EGG-EA-5112

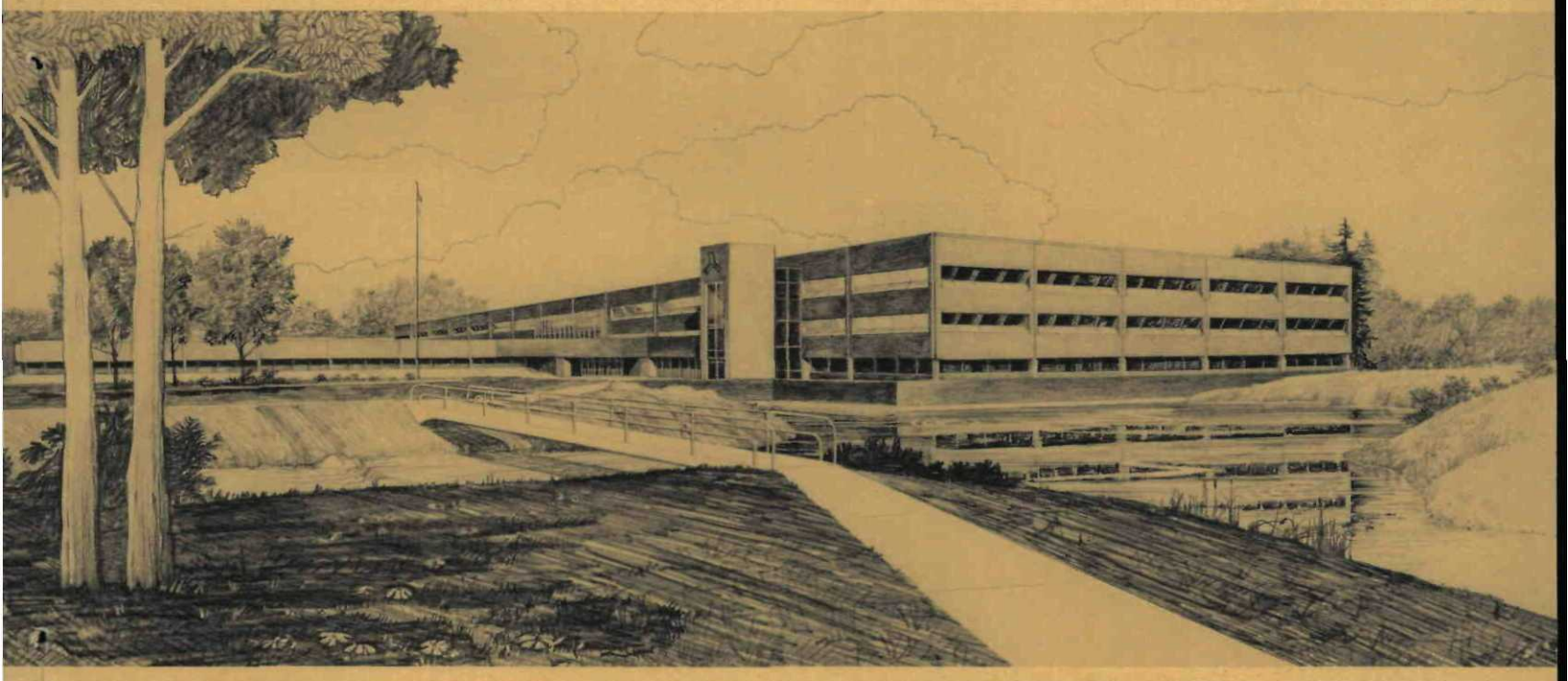
March 1980

ESTIMATORS FOR THE BINOMIAL FAILURE RATE COMMON CAUSE MODEL

MASTER

Corwin L. Atwood

U.S. Department of Energy
Idaho Operations Office • Idaho National Engineering Laboratory



Prepared for the
U.S. Nuclear Regulatory Commission
Under DOE Contract No. DE-AC07-76ID01570
FIN No. A6292

EG&G Idaho

DISCLAIMER

EG&G Idaho, Inc

FORM EG&G-398
(Rev 11-79)

INTERIM REPORT

Accession No. _____

Report No. EGG-EA-5112

Contract Program or Project Title:
Common Cause Data Analysis Program

Subject of this Document:
Estimators for the Binomial Failure Rate Common Cause Model

Type of Document:
Report of Development in Statistical Procedure

Author(s):
Corwin L. Atwood

Date of Document:
March 1980

Responsible NRC Individual and NRC Office or Division:
William Vesely, Jr., Probabilistic Anal. Staff, Nuclear Regulatory Research

This document was prepared primarily for preliminary or internal use. It has not received full review and approval. Since there may be substantive changes, this document should not be considered final.

EG&G Idaho, Inc.
Idaho Falls, Idaho **83415**

Prepared for the
U.S. Nuclear Regulatory Commission
Washington, D.C.
Under DOE Contract No. DE-AC07-76ID01570
NRC FIN No. A6292

INTERIM REPORT

[redacted]

ABSTRACT

In Vesely's binomial failure rate model, a system of m components is hit by random shocks which may cause components simultaneously to fail, each component with equal probability. Individual components may also fail when no shock has occurred. The data possibilities considered are that causes of single failures are identifiable (as shock or not) or not identifiable. Given data from such a system, non-Bayesian and Bayesian point and interval estimators are found for the various quantities of interest. Residual analyses and hypothesis tests are presented for checking the model assumptions. An example is worked out.

ACKNOWLEDGEMENTS

I am grateful to W. E. Vesely for many very helpful comments and suggestions, and to N. D. Cox for valuable discussion. W. J. Suitt wrote the extensive computer coding for performing the work of Section 7.

CONTENTS

FIGURES

TABLES

ESTIMATORS FOR THE BINOMIAL FAILURE RATE COMMON CAUSE MODEL

1. SUMMARY

Vesely[1] proposes a binomial failure rate (BFR) model for modeling common cause failures in a system. In this model, each component of the system has a constant failure rate λ. In addition, a "common cause shock," or "secondary event" in the terminology of reliability analysts, can occur with constant occurrence rate μ. If a shock occurs, the components fail independently of each other, each with probability p.

This paper presents results concerning the BFR model. There are six or seven related, unknown quantities of interest. The exact number depends on whether the causes of single failures are identifiable (i.e., due to shock or just individual failures) or not identifiable. Maximum likelihood estimators are given for all of these quantities. One of the likelihood equations requires numerical solution. Confidence intervals are given, which are sharp for some quantities and conservative for others. Confidence regions, not all rectangular, are given for estimating the quantities simultaneously.

Bayes estimators (posterior modes and means) and Bayes probability intervals are given for all the quantities of interest, as is the three-dimensional probability region for the three basic parameters. All the Bayesian results, except finding the modes, require numerical integration. However, in almost all cases, only integrals with respect to a single variable need to be done numerically. Depending on the parameters of the prior distribution, the integrand may contain an infinite series.

Residual analyses and hypothesis tests are presented for checking the model assumptions. Most of the hypothesis tests require a large sample size.

The results are applied to some of the boiling water nuclear reactor control rod data given by Vesely[1].

2. BASIC CALCULATIONS

Let there be a system with m components operating for time t. (If there are several such systems with the same m, each operating for time t_i, consider them as one system with $t = \Sigma t_i$.) Assume $m > 2$. (In Section 6, where the causes of single failures are identifiable, we will relax this assumption to $m \geq 2$.) Each component has an exponentially distributed lifetime with parameter (failure rate) λ, and the failures are mutually independent. In addition, a common cause shock may hit the system, with the shocks mutually independent and the time between shocks exponentially distributed with parameter μ. If a shock hits the system, the components of the system fail independently of each other, each with probability p. All failures are discovered and repaired as soon as they occur. Define $q = 1 - p$.

Let N_i be the number of occurrences of i simultaneous failures. (We will follow the convention of using capital letters to denote random variables and using the corresponding small letters to denote specific values of the random variables.) Then, basic calculations (see Reference 1 or Mann, Schafer, and Singpurwalla,[2] Section 4.2) yield that

N_i are independent Poisson(λ_i t)

with

$$\lambda_1 = m\lambda + \mu r_1 \tag{2.1}$$

and

$$\lambda_i = \mu r_i, \quad i = 2,.., m.$$

Here, r_i is defined by

$$r_i = \binom{m}{i} p^i q^{m-i}. \tag{2.2}$$

The special expression for λ_1 comes from the fact that a single failure may be simply an individual failure or it may come from a common cause shock that caused only one component to fail. (Reliability analysts would refer to a common cause shock that failed fewer than two components as a "potential common cause.")

The model has been formulated in terms of λ, μ, and p. Also of interest, indeed perhaps of great interest, are λ_1, defined by (2.1), and λ_+, defined by

$$\lambda_+ = \sum_{i=2}^{m} \lambda_i = \mu(1 - r_0 - r_1). \tag{2.3}$$

The quantity λ_+ is the rate of common cause failure occurrences. It must be distinguished from μ, the rate of common cause shocks. A last quantity of interest is μp^k for $2 \leq k \leq m$. This is the rate at which some specific k components fail simultaneously, and is the relevant rate corresponding to a k-element "AND" gate in a fault tree, or to a k-element cut set.

It may be that the causes of single failures are identifiable as either individual failure or common cause shock. Then, N_1 can be decomposed into $N_I + N_C$, where N_I is the number of individual failures, distributed Poisson($m\lambda t$), and N_C is the number of single failures due to common cause shocks, distributed Poisson($\mu r_1 t$). It may be very difficult to extract the necessary information for this decomposition from failure reports. For example, in nuclear industry Licensee Event Reports, the information may simply be unavailable: the cause is reported "unknown." Or, it may be difficult to interpret: is "dirt in valve" a common cause shock that happened to fail only one valve? Finally, the report may be untrustworthy: in one report, 96 control rods failed to insert fully. A possible cause reported was that oil leaked past 96 independent piston seals. We do not believe in such a coincidence, but assume some common cause unrecognized by the reporter. But do we then classify a reported single oil leakage past a piston seal as an individual failure, or as a common cause shock that failed only one component?

There are data sets in which causes of single failures are identifiable. For example, failures due to personnel error may all be regarded as due to common cause shock. However, the major portion of this paper assumes that N_I and N_C are not separately available. Everything simplifies if they are available, so the results are summarized in Section 6.

Since t is fixed, basic calculations show that the probability distribution of a set of failure data depends only on the number of failures, not on the times of the failures (Reference 2, page 180). Define

$$N_+ = \sum_{i=2}^{m} N_i \ .$$

Then

$$P\left[N_1 = n_1,.., N_m = n_m\right]$$

$$= P\left[N_1 = n_1\right] P\left[N_+ = n_+\right] P\left[N_2 = n_2,.., N_m = n_m \middle| N_+ = n_+\right]$$

symbolically written as

$$L = L_1 \times L_2 \times L_3 \ . \tag{2.4}$$

This decomposition is useful because

$$N_1 \sim \text{Poisson}(\lambda_1 t)$$

$$N_+ \sim \text{Poisson}(\lambda_+ t)$$

$$N_2,.., N_m \mid N_+ = n_+ \sim \text{multinomial}(n_+, z_2,.., z_m)$$

where

$$z_i = r_i/(1 - r_o - r_1) \ . \tag{2.5}$$

The multinominal probability L_3 can be written out as

$$P\left[N_2 = n_2, .., N_m = n_m \middle| N_+ = n_+\right] = \frac{n_+!}{n_2! \cdot\cdot n_m!} \prod_{i=2}^{m} z_i^{n_i}$$

$$= \frac{n_+!}{n_2! \cdot\cdot n_m!} \quad \frac{p^s q^{mn_+ - s}}{(1 - q^m - m\, p\, q^{m-1})^{n_+}} \quad \prod_{i=2}^{m} \binom{m}{i}^{n_i} \tag{2.6}$$

where s is defined as

$$s = \sum_{i=2}^{m} i\, n_i ,$$

the total number of components failing in multiple failures.

Note that $2n_+ \leq s \leq mn_+$. Note also that (N_1, N_+, S) is a sufficient statistic.

3. NON-BAYESIAN INFERENCE

3.1 Maximum Likelihood Estimators

It will be most convenient to parametrize the model in terms of (λ_1, λ_+, p). From estimators of these quantities, estimators of λ, μ, and μp^k will also be developed. The parameters λ_1, λ_+, and p are related by (2.1) through (2.3). If we set $\lambda \geq 0$ in (2.1), we obtain that the parameters satisfy the constraint

$$\lambda_1 \geq \lambda_+ \left[m\ p\ q^{m-1} / (1 - q^m - m\ p\ q^{m-1}) \right]. \tag{3.1}$$

This can be abbreviated $\lambda_1 \geq \lambda_+ z_1$, in the notation of (2.5).

From (2.4), the logarithm of the likelihood is

$$\log L = \log L_1 + \log L_2 + \log L_3$$

where

$$\log L_1 = -\lambda_1 + n_1 \log \lambda_1 - \log (n_1!)$$

$$\log L_2 = -\lambda_+ + n_+ \log \lambda_+ - \log (n_+!)$$

and L_3 is given by (2.6) and does not depend on λ_1 or λ_+.

Let us first maximize log L ignoring the constraint (3.1). The maximizing values of λ_1 and λ_+ are

$$\hat{\lambda}_1 = n_1/t$$

$$\hat{\lambda}_+ = n_+/t.$$

The maximizing value of p, denoted $\hat{p}$, must now be found.

If $m = 2$ or if $n_+ = 0$, then L_3 is identically 1. We have assumed throughout that $m > 2$. Now assume also that $n_+ > 0$. Differentiation of $\log L_3$ is straightforward:

$$\frac{\partial \log L_3}{\partial p} = \frac{s}{p\,q} - m\,n_+ \frac{1 - q^{m-1}}{q(1 - q^m - m\,p\,q^{m-1})} . \tag{3.2}$$

It is shown in Section 8.2 that if $s = 2n_+$, then $\hat{p} = 0$. If $s = m\,n_+$, then $\hat{p} = 1$. And otherwise, $q(\partial \log L_3/\partial p)$ is strictly decreasing in p and changes sign between 0 and 1; that is, $\hat{p}$ is the unique solution of

$$s = m\,n_+\,p \frac{1 - q^{m-1}}{1 - q^m - m\,p\,q^{m-1}} . \tag{3.3}$$

If $m = 3$, the smallest value that allows p to be estimated, then (3.3) has an explicit solution: $\hat{p} = 3(s - 2n_+)/(2s - 3n_+)$. For larger m, the solution must be found numerically.

The naive estimate of p would be $s/m\,n_+$, the average proportion of failed components in multiple failure occurrences. We would expect this naive estimate to be biased upwards, since it is based on only those common cause shocks that happen to result in at least two failures. And, in fact, $s/m\,n_+$ is somewhat larger than $\hat{p}$, the MLE. The naive estimator $s/m\,n_+$ can be used as an initial guess for solving (3.3).

It is interesting to note that an estimator based on the conditional first moment of S is the same as $\hat{p}$, because the equation

$$s = E\left[S \middle| N_+ = n_+\right]$$

reduces to (3.3). This can be shown directly or derived from (3.2) and the well-known fact that, under regularity conditions, $E(\partial \log L_3/\partial p) = 0$. The equality of the two estimators is not mere coincidence, but follows from the fact that the truncated binomial distribution is in the exponential family.

If $s/n_+ \gg 2$, then the naive estimate and the MLE are approximately equal. For in this case, the data show many failures per observed shock, indicating that virtually all common cause shocks result in at least two failures. If s/n_+ is close to 2, then the two estimates differ appreciably. The quantity S/N_+ will tend to be large if mp is large and small if mp is small.

Sometimes m is not known exactly. Then it is important to know how $\hat{p}$ varies as a function of m. It is shown in Section 8.1 that the right side of (3.3) is an increasing function of m, for fixed $\hat{p}$. Therefore, as m increases, p must be decreased to preserve equality in (3.3). That is, the MLE $\hat{p}$ decreases as m increases.

Once $\hat{\lambda}_1$, $\hat{\lambda}_+$, and $\hat{p}$ are found, $\hat{\lambda}$ and $\hat{\mu}$ follow from (2.1) through (2.3), and the MLE for μp^k is $\hat{\mu}\hat{p}^k$.

It may be that the above procedure makes $\hat{\lambda}$ negative! This is because log L was maximized without any contraints, whereas it should have been maximized subject to the constraint (3.1); i.e., $\lambda \geq 0$. In this case, proceed as follows.

As a function of λ_1, λ_+, and p, log L has a unique local maximum. Therefore, the local maximum is still unique when log L is written as a function of λ, μ, and p. If this local maximum occurs at $\hat{\lambda} < 0$, then the maximum subject to $\lambda \geq 0$ must occur at $\lambda = 0$. So, set $\lambda = 0$ in (2.4), and set $\hat{\lambda} = 0$. Define

$$N = \sum_{i=1}^{m} N_i \ .$$

Then

$$L = P\left[N = n\right] \ P\left[N_1 = n_1, .., \ N_m = n_m \middle| N = n\right]$$

where

$$N \sim \text{Poisson}(\lambda' t), \qquad \lambda' = \mu(1 - r_o)$$

and

$$P\left[N_1 = n_1, \ldots, N_m = n_m \mid N = n\right]$$

$$= \frac{n!}{n_1! \cdots n_m!} \frac{p^{s'} q^{mn - s'}}{(1 - q^m)^n} \prod_{i=1}^{m} \binom{m}{i}^{n_i}$$

with

$$s' = \sum_{i=1}^{m} i\, n_i \,.$$

The MLE's are found by setting

$$\hat{\lambda}' = n/t$$

and letting $\hat{p}$ be the solution of

$$s' = m\, n\, p/(1 - q^m).$$

This $\hat{p}$ is also the estimator obtained by setting

$$s' = E[S' \mid N = n].$$

It decreases as m increases. Then $\hat{\mu}$ follows from $\hat{\lambda}'$, $\hat{p}$, and the definition of λ', as does $\hat{\mu}\hat{p}^k$.

3.2 Confidence Intervals

3.2.1 Terminology. Since the data are discrete, all confidence regions are necessarily inexact, in the sense that

$$P[\text{region contains parameter}] \geq 1 - \alpha \qquad (3.4)$$

for all values of the parameter, with strict inequality for some values of the parameter. This is a well-known characteristic of discrete data. A $(1 - \alpha)$ confidence region will be called sharp and the inequality (3.4) will be called sharp if the infimum over all values of the parameter satisfies

$$\inf P[\text{region contains parameter}] = 1 - \alpha.$$

[We do not simply say that equality is attained in (3.4) for some p, since attainment of equality depends on whether open or closed confidence intervals are used and whether the parameter space is open or not. Use of the infimum avoids that difficulty.] The confidence region will be called conservative if (3.4) holds but is not sharp.

3.2.2 Intervals for λ_1, λ_+, and p. Sharp confidence intervals for λ_1 and λ_+ are completely standard since N_1 and N_+ are independent Poisson variables. Expressions for the lower and upper limits for λ_1 that are easy to compute from readily accessible tables, are

$$\lambda_{1L} = (\chi^2_{2n_1,\ \alpha/2})/2t$$

$$\lambda_{1U} = (\chi^2_{2(n_1+1),\ 1-\alpha/2})/2t\ .$$

The interval for λ_+ is of the same form, using n_+ instead of n_1. For details, see Johnson and Kotz[3], Vol. 1, Sec. 4.6.2.

A sharp confidence interval (p_L, p_U) for p can be based on the conditional distribution of S given N_+, if $N_+ > 0$. Let p_L and p_U satisfy

$$P\left[S \geq s \middle| N_+ = n_+, p = p_L\right] = \alpha/2 \tag{3.5}$$

$$P\left[S \leq s \middle| N_+ = n_+, p = p_U\right] = \alpha/2 \ . \tag{3.6}$$

Depending on whether s is large or small, it will be more convenient to rewrite one of these equations as

$$P\left[S < s \middle| N_+ = n_+, p = p_L\right] = 1 - \alpha/2 \tag{3.5'}$$

$$P\left[S > s \middle| N_+ = n_+, p = p_U\right] = 1 - \alpha/2 \ . \tag{3.6'}$$

This choice of the confidence interval has a mathematical statistics justification, given in Section 8.2.

To use (3.5) and (3.6), the conditional distribution of S given N_+ must be found. This involves some complexity, and more notation is needed. Let $v = (v_2, .., v_m)$ be any vector of nonnegative integers. Define

$$v_+ = \sum_{i=2}^{m} v_i .$$

Define the set T_k by

$$T_k = \left\{ v \mid v_+ = n_+, \ \sum_{i=2}^{m} i v_i = k \right\} .$$

Let

$$c(v) = \frac{v_+!}{v_2! \cdots v_m!} \prod_{i=2}^{m} \binom{m}{i}^{v_i} .$$

Then for $2n_+ \leq k \leq m\, n_+$,

$$P\left[S = k \mid N_+ = n_+\right] = \sum_{v \in T_k} P\left[N_2 = v_2, \ldots, N_m = v_m \mid N_+ = n_+\right]$$

$$= \frac{p^{m n_+ - k} q^{\quad}}{\left(1 - q^m - m\, p\, q^{m-1}\right)^{n_+}} \sum_{v \in T_k} c(v) \qquad (3.7)$$

from (2.6). The summation in (3.7) can be evaluated for the relevant values of k. Then (3.5) and (3.6) can be solved numerically for p_L and p_U.

The above procedure gives a sharp confidence interval for p, conditional on N_+ having any specified positive value. That is, the probability that the interval contains p, conditional on N_+, is $\geq 1 - \alpha$, and the inequality is sharp. There is no unconditional confidence level, because when $N_+ = 0$ no interval has been defined. To remedy this, when $N_+ = 0$ the interval [0,1] can be used with confidence level 1. Let $I(S,N_+)$ denote the interval for p, depending on S and N_+. Then the unconditional confidence level is

$$P\left[p \in I(S, N_+)\right]$$

$$= \sum_{n_+=0}^{\infty} P\left[p \in I(S,n_+) \mid N_+ = n_+\right] P\left[N_+ = n_+\right]$$

$$\geq \sum_{n_+=1}^{\infty} (1 - \alpha) P\left[N_+ = n_+\right] + 1\, P\left[N_+ = 0\right]$$

$$\geq (1 - \alpha) .$$

The first inequality results from the inexactness of the conditional confidence intervals.

Is the unconditional confidence interval sharp? For any fixed λ_+, $P\left[p \in I(S,N_+)\right]$ is strictly greater than $1-\alpha$. However, as $n_+ \to \infty$, the distribution of S given n_+ approaches a continuous distribution (S given n_+ is asymptotically normal, by the central limit theorem). So for any fixed p, $P\left[p \in I(S,n_+) \middle| N_+ = n_+\right] \to (1-\alpha)$ as $n_+ \to \infty$. Therefore $P\left[p \in I(S,N_+)\right] \to (1-\alpha)$ as $\lambda_+ t \to \infty$, so

$$\inf_{\lambda_+,p} P\left[p \in I(S,N_+)\right] = (1-\alpha).$$

In this sense, the unconditional confidence interval is sharp. This sense of sharpness may not be what the user really wants, since $\lambda_+ t$ may really be moderately small. It may perhaps be possible to shorten the conditional confidence intervals when n_+ is small, and still maintain an unconditional confidence level $(1-\alpha)$. However if this is possible, the details are certainly quite complicated.

3.2.3 Joint Region for $(\lambda_1, \lambda_+, p)$. Confidence intervals for λ_1, λ_+, and p have been given separately. Now consider how to find a confidence region for the three parameters taken together.

Let α_1, α_+, and α_p be between 0 and 1. Let $I_1(N_1)$ denote the $1-\alpha_1$ confidence interval for λ_1, and $I_+(N_+)$ the $1-\alpha_+$ confidence interval for λ_+. Let $I_p(S,N_+)$ be the $1-\alpha_p$ conditional confidence interval for p if $N_+ > 0$, and let it equal [0,1] if $N_+ = 0$. Let us tentatively use the product of the three intervals as a joint confidence region for the three parameters. Then

$$P\left[\lambda_1 \in I_1(N_1),\ \lambda_+ \in I_+(N_+),\ p \in I_p(S,N_+)\right]$$

$$= P\left[\lambda_1 \in I_1(N_1)\right] P\left[\lambda_+ \in I_+(N_+),\ p \in I_p(S,N_+)\right]$$

$$\geq (1-\alpha_1) \sum_{n_+=0}^{\infty} P\left[\lambda_+ \in I_+(n_+),\ p \in I_p(S,n_+) \middle| N_+ = n_+\right] P\left[N_+ = n_+\right].$$

The inequality results from the inexactness of I_1, and is sharp. Now let $\delta(\lambda_+, n_+) = 1$ if $\lambda_+ \in I_+(n_+)$, and let $\delta(\lambda_+, n_+) = 0$ otherwise. The last expression equals

$$(1-\alpha_1) \sum_{n_+=0}^{\infty} \delta(\lambda_+, n_+) \, P\left[p \in I_p\,(S, n_+) \,\middle|\, N_+ = n_+\right] P\left[N_+ = n_+\right]$$

$$\geq (1-\alpha_1) \sum_{n_+=1}^{\infty} \delta(\lambda_+, n_+) \, (1-\alpha_p) \, P\left[N_+ = n_+\right] + (1-\alpha_1) \, \delta(\lambda_+, 0) \, P\left[N_+ = 0\right]$$

$$\geq (1-\alpha_1) \, (1-\alpha_p) \, (1-\alpha_+) \, .$$

As $\lambda_+ t \to \infty$, the last two inequalities approach equality, so the confidence region is sharp.

In fact, this product of the three intervals may be larger than necessary because a portion of it may not satisfy the constraint (3.1). That portion can be deleted as impossible. To visualize this, think of the product of the three intervals as a block with rectagular cross sections. For any p, the cross section is a rectangle parallel to the (λ_1, λ_+) plane. Invocation of (3.1) typically cuts a corner off of the rectangle. Several cross sections for the example of Section 7 are shown in Figure 7.

3.2.4 Joint Regions for Any Two of λ_1, λ_+, and p. If we are interested in two of λ_1, λ_+, and p, the constraint (3.1) does not apply, so the region is simply the product of the two intervals. Work parallel to that of the last section shows that the unconditional confidence level is the product of the two confidence levels for the intervals.

3.2.5 Confidence Regions for Quantities Other than λ_1, λ_+, and p. If we are interested in three independent parameters other than λ_1, λ_+, and p, for example if we want a confidence region for (λ, μ, p), the region for $(\lambda_1, \lambda_+, p)$ can be rewritten in terms of the three desired parameters. The results are not especially neat, but they are straightforward.

It is not apparent how to get sharp confidence intervals for λ, μ, or μp^k, or a sharp two-dimensional region for, say, (λ, μ). One conservative region for (λ, μ) is the set of all (λ, μ) corresponding to any point in the three-dimensional region. Its confidence coefficient is at least that of the three-dimensional region.

A conservative region for μ is based on

$$\mu = \lambda_+ /(1 - q^m - m\, p\, q^{m-1}).$$

If (λ_+, p) is in a confidence region with some confidence level, then the resulting maximum and minimum values of μ form a confidence interval with at least that confidence level. Similarly, a conservative interval for μp^k is based on

$$\mu p^k = \lambda_+ p^k/(1 - q^m - m\, p\, q^{m-1}).$$

The right hand side is monotone in p for $k \geq 2$, so conservative upper and lower values for μp^k can be obtained from the upper and lower values of λ_+ and p.

A conservative one-sided interval for λ can be based on

$$\lambda \leq \lambda_1/m.$$

A conservative two-sided interval can be based on

$$\lambda = \lambda_1/m - \lambda_+\left[pq^{m-1}/(1 - q^m - m\, p\, q^{m-1})\right].$$

This last equation requires maximization or minimization over all three confidence intervals, and has coefficient at least $(1-\alpha_1)(1-\alpha_+)(1-\alpha_p)$. In the example of Section 7, the first method gives a smaller 95% confidence interval.

If the sample size is large, approximate confidence regions can be constructed using the Fisher information matrix. (A reference is Cox and Hinkley,[4] Sec. 9.2.iii.) The Fisher information for λ_1 or λ_+ is t/λ_1 or t/λ_+, respectively. To get the information for p, write the right hand side of (3.2) as

$$\frac{s}{pq} - m\, n_+ \frac{A}{B} .$$

Then

$$I(p) \equiv -E \frac{\partial^2}{\partial p^2} \log L$$

$$= - \frac{E(S)\ (p-q)}{p^2 q^2} + m\, n_+ \frac{B\, A' - A\, B'}{B^2}$$

$$= m\, n_+ \left[\frac{A(q-p)}{pqB} + \frac{B\, A' - A\, B'}{B^2} \right] . \qquad (3.8)$$

The off-diagonal elements of the information matrix are 0. So, if the sample size is large, $\hat{\lambda}_1$, $\hat{\lambda}_+$, and $\hat{p}$ are approximately independent normal with means λ_1, λ_+, and p, and variances λ_1/t, λ_+/t, and $1/I(p)$. From this, approximate confidence regions can be found for any subset of $\{\lambda_1, \lambda_+, p\}$. A conservative interval for λ, μ, or μp^k can be based on the intervals for λ_1, λ_+, and p.

4. BAYESIAN INFERENCE

4.1 Prior and Posterior Distributions

The constraint (3.1) makes the Bayesian problem inherently more complicated than the non-Bayesian one. Any prior distribution must be consistent with (3.1). The posterior distribution cannot be factored neatly as could the likelihood (2.4).

A suitable class of prior distributions must be selected. For greatest usefulness, this class should include the noninformative prior of Box and Tiao[5], Secs. 1.3.4-7. Box and Tiao argue that a noninformative prior for a parameter should be proportional to the square root of the Fisher information for the parameter. Therefore, the (improper) noninformative prior density for λ_1, resp. λ_+, is proportional to $\lambda_1^{-1/2}$, resp. $\lambda_+^{-1/2}$. These distributions are in the class of (improper) gamma distributions. In this paper, λ_1 and λ_2 will have gamma prior distributions, possibly improper.

The information for p is the complicated expression (3.8). Therefore, the noninformative prior density for p is proportional to $I^{1/2}(p)$, the square root of (3.8). It may be approximated in several ways by a beta distribution, as follows. The beta (1/2, 1/2) distribution would be noninformative if the common cause failures were not restricted so that only multiple failures are observable. This is a first approximation to the noninformative prior. Better approximations are obtained by adjusting the parameters of a beta (c,d) distribution. If c = 1/2, then $I^{1/2}(p)$ and the beta density will both be asymptotic to $p^{-1/2}$ for p near 0. Then d may be chosen so that the two distributions have the same mean. For a still better overall approximation, c and d may be chosen so that the two distributions have the same mean and the same variance. Tables 1 and 2 show the correct values c and d for these approximations, for many values of m. Figures 1 through 3 show the noninformative prior cumulative distribution and the three approximations, for m = 5, 20, and 100. In the example considered in Table 7, the third beta approximation appears adequate. In this paper it will be assumed that p has a beta prior distribution, possible one of these approximations to the noninformative distribution.

TABLE 1. BINOMIAL WITH 0 AND 1 TRUNCATED: APPROXIMATELY NONINFORMATIVE d FOR BETA(1/2,d)

m	d
3	0.2893
4	0.3148
5	0.3323
6	0.3455
7	0.3560
8	0.3645
9	0.3716
10	0.3777
12	0.3877
14	0.3956
16	0.4020
18	0.4073
20	0.4119
25	0.4209
30	0.4276
40	0.4372
60	0.4486
80	0.4555
100	0.4602
150	0.4676
200	0.4720
300	0.4773
∞	0.5000

TABLE 2. BINOMIAL WITH 0 AND 1 TRUNCATED: APPROXIMATELY NONINFORMATIVE PARAMETERS FOR BETA(c,d)

m	c	d
3	0.6386	0.3695
4	0.6641	0.4181
5	0.6762	0.4495
6	0.6813	0.4708
7	0.6824	0.4858
8	0.6814	0.4967
9	0.6791	0.5048
10	0.6762	0.5108
12	0.6695	0.5191
14	0.6626	0.5242
16	0.6560	0.5274
18	0.6498	0.5294
20	0.6442	0.5307
25	0.6321	0.5321
30	0.6223	0.5322
40	0.6075	0.5312
60	0.5888	0.5283
80	0.5772	0.5259
100	0.5692	0.5239
150	0.5567	0.5206
200	0.5492	0.5185
300	0.5404	0.5159
∞	0.5000	0.5000

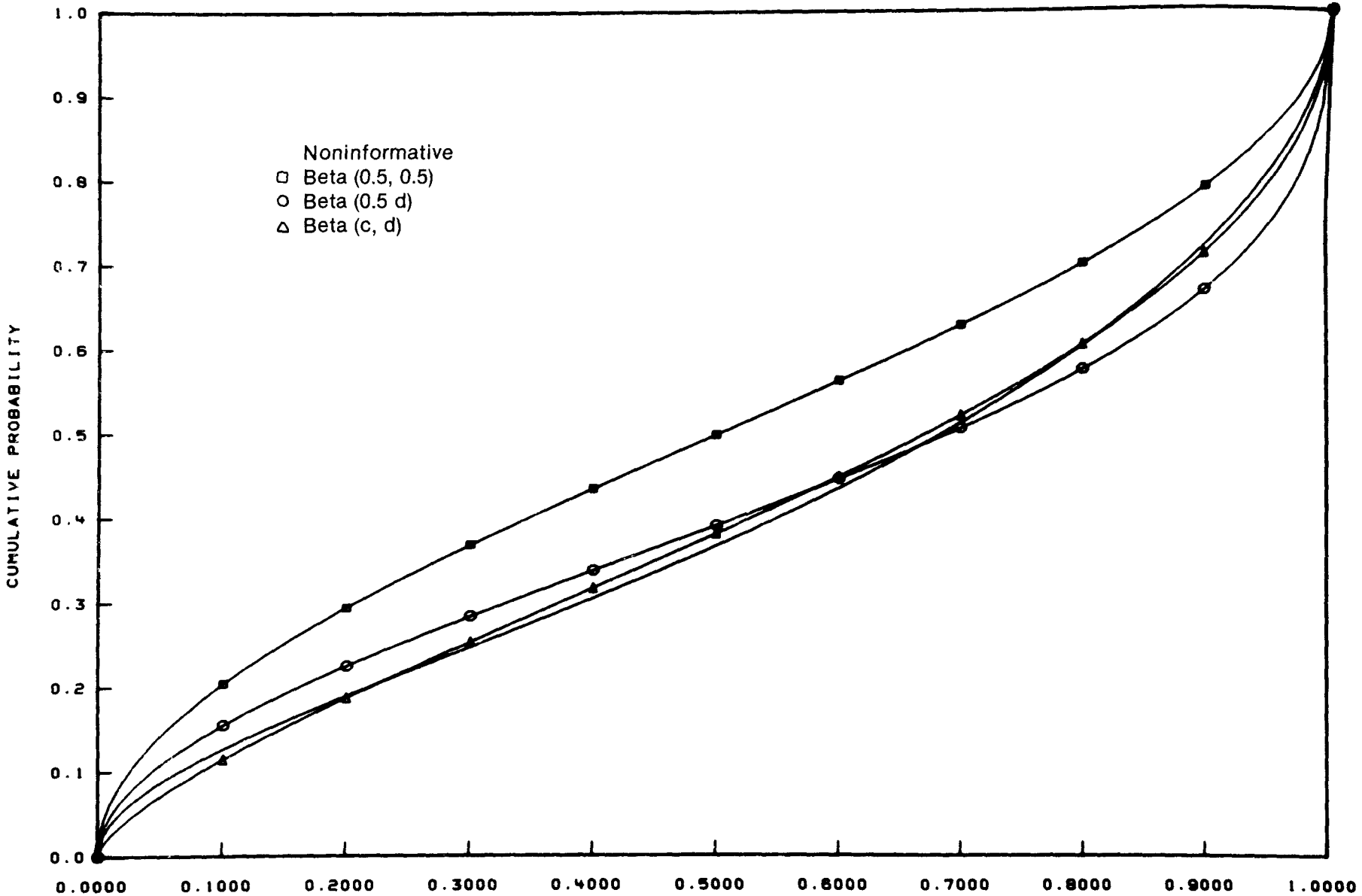


Figure 1. Cumulative distribution functions for noninformative prior and three beta approximations, 0 and 1 truncated, m = 5.

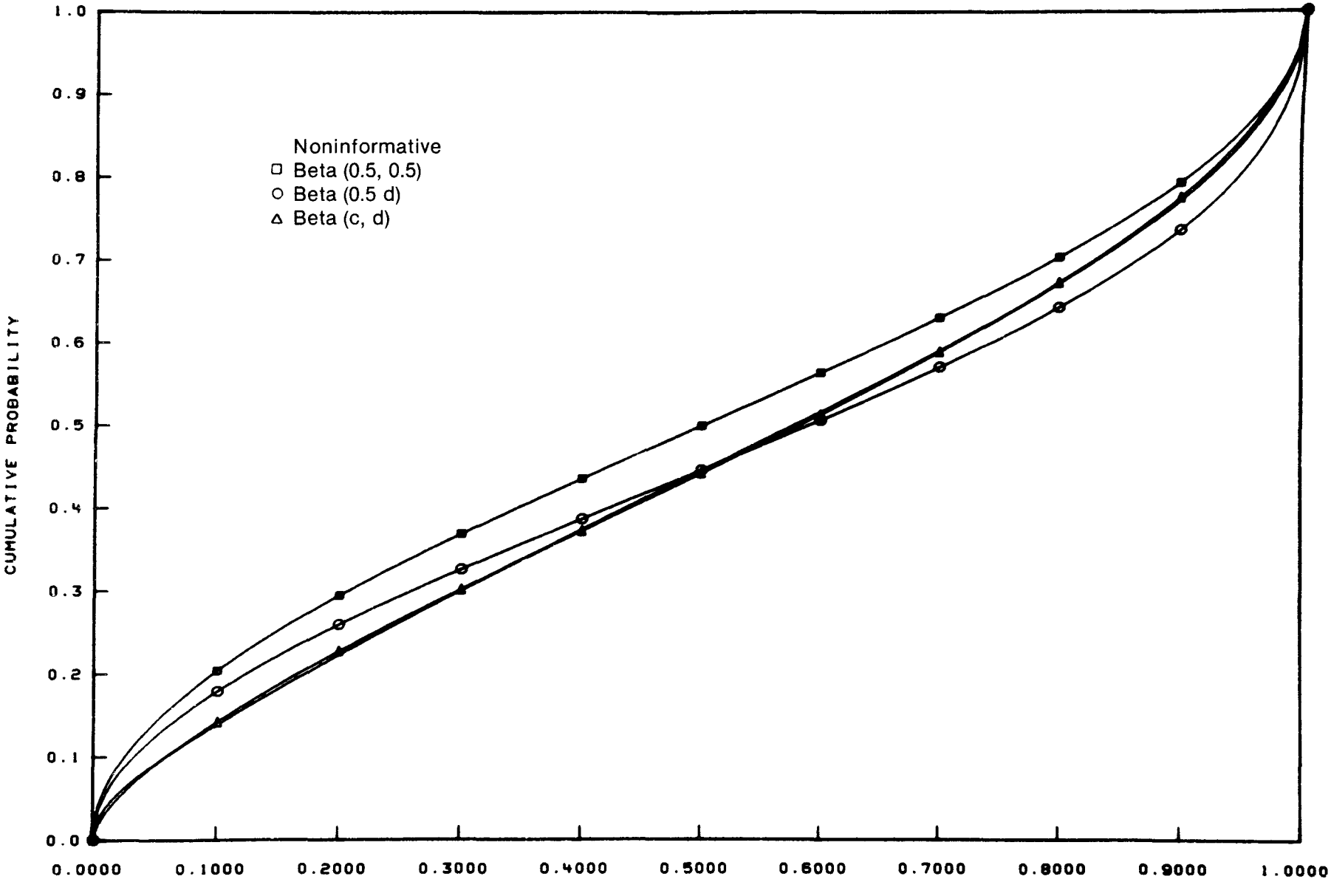


Figure 2. Cumulative distribution functions for noninformative prior and three beta approximations, 0 and 1 truncated, m = 20.

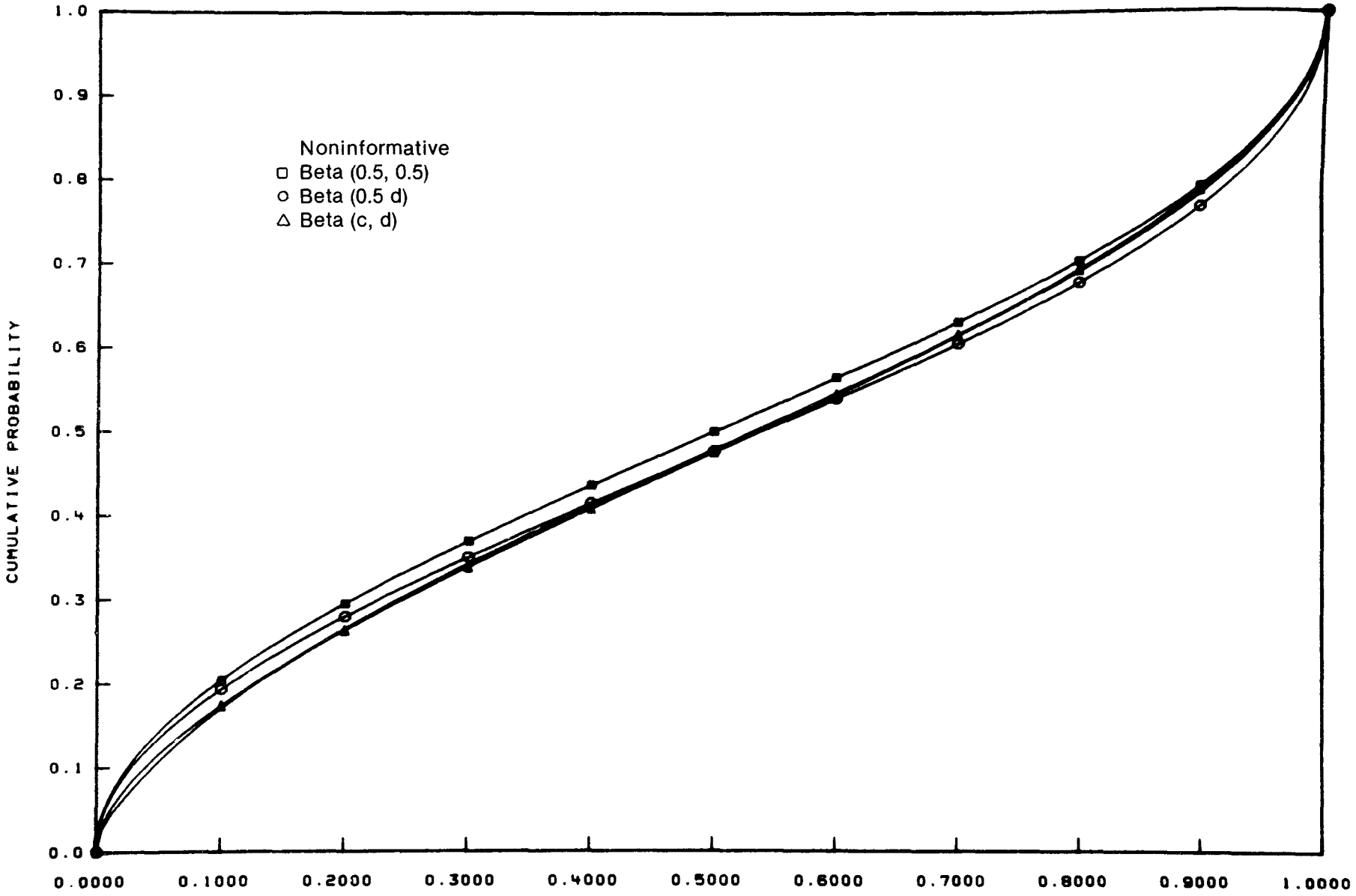


Figure 3. Cumulative distribution functions for noninformative prior and three beta approximations, 0 and 1 truncated, m = 100.

Since λ_1, λ_+, and p must satisfy the constraint (3.1), we will take the joint prior density to be proportional to the product of the individual prior densities in the region where (3.1) is satisfied. (See also Reference 5, pp. 56-58 and Section 1.5.) In summary, the following class of prior densities will be considered:

$$L(\lambda_1, \lambda_+, p)$$

$$= C\ U(\lambda_1, \lambda_+, p)\ e^{-b_1\lambda_1}\ \lambda_1^{a_1-1}\ e^{-b_+\lambda_+}\ \lambda_+^{b_+-1}\ p^{c-1}\ q^{d-1} \qquad (4.1)$$

where C is a constant, and $U(\lambda_1, \lambda_+, p)$ equals 1 where (3.1) holds, and equals 0 elswhere.

The parameters should be restricted as follows: require $a_1 > 0$ and $a_+ > 0$ in order to guarantee that the posterior density, given by (4.2) below, has finite integral for $n_1 \geq 0$ and $n_+ \geq 0$. Require $b_1 \geq 0$ and $b_+ \geq 0$, to force the posterior density to have finite integral for all $t > 0$. Require $c > 0$ and $d > 0$. For, if $c \leq 0$ and $s = 2n_+ + 1$, then the posterior density would be positive at $p = 0$, and if $d \leq 0$ and $s = mn_+ - 1$, the posterior density would be positive at $p = 1$. Either of these possibilities defies any reasonable interpretation of a posterior density, so restrict c and d to prevent it.

If $a_1 = a_+ = 1/2$ and $b_1 = b_+ = 0$, the prior distribution is noninformative for λ_1 and λ_+.

The posterior density of $(\lambda_1, \lambda_+, p)$ is obtained by multiplying (2.4) by (4.1) and adjusting the constant. It will be convenient to write it as follows:

$$L(\lambda_1, \lambda_+, p|\text{data}) = C\ U(\lambda_1, \lambda_+, p)\ V(\lambda_1, \lambda_+, p) \qquad (4.2)$$

where

$$V(\lambda_1, \lambda_+, p) = \frac{(b_1+t)^{a_1+n_1}}{\Gamma(a_1+n_1)} e^{-(b_1+t)\lambda_1} \lambda_1^{a_1+n_1-1}$$

$$\cdot \frac{(b_+ +t)^{a_+ +n_+}}{\Gamma(a_+ +n_+)} e^{-(b_+ +t)\lambda_+} \lambda_+^{a_+ +n_+ -1}$$

$$\cdot \frac{p^{s+c-1} q^{mn_+ -s+d-1}}{\Gamma(1-q^m-mpq^{m-1})^{n_+}} . \qquad (4.3)$$

Here C is that constant such that $\iiint L(\lambda_1, \lambda_+, p \text{ data}) = 1$.

4.2 Bayes Point Estimators

4.2.1 Posterior Modes. One estimate of $(\lambda_1, \lambda_+, p)$ is the mode of the posterior density. If the constraint (3.1) is ignored, the maximizing values for λ_1 and λ_+ are the maximizing values for (4.3),

$$\lambda_1 = (a_1 + n_1 - 1)/(b_1 + t)$$

$$\lambda_+ = (a_+ + n_+ - 1)/(b_+ + t).$$

To find the maximizing p, compute

$$q \frac{\partial}{\partial p} \log L(\lambda_1, \lambda_+, p|\text{data}) = \frac{s + c - 1}{p} - \left\{ c + d - 2 + mn_+ \frac{1-q^{m-1}}{1-q^m-mpq^{m-1}} \right\}.$$

It is shown in Section 8.1 that this quantity either is always positive, or is always negative, or else is decreasing and equals zero at just one point. The possible cases are tabulated in Section 8.1. Therefore, there is a unique posterior mode p. The corresponding values of λ, μ, and μp^k follow from the defining relations (2.1) through (2.3).

Suppose now that the above values of λ_1, λ_+, and p do not satisfy constraint (3.1). Expression (4.3) is unimodal, so a maximum of (4.2) must occur where equality occurs in (3.1). Substitute $\lambda_1=\lambda_+ z_1$ in (4.2) and maximize the result with respect to λ_+, say at λ_+^*. Then

$$L(\lambda_+^* z_1, \lambda_+^*, p|\text{data})$$

$$= C \frac{z_1^{a_1+n_1-1}}{\left[(b_1+t)z_1+(b_+ +t)\right]^{a_1+n_1+a_+ +n_+-2}} \frac{p^{s+c-1} q^{mn_+-s+d-1}}{(1-q^m- mpq^{m-1})^{n_+}}$$

$$= C \frac{p^{s+a_1+n_1+c-2} q^{m(n_+ +n_1+a_1-1)+d-1-(s+a_1+n_1-1)}}{\left[(b_1+t)z_1+(b_+ +t)\right]^{a_1+n_1+a_+ +n_+-2} (1-q^m-mpq^{m-1})^{n_+ +n_1+a_1-1}} .$$

Here C is some constant. This must be maximized with respect to p. If $b_1 = b_+$, the expression simplifies to the following, with $s'=s+a_1+n_1-1$, $n'=n_+ +n_1+a_1-1$:

$$C \frac{p^{s'+c-1} q^{mn'-s'+d-1}}{(1-q^m)^{n'}} \left[\frac{1-q^m}{1-q^m-mpq^{m-1}}\right]^{a_+-1} .$$

This must be maximized numerically with respect to p. (Note, if $c=d=a_1=a_+=1$, then this was the expression which was maximized to get the MLE.)

With this p used to calculate z_1, the maximizing λ_+ is $\lambda_+ = (a_1 + a_+ + n_1 + n_+ - 2)/\left[(b_1 + t)z_1 + (b_+ + t)\right]$, and then $\lambda_1 = \lambda_+ z_1$.

4.2.1 Posterior Means. A more difficult estimate to obtain is the triple of posterior means. The difficulty arises from having to integrate (4.3) over the region satisfying (3.1) rather than over the region $\lambda_1 \geq 0$, $\lambda_+ \geq 0$, $0 \leq p \leq 1$.

The integrals needed are usually of the following forms. For $A > 0$ and $B > 0$, define

$$I(A,B,x) = \frac{B^A}{\Gamma(A)} \int_0^x e^{-Bt}\ t^{A-1}\ dt \qquad (4.4)$$

the gamma(A,B) cumulative distribution function. Note that $I(A,B,\infty) = 1$ for any A and B. Now assume that A, B, A', and B' are all positive, and define

$$J(A,B,A',B',x) = \int_0^x \int_s^\infty \frac{B^A}{\Gamma(A)} e^{-Bt}\ t^{A-1}\ \frac{(B')^{A'}}{\Gamma(A')} e^{-B's}\ s^{A'-1}\ dt\ ds. \qquad (4.5)$$

In Section 8.3, formulas are given for evaluating I(A,B,x) and J(A,B,A',B',x) as finite sums or infinite series. The formulas are simplest if A and A' are integers, somewhat harder to evaluate if A and A' are integers plus 1/2, and hardest to evaulate if A and A' are arbitrary. In practice, the formulas for evaluating I and J should be written into computer subprograms.

To express the results compactly, let us also define

$$f(p) = C_0 \frac{p^D q^E}{(1-q^m-mpq^{m-1})^F} \qquad (4.6)$$

where

$D = s+c-1$

$E = mn_+-s+d-1$

$F = n_+$

C_0 = normalizing constant.

Let

$A_1 = a_1+n_1$

$A_+ = a_++n_+$

$$B_1 = (b_1+t)z_1$$

$$B_+ = b_++t \ .$$

With this notation defined, the results are easily expressed. Let us begin by obtaining $L(p|\text{data})$. This is done by integrating (4.2), first with respect to λ_1 (making the change of variables $u = \lambda_1/z_1$), and then with respect to λ_+. Integration is over the region satisfying (3.1). Then

$$L(p|\text{data}) = f(p)\ J(A_1,\ B_1,\ A_+,\ B_+,\ \infty). \tag{4.7}$$

The quantity J depends on p through B_1. The constant C_0 in f(p) is such that $\int L(p|\text{data})\ dp = 1$. It must be found numerically.

Once C_0 is found, $p\ L(p|\text{data})$ can be integrated numerically to get the posterior mean of p. To find the mean of λ_1, λ_+, μ, or μp^k, multiply (4.2) by the appropriate quantity and integrate. Integration with respect to λ_1 and λ_+ resembles the integration to obtain (4.7), and gives an expression of the form

$$C_1\ f(p)\ J(A_1,\ B_1,\ A_+,\ B_+,\ \infty) \tag{4.8}$$

with redefined values for some of the constants. The changes in the constants are

for λ_1 $\quad A_1 = a_1 + n_1 + 1$

for λ_+ $\quad A_+ = a_+ + n_+ + 1$

for μ $\quad A_+ = a_+ + n_+ + 1, \quad F = n_+ + 1$

for μp^k $\quad A_+ = a_+ + n_+ + 1, \quad D = s+c-1+k, \quad F = n_+ + 1.$

The multiplicative constant C_1 is $(a_1+n_1)/(b_1+t)$ for λ_1, and $(a_++n_+)/B_+$ for λ_+, μ, and μp^k. The constant C_0 remains the same throughout, the

value needed to make $\int L(p|\text{data})dp = 1$. The desired posterior mean is found by integrating (4.8) numerically with respect to p.

The posterior mean of λ is given by

$$E\left[\lambda_1/m\right] - E\left[\lambda_+ \; p \; q^{m-1}/(1 - q^m - m \; p \; q^{m-1})\right].$$

$E \; \lambda_1$ has just been found. The second term is found in the same way, using

$$A_+ = a_+ + n_+ + 1, \quad D = s + c,$$

$$E = mn_+ - s + d + m - 2, \quad F = n_+ + 1,$$

and multiplicative factor $C_1 = (\alpha_+ + n_+)/B_+$.

4.3 Bayes Probability Regions

4.3.1 Joint Region for Three Parameters. A three-dimensional region R is sought such that $P\left[(\lambda_1, \lambda_+, p) \in R|\text{data}\right]$ equals some specified value. There are many such regions R. One approach is to try to treat the three parameters one at a time, so that R will, to some extent, resemble the confidence region of Section 3. Another approach, following Box and Tiao[5], Sec. 2.8, is to choose R such that the posterior density is greater at any point inside R than at any point outside R. This is called the highest posterior density (HPD) region. While this second approach is natural in principle, it requires numerical triple integration, so is awkward in practice. Moreover, if such a region were found it would be shaped like a highly distorted ellipsoid and so would be difficult to describe. In particular, it would not be defined by any simple set of equations. For these reasons, the first approach is the only one which will be followed for a three-dimensional region.

The posterior density of p is given by (4.7), so for any desired α_p between 0 and 1, an interval I_p can be found numerically such that $P\left[p \in I_p|\text{data}\right] = 1 - \alpha_p$. This interval may be the interval with equal tail probabilities, or the HPD interval for p, or another interval.

Now use

$$L(\lambda_+|p, \text{data}) = \int L(\lambda_1, \lambda_+, p|\text{data})\, d\lambda_1 / L(p|\text{data})$$

integrating over λ_1 satisfying (3.1), to obtain

$$L(\lambda_+|p, \text{data}) = \frac{B_+^{A_+}}{\Gamma(A_+)} e^{-B_+\lambda_+} \lambda_+^{A_+-1} \left[1 - I(A_1, B_1, \lambda_+)\right] / J(A_1, B_1, A_+, B_+, \infty).$$

The constants A_1, B_1, A_+, and B_+ are as defined below (4.6). Therefore, for any p, an interval $I_+(p)$ can be found numerically such that $P\left[\lambda_+ \in I_+(p)|p, \text{data}\right] = 1 - \alpha_+$. This interval may be such that there is equal probability (conditional on p and the data) that λ_+ is on each side of $I_+(p)$. Or, the interval may be the conditional analogue of an HPD interval, chosen now so that $L(\lambda_+|p, \text{data})$ is highest in $I_+(p)$. Or the interval may be chosen in some other way.

Finally, from (4.2), the conditional density of λ_1 is

$$L(\lambda_1|\lambda_+, p, \text{data}) = \frac{(b_1+t)^{A_1}}{\Gamma(A_1)} e^{-\lambda_1(b_1+t)} \lambda_1^{A_1-1} / \left[1 - I(A_1, B_1, \lambda_+)\right]$$

for $\lambda_1 \geq \lambda_+ z_1$.

So an interval $I_1(\lambda_+, p)$ can be found such that
$P\left[\lambda_1 \in I_1(\lambda_+, p)|\lambda_+, p, \text{data}\right] = 1 - \alpha_1$.

The resulting three-dimensional region has posterior probability $(1-\alpha_1)\ (1-\alpha_+)\ (1-\alpha_p)$.

If desired, this three-dimensional region can be re-expressed as a region for (λ, μ, p).

4.3.2 Intervals for Single Parameters. Let us find a probability interval for μp^k. For any $c \geq 0$,

$$P\left[\mu p^k \leq c \middle| \text{data}\right] = P\left[\frac{\lambda_+ p^k}{(1-r_0-r_1)} \leq c \middle| \text{data}\right]$$

$$= P\left[\lambda_1 \geq z_1 \lambda_+,\ \lambda_+ \leq c(1-r_0-r_1)/p^k,\ 0 < p < 1 \middle| \text{data}\right].$$

Denote $(1-r_0-r_1)/p^k$ by $u(p)$. Make the change of variables $t = \lambda_1/z_1$. Then, using the notation (4.5) and (4.6), the above probability equals

$$\int_0^1 J\left[A_1,\ B_1,\ A_+,\ B_+,\ cu(p)\right] f(p)\ dp. \tag{4.9}$$

The expression J may be evaluated using (8.10) through (8.13). Even if the evaluation requires summing an infinite series, this is probably faster than performing a double or triple integral numerically. So an interval $[c,d]$ can be found numerically such that $P\left[c \leq \mu p^k \leq d \middle| \text{data}\right]$ equals any desired value $1-\alpha$ between 0 and 1. The interval can be chosen so that

$$P\left[\mu p^k < c \middle| \text{data}\right] = P\left[\mu p^k > d \middle| \text{data}\right] = \alpha/2.$$

This gives the interval with equal tail probabilities. If instead the HPD interval is desired, the posterior density of μp^k must be found by differentiating (4.9) with respect to c. By the definitions (4.5) and (4.4),

$$L_{\mu p^k}(c|\text{data}) = \int_0^1 \frac{\partial}{\partial c} J\left[A_1,\ B_1,\ A_+,\ B_+,\ cu(p)\right] f(p)$$

$$= \int_0^1 \frac{B_+^{A_+}}{\Gamma(A_+)} e^{-B_+ cu(p)} [cu(p)]^{A_+} \left[1 - I(A_1,\ B_1,\ cu(p))\right] u(p)\ f(p)\ dp.$$

Evaluation of this density requires use of one of (8.5) through (8.7) followed by numerical integration, but when this is done the HPD interval for μp^k can be found.

An interval for μ follows from the above by setting $k = 0$. An interval for λ_+ is obtained by replacing $u(p)$ by 1 in (4.9). An interval for λ_1 can be found by using

$$P\left[\lambda_1 \leq c \,|\, \text{data}\right]$$

$$= P\left[z_1 \lambda_+ \leq \lambda_1 \leq c,\ 0 < p < 1 \,|\, \text{data}\right]$$

$$= P\left[\lambda_+ \leq c/z_1 \,\middle|\, \text{data}\right]$$

$$- P\left[\lambda_1 > c,\ \lambda_+ \leq c/z_1,\ 0 < p < 1 \,|\, \text{data}\right]$$

$$= \int_0^1 J(A_1, B_1, A_+, B_+, c)\, f(p)\, dp$$

$$- \int_0^1 \left[1 - I(A_1, B_1, c/z_1)\right] I(A_+, B_+, c/z_1)\, f(p)\, dp.$$

Numerical integration can be used to evaluate this.

An interval for λ can be found using

$$P[\lambda \leq c \,|\, \text{data}]$$

$$= P\left[\lambda_1 - \lambda_+ z_1 \leq mc \,\middle|\, \text{data}\right]$$

$$= 1 - P\left[\lambda_1/z_1 > \lambda_+ + mc/z_1,\ 0 < p < 1 \,\middle|\, \text{data}\right]$$

$$= 1 - \int_0^1 \int_0^\infty \left[1 - I(A_1, B_1, \lambda_+ + mc/z_1)\right] \frac{B_+^{A_+}}{\Gamma(A_+)} e^{-B_+ \lambda_+} \lambda_+^{A_+ - 1} f(p)\, d\lambda_+\, dp.$$

This must be evaluated by numerical double integration.

Two-dimensional probability regions for pairs of the parameters will not be given.

5. DIAGNOSTIC CHECKS ON THE MODEL

5.1 Poisson Parameters

Suppose that there are I "sources" of data. These sources could be plants, vendors, time periods, etc. Let us investigate whether the sources all have the same values of λ_1 or λ_+. The methods below do not use the binomial distribution assumptions at all. The only assumption is that the single failures and multiple failures form two Poisson processes.

5.1.1 Graphical Methods. Denote the λ_+ corresponding to the ith source by λ_{+i}. Suppose that all the sources but the ith have a common λ_+, denoted $\lambda_{+,-i}$. Let N_{+i} and t_i be the number of multiple failures and the operating time for the ith source.

To test the null hypothesis $\lambda_{+i} = \lambda_{+,-i}$, a uniformly most powerful similar test is based on N_{+i} given N_+. The conditional distribution of $N_{+i} \mid N_+ = n_+$ is binomial$(n_+, t_i/t)$ under the null hypothesis. (These assertions are all shown in Cox and Hinkley[4], pp. 136-7.)

One can therefore examine either residuals or the corresponding significance levels. The ith standardized residual is defined as

$$R_i = \frac{N_{+i} - n_+ t_i/t}{\left[n_+ \frac{t_i}{t}\left(1 - \frac{t_i}{t}\right)\right]^{1/2}} .$$

If λ_{+i} is the same for all i, then for a given n_+, the R_i's all have mean 0 and variance 1. Define the ith significance level α_i as

$$\text{Min}\left(2P\left[N_{+i} \geq n_{+i} \mid N_+ = n_+\right],\ 2P\left[N_{+i} \leq n_{+i} \mid N_+ = n_+\right],\ 1\right).$$

That is, the ith significance level is $2P\left[N_{+i} \geq n_{+i} \mid N_+ = n_+\right]$ if the observed n_{+i} is greater than the median, and it is similarly defined if n_{+i} is less than the median or equal to the median. The tail probability

is doubled to reflect the fact that N_{+i} could be, a priori, either large or small.

Outliers correspond to large values of R_i or to small values of α_i. The standardized residuals can be graphed, but they have a skewed distribution, so a glance at the plot will not necessarily pick out the values which are significantly large or small. For this, the individual significance levels are preferable.

These methods can be used to identify sources for which λ_{+i} seems to be substantially different from the others. They cannot be used for an overall test of whether all the λ_{+i}'s are equal, for if the number of sources is large, then random variability alone will produce some apparently extreme values of N_{+i}. Testing is discussed below.

For investigating λ_1, exactly the same techniques work, replacing N_{+i} and n_+ by N_{1i} and n_1.

5.1.2 Hypothesis Tests. Let us test

H_0: $N_{+i} \sim \text{Poisson}(\lambda_+ t_i)$, $i = 1,..,I$

H_1: $N_{+i} \sim \text{Poisson}(\lambda_{+i})$, $i = 1,..,I$, with no relation among the λ_{+i}'s.

If the sample size n_+ is large, the generalized likelihood ratio test can be used. (See Mood, Graybill, and Boes[6], Sec. IX.5.1.) The test statistic is

$$-2 \log \Lambda = 2 \sum_{i=1}^{I} N_{+i} \log\left(\frac{N_{+i}}{N_+} \frac{t}{t_i}\right). \qquad (5.1)$$

The null hypothesis should be rejected if (5.1) is greater than the $1-\alpha$ point of a χ^2 distribution with $I-1$ degrees of freedom.

This is exactly the same test as would be used to test

$$H_0: \ (N_{+1},..,N_{+I}) \big| n_+ \sim \text{multinomial}(n_+, \ t_1/t,..,t_I/t). \tag{5.2}$$

If n_+ is not large, the test statistic (5.1) can still be used, but since λ_+ is unknown the critical point should be derived from the multinomial distribution (5.2). The exact distribution can be calculated for small I and n_+ by using the program POLYPOW, presented by Atwood[7]. This works because the multinomial probabilities are the terms in the algebraic expansion of

$$(t_1/t + \ . \ . \ . \ + t_I/t)^{n_+} \ ,$$

and POLYPOW finds these terms.

The X^2 test statistic

$$\sum \frac{(N_{+i} - n_+ t_i \ /t)^2}{n_+ t_i/t}$$

could be used instead of the likelihood ratio test statistic.

For studying λ_1, simply replace N_{+i} and N_+ by N_{1i} and N_1.

5.2 Binomial Parameter

Suppose again that there are I sources of data, which may be plants, etc., as before, but may now also be the n_+ individual observations of multiple failures. Let us investigate whether the I sources have the same p, and whether the data as a whole seem consistent with the assumption of a binomial distribution.

5.2.1 Graphical Methods. Let the random variable X be the number of failed components in a multiple failure. (We have been assuming that X is truncated binomial(m,p), truncated so that $X \geq 2$.) Let S_i denote S (i.e., the total number of failed components in multiple failures) based on observations from only the ith source. So, conditional on n_{+i}, S_i is a sum of n_{+i} independent observations of X. Denote $S - S_i$ by S_{-i}. Then, conditional on n_{+i} and $n_{+,-i}$, $S_i/n_{+i} - S_{-i}/n_{+,-i}$ has expectation 0 and variance

$$\left[1/n_{+i} + 1/n_{+,-i}\right] \text{ var } X.$$

Define

$$R_i = \left[S_i/n_{+i} - S_{-i}/n_{+,-i}\right] / \left[1/n_{+i} + 1/n_{+,-i}\right]^{1/2} . \qquad (5.3)$$

If all sources correspond to the same distribution of X (in particular, if all sources obey the BFR assumptions with the same p), then, conditional on the n_{+i}'s, the R_i are identically distributed with mean 0 and variance = var X. If X is truncated binomial(m,p) with $X \geq 2$, then direct calculation yields

$$EX = \frac{m\,p(1 - q^{m-1})}{1 - q^m - mpq^{m-1}}$$

$$EX(X-1) = \frac{m\,(m-1)\,p^2}{1 - q^m - mpq^{m-1}}$$

and the variance of X is

$$\sigma_X^2 = EX(X-1) - EX(EX-1).$$

This may be estimated by substituting some estimate for p. The standardized residuals are then obtained by dividing R_i by $\hat{\sigma}_X$, the estimate of σ_X.

They are not independent, but a plot of i versus $R_i/\hat{\sigma}_X$ will help point out sources with unusually high or low p.

The above method was used to identify sources with high or low numbers of failures. The following method investigates whether the distribution within a source is really truncated binomial. To simplify the notation, assume that there is a single data source, with n_+ multiple failure occurrences. For $2 \leq i \leq m$, conditional on n_+, the number of instances of i failures, N_i, is binomial(n_+,z_i), with z_i defined by (2.5) and (2.2). Let $\hat{z}_i$ be the estimate based on the MLE $\hat{p}$, i.e.,

$$\hat{z}_i = \binom{m}{i} \hat{p}^i \hat{q}^{m-i}/(1 - \hat{q}^m - m\hat{p}\hat{q}^{m-1}) .$$

Then

$$U_i = \frac{N_i - n_+ \hat{z}_i}{\left[n_+ \hat{z}_i(1 - \hat{z}_i)\right]^{1/2}} \tag{5.4}$$

has mean and variance approximately 0 and 1. Large or small values, or strong patterns, indicate that the distribution is not truncated binomial.

5.2.2 Hypothesis Tests. Let us test the null hypothesis that X is a truncated binomial(m,p) random variable truncated such that $X \geq 2$. Under this hypothesis, $P[X = i] = z_i$, in the notation of (2.5), for i=2,..,m.

Possible alternate hypotheses are that p varies from source to source, and that the distribution of X is the same for all sources but is not truncated binomial(m,p).

If the sample size is large, the generalized likelihood ratio test can be used. For testing against the alternate hypothesis that the sources are BFR but with possibly different p's, the test statistic is

$$-2 \log \Lambda = -2 L_o + 2 \sum_{i=1}^{I} L_i$$

where

$$L_0 = s \log (\hat{p}/\hat{q}) + n_+ \log \left[\hat{q}^m/(1 - \hat{q}^m - m\,\hat{p}\,\hat{q}^{m-1})\right] \tag{5.5}$$

and

$$L_i = s_i \log (\hat{p}_i/\hat{q}_i) + n_{+i} \log \left[\hat{q}_i{}^m/(1 - \hat{q}_i{}^m - m\,\hat{p}_i\,\hat{q}_i{}^{m-1})\right]. \tag{5.6}$$

Here $\hat{p}$ is the MLE based on all the data, and $\hat{p}_i$ is the MLE based on only the ith source. Under H_0, $-2 \log \Lambda$ is asymptotically $\chi^2(I-1)$.

To test against the alternate hypothesis that X is not truncated binomial(m,p), the test statistic is

$$-2 \log \Lambda = 2 \sum_{i=2}^{m} n_i \log (n_i/(n_+ \hat{z}_i))$$

where $\hat{z}_i$ was defined above (5.4). Under H_0, the asymptotic distribution of $-2 \log \Lambda$ is $\chi^2(m-4)$. Another possible test in this case is the ordinary chi-square goodness of fit test. The test statistic is

$$\sum_{i=2}^{m} (n_i - n_+\hat{z}_i)^2/n_+\hat{z}_i$$

and its asymptotic distribution under H_0 is $\chi^2(m-4)$. If the sample size is only moderate rather than large, then the cells can be grouped and the degrees of freedom adjusted in the usual way.

6. RESULTS IF CAUSES OF SINGLE FAILURES ARE IDENTIFIABLE

6.1 Definitions

The data consist of $(n_I, n_C, n_2,.., n_m)$, where n_I is the number of single failures that were not due to common cause shocks, and n_C is the number of single failures due to common cause shocks. Now m may be as small as 2. Define

$$n_+' = n_C + n_+$$

$$\lambda_+' = \mu r_1 + \lambda_+ = \mu(1 - r_o)$$

$$s' = n_C + s$$

$$z_i' = r_i/(1 - r_o) \quad \text{for } i = 1,.., m.$$

Then

$$P\left[N_I = n_I, N_C = n_C, N_2 = n_2,.., N_m = n_m\right]$$
$$= P\left[N_I = n_I\right] P\left[N_+' = n_+'\right] P\left[N_C = n_C, N_2 = n_2,.., N_m = n_m | N_+' = n_+'\right]$$

where

$$N_I \sim \text{Poisson}(\lambda mt)$$

$$N_+' \sim \text{Poisson}(\lambda_+' t)$$

$$N_C,.., N_m | N_+' = n_+' \sim \text{multinomial}(n_+', z_1',.., z_m').$$

Inference follows the pattern of Sections 3 through 5, but is much simpler because the most easily estimable parameters, λ, λ_+', and p, are not constrained by any analogue of (3.1).

6.2 Non-Bayesian Inference

The Poisson parameters λ and λ_+' can be estimated, both by MLE's and by confidence intervals, in the standard way. The MLE for p is the solution of

$$s' = m\, n_+'\, p/(1 - q^m).$$

[In the special case m = 2, the solution becomes simply $\hat{p} = 2(s'-n_+')/s'$.] Maximum likelihood estimates for any other parameters--λ_1, λ_+, μ, or μp^k--follow from substitution of λ, λ_+, and p into the appropriate defining equations involving the desired quantities.

If $n_+' > 0$, a confidence interval for p can be based on the conditional distribution of S' given N_+', using analogues of (3.5) and (3.6). This conditional distribution is now given. Let $v = (v_1,.., v_m)$ be any vector of nonnegative integers. Define

$$v_+' = \sum_{i=1}^{m} v_i$$

Define the set T_k by

$$T_k = \left\{ v \,\middle|\, v_+' = n_+',\ \sum_{i=1}^{m} i v_i = k \right\}.$$

Let

$$c(v) = \frac{v_+'!}{v_1'!\cdots v_m'!} \prod_{i=1}^{m} \binom{m}{i}^{v_i}.$$

Then, for $n_+ \le k \le m\, n_+$,

$$P\left[S' = k \middle| N_+' = n_+'\right] = \sum_{v \in T_k} P\left[N_C = v_1,\ N_2 = v_2, \ldots,\ N_m = v_m \ N_+' = n_+'\right]$$

$$= \frac{p^{s'} q^{mn_+' - s'}}{(1 - q^m)^{n_+'}} \sum_{v \in T_k} c(v) \ .$$

If the interval [0,1] is used when $N_+' = 0$, then the procedure has unconditional confidence level $(1-\alpha)$.

A joint confidence region for more than one of λ, λ_+', and p is simply the product of the confidence intervals. Conservative intervals for μ and μp^k can be based on the intervals for λ_+' and p in the obvious ways. For λ_1 and λ_+, conservative intervals can be based on the intervals for λ, λ_+', and p. Alternatively, the confidence intervals of Section 3.2 can still be used for λ_1 and λ_+.

If n_I and n_+' are large, then $\hat{\lambda}$, $\hat{\lambda}_+'$, and $\hat{p}$ are asymptotically independent normal with means λ, λ_+', and p, and with variances λ/mt, λ_+'/t and $1/I(p)$, where

$$I(p) = \frac{mn_+'}{pq} \frac{1 - q^m - mpq^{m-1}}{(1 - q^m)^2} \ . \tag{6.1}$$

Approximate confidence intervals can be based on this asymptotic distribution.

6.3 Bayesian Inference

We will consider the class of prior distributions: $\lambda \sim \Gamma(a,b)$, $\lambda_+' \sim \Gamma(a_+', b_+')$, $p \sim \text{beta}(c,d)$, with λ, λ_+' and p independent. If $a=a_+'=1/2$, $b=b_+'=0$, then the (improper) priors are noninformative for λ and λ_+'. The noninformative prior advocated by Box and Tiao[5] for p is proportional to $I^{1/2}(p)$, with I(p) given in (6.1). If $c=d=1/2$, the beta(c,d) distribution approximates the noninformative prior. Tables 3 and 4 give values of c and d that provide better approximations, and Figures 4 through 6 show these cumulative distributions for m = 5, 20, and 100.

TABLE 3. BINOMIAL WITH 0 TRUNCATED: APPROXIMATELY NONINFORMATIVE d FOR BETA(1/2, d)

m	d
2	0.3541
3	0.3776
4	0.3923
5	0.4027
6	0.4105
7	0.4167
8	0.4217
9	0.4260
10	0.4296
12	0.4355
14	0.4401
16	0.4439
18	0.4470
20	0.4496
25	0.4549
30	0.4588
40	0.4642
60	0.4708
80	0.4748
100	0.4775
150	0.4817
200	0.4842
300	0.4872
∞	0.5000

TABLE 4. BINOMIAL WITH 0 TRUNCATED: APPROXIMATELY NONINFORMATIVE PARAMETERS FOR BETA(d,0)

m	c	d
2	0.5902	0.4179
3	0.6051	0.4569
4	0.6101	0.4786
5	0.6108	0.4919
6	0.6096	0.5004
7	0.6075	0.5063
8	0.6050	0.5103
9	0.6024	0.5132
10	0.5998	0.5154
12	0.5949	0.5181
14	0.5904	0.5197
16	0.5864	0.5205
18	0.5828	0.5210
20	0.5795	0.5212
25	0.5728	0.5211
30	0.5675	0.5207
40	0.5597	0.5196
60	0.5497	0.5177
80	0.5436	0.5162
100	0.5394	0.5151
150	0.5327	0.5132
200	0.5287	0.5120
300	0.5240	0.5106
∞	0.5000	0.5000

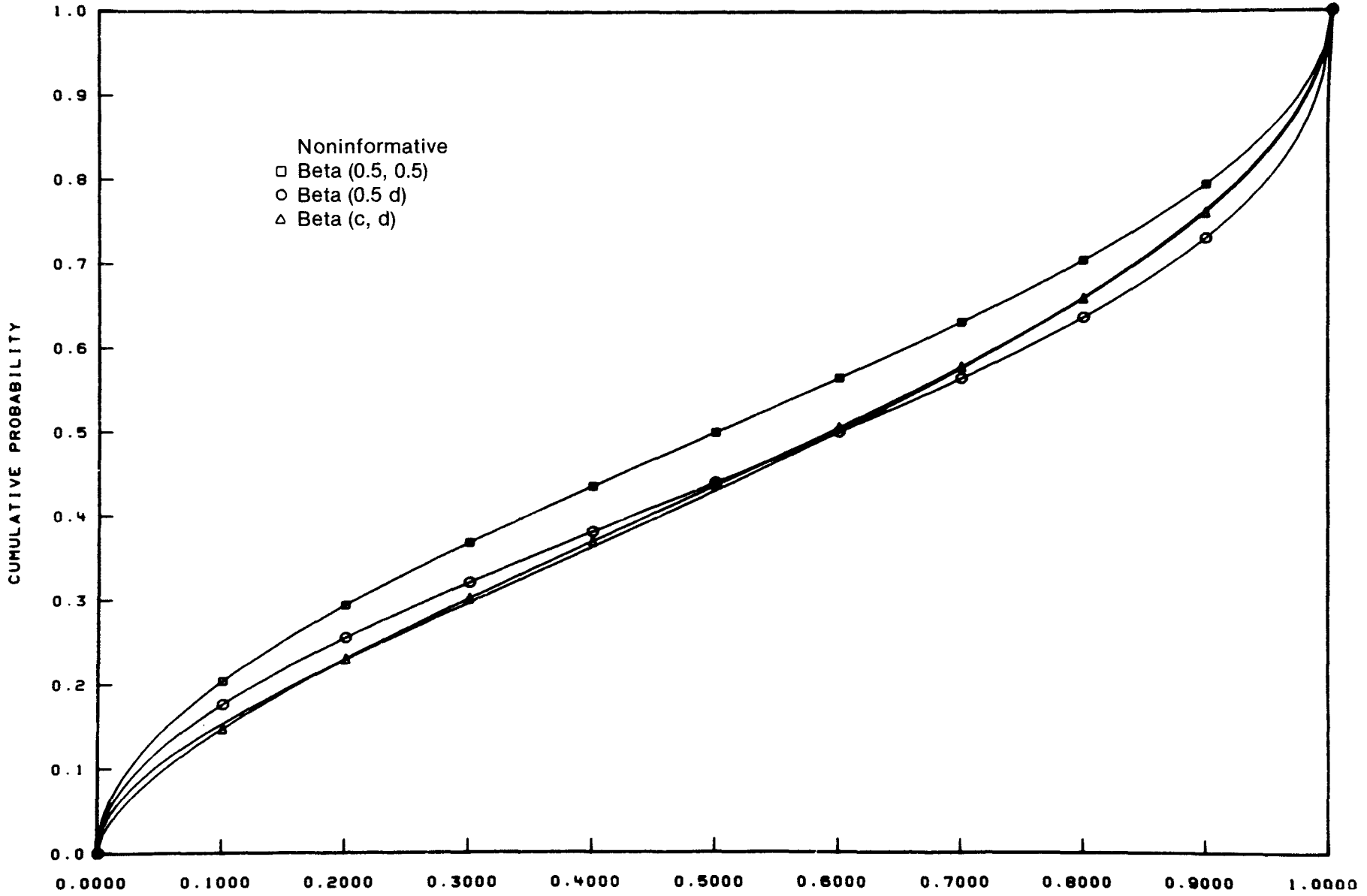


Figure 4. Cumulative distribution functions for noninformative prior and three beta approximations, 0 truncated, m = 5.

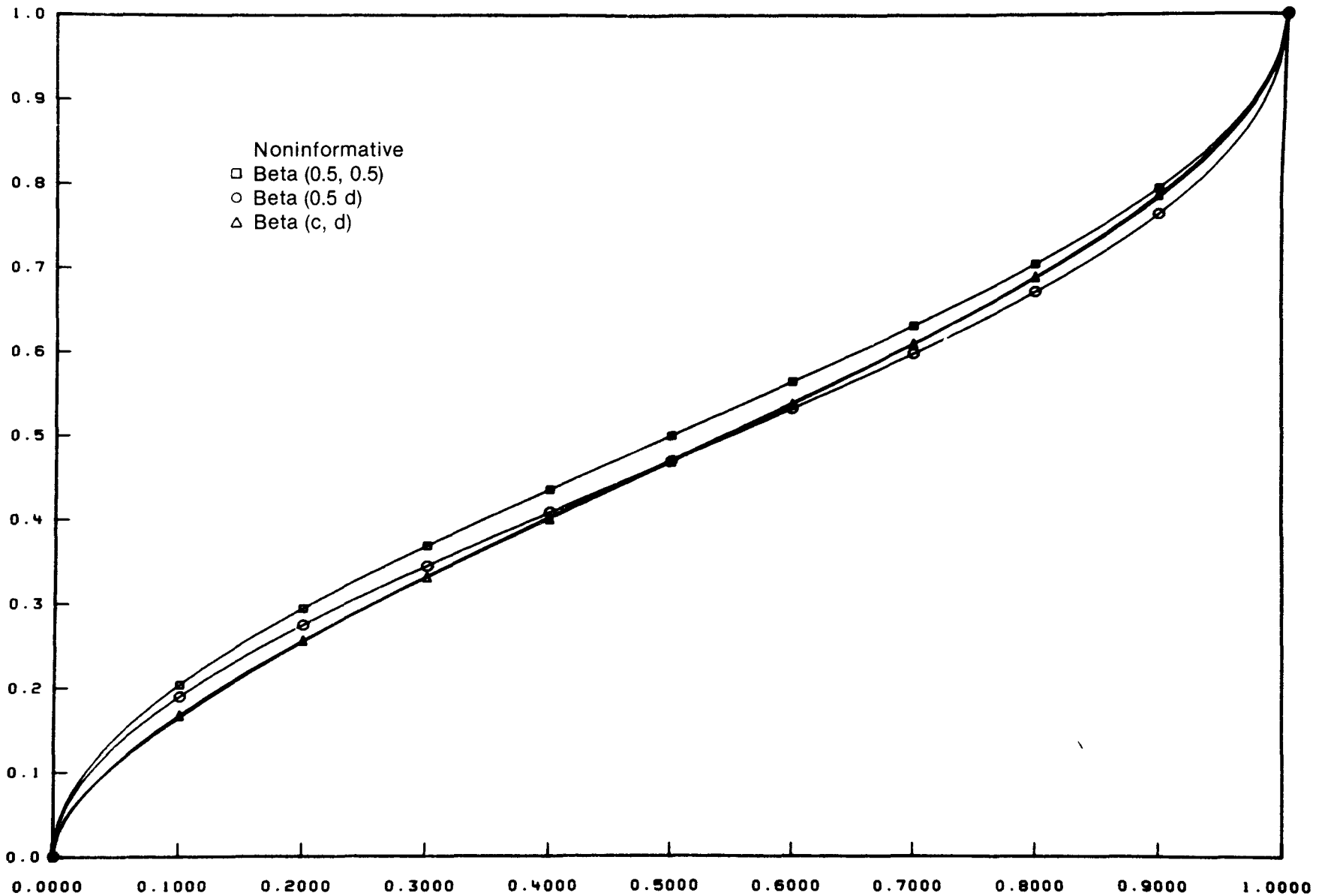


Figure 5. Cumulative distribution functions for noninformative prior and three beta approximations, 0 truncated, m = 20.

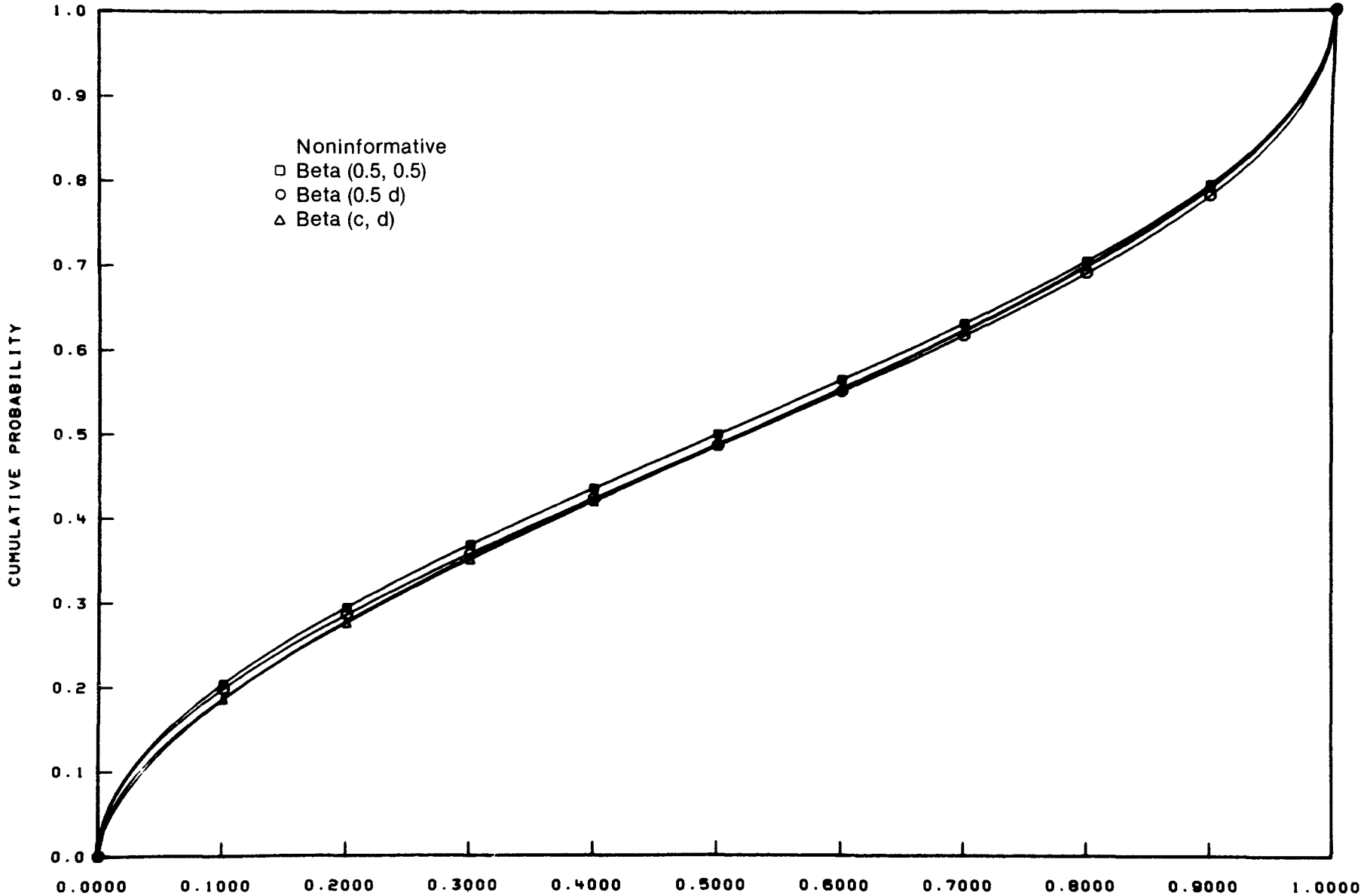


Figure 6. Cumulative distribution functions for noninformative prior and three beta approximations, 0 truncated, m = 100.

The posterior density factors into the product

$$L(\lambda, \lambda_+', p|\text{data}) = L(\lambda|\text{data})\, L(\lambda_+'|\text{data})\, L(p|\text{data})$$

where

$$\lambda|\text{data} \sim \Gamma(a + n_I,\ b + mt)$$

$$\lambda_+'|\text{data} \sim \Gamma(a_+' + n_+',\ b_+' + t)$$

and

$$L(p|\text{data}) = C\, \frac{n_+'!}{n_C'!, \ldots, n_m'!} \binom{m}{1}^{n_C} \prod_{i=2}^{m} \binom{m}{i}^{n_i} \frac{p^{s'+c-1}\, q^{mn_+'-s'+d-1}}{(1-q^m)^{n_+'}}\ .$$

Bayes point and interval estimation of λ, λ_+', and p are routine, using numerical integration to treat p. Point estimates of λ_1, λ_+, μ, and μp^k corresponding to the mode of the posterior density follow from the equations relating the parameters, and the posterior means of these quantities are obtained by straightforward integration.

A probability interval for μp^k can be found as in Section 4.3.2:

$$P\left[c \le \mu p^k \le d \mid \text{data}\right]$$

$$= \int_0^1 P\left[c(1-q^m)/p^k \le \lambda_+' \le d\ (1-q^m)/p^k \mid p, \text{data}\right] L(p|\text{data})\ dp$$

$$= \int_0^1 P\left[cu(p) \le \lambda_+' \le d\ u(p) \mid p, \text{data}\right] L(p|\text{data})\ dp$$

with $u(p) = (1-q^m)/p^k$. In the notation of (4.4), this equals

$$\int_0^1 \left[I(A_+', B_+', d\,u(p)) - I(A_+', B_+', c\,u(p)) \right] L(p|data)\, dp$$

with $A_+' = a_+' + n_+'$ and $B_+' = b_+' + t$. The integrand may be evaluated using (8.6) through (8.8), and the integral may be evaluated numerically.

For μ, the method is the same, with k=0. For

$$\lambda_+ = \lambda_+' \, (1 - q^m - mpq^{m-1})/(1 - q^m),$$

the method is the same as for μp^k, but now set $u(p) = (1-q^m)/(1-q^m-mpq^{m-1})$.

For

$$\lambda_1 = m\lambda + \lambda_+' \, (mpq^{m-1})/(1-q^m),$$

the method is similar, but integration is with respect to λ and λ_+'. The details are messy and will not be given. If N_I and N_C are identifiable, then λ_1 is a less natural parameter than λ and may not be of interest to the user.

6.4 Diagnostic Checks

The checks for the Poisson parameters in Section 6 all carry over if N_{+i}, n_+, λ_{+i}, and λ_+ are replaced by N_{+i}', n_+', λ_{+i}', and λ_+', and if N_{1i}, n_1, λ_{1i}, and λ_1 are replaced by N_{Ii}, n_I, λ_i, and λ.

The checks for the binomial parameter p carry over under the following translation. Now, X takes values from 1 to m. Primes should be given to S, S_i, S_{-i}, n_{+i}, $n_{+,-i}$, and $\hat{z}_i$. Equations (5.5) and (5.6) must be replaced by

$$L_o = s' \log (\hat{p}/\hat{q}) + n_+' \log \left[\hat{q}^m/(1-\hat{q}^m)\right]$$

$$L_i = s_i' \log (\hat{p}_i/\hat{q}_i) + n_{+i}' \log \left[\hat{q}_i^m/(1-\hat{q}_i^m)\right].$$

To test against the alternate hypothesis that X is not truncated binomial, the likelihood ratio test statistic is

$$-2 \log \Lambda = 2n_C \log \left[n_C/(n_+' \hat{z}_1')\right] + 2 \sum_{i=2}^{m} n_i \log \left[n_i/(n_+' \hat{z}_i')\right]$$

where

$$\hat{z}_i' = \binom{m}{i} \hat{p}^i \hat{q}^{m-i}/\left[1 - \hat{q}^m\right].$$

Under H_o, the asymptotic distribution is $X^2(m-3)$. The X^2 test statistic

$$\frac{(n_C - n_+' \hat{z}_1')^2}{n_+' \hat{z}_1'} + \sum_{i=2}^{m} \frac{(n_i - n_+' \hat{z}_i')^2}{n_+' \hat{z}_i'}$$

may be used instead of the likelihood ratio test statistic.

7. ILLUSTRATIVE EXAMPLE

7.1 Data Used

Vesely[1] gives data from 20 U.S. commercial boiling water reactors. The components are control rods, and a failure may be defined in at least two ways: failure to insert past notch 04, or any failure to meet technical specifications. The first definition includes failures that are serious enough to affect safety. The second also includes incidents such as slow rod insertion.

This data set is used for illustrative purposes only! Much more data has become available since the data set of Reference 1 was published. The treatment presented here is also naive in that it ignores the event descriptions given in the reports. These descriptions might suggest qualitative differences among the events, which would lead a careful analyst to consider portions of the data separately. Finally, the diagnostic checks point out inhomogeneity among the plants, and one failure occurrence which is a clear outlier. These are not investigated further here. A more thorough analysis of much more extensive data is now underway at the Idaho National Engineering Laboratory. Those who are interested in the answers, not merely in the method, must refer to the INEL reports that will appear. The numerical results given below serve only to illustrate the method.

The model of this paper assumes that m, the number of components, is constant. In Reference 1 the number of rods in a plant varies from 32 to 185. We will consider plants with 177 or 185 rods. The data for these eight plants are summarized in Table 5. The effect of pooling data with these two values of m will be discussed below.

The causes of single failures will be considered as not identifiable.

TABLE 5. SUMMARY OF FAILURE DATA

Reactor	Number of Rods (m)	Total Operating Months (t)[a]	Failures to Insert Past Notch 04	All Incidents
Dresden 2	177	41	$n_3=1$	$n_{96}=1$
				$n_2=2$
				$n_1=1$
Dresden 3	177	61	$n_1=1$	$n_3=1$
Quad-Cities 1	177	51	None	None
Quad-Cities 2	177	49	None	None
Peach Bottom 2	185	30	None	None
Peach Bottom 3	185	25	None	None
Browns Ferry 1	185	20	None	None
Browns Ferry 2	185	13	None	None
		290		

a. Operating time in hours assumes 720 hours per month.

7.2 Analysis of Failures to Insert Past Notch 04

For this data set, $n_1 = 1$, $n_+ = 1$, $s = 3$, $t = 208800$ hours, and estimates will be calculated using both $m = 177$ and $m = 185$. Estimates will be found for λ_1, λ_+, λ, μ, p, and μp^2. The last quantity is an example of μp^k.

Maximum likelihood estimates and 95% confidence intervals are given in Table 6. All the confidence intervals are two-sided (probability 0.025 for

TABLE 6. MAXIMUM LIKELIHOOD ESTIMATES AND 95% CONFIDENCE INTERVALS

	m = 177		m = 185	
Parameter	MLE	Confidence Interval	MLE	Confidence Interval
p	0.0122	(0.00043, 0.0483)	0.0117	(0.00041, 0.0462)
$10^6 \times \lambda_1$	4.79	(0.121, 26.7)	4.79	(0.121, 26.7)
$10^6 \times \lambda_+$	4.79	(0.121, 26.7)	4.79	(0.121, 26.7)
$10^8 \times \lambda$	1.65	(0, 12.8)	1.58	(0, 12.8)
$10^6 \times \mu$	7.51	(0.061, 42248)	7.51	(0.061, 42468)
$10^9 \times \mu p^2$	1.12	(0.0040, 87.4)	1.02	(0.0037, 80.1)

each tail), except for the interval for λ. The interval for λ is the conservative one-sided interval based on $\lambda \leq \lambda_1/m$ (probability 0.05 for the upper tail). The two-sided interval for λ that is based on λ_1, λ_+, and p, and is described in Section 3.2.5, turns out also to have its lower end point at 0, and the interval is strictly larger than the interval of Table 6. The intervals for μ and μp^2 are based on two-sided intervals for λ_+ and p with $(1-\alpha)(1-\alpha_p) = 0.95$ and $\alpha_+ = \alpha_p$.

Bayesian point and interval estimates are given in Table 7. The intervals all have posterior probability 0.025 in each tail. Highest posterior density intervals are not shown because the computer programming is not yet complete. An interval for λ is not given because of the lengthy computation required. The prior distribution is noninformative for λ_1 and λ_+. The first two portions of the table use a beta prior distribution for p that is approximately noninformative. The third portion uses a beta(0.5, 0.5) distribution.

A comparison of the first and third sections of Table 7 shows that the entries change very little, whether the prior beta distribution has

TABLE 7. BAYESIAN POINT ESTIMATES AND 95% INTERVALS

Parameter	Posterior Mode	Posterior Mean	Interval
	m = 177, p ~ beta(0.552, 0.519)		
p	0.00769	0.0189	(0.00418, 0.0439)
$10^6 \times \lambda_1$	2.39	8.29	(0.923, 23.7)
$10^6 \times \lambda_+$	2.39	6.07	(0.411, 19.6)
$10^8 \times \lambda$	0.156	4.05	--
$10^6 \times \mu$	6.06	9.01	(0.626, 29.3)
$10^9 \times \mu p^2$	0.358	3.55	(0.0836, 18.4)
	m = 185, p ~ beta(0.551, 0.519)		
p	0.00734	0.0180	(0.00399, 0.0420)
$10^6 \times \lambda_1$	2.39	8.30	(0.923, 23.7)
$10^6 \times \lambda_+$	2.39	6.07	(0.411, 19.6)
$10^8 \times \lambda$	0.145	3.88	--
$10^6 \times \mu$	6.08	9.01	(0.626, 29.3)
$10^9 \times \mu p^2$	0.327	3.25	(0.0765, 16.8)
	m = 177, p ~ beta(0.5, 0.5)		
p	0.00710	0.0186	(0.00404, 0.0435)
$10^6 \times \lambda_1$	2.40	8.33	(0.937, 23.8)
$10^6 \times \lambda_+$	2.40	6.04	(0.407, 19.5)
$10^8 \times \lambda$	0.00	4.05	--
$10^6 \times \mu$	6.68	9.08	(0.630, 29.6)
$10^9 \times \mu p^2$	0.337	3.46	(0.0804, 18.0)

approximately noninformative parameters or parameters 0.5 and 0.5. As shown in Figures 1 through 3, the approximately noninformative beta distribution is "closer" to the noninformative prior than to the beta(0.5, 0.5) distribution. It therefore seems unlikely that the table entries would change much if the prior distribution were the noninformative prior based on the information matrix. This indicates that the beta approximation to the noninformative prior is adequate. Investigation of this point when m is small will be carried out at a later date.

In Section 3.1, it was mentioned that the MLE p decreases as m increases, so a more conservative (larger) estimate p will result from using $m = 177$, rather than $m = 185$. Tables 6 and 7 show that in this example, it is more conservative by any criterion to use $m = 177$ rather than $m = 185$, since doing so gives larger estimates of λ, p, and μp^2, and identical or virtually identical estimates of λ_1, λ_+, and μ. The maximum likelihood estimates and posterior means for p and for λ are approximately proportional to $1/m$.

In every case, the maximum likelihood estimate lies between the Bayes posterior mode and the Bayes posterior mean. In problems with a single binomial or Poisson parameter, it can be shown directly that this relation holds when the noninformative prior is used. It is interesting that it also holds for every parameter in the present multiparameter problem.

For λ_1, λ_+, and p, the Bayes intervals are shorter than the corresponding confidence intervals. In problems with a single binomial or Poisson parameter, it can be shown directly that the Bayesian interval based on the noninformative prior must be strictly shorter than the corresponding confidence interval. This is because confidence intervals based on discrete data are inexact, with the (unknown) true confidence level being greater than or equal to the nominal level. In the present multiparameter problem, the same relation holds between the sizes of the confidence interval and the Bayes interval, presumably for the same reason. For μ and μp^2, the Bayes intervals are also shorter than the confidence intervals. This is due both to the reason just mentioned and to the fact

that these confidence intervals are conservative rather than sharp. The Bayes interval for λ is not shown because it requires lengthy computation.

Figure 7 shows a 90% confidence region for (λ_+, λ_1), given three values of p. The values of p are the MLE and the ends of the 95% confidence interval for p. For each p, the portion of the square which is above the corresponding line is that portion satisfying the constraint (3.1). Figures 8 through 11 show 90% Bayes regions for (λ_+, λ_1), given various values of p. The values of p are the posterior mode, the posterior mean, and the end points of the 95% probability interval for p. The Bayesian and non-Bayesian regions are not directly comparable, since they assume different values of p. However, in all the figures, the estimates of λ_+ and λ_1 are positively correlated, due to the constraint (3.1), and the correlation is strongest for small p.

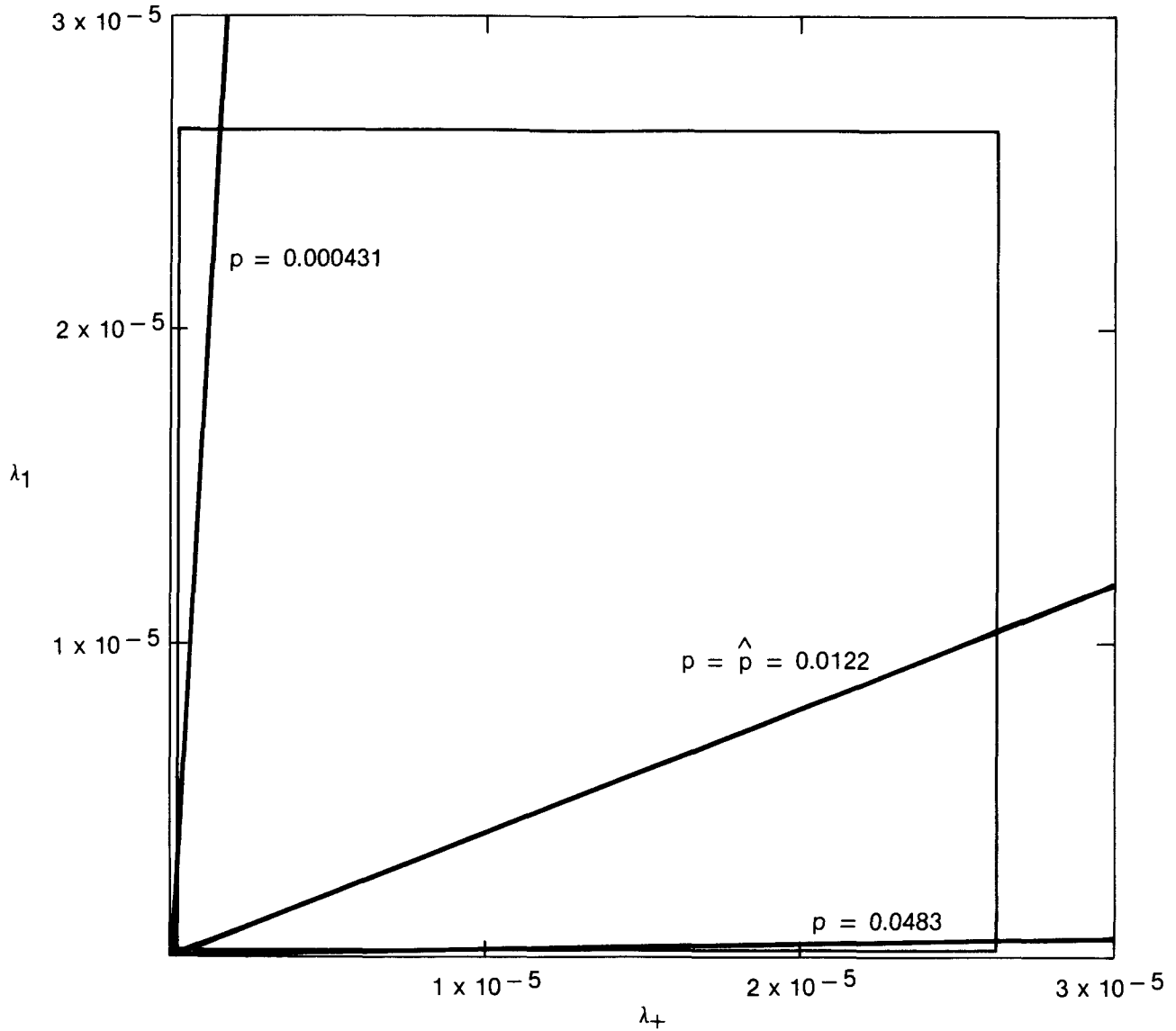


Figure 7. Ninety percent confidence region for (λ_+, λ_1), given p. For each p, the region is that portion of the rectangle above the line corresponding to p.

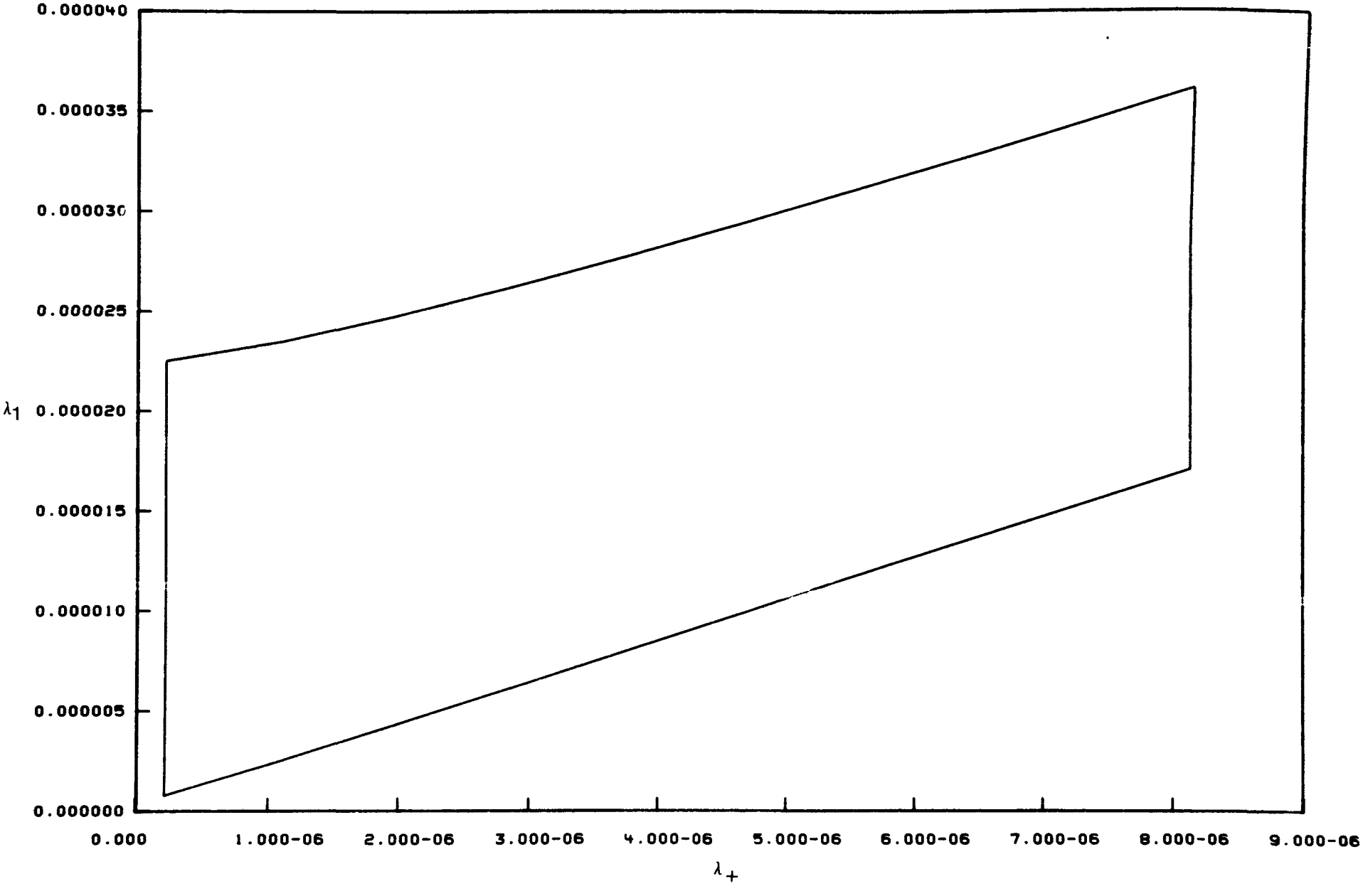


Figure 8. Ninety percent Bayes probability region for (λ_+, λ_1), given p = 0.00418 = lower end of interval for p.

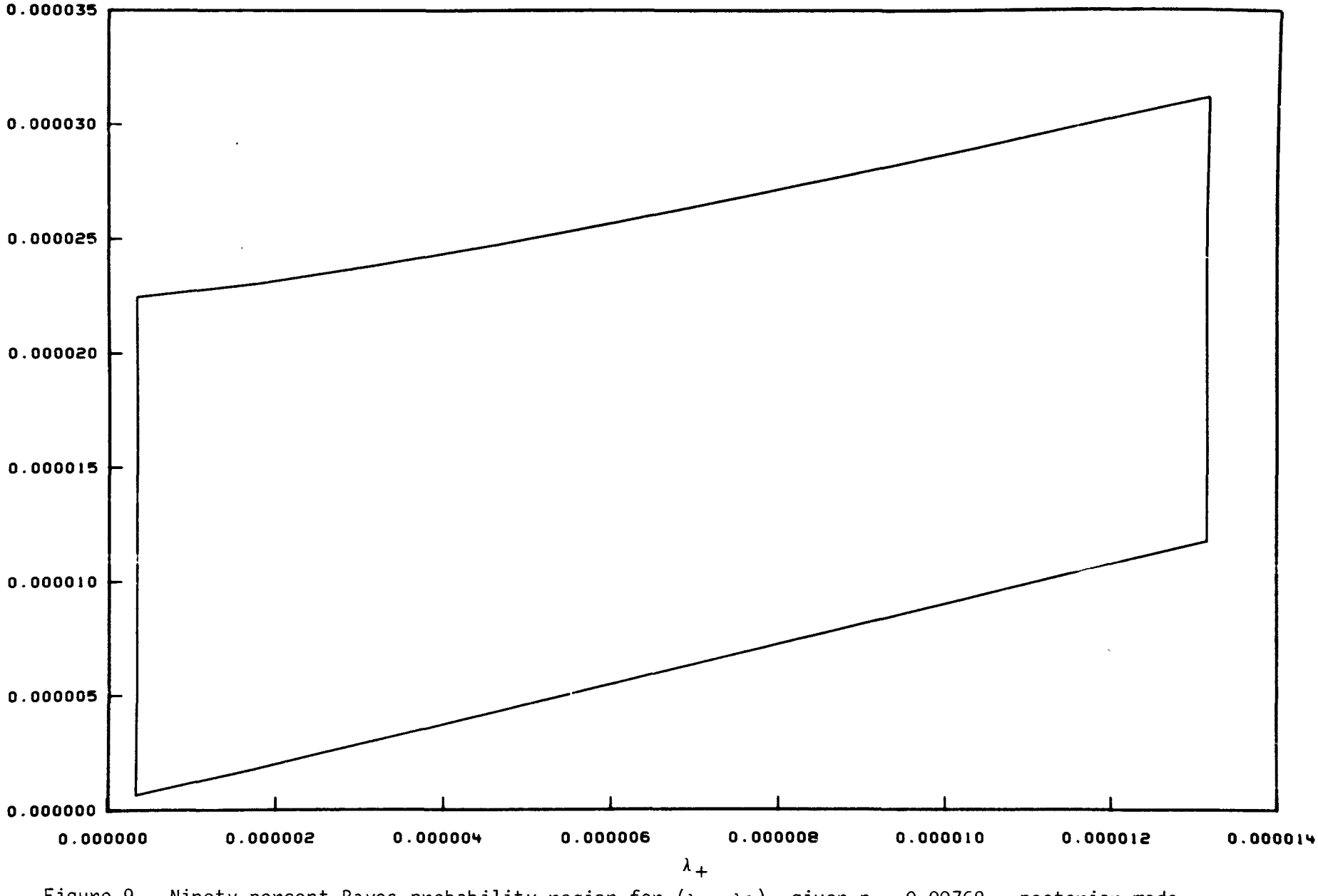


Figure 9. Ninety percent Bayes probability region for (λ_+, λ_1), given p = 0.00769 = posterior mode.

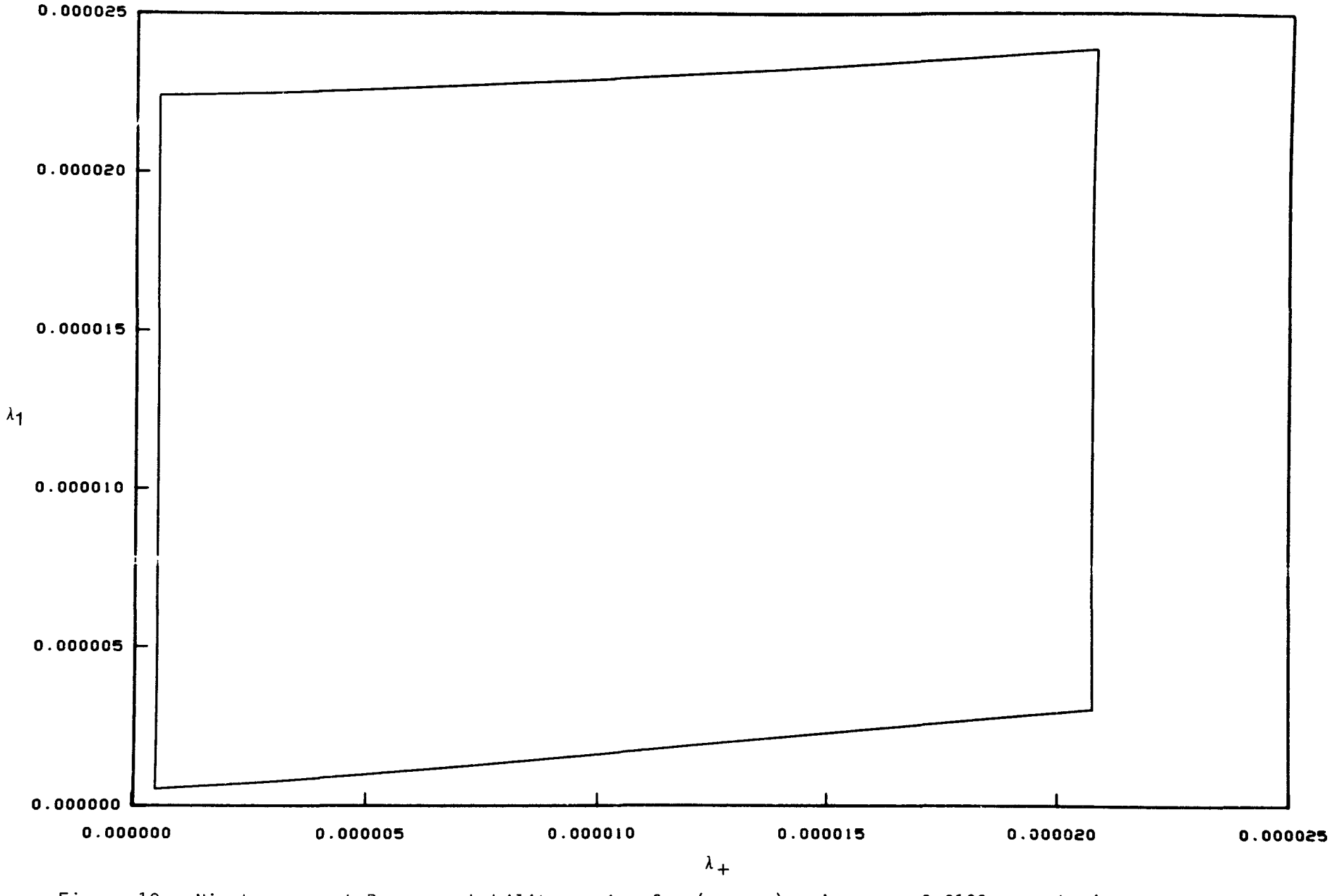


Figure 10. Ninety percent Bayes probability region for (λ_+, λ_1), given p = 0.0189 = posterior mean.

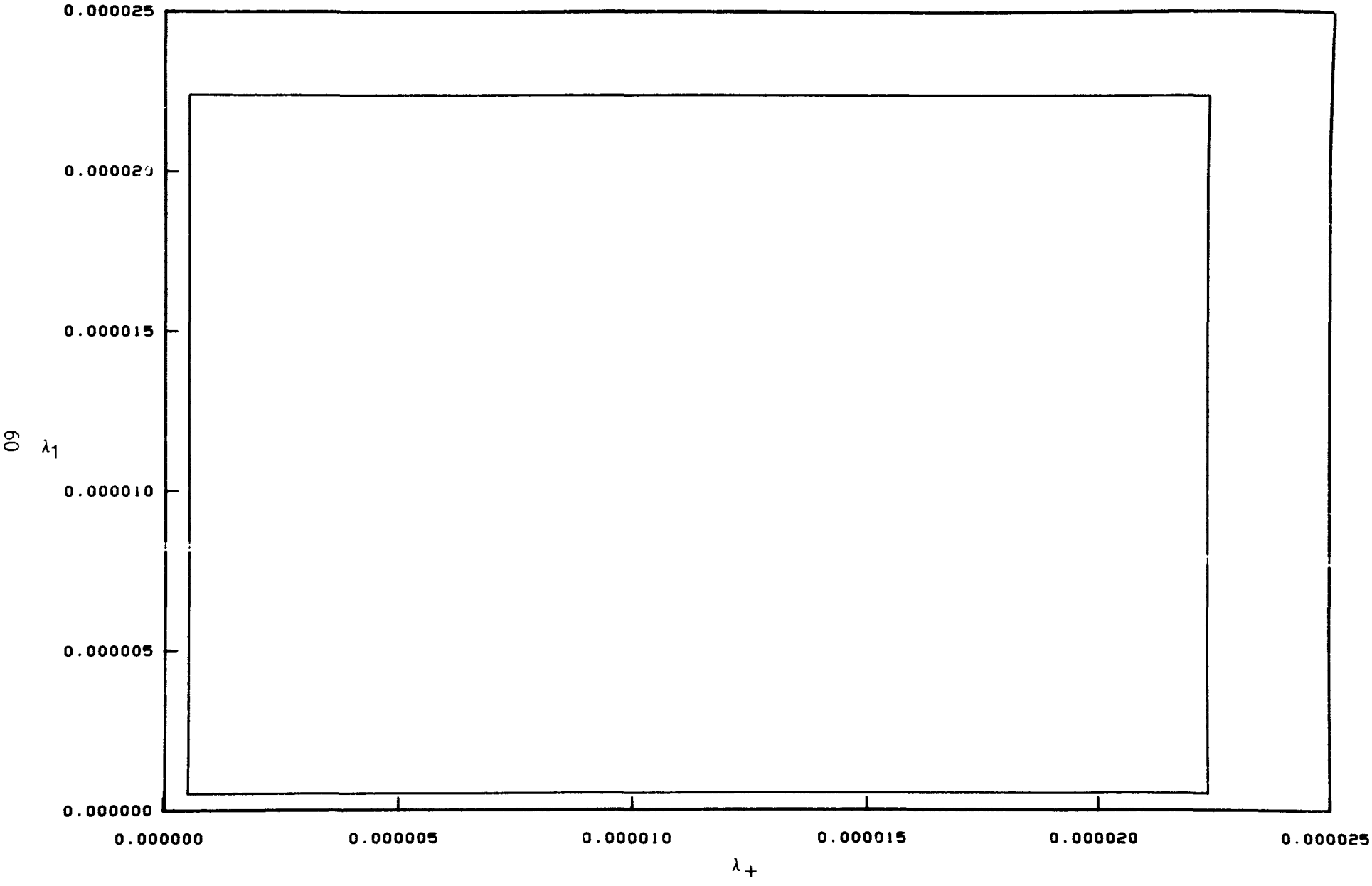


Figure 11. Ninety percent Bayes probability region for (λ_+, λ_1), given p = 0.0439 = upper end of interval for p.

7.3 Diagnostic Checks

In order to have enough data to illustrate the methods for diagnostic checks, we will consider all five failure incidents.

Let us combine single and multiple failures, and investigate $\lambda_1 + \lambda_+$. Table 8 shows the standardized residuals and significance levels for the eight plants, and for the parameter $\lambda_1 + \lambda_+$. The first plant shows a small significance level, indicating that the plant seems to be anomalous. However, one would expect random variability alone to produce some apparent anomalies among many plants. To test whether the plants all have the same value of $\lambda_1 + \lambda_+$, the likelihood ratio statistic (5.1) can be calculated to be 13.76. If all the plants have the same value of $\lambda_1 + \lambda_+$, then

$$P(-2 \log \Lambda \geq 13.76 | N_1 + N_+ = 5) = 0.016.$$

This probability is exact rather than asymptotic, and indicates that the plants did not have the same value of $\lambda_1 + \lambda_+$.

TABLE 8. STANDARDIZED RESIDUALS AND SIGNIFICANCE LEVELS FOR $\lambda_1 + \lambda_+$

Plant	t_i	Observed Number of Failures	Standardized Residual	Significance Level
Dresden 2	41	4	4.23	0.0035
Dresden 3	61	1	-0.06	1.0
Quad-Cities 1	51	0	-1.03	0.76
Quad-Cities 2	49	0	-1.01	0.79
Peach Bottom 2	30	0	-0.76	1.0
Peach Bottom 3	25	0	-0.61	1.0
Browns Ferry 1	20	0	-0.69	1.0
Browns Ferry 2	13	0	-0.48	1.0

As an illustration of diagnostic checks for the assumptions about p, let us consider the four multiple failure occurrences as the "sources." Then the values of n_{+i} and $n_{+,-i}$ are 1 and 3 for each source. For this data set, the MLE $\hat{p}$ is 0.14548. This agrees to five places with the estimate s/mn_+, which would be obtained if the observed data were treated as binomial rather than truncated binomial. Since the truncation effect is negligible, an approximate 95% confidence interval for p has upper end at

$$\hat{p} + 2\left[\hat{p}\hat{q}/(mn_+)\right]^{1/2} = 0.1720.$$

The standardized residuals from (5.3) are given in Table 9, based on $p = 0.1455$ and $p = 0.1720$. Plausible magnitudes of standardized residuals are, say, less than 3. The magnitudes in Table 9 are so much greater than 3, whichever p is used, that it seems conclusive that the data do not come from a single BFR model.

TABLE 9. STANDARDIZED RESIDUALS FOR p
Each multiple failure treated as data source, using formula (5.3)

Number of Failed Components ($=S_i$)	$R_i/\hat{\sigma}_x$	
	p = 0.1455	p = 0.1720
96	17.29	16.16
2	-5.85	-5.46
2	-5.85	-5.46
3	-5.60	-5.23
103		

Even more dramatic results are seen if the data are treated as coming from a single source, and U_i is calculated from (5.4). The values are summarized in Table 10. Truly enormous quantities are obtained because of observed failure numbers with extremely small estimated probabilities.

TABLE 10. STANDARDIZED RESIDUALS FOR BINOMIAL DISTRIBUTION
Data treated as coming from single source, using formula (5.4)

Number of Failed Components i	Number of Occurrences N_i	$U_i = (N_i - n_+ \hat{z}_i)/\left[n_+ \hat{z}_i (1-\hat{z}_i)\right]^{1/2}$
2	2	51914.0
3	1	8237.0
4	0	-0.000
5	0	-0.001
. . .	. . .	. . .
24	0	-0.595
25	0	-0.608
26	0	-0.607
27	0	-0.591
. . .	. . .	. . .
95	0	-0.000
96	1	5.72×10^{16}
97	0	-0.000
. . .	. . .	. . .
177	0	-0.000

In fact, in this case, it is possible to test

H_0: the four multiple failures come from a single BFR model.

Under H_0, given that 103 components fail in four occurrences, the conditional distribution of the four failure counts is multinomial(103, 1/4, 1/4, 1/4, 1/4). The distribution can then be found subject to the additional condition that each failure occurrence involves at least two failed components. Somewhat tedious but direct combinatorial calculations yield

$$P\left[-2 \log \Lambda \geq \text{observed} \middle| H_0,\ n_+ = 4,\ s = 103,\ \text{each failure count} \geq 2\right]$$

$$\doteq 5 \times 10^{-49} .$$

Therefore, the occurrence with 96 failures does not come from the same BFR model as the other failure occurrences.

8. TECHNICAL DETAILS

8.1 Maximization of Likelihoods

Consider

$$L(p) = C \frac{p^{s+c-1} q^{mn_+ - s + d - 1}}{(1 - q^m - mpq^{m-1})^{n_+}} .$$

If $c > 0$ and $d > 0$, this is the posterior density of p from a beta prior. If $c = d = 1$, it is the likelihood.

If $m = 2$ or $n_+ = 0$, then s can only take the one value $2n_+$, and the data contain no information about p. So assume $m > 2$ and $n_+ > 0$.

We will show that L(p) has a unique maximum (possibly infinite), at some p_0 which is located according to the following seven cases and subcases.

If	Then
$2n_+ < s < m\, n_+$	$0 < p_0 < 1$
$s = m\, n_+$	
$d > 1$	$0 < p_0 < 1$
$d \leq 1$	$p = 1$
$s = 2n_+$	
$c > 1$	$0 < p_0 < 1$
$c < 1$	$p_0 = 0$
$c = 1$	
$-\frac{n_+(m-2)}{3} - d + 1 > 0$	$0 < p_0 < 1$
$-\frac{n_+(m-2)}{3} - d + 1 \leq 0$	$p_0 = 0$

Moreover, if $0 < p_0 < 1$, it is the value at which expression (8.1) equals zero.

Since s is an integer and $c > 0$, $s + c - 1 - 2n_+$ can be negative if and only if $s = 2n_+$ and $c < 1$. In this case $L(p)$ is finite for $0 < p \leq 1$ and $L(p) \to \infty$ as $p \to 0$. So $p_0 = 0$.

Assume now that $s + c - 1 - 2n_+ \geq 0$. Let us show that $(\partial/\partial p)L(p)$ is zero at at most one point, and that the sign change can only be from positive to negative. The cases then follow from considering whether $(\partial/\partial p)L(p)$ changes sign or not. Since $(\partial/\partial p)L(p)$ has the same sign as $q(\partial/\partial p)\log L(p)$, it is sufficient to show that

$$q\frac{\partial}{\partial p}\log L(p) = \frac{s+c-1}{p} - \left\{c+d-2 + mn_+ \frac{1-q^{m-1}}{1-q^m-mpq^{m-1}}\right\} \tag{8.1}$$

is strictly decreasing in p for $0 < p < 1$. Add the constant c+d-2 to both sides of (8.1) and write the result as

$$\frac{s+c-1-2n_+}{p} + n_+\left\{\frac{2}{p} - m\frac{1-q^{m-1}}{1-q^m-mpq^{m-1}}\right\}. \tag{8.2}$$

Since $s+c-1-2n_+ \geq 0$ by assumption, the first term of (8.2) is decreasing in p. Therefore it is sufficient to show that the expression in brackets is strictly decreasing in p for $0 < p < 1$ and $m > 2$.

The derivative of the expression in brackets in (8.2) is

$$\frac{m(m-1)\,p^2\,q^{m-2}\,(1-q^m-mp) - 2(1-q^m-mpq^{m-1})^2}{p^2\,(1-q^m-mpq^{m-1})^2}.$$

Denote the numerator by $A_m(p)$. It is not obvious that $A_m(p)$ is negative, and we will show it by obtaining successively simpler expressions

$B_m(p)$, $C_m(p)$, and $D_m(p)$. The outline below omits the tedious but direct algebra. Each assertion is verified in the obvious way.

$$A_2(p) = 0, \quad \text{for } 0 < p < 1.$$

$$\begin{aligned} &A_{m+1}(p) - A_m(p) \\ &= -mp^2q^{m-2}\left\{6 - (4m+6)p + (m+1)^2p^2 \right. \\ &\left. + q^m\left[-6 - (2m-6)p + (m-1)p^2\right]\right\} \\ &= -mp^2q^{m-1}\, B_m(p) \end{aligned}$$

defining $B_m(p)$. Setting $m=2$ and $q=1-p$ yields

$$B_2(p) = p^4 \quad > 0, \quad \text{for } 0 < p < 1.$$

$$B_{m+1}(p) - B_m(p) = p\left\{-4 + (2m+3)p + q^m\left[4 + (2m-3)p - mp^2\right]\right\} = p\, C_m(p).$$

$$C_2(p) = p^3(5 - 2p) \quad > 0, \quad \text{for } 0 < p < 1.$$

$$C_{m+1}(p) - C_m(p) = p\left\{2 + q^m\left[-2 - 2mp + (m+1)\, p^2\right]\right. \quad = p\, D_m(p).$$

$$D_2(p) = p^2(9 - 10p + 3p^2) \quad > 0, \quad \text{for all } p.$$

$$D_{m+1}(p) - D_m(p) = p^2q^m\left[2m+3 - (m+2)p\right] \quad > 0, \quad \text{for } m \geq 2,\ 0 < p < 1.$$

Working up the above lines, we obtain successively $D_m(p) > 0$ for $m \geq 2$, $0 < p < 1$, then $C_m(p) > 0$ for $m \geq 2$, $0 < p < 1$, then $B_m(p) > 0$ for $m \geq 2$, $0 < p < 1$, and finally $A_m(p) < 0$ for $m > 2$, $0 < p < 1$. This shows that (8.2), and therefore (8.1), is strictly decreasing. So $L(p)$ is maximized at a unique p_0.

To locate p_0, consider first (8.1) as $p \to 1$. The limit is

$s - mn_+ - d + 1,$

which is negative if and only if

$s < mn_+$

or

$s = mn_+$ and $d > 1$.

So these are the cases for which $p_0 < 1$.

Now consider (8.1) as $p \to 0$. Since

$$1 - q^{m-1} = (m-1)p - \binom{m-1}{2}p^2 + O(p^3)$$

and

$$1 - q^m - mpq^{m-1} = \binom{m}{2}p^2 q^{m-2} + \binom{m}{3}p^3 + O(p^4),$$

we obtain after manipulation, that as $p \to 0$ (8.1) equals

$$\frac{s + c - 1 - 2n_+}{p} - \frac{n_+(m-2)}{3} - (c+d-2) + O(p). \qquad (8.3)$$

Recall that $s+c-1-2n_+$ is assumed to be ≥ 0. Expression (8.3) is positive if

$s > 2n_+$

or

$s = 2n_+$, $c > 1$

or

$s = 2n_+$, $c=1$, $-n_+(m-2)/3 - d + 1 > 0$.

So these are the cases for which $p_0 > 0$.

All the assertions made at the start of this section about the maximization of $L(p)$ have now been proved.

We now verify the claim made in Section 3.1 that the right side of (3.3) is increasing in m. Let

$$G(m) = mp \frac{1 - q^{m-1}}{1 - q^m - mpq^{m-1}} .$$

To show that $G(m)$ is increasing in m, let us show that $G(m+1) - G(m)$ is positive for $p > 0$. Algebraic manipulation shows that $G(m+1) - G(m)$ is positive if and only if

$$(m+1)(1-q^m)(1-q^m - mpq^{m-1}) - m(1-q^{m-1})(1-q^{m+1} - (m+1)pq^m)$$

is positive. This quantity equals

$$(1-q^m)^2 - m^2p^2q^{m-1} = A$$

defining A. At $p = 0$, A is zero. The derivative $\partial A/\partial p$ equals

$$mq^{m-2}\left[2q(1-q^m) - 2mpq + m(m-1)p^2\right] = mq^{m-2} B,$$

defining B. At $p = 0$, B and its first derivative are both zero. The second derivative is

$$2m(m+1)(1-q^{m-1}) > 0.$$

Therefore, B is positive for all $p > 0$. Therefore, so is A, and therefore, $G(m+1) > G(m)$ for $p > 0$.

8.2 Confidence Interval for p

In Section 3.2, a confidence interval for p was given, based on the conditional distribution of S given N_+. The choice of this particular confidence interval is now justified. The argument uses the properties of similar tests and Neyman structure (Ferguson[8], pp. 226-7, or Reference 4, pp. 134-5) and monotone likelihood ratios (Reference 6, p. 423, or Reference 8, p. 208).

By the factorization (2.4), the problem may be formulated in terms of (λ_+, p) and the sufficient statistic (N_+, S), ignoring λ_1 and N_1. Suppose that we were testing

$$H_0: \quad p \leq p_0$$

$$H_1: \quad p > p_0$$

for some p_0 and some desired level α. A "similar" test is one with

$$P\left[\text{reject } H_0 \middle| p = p_0\right] = \alpha$$

regardless of the value of λ_+. On the boundary

$$\left\{(\lambda_+, p) \middle| p = p_0\right\}$$

N_+ is sufficient for λ_+. So, a test such that

$$P\left[(S, N_+) \text{ in critical region} \middle| N_+ = n_+\right] = \alpha \quad \text{for all } n_+ \tag{8.4}$$

has "Neyman structure." Since a Poisson random variable is boundedly complete, every similar test has Neyman structure, and any test which is uniformly most powerful (UMP) among tests satisfying (8.4) is UMP similar.

The conditional distribution of S given N_+ is written in (3.7). So the likelihood ratio is

$$P\left[S = k \mid N_+ = n_+;\ p_1\right] / P\left[S = k \mid N_+ = n_+;\ p_2\right]$$

$$= \left(\frac{p_2}{q_2}\,\frac{q_1}{p_1}\right)^k \frac{q_2^{mn_+}\left(1 - q_1^{\,m} - mp_1q_1^{\,m-1}\right)^{n_+}}{q_1^{mn_+}\left(1 - q_2^{\,m} - mp_2p_2^{\,m-1}\right)^{n_+}}$$

which is monotone increasing in k for fixed p_1 and p_2, $p_1 < p_2$. Therefore, a UMP test among tests satisfying (8.4) is to reject H_0 if and only if S is greater than some $c(n_+)$, with possible randomization if $S=c(n_+)$. This test is UMP similar.

Since a UMP similar test of a one-sided hypothesis has a one-sided critical region, good tests of two-sided hypotheses have two-sided critical regions, and one convenient such test assigns equal probabilities to the two tails. The corresponding confidence interval is given by (3.5) and (3.6).

8.3 Integrals

Assume that $A > 0$ and $B > 0$. We repeat definition (4.4) here.

$$I(A,\ B,\ x) = \frac{B^A}{\Gamma(A)} \int_0^x e^{-Bt}\ t^{A-1}\ dt \tag{4.4}$$

Then $I(A,B,\infty) = 1$ for any A and B. For $x < \infty$ the following three results hold. [See also Johnson and Kotz[3], Vol. 2, Ch. 17, equations (23.1) through (24).]

If A is a positive integer,

$$I(A,\ B,\ x) = 1 - e^{-Bx} \sum_{j=0}^{A-1} \frac{(Bx)^j}{j!}\ . \tag{8.5}$$

If $A = k + 1/2$ for integer $k \geq 0$,

$$I(A, B, x) = -e^{-Bx} \sum_{j=0}^{k-1} \frac{(Bx)^{j+1/2}}{\Gamma(j+3/2)} + 2\,\Phi\left[(2Bx)^{1/2}\right] - 1 \qquad (8.6)$$

where Φ is the standard normal cumulative distribution function.

For arbitrary $A > 0$,

$$I(A, B, x) = e^{-Bx} \sum_{j=0}^{\infty} \frac{(Bx)^{A+j}}{\Gamma(A+j+1)} \quad . \qquad (8.7)$$

To prove (8.5) and (8.6), integrate by parts and use induction. The proof of (8.6) is completed by observing that $I(1/2, B, x) = P[X \leq x]$ where $X \sim \Gamma(1/2, B)$. But this equals $P[2BX \leq 2Bx]$ where $2BX \sim \chi^2(1)$. Finally, this equals $2\Phi\left[(2Bx)^{1/2}\right] - 1$. To prove (8.7), observe that the two sides have identical derivatives and are equal at x=0.

Assume now that A, B, A', and B' are all positive. Define

$$J(A, B, A', B', x)$$

$$= \int_0^x \int_s^\infty \frac{B^A}{\Gamma(A)}\, e^{-Bt} t^{A-1}\, \frac{(B')^{A'}}{\Gamma(A')}\, e^{-B's} s^{A'-1} dt\, ds. \qquad (4.5)$$

Let

$$W(A, B, A', B') = \left(\frac{B}{B+B'}\right)^A \left(\frac{B'}{B+B'}\right)^{A'} \frac{\Gamma(A+A')}{\Gamma(A+1)\,\Gamma(A')} \quad .$$

Note the slight asymmetry of A and A' in W. The following four results hold.

If A is an integer,

$$J(A, B, A', B', x) = \sum_{j=0}^{A-1} W(j, B, A', B')\, I(A'+j, B+B', x). \tag{8.8}$$

If A' is an integer,

$$J(A, B, A', B', x)$$

$$= I(A, B, x) + I(A', B', x) - I(A, B, x)\, I(A', B', x)$$

$$- \sum_{j=0}^{A'-1} W(j, B', A, B)\, I(A+j, B+B', x). \tag{8.9}$$

If A and A' are arbitrary positive numbers,

$$J(A, B, A', B', x)$$

$$= I(A', B', x)$$

$$- \sum_{j=0}^{\infty} W(A+j, B, A', B')\, I(A+A'+j, B+B', x) \tag{8.10}$$

$$= I(A', B', x)\,[1 - I(A, B, x)]$$

$$+ \sum_{j=0}^{\infty} W(A'+j, B', A, B)\, I(A+A'+j, B+B', x). \tag{8.11}$$

Equation (8.8) follows from (4.5) and use of (8.5). So does (8.9), after reversing the order of integration in (4.5). Equations (8.10) and (8.11) follow from (8.7), using both possible orders of integration in (4.5).

Of special interest is $J(A, B, A', B', \infty)$. If A or A' is an integer, then (8.8) and (8.9) give finite sums for $J(A, B, A', B', \infty)$. If A and A' are both integers plus 1/2, then there is also a finite expression for J(A, B, A', B', ∞):

If $A=k+1/2$ and $A'=k'+1/2$, for nonnegative integers k and k', then

$$J(A, B, A', B', \infty)$$

$$= 1 + \sum_{j=0}^{k-1} W(j+1/2, B, A', B') - \sum_{j=0}^{k'-1} W(j+1/2, B', 1/2, B)$$

$$- \frac{2}{\pi} \arctan\left[(B/B')^{1/2}\right]. \qquad (8.12)$$

To prove this last assertion, use (8.6) on the inner integral of (4.5). Note that $I(A, B, \infty) = 1$ for any A and B, and obtain

$$J(A, B, A', B', \infty)$$

$$= 2 + \sum_{j=0}^{k-1} W(j+1/2, B, A', B') - 2\int_0^\infty \Phi\left[(2Bs)^{1/2}\right] \frac{(B')^{A'}}{\Gamma(A')} e^{-B's} s^{A'-1} ds.$$

Integrate the last integral by parts k' times, obtaining

$$J(A, B, A', B', \infty)$$

$$= 2 + \sum_{j=0}^{k-1} W(j+1/2, B, A', B') - \sum_{j=0}^{k'-1} W(j+1/2, B', 1/2, B)$$

$$- 2\int_0^\infty \Phi\left[(2Bs)^{1/2}\right] \frac{(B')^{1/2}}{\Gamma(1/2)} e^{-B's} s^{-1/2} ds.$$

The last integral equals $P\left[T \leq (2BS)^{1/2}\right]$, where $T \sim N(0,1)$ and $S \sim \Gamma(1/2, B')$. The distribution of $2B'S$ is $\chi^2(1)$, so the integral equals $P\left[T \leq (B/B')^{1/2}|Z|\right]$, where T and Z are independent N(0,1). This probability is $\left\{1 + \frac{2}{\pi} \arctan\left[(B/B')^{1/2}\right]\right\}/2$, by the spherical symmetry of the bivariate normal distribution. This proves (8.12).

If $J(A, B, A', B', \infty)$ must be evaluated for arbitrary A and A', then one of (8.10) and (8.11) may converge much faster than the other. To decide which series to use, define

$Q = B/(B+B')$

$Q' = B'/(B+B')$.

The ratio of term j+1 of the infinite series to term j is

$$\frac{A + A' + j}{A + j + 1} Q$$

for (8.10), and

$$\frac{A + A' + j}{A' + j + 1} Q'$$

for (8.11). These expressions are monotone in j (decreasing if $A > 1$ and $A' > 1$), so are bounded by

$$\max\left(1, \frac{A + A'}{A + 1}\right) Q \tag{8.13}$$

and

$$\max\left(1, \frac{A + A'}{A' + 1}\right) Q' \ . \tag{8.14}$$

At least one of (8.13) and (8.14) is less than 1. For they are strictly bounded, respectively, by

$$\frac{Q}{A/(A + A')} \tag{8.15}$$

and

$$\frac{Q'}{A'/(A + A')} = \frac{1 - Q}{1 - A/(A + A')} \tag{8.16}$$

and the numerator of (8.15) is greater than the denominator if and only if the numerator of (8.16) is less than the denominator. So, reasonably fast convergence is assured by using (8.10) if (8.15) < (8.16) and using (8.11) otherwise.

9. REFERENCES

1. W. E. Vesely, "Estimating Common Cause Failure Probabilities in Reliability and Risk Analyses: Marshall-Olkin Specializations," Nuclear Systems Reliability Engineering and Risk Assessment, Philadelphia: SIAM, 1977, pp 314-341.

2. N. R. Mann, R. E. Schafer, and N. D. Singpurwalla, Methods for Statistical Analysis of Reliability and Life Data, New York: John Wiley & Sons, Inc., 1974.

3. N. L. Johnson and S. Kotz, Distributions in Statistics, New York: John Wiley & Sons, Inc., Vol. 1, Discrete Distributions, 1969, Vol. 2, Continuous Univariate Distributions - I, 1970.

4. D. R. Cox and D. V. Hinkley, Theoretical Statistics, London: Chapman and Hall Ltd., 1974.

5. G. E. P. Box and G. C. Tiao, Bayesian Inference in Statistical Analysis, Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1973.

6. A. M. Mood, F. A. Graybill, and D. C. Boes, Introduction to the Theory of Statistics, Third Edition, New York: McGraw-Hill, Inc., 1974.

7. C. L. Atwood, "Teaching Fortran to do Algebra, with Applications to Moment Computations," 1978 Proceedings of the Statistical Computing Section, American Statistical Association, Washington, D. C., pp 139-140.

8. T. S. Ferguson, Mathematical Statistics: A Decision Theoretic Approach, New York: Academic Press, 1967.