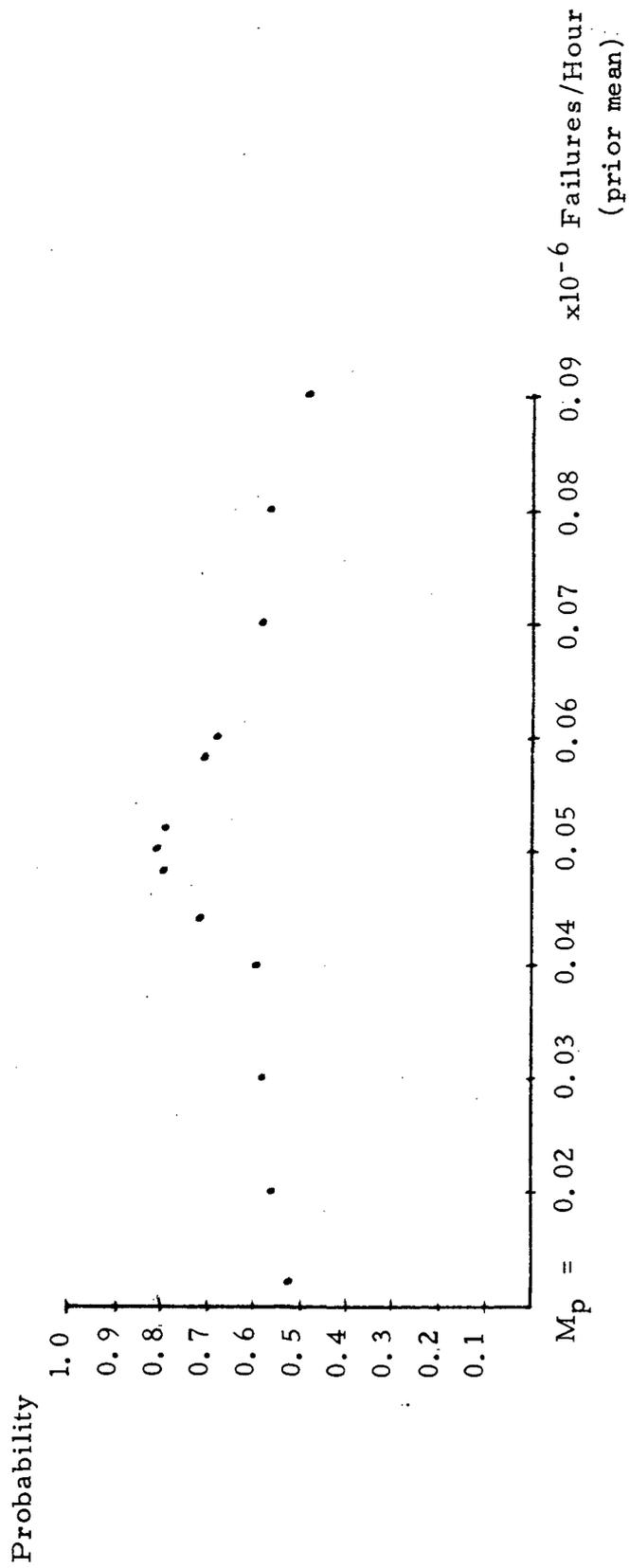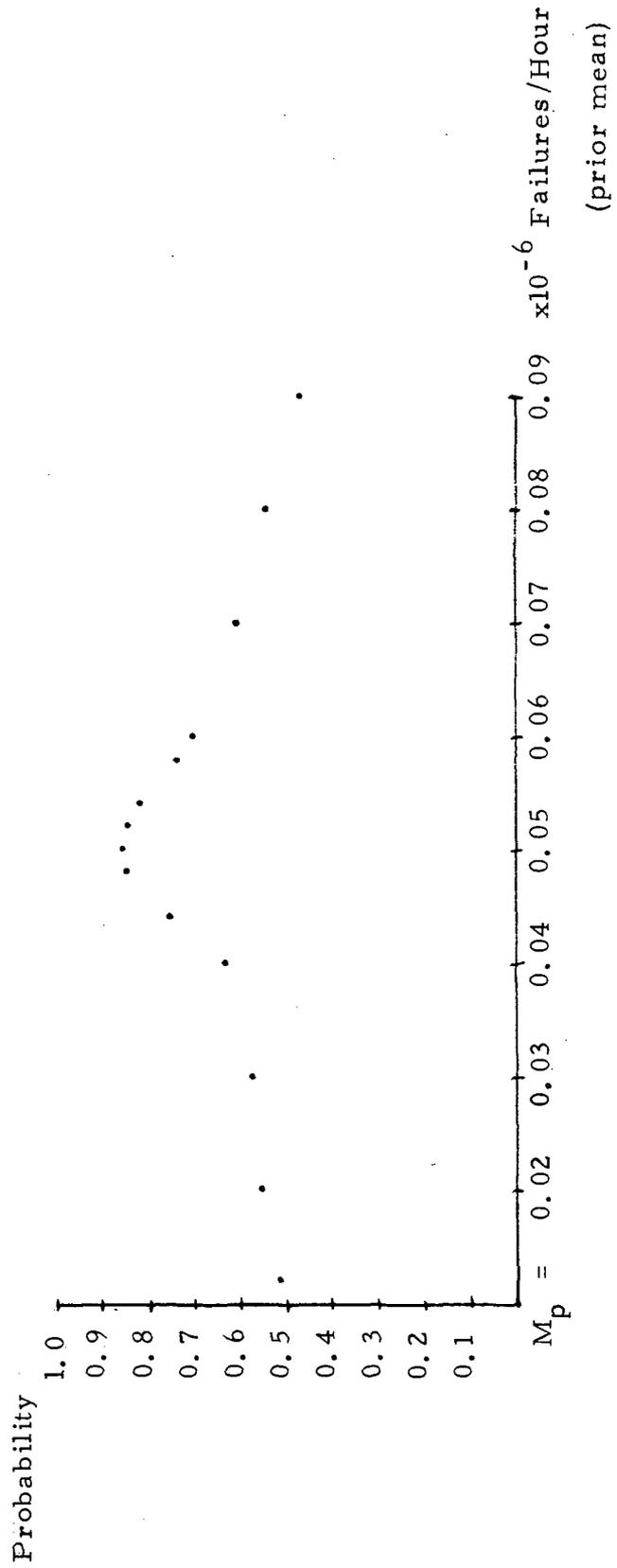Fig. 4. Probability of Bayesian estimator being better than classical estimator, K = 0.008

JPL Technical Memorandum 33-614   35

Simulation Rate = 0.05×10^{-6} Failures

Data = 10

Trials = 500

K = Variance Factor = 0.001

Probability

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

M_p = 0.02   0.03   0.04   0.05   0.06   0.07   0.08   0.09 $\times 10^{-6}$ Failures/Hour

(prior mean)

Fig. 5.   Probability of Bayesian estimator being better than classical estimator, K = 0.001
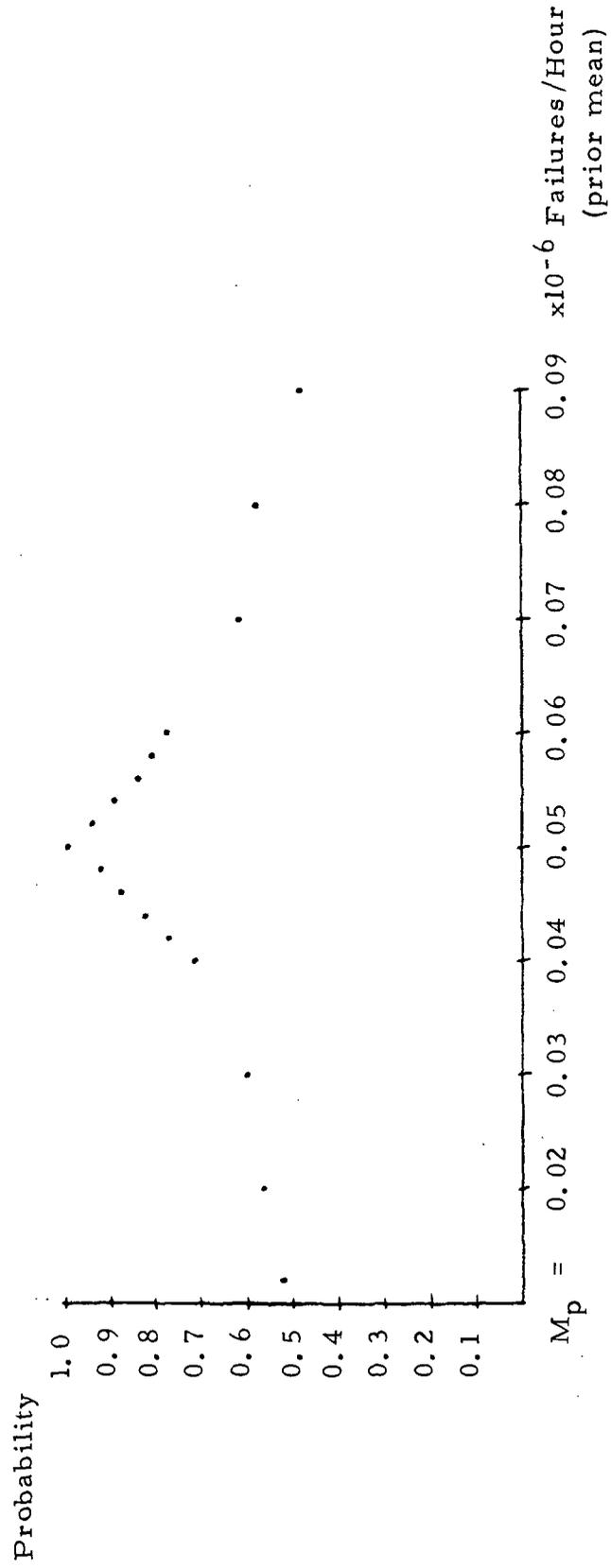
Fig. 6. Probability of Bayesian estimator being better than classical estimator, K = 0.003

Simulation Rate = 0.05×10^{-6} Failures/Hour

Data = 10

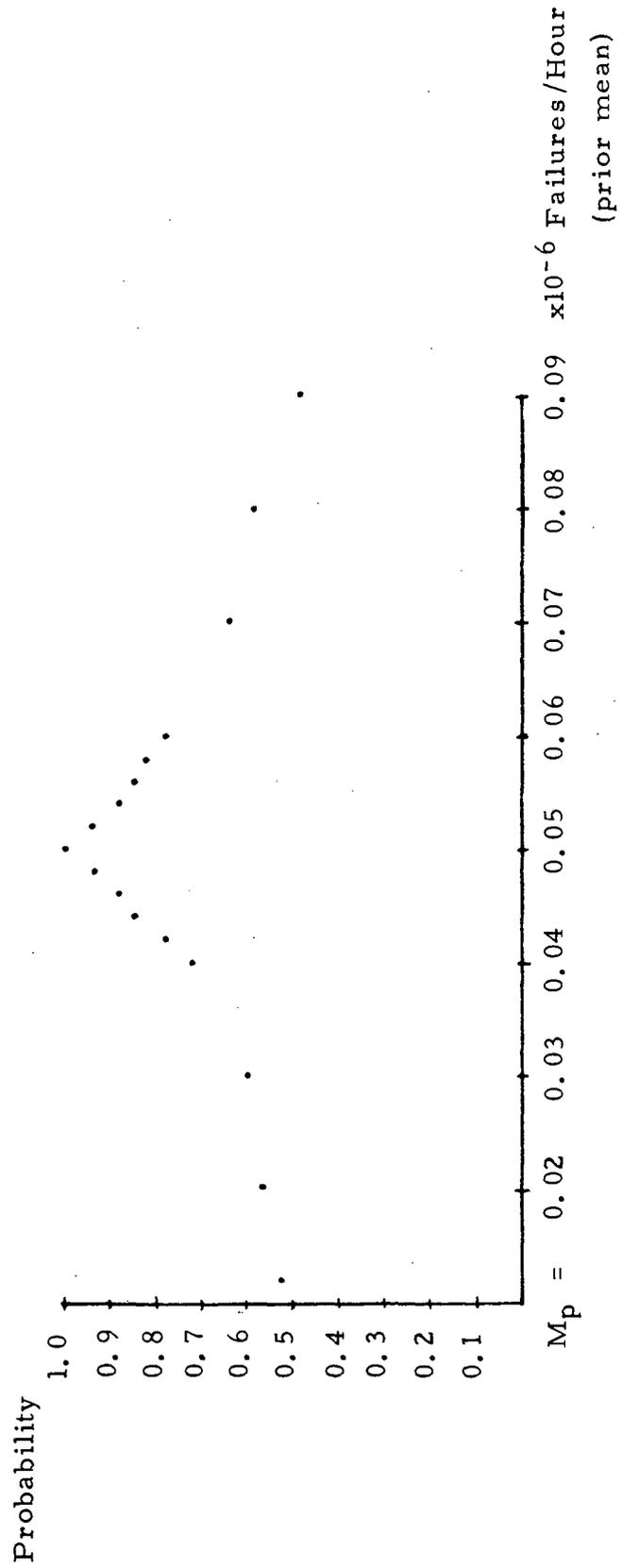Trials = 500

K = Variance Factor = 0.01

Fig. 7. Probability of Bayesian estimator being better than classical estimator, K = 0.01

Simulation Rate = 0.05×10^{-6} Failures/Hour

Data = 10

Trials = 500

K = Variance Factor = 0.3

Probability

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

M_p = 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 $\times 10^{-6}$ Failures/Hour
(prior mean)

Fig. 8. Probability of Bayesian estimator being better than classical estimator, K = 0.3

Simulation Rate = 0.05×10^{-6} Failures/Hour

Data = 10

Trials = 500

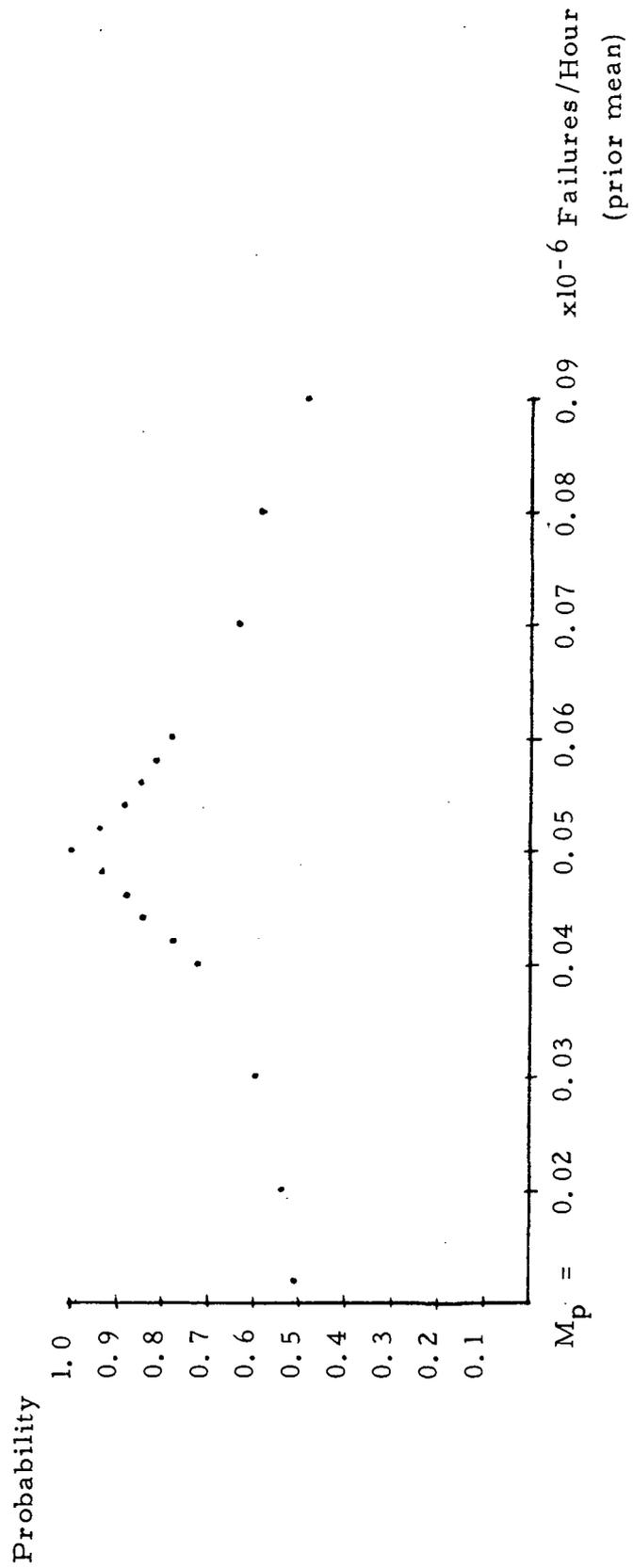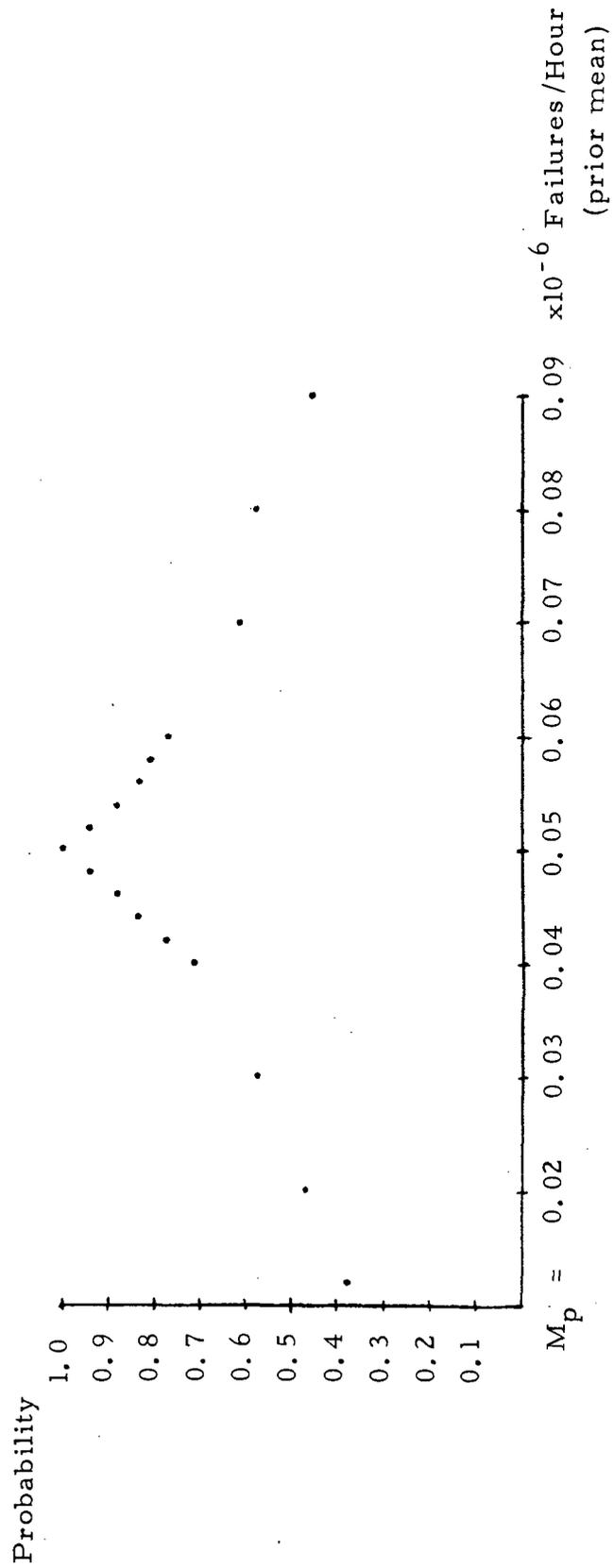K = Variance Factor = 0.9

Fig. 9. Probability of Bayesian estimator being better than classical estimator, K = 0.9

Simulation Rate = 0.05x10^{-6} Failures/Hour

Data = 10

Trials = 500

K = Variance Factor = 2.0

Fig. 10. Probability of Bayesian estimator being better than classical estimator, K = 2.0
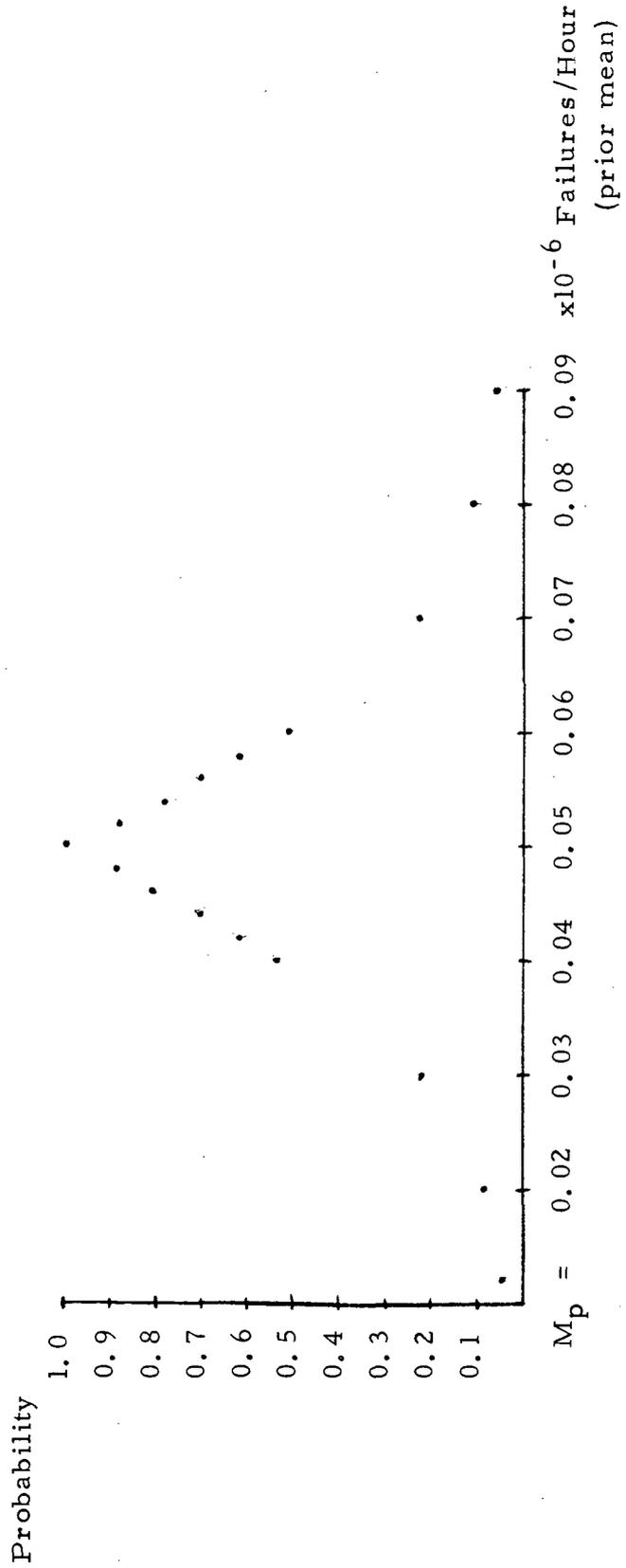
Fig. 11. Probability of Bayesian estimator being better than classical estimator, K = 20.0

Fig. 12. True failure rate, classical estimator, and Bayesian estimator with respect to number of simulated trials: $\lambda^* = 0.05 \times 10^{-6}$, $V_p = 0.34 \times 10^{-14}$, $D_p = 0.58 \times 10^{-7}$ failures/hour

Fig. 13. True failure rate, classical estimator, and Bayesian estimator with respect to number of simulated trials: $\lambda^* = 0.05 \times 10^{-6}$, $V_p = 0.14 \times 10^{-16}$, $D_v = 0.37 \times 10^{-8}$ failures/hour

$\lambda^* = .05 \times 10^{-6}$

$C = CLASSICAL\ ESTIMATOR$

$* = BAYESIAN\ ESTIMATOR$

INCREASING DATA →

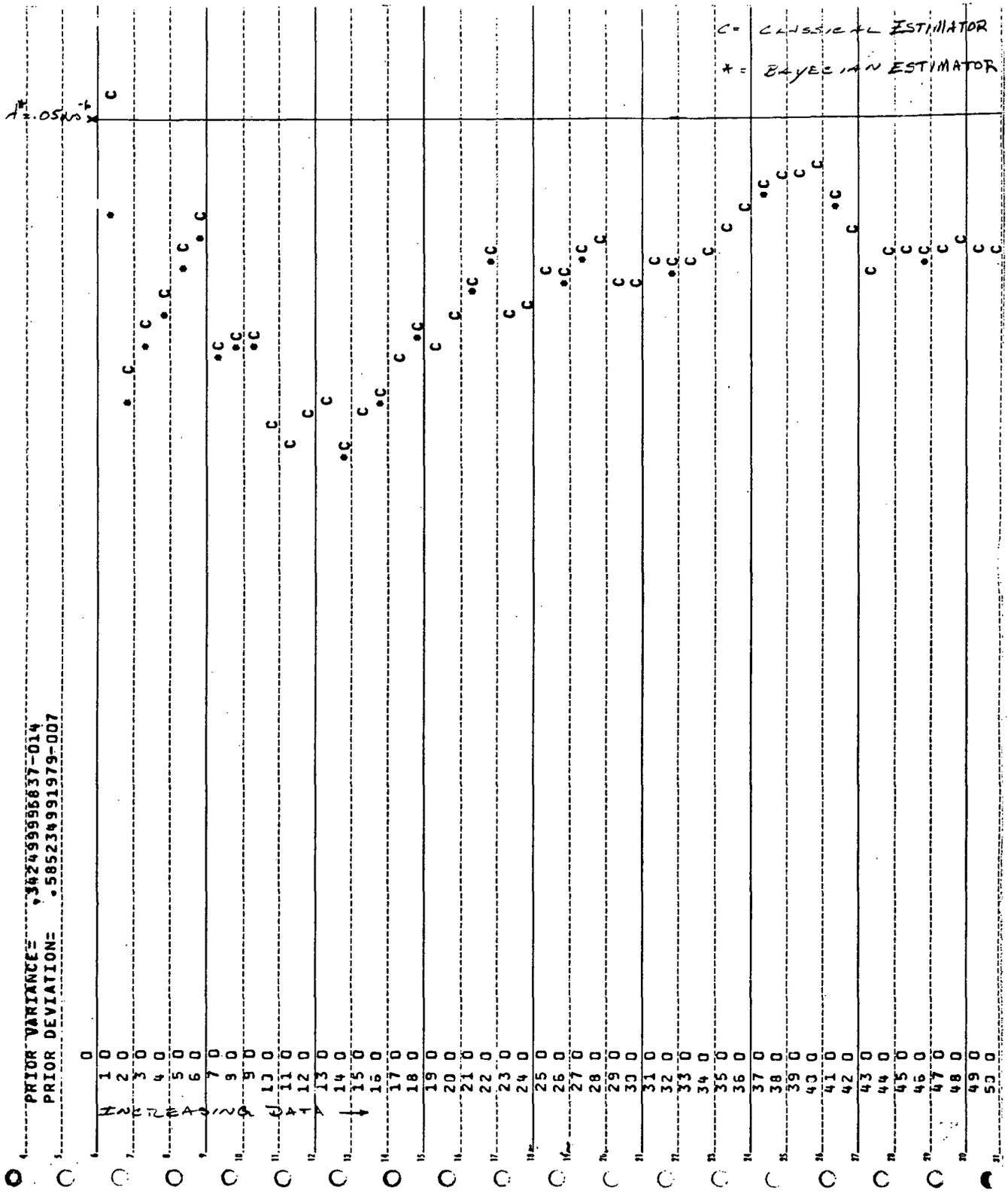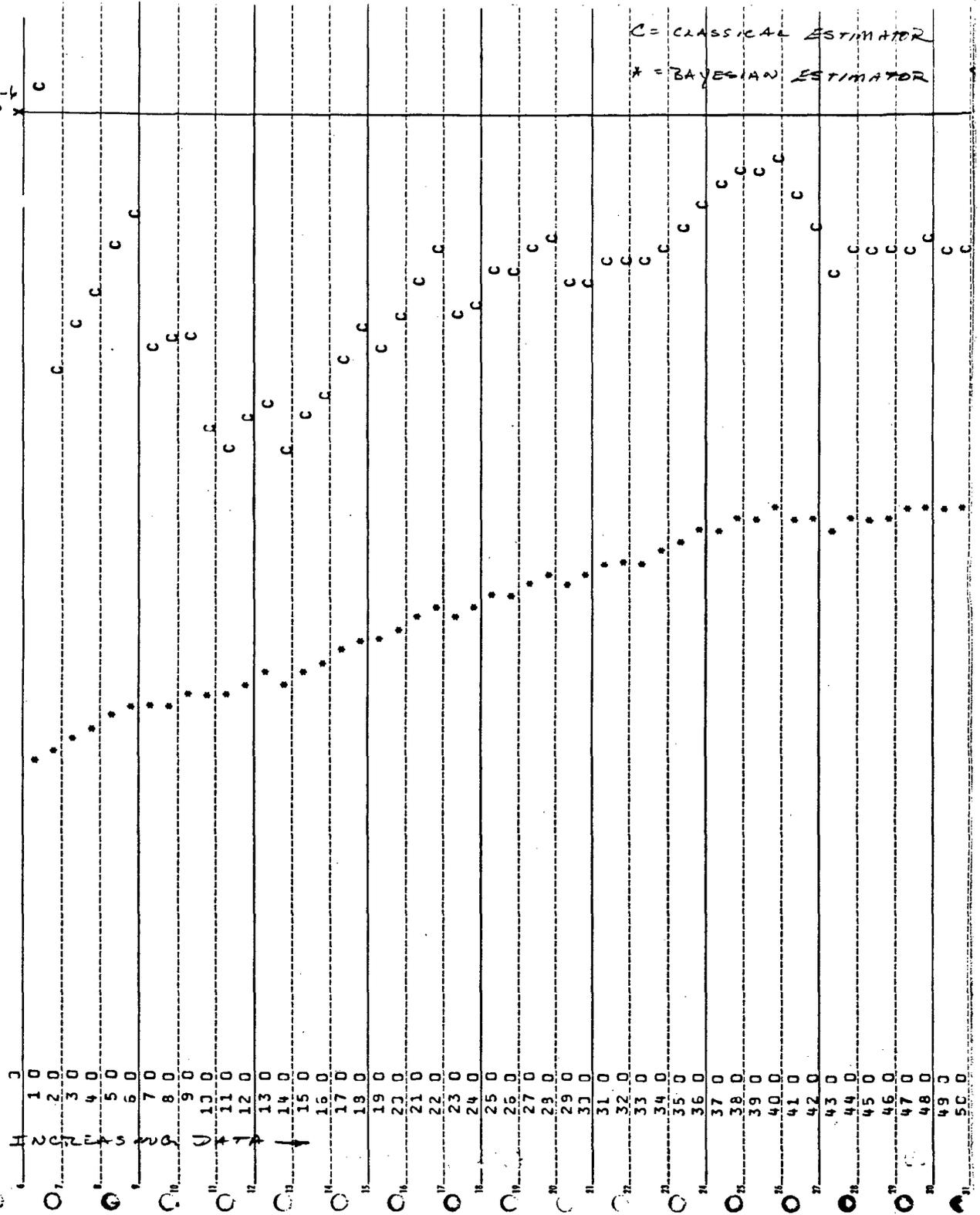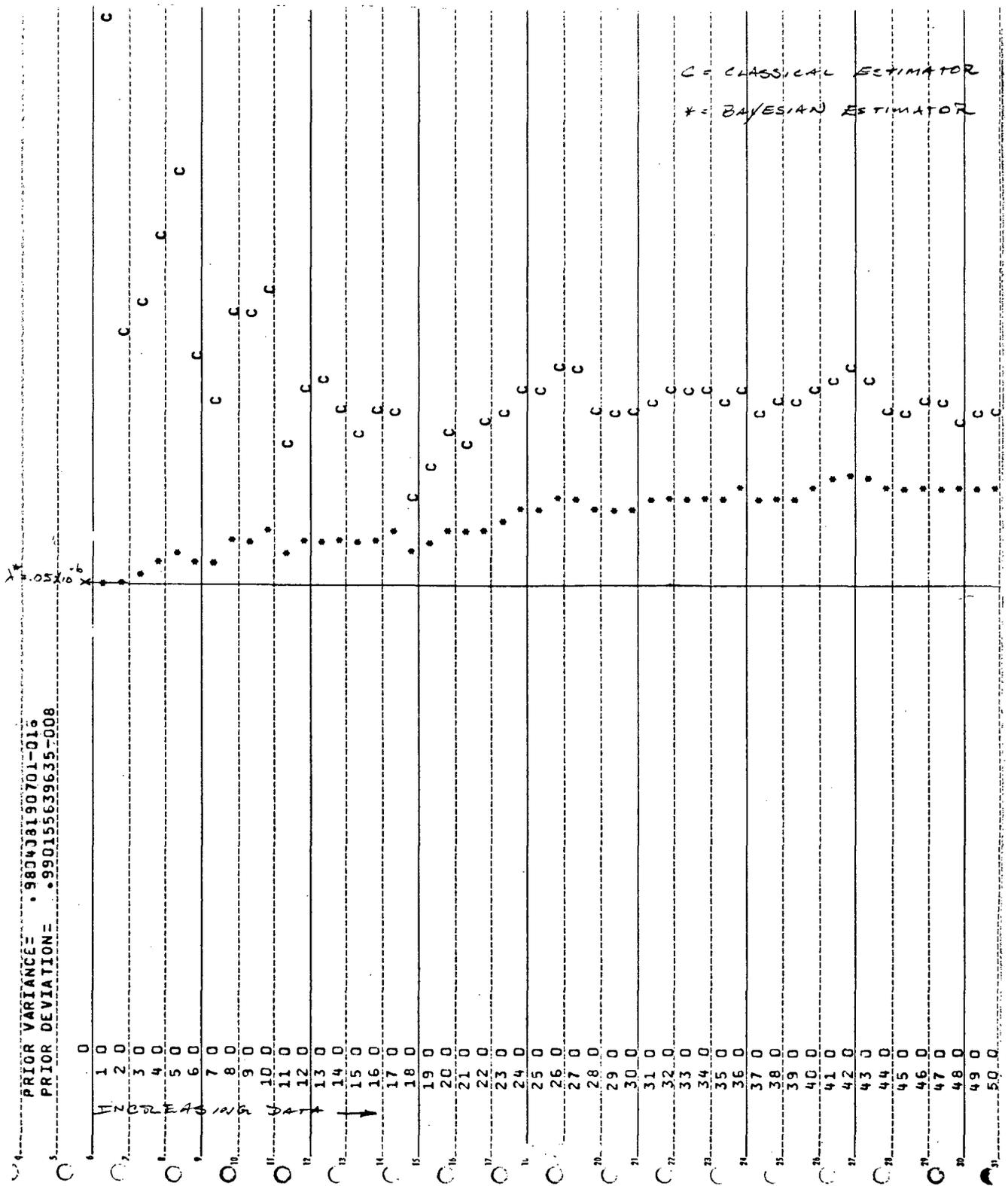Fig. 14. True failure rate, classical estimator, and Bayesian estimator with respect to number of simulated trials: $\lambda^* = 0.05 \times 10^{-6}$, $V_p = 0.98 \times 10^{-16}$, $D_p = 0.99 \times 10^{-8}$ failures/hour

DERIVATION OF FAILURE-RATE DENSITY FUNCTION

Define a failure-rate-estimator (λ) as a function of times-to-failure (i.e. cumulative)

$$\lambda = h\ (t')$$

where $h\ (t') = r/t'$

(A-1)   therefore $\lambda = r/t'$

where $t' = \sum_{i=1}^{\lambda} t_i$

and r = total number of failures occurring at rate μ, μ is the unknown true failure-rate.

From Reference 7, a relation between sums of exponentially distributed random variables and "r" failures occurring at rate "μ" is derived. The resulting probability density function is:

(A-2)
$$f(t') = \begin{cases} \dfrac{\mu^{\lambda}}{\Gamma(\lambda)} (t')^{\lambda-1} e^{-\mu t'} & t' \geq 0 \\ 0 & t' < 0 \end{cases}$$

This is the density for times-to-failure ($\sum_{i=1}^{\lambda} t_i$) to rth event occurring at rate μ .

Next, we find, the related density function for μ using Equation A-2 and a change of variable.

The density function for λ is denoted by $g(\lambda)$

(A-3) $\quad g(\lambda) = \dfrac{\mu^{n}}{\Gamma(n)} (t')^{n-1} e^{-\mu t'} \left(\dfrac{dt'}{d\lambda}\right)$

but $\quad t' = \dfrac{n}{\lambda} \quad$ so

(A-4) $\quad g(\lambda) = \dfrac{(\mu n)^{n}}{\Gamma(n)} \left(\dfrac{1}{\lambda}\right)^{n+1} e^{-\left(\frac{\mu n}{\lambda}\right)}$

this is an inverted gamma density.

Proof that $g(\lambda)$ is a density function

Let $\quad I = \displaystyle\int_{\lambda=0}^{\infty} \dfrac{(\mu n)^{n}}{\Gamma(n)} \left(\dfrac{1}{\lambda}\right)^{n+1} e^{-\left(\frac{\mu n}{\lambda}\right)} d\lambda$

Substituting $\quad x = \dfrac{1}{\lambda}$

$I = \dfrac{-(\mu n)^{n}}{\Gamma(n)} \displaystyle\int_{x=\infty}^{0} \dfrac{x^{n+1} e^{-(\mu n)x}}{x^{2}} dx$

$$I = \frac{(\mu\lambda)^{r}}{\Gamma(\lambda)} \int_0^{\infty} \chi^{\lambda-1} \, e^{-(\mu\lambda)\chi} \, d\chi$$

but
$$\int_0^{\infty} \chi^{\lambda-1} \, e^{-(\mu\lambda)\chi} \, d\chi = \frac{\Gamma(\lambda)}{\left[(\mu\lambda-1)+1\right]}$$

therefore

$$I = \frac{(\mu\lambda)^{r}}{\Gamma(\lambda)} \; \frac{\Gamma(\lambda)}{(\mu\lambda)^{r}} = 1$$

However, the converse situation is what we wish to examine. That is to say, we wish to have an expectation of the value of the true rate (μ) given the estimator (λ) and the corresponding failures (r). The derivations which follow are directed at this converse situation. A gamma density assumption will be used for the description of the density function for μ. It will be shown that this approach yields the same result as the assumptions and derivations of section 3.

DERIVATION OF CONVERSE INVERTED GAMMA FUNCTION

$$g(\lambda \mid \mu, r) = \frac{(\mu r)^r}{\Gamma(r)} \left(\frac{1}{\lambda}\right)^{r+1} e^{-\left(\frac{\mu r}{\lambda}\right)} \qquad \text{from Equation A-4}$$

Assume that the true failure-rate (denoted by μ) has a prior

gamma density with parameters α, ϕ. See Equation A-5

$$(A\text{-}5) \quad f(\mu) = \frac{\phi^{\alpha}}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\phi\mu}$$

Using Bayes rule, the converse function is given by Equation A-6

$$(A\text{-}6) \quad h(\mu \mid \lambda, r) = \frac{f(\mu) \cdot g(\lambda \mid \mu, r)}{\int_{\mu=0}^{\infty} f(\mu) \cdot g(\lambda \mid \mu, r) \, d\mu}$$

$$\text{where} \quad f(\mu) \cdot g(\lambda \mid \mu, r) = \frac{(\mu r)^r \phi^{\alpha}}{\Gamma(r) \Gamma(\alpha)} \frac{\mu^{\alpha-1}}{\lambda^{r+1}} e^{-\left(\phi + \frac{r}{\lambda}\right)\mu}$$

Then

$$h(\mu \mid \lambda, \kappa) = \cfrac{\dfrac{\lambda^{\kappa} \phi^{\alpha}}{\lambda^{\kappa+1} \Gamma(\kappa) \Gamma(\alpha)} \mu^{(\alpha+\kappa)-1} e^{-\left(\phi + \frac{\kappa}{\lambda}\right)\mu}}{\dfrac{\lambda^{\kappa} \phi^{\alpha}}{\lambda^{\kappa+1} \Gamma(\kappa) \Gamma(\alpha)} \displaystyle\int_{\mu=0}^{\infty} \mu^{(\alpha+\kappa)-1} e^{-\left(\phi + \kappa/\lambda\right)\mu} d\mu}$$

Let $\quad y = \left(\phi + \dfrac{\kappa}{\lambda}\right)\mu \qquad\qquad$ in the integral. Then

$$dy = \left(\phi + \dfrac{\kappa}{\lambda}\right) d\mu$$

$$h(\mu \mid \lambda, \kappa) = \cfrac{\mu^{(\alpha+\kappa)-1} e^{-\left(\phi + \frac{\kappa}{\lambda}\right)\mu}}{\dfrac{1}{\phi + \frac{\kappa}{\lambda}} \displaystyle\int_{y=0}^{\infty} \dfrac{y^{(\alpha+\kappa)-1}}{\left(\phi + \frac{\kappa}{\lambda}\right)^{(\alpha+\kappa)-1}} e^{-y} \underbrace{\left(\phi + \frac{\kappa}{\lambda}\right) d\mu}_{dy}}$$

$$h(\mu \mid \lambda, \kappa) = \cfrac{\mu^{(\alpha+\kappa)-1} e^{-\left(\phi + \kappa/\lambda\right)\mu}}{\dfrac{1}{\left(\phi + \kappa/\lambda\right)} \displaystyle\int_{y=0}^{\infty} y^{(\alpha+\kappa)-1} e^{-y} dy}$$

$$h(\mu \mid \lambda, \kappa) = \cfrac{\mu^{(\alpha+\kappa)-1} e^{-\left(\phi + \kappa/\lambda\right)\mu}}{\Gamma(\alpha+\kappa) \Big/ \left(\phi + \kappa/\lambda\right)^{\alpha+\kappa}}$$

$$h(\mu \mid \lambda, \kappa) = \dfrac{\left(\phi + \kappa/\lambda\right)^{\kappa+\alpha}}{\Gamma(\kappa+\alpha)} \left(\mu+\kappa\right)^{-1} e^{-\left(\phi + \kappa/\lambda\right)\mu}$$

(converse inverted Gamma density)

PROOF THAT $h(\mu|\lambda, r)$ IS DENSITY FUNCTION

$$\int_{\mu=0}^{\infty} h(\mu|\lambda, r)\, d\mu = \frac{\left(\rho + \frac{r}{\lambda}\right)^{\alpha + r}}{\Gamma(r+\alpha)} \int_{\mu=0}^{\infty} \mu^{(r+\alpha)-1}\, e^{-\left(\rho + \frac{r}{\lambda}\right)\mu}\, d\mu$$

Let
$$y = \left(\rho + \frac{r}{\lambda}\right)\mu$$
$$dy = \left(\rho + \frac{r}{\lambda}\right) d\mu$$

$$\int_{\mu=0}^{\infty} h(\mu|\lambda, r) = \frac{\left(\rho + \frac{r}{\lambda}\right)^{\alpha + r}}{\Gamma(r+\alpha)} \int_{y=0}^{\infty} \frac{1}{\left(\rho + \frac{r}{\lambda}\right)^{(\alpha + r)}}\, y^{(\alpha + r)-1}\, e^{-y}\, dy$$

$$= \frac{\left(\rho + \frac{r}{\lambda}\right)^{\alpha + r}}{\Gamma(\alpha + r)} \cdot \frac{\Gamma(\alpha + r)}{\left(\rho + \frac{r}{\lambda}\right)^{\alpha + r}} = 1$$

$$h(\mu \mid \lambda, n) = K \, \mu^{(\alpha+n)-1} \, e^{-(\varphi + n/\lambda)\mu}$$

where $K = \dfrac{\left(\varphi + n/\lambda\right)^{\alpha+n}}{\Gamma(\alpha+n)}$

$$E(\mu) = \int_{\mu=0}^{\infty} \mu \, h(\mu \mid \lambda, n) \, d\mu = K \int_{\mu=0}^{\infty} \mu^{\alpha+n} \, e^{-(\varphi + n/\lambda)\mu} \, d\mu$$

Integrating by parts

$$E(\mu) = K \left\{ \frac{\alpha+n}{\varphi + n/\lambda} \int_{0}^{\infty} \mu^{(\alpha+n)-1} \, e^{-(\varphi + n/\lambda)\mu} \, d\mu \right\}$$

Let $\quad y = \left(\varphi + n/\lambda\right)\mu$

$\qquad dy = \left(\varphi + n/\lambda\right) d\mu$

$$E(\mu) = K \left\{ \frac{\alpha+n}{\left(\varphi + n/\lambda\right)^{\alpha+n+1}} \int_{y=0}^{\infty} y^{(\alpha+n)-1} \, e^{-y} \, dy \right\}$$

$$E(\mu) = \frac{\left(\varphi + n/\lambda\right)^{\alpha+n}}{\Gamma(\alpha+n)} \cdot \frac{\Gamma(\alpha+n)}{\left(\varphi + n/\lambda\right)^{\alpha+n+1}}$$

$$E(\mu) = \frac{\alpha+n}{\varphi + n/\lambda} \qquad \begin{array}{l}\text{expected value of converse inverted} \\ \text{gamma density function}\end{array}$$

APPENDIX B

GAMMA ITERATION DERIVATIONS

From the text, it is seen that the general formulation of the Bayesian application is given by Equation 17. This equation is reproduced here for convenience:

$$f(\lambda|F) = \frac{K/F! \, e^{-\lambda t} (\lambda t)^F \, \lambda^{\alpha-1} \, e^{-(\lambda/\beta)}}{\int_{\lambda=0}^{\infty} K/F! \, e^{-\lambda t} (\lambda t)^F \, \lambda^{\alpha-1} \, e^{-(\lambda/\beta)} \, d\lambda}$$

Suppose that initial or prior information is represented by failures F_o, and time t_o. The gamma form of this prior information becomes:

$$f_o(\lambda|F_o, t_o) = k_o \, \lambda^{\alpha-1} \, e^{-(\lambda/\beta)}$$

where $K_o = \dfrac{1}{\beta^{\alpha} \, \Gamma(\alpha)}$

Suppose that new data of the form F_1, t_1 becomes available. A substitution of this data into Equation 17 yields:

$$f(\lambda|F_1, t_1) = \frac{K_o/F_1! \, e^{-\lambda t_1} (\lambda t_1)^{F_1} \, \lambda^{\alpha-1} \, e^{-(\lambda/\beta)}}{\int_{\lambda=0}^{\infty} \frac{K_o}{F_1!} \, e^{-\lambda t_1} (\lambda t_1)^{F_1} \, \lambda^{\alpha-1} \, e^{-(\lambda/\beta)} \, d\lambda}$$

Simplifying,

$$f(\lambda \mid F_1, t_1) = \frac{\lambda^{(\alpha + F_1) - 1} \, e^{-\lambda(1/\beta + t_1)}}{\int_{\lambda=0}^{\infty} \lambda^{(\alpha + F_1) - 1} \, e^{-\lambda(1/\beta + t_1)} \, d\lambda} \qquad (B-1)$$

The denominator of Equation B-1 can be interpreted conveniently using a change of variable.

Let $\quad y = \lambda \left(\frac{1}{\beta} + t_1 \right)$

then $\quad dy = \left(\frac{1}{\beta} + t_1 \right) d\lambda$

Making this substitution, Equation B-1 becomes:

$$f(\lambda \mid F_1, t_1) = \frac{\lambda^{(\alpha + F_1) - 1} \, e^{-\lambda(1/\beta + t_1)}}{\dfrac{1}{\left(\frac{1}{\beta} + t_1 \right)^{\alpha + F_1}} \int_{y=0}^{\infty} y^{(\alpha + F_1) - 1} \, e^{-y} \, dy} \qquad (B-2)$$

The integral left in the denominator is the gamma function. This integral can be represented by the following notation:

$$\int_{y=0}^{\infty} y^{(\alpha + F_1) - 1} \, e^{-y} \, dy = \Gamma(\alpha + F_1)$$

With the appropriate substitutions and simplifications, Equation B-2

(i.e., Equation 19 of the text) becomes:

$$f(\lambda \mid F_1, t_1) = k_1 \, \lambda^{(\alpha + F_1) - 1} \, e^{-\lambda \left(\frac{1}{\beta} + t_1 \right)}$$

where $K_1 = \dfrac{\left(\frac{1}{\beta} + t_1 \right)^{\alpha + F_1}}{\Gamma(\alpha + F_1)}$

Successive applications of the preceding rationale yield Equations
20 and 21 of the text.

APPENDIX C

Derivations for mean, variance and deviation of the Gamma Function given the definition (Equation 22 are presented below):

Mean Derivations

$$E(\lambda) \equiv \int_{\lambda} \lambda \, f(\lambda) \, d\lambda \qquad\qquad \text{(c-1)} \qquad 0 \leq \lambda \leq \infty$$

Substitute Equation 18

then

$$E(\lambda) = K_0 \int_{\lambda} \lambda^{\alpha} e^{-\lambda/\beta} \, d\lambda$$

where $K_0 = \dfrac{1}{\beta^{\alpha} \, \Gamma(\alpha)}$

Integrating by parts yields:

$$E(\lambda) = K_0 \, \alpha\beta \int_{0}^{\infty} \lambda^{\alpha-1} e^{-\lambda/\beta} \, d\lambda$$

Changing variable, let $y = \lambda/\beta$

then

$$E(\lambda) = K_0 \, \alpha\beta^{\alpha+1} \int_{y=0}^{\infty} y^{\alpha-1} e^{-y} \, dy$$

$$E(\lambda) = K_0 \, \alpha \, \mathscr{e}^{\alpha+1} \, \overline{\Gamma}(\alpha)$$

Substituting $K_0 = \dfrac{1}{\mathscr{e}^{\alpha} \, \overline{\Gamma}(\alpha)}$ $E(\lambda)$ becomes:

With the incorporation of new data F, and t, $f(\lambda)$ becomes

$f(\lambda \mid F_1, t_1)$ as noted in Appendix B.

Using the Equation 19 and the definition of expected value, we get:

$$E_1(\lambda) = K_1 \int_{\lambda=0}^{\infty} \lambda^{\alpha+F_1} \, \mathscr{e}^{-\lambda\left(\frac{1}{\beta}+t_1\right)} \, d\lambda$$

where $K_1 = \dfrac{\left(\frac{1}{\beta}+t_1\right)^{\alpha+F_1}}{\overline{\Gamma}(\alpha+F_1)}$

Integrating,

$$E_1(\lambda) = k_1 \, \frac{\alpha + F_1}{1/\beta + t_1} \int_{\lambda=0}^{\infty} \lambda^{(\alpha + F_1)-1} \, e^{-\lambda(1/\beta + t_1)} \, d\lambda$$

Changing variable i.e., $y = \lambda(1/\beta + t_1)$

$$E_1(\lambda) = k_1 \, \frac{\alpha + F_1}{1/\beta + t_1} \, \frac{1}{(1/\beta + t_1)^{\alpha + F_1}} \int_{y=0}^{\infty} y^{(\alpha + F_1)-1} \, e^{-y} \, dy$$

But

$$\int_{y=0}^{\infty} y^{(\alpha + F_1)-1} \, e^{-y} \, dy = \Gamma(\alpha + F_1)$$

therefore

$$E_1(\lambda) = \frac{\alpha + F_1}{1/\beta + t_1}$$

Successive applications and an induction argument using the notation developed will yield Equation 24 found in the text.

$$E_m(\lambda) = \frac{\alpha + \sum_{i=1}^{m} F_i}{1/\beta + \sum_{i=1}^{m} t_i}$$

The variance for a continuous density function f (λ) is given by:

$$E\left[(\lambda - E(\lambda))^2\right] = \int_\lambda (\lambda - E(\lambda))^2 \, f(\lambda) \, d\lambda$$

Let $E(\lambda) \equiv \mu$

then

$$E\left[(\lambda - \mu)^2\right] = \int_\lambda \left(\lambda^2 + \mu^2 - 2\lambda\mu\right) f(\lambda) \, d\lambda$$

$$E\left[(\lambda - \mu)^2\right] = \int_\lambda \lambda^2 \, f(\lambda) \, d\lambda \; - \; \mu^2$$

From Equation 18, f(λ) = $K_0 \, \lambda^{\alpha-1} \, e^{-\lambda/\beta}$

where $K_0 = \dfrac{1}{\beta^\alpha \, \Gamma(\alpha)}$

Therefore

$$E\left[(\lambda - \mu)^2\right] = K_0 \int_{\lambda=0}^{\infty} \lambda^{\alpha+1} \, e^{-\lambda/\beta} \, d\lambda \; - \; \mu^2$$

In changing variable, let $y = \lambda/\beta$

$$dy = \frac{d\lambda}{\beta}$$

Integrating twice (by parts) yields the following equation.

$$E\left[(\lambda - \mu)^2\right] = K_0 \beta^{\alpha+2}(\alpha+1)\alpha \int_{y=0}^{\infty} y^{\alpha-1} e^{-y} dy - \mu^2$$

where $\displaystyle\int_{y=0}^{\infty} y^{\alpha-1} e^{-y} dy$ is the Gamma function $\Gamma(\alpha)$

Combining terms and simplifying

$$E\left[(\lambda - \mu)^2\right] = \alpha \beta^2$$

Equivalently

$$E\left[(\lambda - \mu)^2\right] = \frac{\alpha}{(1/\beta)^2}$$

$$E\left[(\lambda - E_1(\lambda))^2\right] = \int_\lambda (\lambda - E_1(\lambda))^2 \, f_1(\lambda \mid F_1) \, d\lambda = \frac{\alpha + F_1}{\left(\frac{1}{\beta} + t_1\right)^2}$$

Where $\quad E_1 = \dfrac{\alpha + F_1}{\frac{1}{\beta} + t_1}$

$$K_1 = \left[\left(\frac{1}{\beta} + t_1\right)^{\alpha + F_1}\right] \Big/ \Gamma(\alpha + F_1)$$

$$E\left[(\lambda - E_m(\lambda))^2\right] = \int_\lambda (\lambda - E_m(\lambda))^2 \, f_m(\lambda \mid F_m) \, d\lambda = \frac{\alpha + \sum\limits_{i=1}^{m} F_i}{\left(\frac{1}{\beta} + \sum\limits_{i=1}^{m} t_i\right)^2}$$

Where

$$E_m(\lambda) = \frac{\alpha + \sum\limits_{i=1}^{m} F_i}{\frac{1}{\beta} + \sum\limits_{i=1}^{m} t_i}$$

$$K_m = \left[\left(\frac{1}{\beta} + \sum\limits_{i=1}^{m} t_i\right)^{\alpha + \sum\limits_{i=1}^{m} F_i}\right] \Big/ \Gamma\left(\alpha + \sum\limits_{i=1}^{m} F_i\right)$$

This process may be continued in the above manner. The details involve a change of variable for integration and substitution of the appropriate constants. An induction argument yields Equation 25 of the text.

APPENDIX D

COMPARISON OF DISCRETE BAYESIAN TO THE CONTINUOUS
BAYESIAN FAILURE RATE ESTIMATION

1. Introduction

This comparison was prompted by the methods and formulations

contined in Reference 8. The Bayesian formulation presented

in this reference quantifies prior information discretely. The

complete Bayesian formulation is developed via an assumed Weibull

hazard-rate function. The discrete Bayesian formulation takes the

following form:

$$P(A|B)_i = \frac{P(B|A)_i \, P(A)_i}{\sum\limits_{i=1}^{m} P(B|A)_i \, P(A)_i} \qquad i = 1, 2, 3, \cdots, m$$

where i = a particular cell

n = number of cells

$P(A)_i$ = probability of event A (prior for cell$_i$)

$P(B|A)_i$ = probability of event B given A for cell$_i$

$P(A|B)_i$ = probability of event A given B for cell$_i$

A three parameter Weibull assumption yields the following

form for P (B|A):

$$P(B|A)_i = \frac{\left\{ \frac{1}{\theta_i} \left[(t_2 - \varepsilon_i)^{m_i} - (t_1 - \varepsilon_i)^{m_i} \right] \right\}^R \exp\left\{ -\frac{1}{\theta_i} \left[(t_2 - \varepsilon_i)^{m_i} - (t_1 - \varepsilon_i)^{m_i} \right] \right\}}{R!}$$

Where θ = Scale parameter (i.e.; failure rate = $1/\theta$)

M = Shape parameter

E = Position parameter

R = Number of failures in the time interval

II. Experiment

Development of a discrete form of a Bayesian technique provided a significant departure from the continuous technique treated in the text. It was felt that a comparison experiment should be conducted. This experiment was designed to simultaneously compare three quantities the classical failure-rate-estimator and the expected values (means) for both the continuous and discrete Bayesian formulations.

The standard exponential assumption is a special case of the Weilbull. Therefore, the appropriate parameter substitutions for the Weibull yielded the exponential function. The discrete computation algorithm was developed for an arbitrary number of failure (m) and sum of failure times (Σt). The posteriori probability for a given cell $(i = 1, 2, 3, \cdots, m)$ is given by the following equation:

$$P(A|B)_i = \frac{P(A)_i \left(1/\theta_i\right)^m \exp\left(-\left[1/\theta_i \sum_{j=1}^{m} t_j\right]\right)}{\sum_{i=1}^{m} \left\{P(A)_i \left(1/\theta_i\right)^m \exp\left(-\left[1/\theta_i \sum_{j=1}^{m} t_j\right]\right)\right\}}$$

The method of mathematical expectation (section 4.1) was used to derive the failure-rate values used for comparison. Discrete and

continuous prior distributions were made equivalent by equating the first and second moments of the respective priors (section 5). Computer-simulated times-to-failure were used as the comparative data base. Several prior variations were then considered. The expected value comparison was then conducted at each of thirty simulated data points.

III. Results

The results of this comparative simulation show no substantial difference between the expected values of the respective continuous and discrete Bayesian forms. This implies that the results reported in section 6 also hold for the discrete Bayesian when compared with the classical estimation technique. There are more subtle ways in which the continuous and discrete Bayesian forms differ. These differences are summarized below.

A.   Discrete Bayesian Advantages over Continuous Gamma

1.   Prior and subsequent distributions can represent multi-modal situations.

2.   Formulation of the prior distribution with respect to limited information may be more compatible with "engineering judgment."

3.   The discrete Bayesian form appears to be the only way in which the increasing failure rate can be treated.

4.   Complexity due to selection of mean and variance values and related distribution constants is eliminated.

B. Discrete Bayesian Disadvantages

    1.   The posteriori of the discrete Bayesian can never achieve a value not given by the prior distibution. Therefore, if the prior is in error (i.e. the prior density is not positive at the value, the discrete Bayesian will not converge to the true value with the incorporation of data.

    2.   The refinement of the discrete posteriori estimation process is dependent upon the cell size.