AD-A183 077   BAYESIAN FAILURE PROBABILITY MODEL SENSITIVITY STUDY   1/1
                (U) BOOZ-ALLEN AND HAMILTON INC BETHESDA MD   30 MAY 86
                NCS-TIB-86-18 DCA100-82-C-0034

UNCLASSIFIED   F/G 25/2   NL

E ND
9 87
DT IC

NCS TIB 86-10

NATIONAL COMMUNICATIONS SYSTEM

BAYESIAN FAILURE PROBABILITY MODEL SENSITIVITY STUDY

MAY 30, 1986

AUG 1 1 1987

A

OFFICE OF THE MANAGER
NATIONAL COMMUNICATIONS SYSTEM
WASHINGTON, D.C. 20305

AD-A183877

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
Unclassified	
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
	Unlimited Distribution
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
NCS TIB 86-10	

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Booz, Allen & Hamilton		

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
4330 East-West Highway Bethesda, MD 20814-4455	

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
National Communications System	NCS-TS	DCA100-82-C-0034

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
Office of Technology & Standards Washington, DC 20305-2010	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Bayesian Failure Probability Model Sensitivity Study

12. PERSONAL AUTHOR(S)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM _____ TO _____	30 May 1986	69

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	High-altitude Electromagnetic Pulse (HEMP)

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

The Office of the Manager, National Communications System (OMNCS) has developed a system-level approach for estimating the effects of High-Altitude Electromagnetic Pulse (HEMP) on the connectivity of telecommunications networks. This approach incorporates a Bayesian statistical model which estimates the HEMP-induced failure probabilities of telecommunications switches and transmission facilities. The purpose of this analysis is to address the sensitivity of the Bayesian model. This is done by systematically varying two model input parameters--the number of observations, and the equipment failure rates. Throughout the study, a non-informative prior distribution is used. The sensitivity of the Bayesian model to the non-informative prior distribution is investigated from a theoretical mathematical perspective.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dennis Bodson	202-692-2124	NCS-TS

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

NATIONAL COMMUNICATIONS SYSTEM

BAYESIAN FAILURE PROBABILITY MODEL SENSITIVITY STUDY

MAY 30, 1986

AUG 1 1 1987

A

OFFICE OF THE MANAGER
NATIONAL COMMUNICATIONS SYSTEM
WASHINGTON, D.C. 20305

TABLE OF CONTENTS

RE: Classified References, Distribution
Unlimited
No change per Ms. Janet Orndorff, NCS/TS

DTIC
COPY
INSPECTED
6

A-1

I N D E X   O F   E X H I B I T S

INDEX OF EXHIBITS
(Continued)

EXECUTIVE SUMMARY

The Office of the Manager, National Communications System (OMNCS) has developed a system-level approach for estimating the effects of High-Altitude Electromagnetic Pulse (HEMP) on the connectivity of telecommunications networks. This approach incorporates a Bayesian statistical model which estimates the HEMP-induced failure probabilities of telecommunications switches and transmission facilities. The OMNCS has received comments from members of the EMP community recommending that sensitivity studies be conducted on the parameters employed by the Bayesian model. The recommendation is addressed in this report.

A major input to the Bayesian model is simulated HEMP equipment test data. In a typical test program, each equipment tested is exposed to a range of HEMP stress levels. The sole criterion of each HEMP exposure for purposes of this analysis is whether the equipment is operable following exposure. Based on the results of the HEMP tests and a non-informative prior distribution, equipment failure Probability Density Function (PDF) and Cumulative Distribution Function (CDF) curves are developed for each tested equipment. The PDF and CDF curves provide the basis by which the system-level model probabilistically estimates whether a network switch or transmission facility survives an exposure to HEMP.

The non-informative prior distribution is selected because it represents an unbiased view of interpreting experimental data (reference 5). That is, all possible experimental outcomes are treated equally prior to observing the data. This characteristic makes the non-informative prior ideal for interpreting the EMP test data, because before the data are interpreted, nothing is really known about the EMP effects on the equipments. The non-informative prior allows the data to have the greatest effect possible in producing the posterior estimates of the failure probability distribution. Thus the non-informative prior is considered the optimal prior for this application.

The purpose of this analysis is to address the sensitivity of the Bayesian model. This is done by systematically varying two model input parameters--the number of observations, and the equipment failure rates. Throughout the study, a non-informative prior distribution is used. The sensitivity of the Bayesian model to the non-informative prior distribution is investigated from a theoretical mathematical perspective. Exhibit ES-1 illustrates the effects of varying the number of observations for an equipment that fails at an observed rate of 50 percent during testing at a HEMP stress level. The PDF

EXHIBIT ES-1
Probability Density Functions: 50 Percent
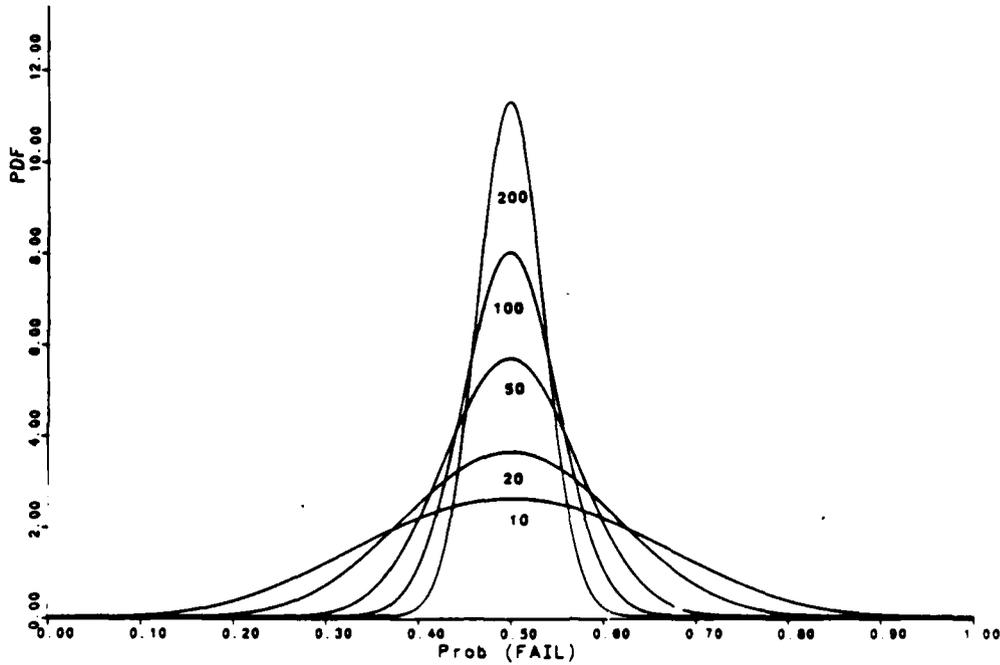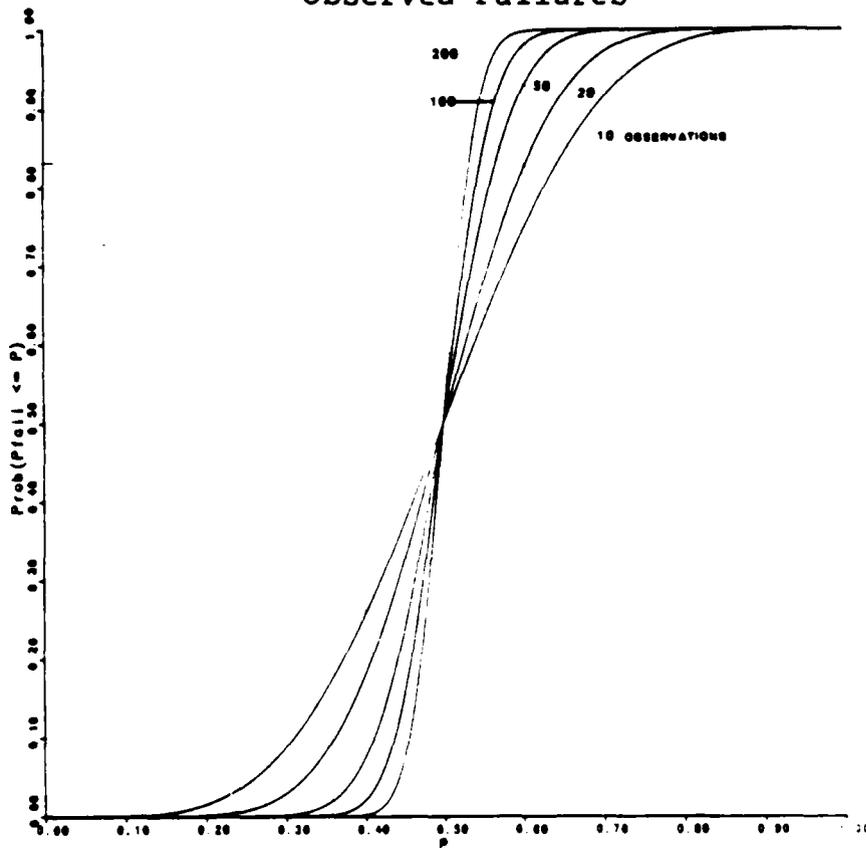## Observed Failures

EXHIBIT ES-2
Cumulative Distribution Functions: 50 Percent
Observed Failures

curves indicate that by increasing the number of observations, the standard deviation around the mean (50 percent) decreases. In other words, curves for large observation sizes are more densely centered around the mean than curves with lower observation sizes. Therefore, higher observation sizes have the effect of providing greater confidence in the results of the tests. This phenomenon is illustrated with a comparison of the 20 and 100 observation curves, implying 10 and 50 equipment failures, respectively. For the PDF curve with 100 observations, it is seen that the entire curve lies within the range of 40-60 percent failure probability. In contrast, the curve with 20 observations has its PDF curve ranging between 20-80 percent failure probability.

Exhibit ES-2 illustrates the CDF curves, which are the integrals of the functions shown in Exhibit ES-1. Referring to the same observation sizes of 20 and 100, it is seen that with 20 observations one is 83 percent confident that the failure rate is below 60 percent, while with 100 observations the confidence level is 98 percent. Therefore, by choosing a large enough sample size (number of observations), one has the ability to predict with a high confidence level the actual equipment failure probabilities.

Exhibit ES-3 illustrates the PDF curves for observed failure rates of 10, 20, and 50 percent with a sample size of 20. The PDFs for the same failure rates are presented in Exhibit ES-4 for a sample size of 200. Note that the curves in Exhibit ES-4 are more similar in shape to one another than those in Exhibit ES-3. This implies that with large sample sizes, the effect of variation in the observed failure rate is to translate the location of the PDF curve over the X-axis. This phenomenon gives credence to the selection of the non-informative prior distribution, for it is solely the observed failure rate (at large observation sizes) which dictates the PDF curve position.

Results of this study indicate the validity of selecting the non-informative prior distribution. As the number of observations increases, the effects of the prior are minimized and the effects of the data observations dominate. In addition, with larger sample sizes, one can predict with greater confidence the failure probability of an equipment. As indicated in Exhibit ES-5, roughly 150 observations are needed, with no failures, to obtain a confidence level of 90 percent that the actual failure rate of the equipment is below 1 percent. Use of the Bayesian approach in such a manner can be a very valuable tool for EMP test planning. It can provide insight into how many observations are needed to obtain a particular confidence level.

EXHIBIT ES-3
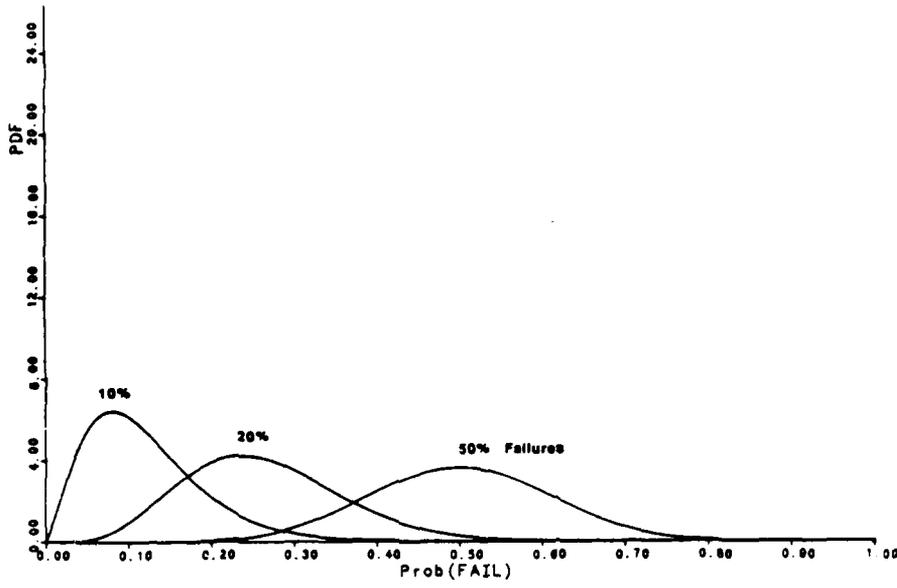Probability Density Functions: Sample Size of 20

EXHIBIT ES-4
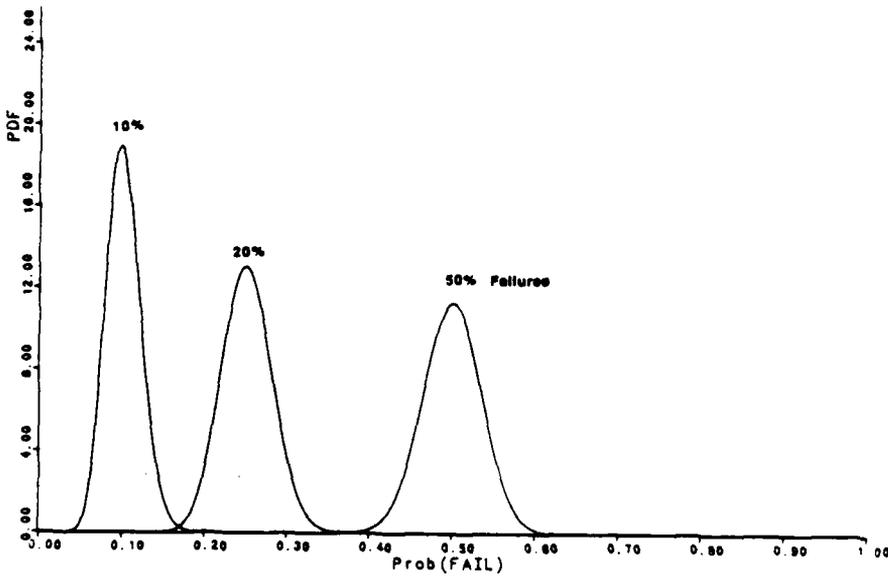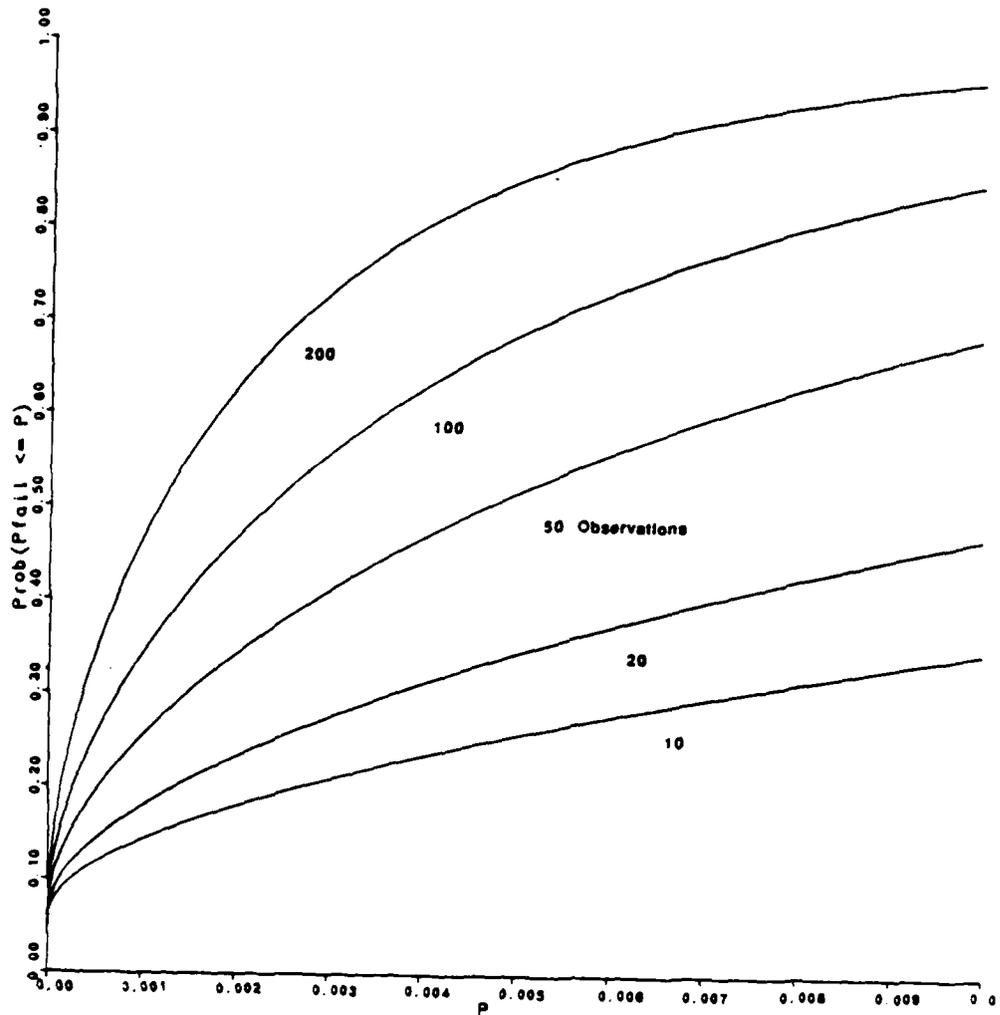Probability Density Functions: Sample Size of 200

Cumulative Distribution Functions: No Observed Failures

1.0   INTRODUCTION

1.0   INTRODUCTION

This report presents a sensitivity study of a statistical model used by the the Office of the Manager, National Communications System (OMNCS) to estimate failure probability distributions of telecommunications switches and transmission facilities subjected to High-Altitude Electromagnetic Pulse (HEMP). The model is based on an application of Bayesian statistical theory, and it uses the results of HEMP testing programs as input data. The outputs of this Bayesian model support estimates of the effects of HEMP on the connectivity of major telecommunications networks used for National Security Emergency Preparedness (NSEP) missions. This study will help explain the meaning and limitations of such network connectivity analyses by quantifying the sensitivity of the Bayesian model outputs to variations in its inputs.

1.1   BACKGROUND

The OMNCS has been directed by the National Security Council to identify actions necessary to mitigate the potential effects of HEMP on the Nation's telecommunications assets supporting NSEP requirements. In response, the OMNCS has established an EMP mitigation program with the objective of identifying and, where appropriate, mitigating the effects of HFMP on NSEP telecommunications capabilities. This program includes coordinated physical testing of selected telecommunications equipment types in simulated HEMP environments, combined with computer-based quantitative analyses of HEMP effects on telecommunications networks. The focus of testing and network analysis for this program to date has been the Public Switched Network (PSN), the basis for a substantial portion of the Nation's NSEP telecommunications capabilities. In addition, special attention has been given to the estimation of the effects of HEMP on the OMNCS-sponsored Nationwide Emergency Telecommunications System (NETS). The goal of these HEMP-related activities is to develop a cost-effective, comprehensive HEMP mitigation strategy, focused on the PSN and including NETS. When combined with other OMNCS initiatives, such a strategy could significantly enhance the ability to maintain and reconstitute NSEP telecommunications following a nuclear attack against the United States that includes high-altitude exoatmosphere detonations.

To date, the OMNCS has performed extensive HEMP testing of selected telecommunications equipment types deemed important to fulfilling NSEP requirements. Recently these equipments have included the T1 digital cable transmission system, the FT3C multi-mode fiber optic transmission system, the AT&T Technologies 5ESS switching system, and a prototype OMNCSsponsored NETS call control module cabinet. Other types of equipment have been HEMP-tested by agencies other than the OMNCS in earlier test programs. These equipments include the AT&T 1ESS switching system, the TD-2 digital microwave transmission system, and the L4 analog cable transmission system. An equipment HEMP effects evaluation based on the testing to date is presented in Reference 1. While not all types of PSN switching and transmission systems have been tested, the knowledge that has been obtained to date provides a basis for studying the potential effects of HEMP on the entire PSN and those elements of the PSN selected for use in NETS.

Reference 2 introduces the methodology used by the OMNCS to estimate HEMP effects on network connectivity. As shown in Exhibit 1-1, the methodology is based on a Monte Carlo network-level analysis approach, with input data consisting of network topology information and network element estimated failure probability distributions. The required failure probability distributions come from the Bayesian model described in this report. Reference 3 presents the results of applying the network-level analysis approach to a major portion of the PSN. The network analyzed in that report consists of those PSN switches and transmission facilities which comprised the AT&T toll network prior to divestiture. Finally, Reference 4 presents the results of a network-level analysis of estimated HEMP effects on NETS.

1.2 PURPOSE

This report focuses on the sensitivity of the Bayesian failure model outputs to variations in its inputs; it does not address the resultant sensitivity of network-level analysis results to variations in the Bayesian model inputs. Such network-level sensitivity studies are being conducted by the OMNCS in a related effort, and will incorporate the results of this study.

The sensitivity study in this report is in direct response to comments received by the OMNCS from members of the EMP analysis community concerned with interpreting the outputs of the Bayesian model. Specific questions have

EXHIBIT 1-1
Network-Level HEMP Effects Analysis Approach

```
┌─────────────────────────────────┐
│   ╭───────────────────╮         │
│   │   EQUIPMENT        │         │     1. FAILURE MODEL
│   │   EMP TEST         │         │
│   │   DATA             │         │
│   ╰───────────────────╯         │
│            │                     │
│            ▼                     │
│   ┌───────────────────┐         │
│   │   EQUIPMENT        │         │
│   │   FAILURE          │         │
│   │   PROBABILITY      │         │
│   │   MODEL            │         │
│   └───────────────────┘         │
│            │                     │
└────────────┼─────────────────────┘
```

```
┌────────────────────────────────────────┐
│            ▼                             │
│   ╭───────────────────╮                 │
│   │   EQUIPMENT        │                 │
│   │   FAILURE          │                 │
│   │   PROBABILITY      │                 │
│   │   DISTRIBUTIONS    │                 │
│   ╰───────────────────╯                 │
│            │                             │   2. NETWORK MODEL
│            ▼                             │
│  ╭──────────╮    ┌──────────────┐       │
│  │ NETWORK  │───▶│  NETWORK     │       │
│  │TOPOLOGIES│    │  CONNECTIVITY│       │
│  ╰──────────╯    │  ANALYSIS    │       │
│                  │  MODEL       │       │
│                  └──────────────┘       │
│                         │                │
│                         ▼                │
│                  ╭──────────────╮       │
│                  │  NETWORK     │       │
│                  │  EMP         │       │
│                  │  EFFECTS     │       │
│                  │  ESTIMATES   │       │
│                  ╰──────────────╯       │
└────────────────────────────────────────┘
```
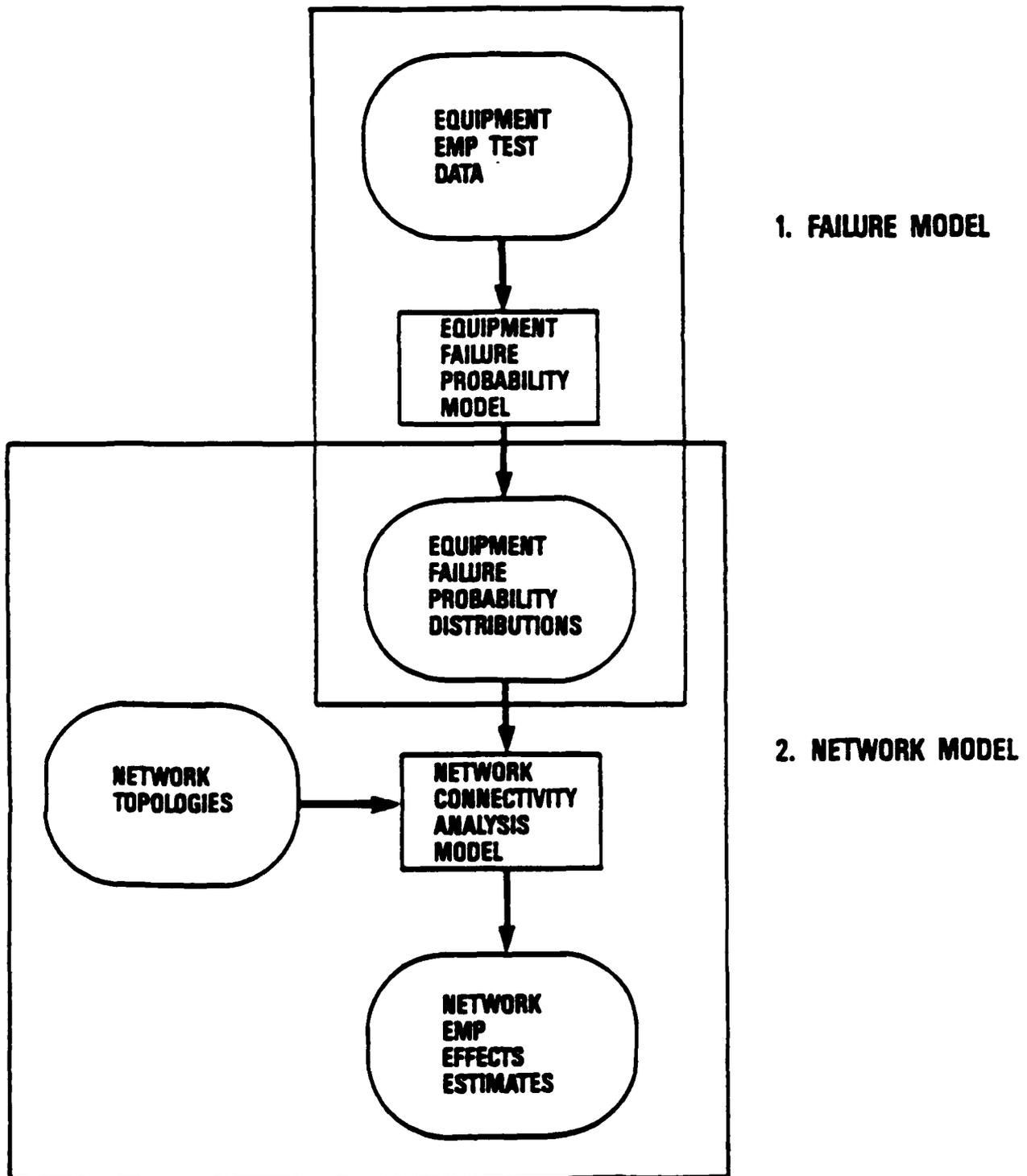
been raised regarding the response of the output posterior failure probability distribution to the assumed prior distribution. The rationale and consequences of selecting the "noninformative" prior distribution for the Bayesian application are offered in response to these concerns.

1.3 APPROACH

Three basic inputs to the Bayesian failure probability model, as shown in Exhibit 1-2, are used here. In addition to an assumed prior distribution, these inputs relate to test sample sizes and numbers of observed failures from HEMP-induced equipment failure testing. The approach used in this study to investigate model sensitivity is to address the three types of input information separately. Thus, if there is specific concern with, for instance, sensitivity of estimated failure probability distribution to test sample size, the sensitivity results for that input parameter can be referenced independently of the other two inputs.

For the two discrete integer model inputs--sample size and number of failures--sensitivity is studied parametrically. That is, fixed values are selected for each input, and the resultant model outputs are plotted and interpreted. No attempt is made to compare the fixed values used in this parametric analysis to actual test data observed in the test programs conducted by the OMNCS or other agencies. Thus, this sensitivity study is generally applicable and is not constrained by specific testing.

The third Bayesian model input--the assumed prior distribution of failure probability--is studied by showing the theoretical effect of the selected noninformative prior distribution on the output posterior distribution. An alternative approach postulating prior distributions other than the noninformative prior distribution and observing the corresponding posterior distributions response has been suggested. Such an approach has been rejected for this study because, even though alternative prior distributions are available for actual analyses, no qualitative or quantitative basis exists for selecting a prior other than that which is least informative relative to the data. Thus, this study presents the theoretic basis for deriving the mathematical representation of the noninformative prior distribution, and justifies its use on theoretical grounds. The consequences of using the

EXHIBIT 1-2

Bayesian Failure Probability Model:
Functional Flow Diagram

```
   ┌─────────────┐   ┌──────────────────┐   ┌──────────────┐
   │   PRIOR     │   │     # OF         │   │    # OF      │
   │DISTRIBUTION │   │  OBSERVATIONS    │   │  OBSERVED    │
   │             │   │ (SAMPLE  SIZE)   │   │  FAILURES    │
   └─────────────┘   └──────────────────┘   └──────────────┘
            \                │                 /
             \               │                /
              \              ▼               /
           ┌──────────────────────────────┐
           │          BAYESIAN            │
           │          FAILURE             │
           │        PROBABILITY           │
           │           MODEL              │
           └──────────────────────────────┘
                         │
                         ▼
                 ┌─────────────────┐
                 │    FAILURE      │
                 │  PROBABILITY    │
                 │ DISTRIBUTIONS   │
                 └─────────────────┘
```
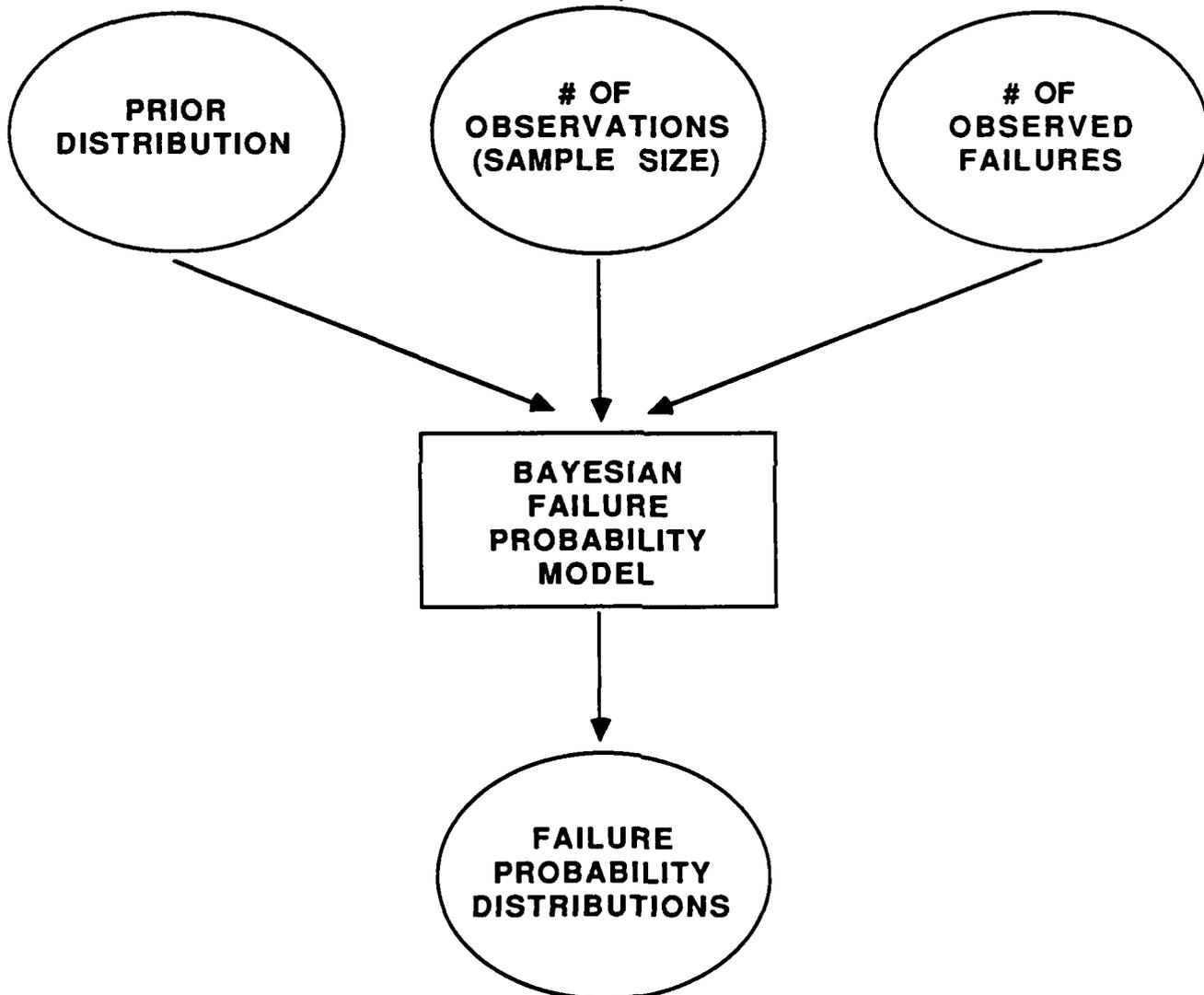
noninformative prior distribution are shown graphically as they affect the shape of the output posterior distributions for different test conditions.

1.4 ORGANIZATION

This report describes the Bayesian model, presents the results of the three individual sensitivity studies corresponding to the three different model inputs, and summarizes the findings of the studies for consolidated reference. An appendix shows a listing of the computer code which implements the Bayesian model, followed by the list of references cited throughout this report.

2.0 BAYESIAN METHODOLOGY

2.0 BAYESIAN METHODOLOGY

This chapter presents the rationale and mathematical derivation for the Bayesian statistical interpretation of test data to estimate EMP-induced telecommunications equipment failure probabilities. The numerical approximations used to actually calculate the Bayesian posterior failure probability distributions are presented, and the computer code which implements the numerical approximations is introduced. This chapter draws heavily upon information presented in Reference 5, including some verbatim passages.

2.1 RATIONALE

The use of data from EMP tests of telecommunications equipment to estimate the EMP-induced failure probabilities of those equipments can be treated as a problem in statistical inference. As a statistical inference tool, Bayesian analysis provides an approach for incorporating test data and assumptions together into a single quantitative analysis methodology to infer information indicated by the data. A prior distribution, which is supposed to represent what is known about unknown parameters before data is available, plays an important role in Bayesian analysis. Such a distribution can be used to represent prior knowledge or ignorance about the parameter under study--in this case, failure probability.

In problems of scientific inference where empirical test data are available, it is often desirable for all estimates regarding the characteristics of the parameter under study to be based on the available data. Consequently, it is often appropriate to conduct the analysis as if a state of relative ignorance existed a priori. In this analysis, a "noninformative" prior distribution is used. The aim is to obtain an inference which would be appropriate for an unprejudiced observer. It should be acknowledged that, even within the statistical community, there is some uneasiness about the use of prior distributions, which is often associated with the fear that the prior may dominate and distort the information contained in the empirical data. By careful choice of an inferential parametric model structure and an appropriate noninformative prior, Bayesian analysis can produce the reverse of what is feared.

2.2 DERIVATION

The use of Bayesian inference to characterize EMP-induced equipment failure probabilities is based on an application of Bayes' Theorem. Suppose that y is an observation whose conditional probability distribution $f(y:p)$ depends on the value of a parameter p. Suppose also that p itself has a probability distribution $f(p)$. Then, $f(y:p)f(p) = f(y,p) = f(p:y)f(y)$. Here, $f(y,p)$ is the joint probability distribution for y and p, $f(p:y)$ is the conditional probability distribution for p dependent upon y, and $f(y)$ is the probability distribution for y. Given the observed data y, the conditional distribution of p is

$$f(p:y) = \frac{f(y:p)\ f(p)}{f(y)}. \qquad (2.1)$$

Also, if p is continuous, $f(y)$ can be written as

$$f(y) = E\ [f(y:p)] = C^{-1} = \int f(y:p)\ f(p)\ dp$$

where the integral is taken over the admissible range of p, and where $E[f(y:p)]$ is the mathematical expectation of $f(y:p)$ with respect to the distribution $f(p)$. The quantity C is merely a "normalizing" constant necessary to ensure that the distribution $f(p:y)$ integrates to one.

Equation 2.1 can thus be written alternatively as

$$f(p:y) = C\ f(y:p)\ f(p). \qquad (2.2)$$

The statement of (2.1) or its equivalent (2.2) is usually referred to as Bayes' theorem. In this expression, $f(p)$, which tells what is known about p without knowledge of the data, is called the prior distribution of p, or the distribution of p a priori. Correspondingly, $f(p:y)$, which tells what is known about p given knowledge of the data, is called the posterior distribution of p given y, or the distribution of p a posteriori.

Now, given the data y, $f(y:p)$ in equation (2.2) may be regarded as a function not of y but of p. When so regarded, it is called the likelihood function of p for given y and can be written $L(p:y)$. Bayes formula can thus be written as:

$$f(p:y) = L(p:y)\ f(p). \qquad (2.3)$$

In other words, Bayes' theorem says that the probability distribution for the parameter p posterior to the data y

is proportional to the product of the distribution for p prior to the data and the likelihood for p given y. That is,

posterior distribution $\propto$ likelihood x prior distribution.

The likelihood function L(p:y) plays a very important role in Bayes' formula. It is the function through which the data y modifies prior knowledge of p. It can therefore be regarded as representing the information about p coming from the data.

The likelihood function is defined up to a multiplicative constant, that is, multiplication by a constant leaves the likelihood unchanged. This is in accord with the role it plays in Bayes' formula, since multiplying the likelihood function by an arbitrary constant will have no effect on the posterior distribution of p. The constant will cancel upon normalizing the product on the right hand side of equation (2.3). Only the relative value of the likelihood function is important. Thus, much of the derivation which follows uses proportionality statements rather than equations for the sake of simplicity. Also, when the integral $\int L(p:y)dp$, taken over the admissible range of p, is finite, then occasionally it will be convenient to refer to the quantity

$$\frac{L(p:y)}{\int L(p:y)dp}.$$

This expression is called the standardized likelihood, that is, the likelihood scaled so that the area under its curve is one.

The contribution of the prior distribution in helping to determine the posterior distribution of a parameter p is dependent on its sharpness or flatness in relation to the sharpness or flatness of the likelihood with which it is combined. For example, after a single test data observation, the likelihood might not be sharply peaked relative to a particular prior distribution. The prior distribution would therefore be influential in determining the posterior distribution. On the other hand, after multiple observations, the prior distribution might be relatively flat compared with the likelihood function. Such a prior would therefore not be very influential in deciding the corresponding posterior distribution of p. It would be said that, after multiple observations, the prior is dominated by the likelihood. Exhibit 2-1

EXHIBIT 2-1
Dominant Likelihood and Prior Distributions

Likelihood

Prior Distribution

p →

(a) Dominant Likelihood

Prior Distribution

Likelihood

p →
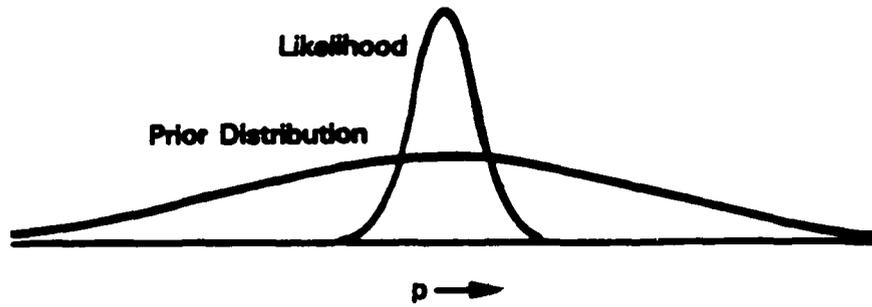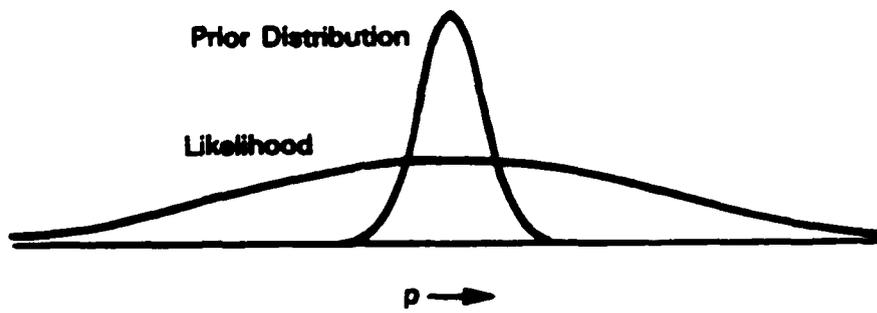
(b) Dominant Prior

illustrates inferential analyses in which the likelihood
dominates the prior (a) and vice versa (b).

It is often appropriate to analyze data from scien-
tific investigations on the assumption that the likelihood
dominates the prior. Two reasons for this are:

(1) A scientific investigation, such as controlled
     EMP testing of telecommunications equipment, is
     not usually undertaken unless information
     supplied by the investigation is likely to be
     considerably move precise or valid than infor-
     mation already available. In brief, expensive
     testing and analysis is not usually undertaken
     unless it is likely to increase knowledge by a
     substantial amount.

(2) Even if strong prior beliefs are held about the
     value of a parameter, nevertheless, in reporting
     analysis results it would usually be appropriate
     and most convincing if the data were analyzed
     against a reference prior which is dominated by
     the likelihood. Then, irrespective of what any-
     one believed at the outset, the posterior dis-
     tribution would represent what someone who a
     priori knew very little about the parameter
     should believe in light of the observed data.

The term reference prior, introduced above, refers to
a prior which is convenient to use as a standard. In
principle, a reference prior may or may not be dominated
by the likelihood. In general, a prior which is dominated
by the likelihood is one which does not change very much
over the region in which the likelihood is appreciable,
and does not assume large values outside the range, as in
figure (a) of Exhibit 2-1. A prior distribution which has
these properties is referred to as a locally uniform
prior. For such a prior distribution, the result from
Bayes' formula can be approximated by substituting a con-
stant for the prior distribution. Thus, for a locally
uniform prior, the posterior distribution is approximately
numerically equal to the standardized likelihood.

At this point, an argument is presented for choosing a
particular metric, in terms of which, a locally uniform
prior can be regarded as noninformative about the param-
eters. It is important to bear in mind that one can never
be in a state of complete ignorance; further, the state-
ment "knowing little a priori" can only have meaning
relative to the information provided by an experiment.
For example, prior knowledge may be substantial compared

with the information from just a single experimental
observation, but it would likely be noninformative rela-
tive to that from a very large number of observations. A
prior distribution is supposed to represent knowledge
about a parameter before the outcome of a projected
experiment is known. Thus, the main issue is how to
select a prior which provides little information relative
to what is expected to be provided by the intended
experiment.

In general, suppose it is possible to express the
unknown parameter p in terms of a metric q(p) so that the
corresponding likelihood is data translated. This means
that the likelihood curve for g(p) is completely deter-
mined a priori except for its location, which depends on
the data yet to be observed. Then, to say little is known
a priori relative to what the data will show, may be ex-
pressed by saying one is almost equally willing to accept
one value of g(p) as another. This state of indifference
may be expressed by taking q(p) to be locally uniform a
priori, and the resulting prior distribution is called
noninformative for g(p) with respect to the data. In
general, if the noninformative prior is locally uniform in
g(p), then the corresponding noninformative prior is
locally proportional to dg/dp, assuming the transformation
is one to one (Reference 5).

For the case of n independent trials, in each of which
the probability of failure is p, a binomial model for the
number of observed failures may be employed. The prob-
ability of y successes in n trials is given by the
binomial distribution:

$$\text{Prob } (y:p) = \frac{n!}{y!(n - y)!} \, p^y(1 - p)^{n-y}, \quad y = 0, 1, \ldots, n,$$

so that the likelihood is:

$$L(p:y) \propto p^y(1 - p)^{n-y}.$$

For the binomial distribution, a transformed metric
for which the likelihood curve is very nearly data trans-
lated, and for which a locally uniform prior distribution
is nearly noninformative is given by:

$$q(p) = \sin^{-1}\sqrt{p}.$$

(Chapter 3.0 shows why this particular metric meets
the requirements described above for producing a noninfor-
mative prior distribution.) This, in turn, implies that

the corresponding nearly noninformative prior for p is proportional to:

$$f(p) \propto \frac{dq}{dp} \propto [p(1 - p)]^{-1/2}.$$

If this approximately noninformative prior is employed, then the posterior distribution for p given y becomes proportional to:

$$f(p:y) \propto p^{y - 1/2}(1 - p)^{n - y - 1/2}, \quad 0 < p < 1.$$

After substitution of the appropriate normalizing constant, the corresponding posterior distribution for p is the beta distribution, whose probability density function is given by:

$$f(p:y) = \frac{\Gamma(n + 1)}{\Gamma(y + 1/2) \Gamma(n-y + 1/2)} p^{y-1/2} (1-p)^{n-y-1/2}, \quad 0 < p < 1 \quad .$$

The corresponding cumulative distribution function is therefore:

$$F_p (x:n,y) = \text{Prob} (0 < p < x)$$

$$= \frac{\Gamma(n + 1)}{\Gamma(y + 1/2) \Gamma(n-y + 1/2)} \int_0^x p^{y-1/2} (1-p)^{n-y-1/2} dp$$

for $0 < x < 1$.

This cumulative distribution function for p describes the Bayesian posterior distribution for the actual probability of equipment failure given that y failures were observed in n trials, assuming a noninformative prior distribution for failure probability.

2.3 NUMERICAL APPROXIMATION

The posterior cumulative distribution function for failure probability is the basic output of the Bayesian model described here. To calculate probability values using this function, a numerical approximation is employed. The numerical approximation begins by noting that the cumulative distribution function above has the form of an incomplete Beta function, $I_x(a,b)$, with

    a = y + 1/2
    b = n - y + 1/2
    y = number of observed failures
    n = total observations (sample size)

2-7

because

$$I_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}\,dt \qquad 0 < x < 1.$$

The incomplete Beta function is described in Reference 6, Equation 26.5.1. This function can be approximated numerically using the cumulative distribution function, P(z), of the standard normal distribution as follows (from Reference 6, Equation 26.5.21):

$$I_x(a,b) = P(z)$$

where:

$$z = \frac{3[v(1-1/9b) - w(1-1/9a)]}{[v^2/b + w^2/a]^{1/2}}$$

$$v = (bx)^{1/3}$$

$$w = [a(1-x)]^{1/3}.$$

This approximation to the incomplete Beta function is accurate to within $\pm 5 \times 10^{-3}$ if $a+b > 6$.

The cumulative distribution function, P(z), of the standard normal distribution can be approximated numerically through a polynomial function as follows (Reference 6, Equation 26.2.19):

$$P(z) = 1 - (1/2)(1 + d_1 z + d_2 z^2 + d_3 z^3 + d_4 z^4 + d_5 z^5 + d_6 z^6)^{-16}$$

where:
$\quad d_1 = .0498673470$
$\quad d_2 = .0211410061$
$\quad d_3 = .0032776263$
$\quad d_4 = .0000380036$
$\quad d_5 = .0000488906$
$\quad d_6 = .0000053830.$

This approximation to the standard normal cumulative distribution function is accurate to within $\pm 1.5 \times 10^{-7}$.

Appendix A lists the computer code that actually calculates the Bayesian posterior failure probability cumulative distribution functions and approximates their associated probability density functions. The basic inputs to the model are:

n = The number of EMP test shots to which a piece of telecommunications equipment is subjected at a particular EMP stress level, or within a range of stress levels.

y = The number of failures observed out of the n observations above.

For a particular analysis, the decision of what constitutes an observation and a failure would have to be made. Also, although the OMNCS uses this model to characterize EMP-induced telecommunications equipment failure probabilities, the statistical methodology described above is applicable beyond the EMP context.

The output of this Bayesian model is a probability density function and a cumulative distribution function for failure probability corresponding to a particular (n,y) pair. For example, if a test is conducted in which there are 25 observations with exactly one resulting in a failure, the output Bayesian probability density function would be as shown in Exhibit 2-2. The function shows that the actual failure probability (x-axis) of the unit under test is most likely near 0.04 (where the curve peaks). However, there is some possibility that the actual failure probability is lower or higher than 0.04, and may be as high as 0.25 (where the curve approaches zero on the x-axis). The probability density function gives a graphic indication of the most likely value of the actual failure probability, near where the curve peaks. Also, the width of the curve indicates graphically how likely the actual value of the parameter is near the curve's location of central tendency. However, the probability density function does not directly allow numerical probabilistic statements regarding the actual value of the parameter. Instead, the cumulative distribution function is used.

Exhibit 2-3 shows the cumulative distribution function corresponding to the case of 25 observations and one failure. The y-axis gives the probability that the actual failure probability is less than the indicated value on the x-axis. In this example, there is a probability of 0.50 (y-axis) that the actual failure probability is less than 0.05 (x-axis). The actual failure probability is almost certainly less than 0.25 in this example, where the cumulative distribution function approaches 1.0.

EXHIBIT 2-2
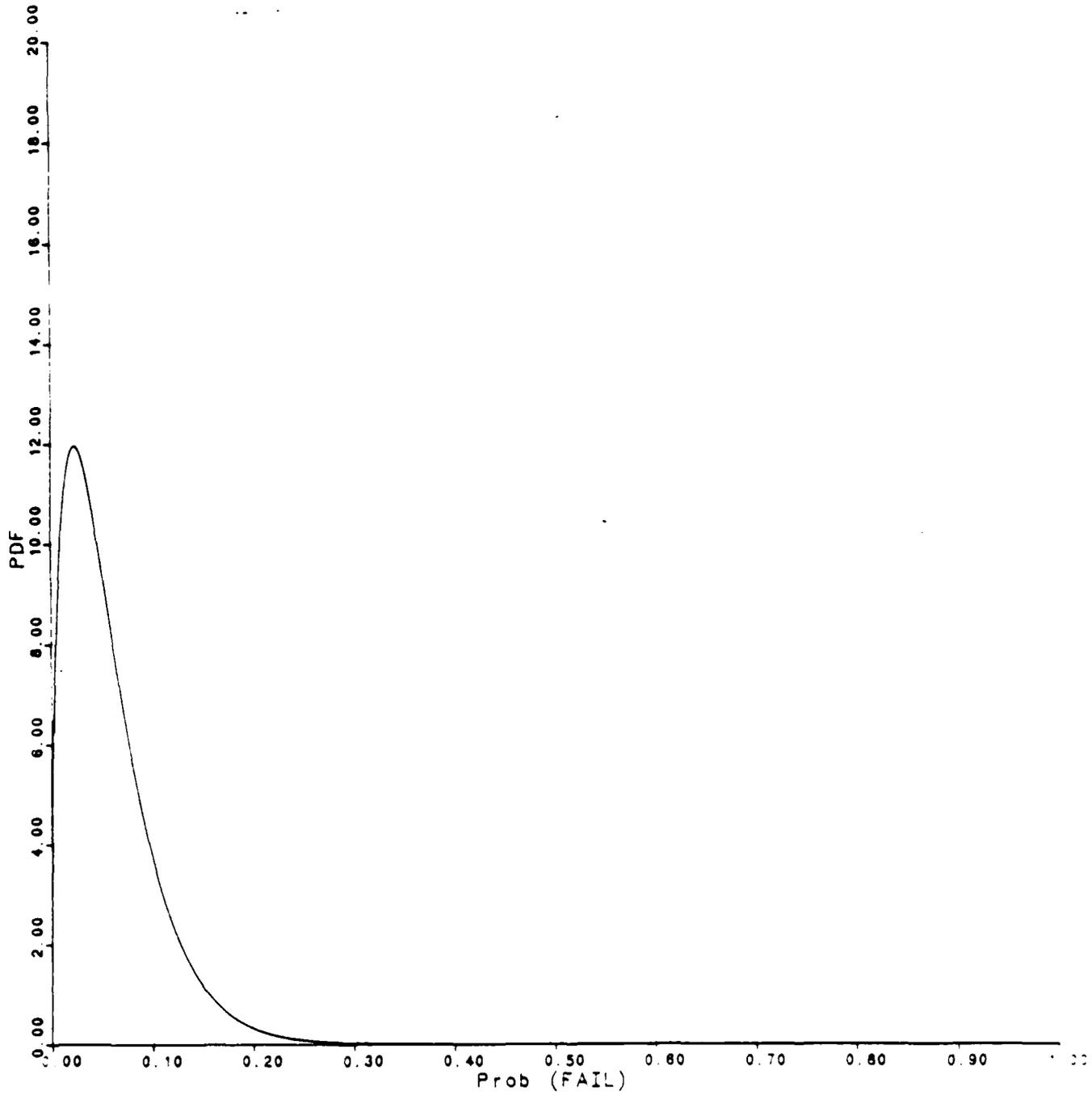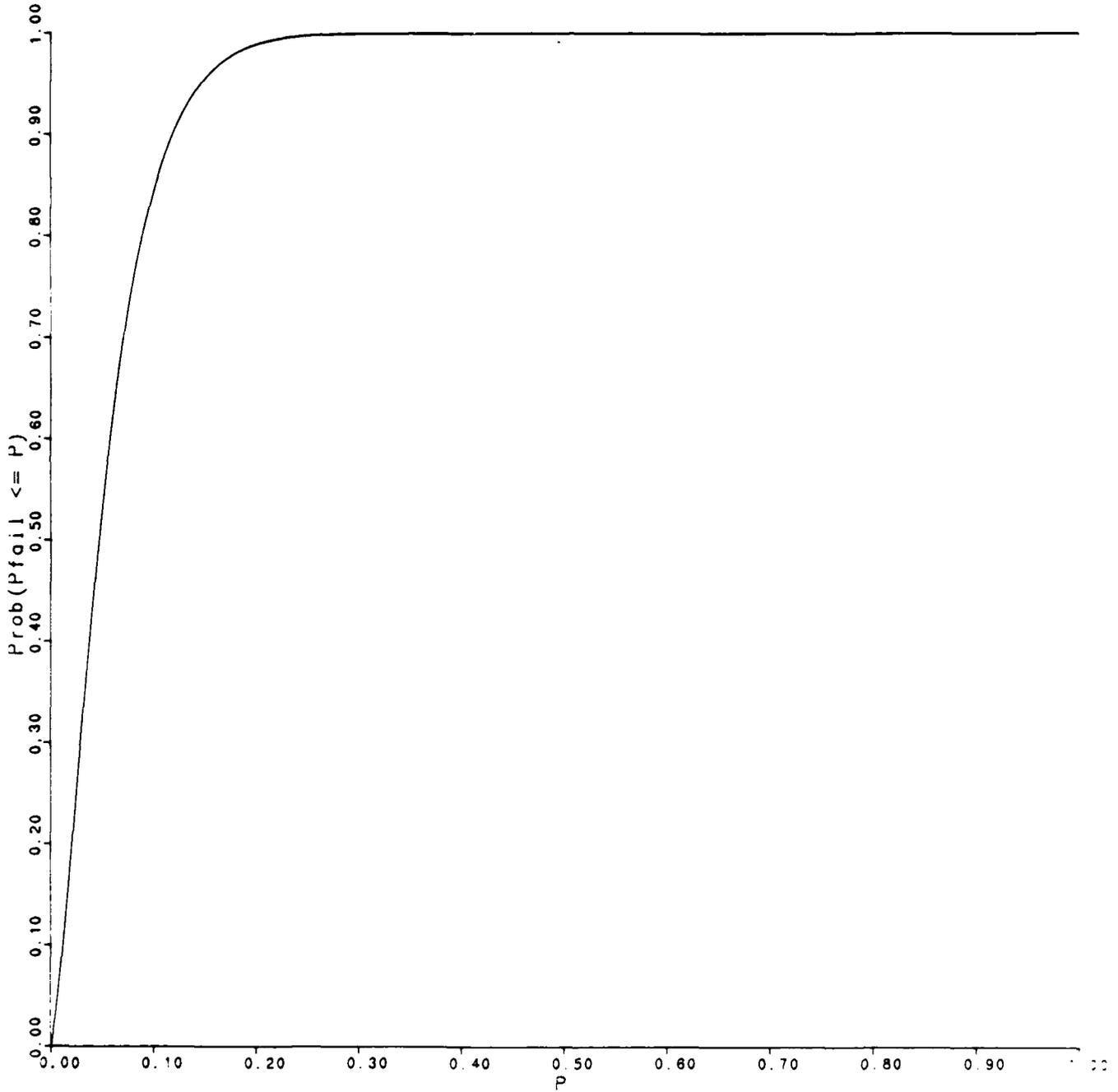Example Bayesian Probability Density Function

EXHIBIT 2-3
Example Cumulative Distribution Function

3.0 PRIOR DISTRIBUTION

3.0 PRIOR DISTRIBUTION

In this chapter, the prior distribution used to derive
the Bayesian posterior probability distribution in the
previous chapter is shown to meet the requirements for a
noninformative prior. The approach used herein to inves-
tigate the sensitivity of the Bayesian posterior distri-
bution to the prior was selected over simple examination
of how the posterior distribution responds to alternate
priors. There is no basis for selecting alternate priors
with justification stronger than that for the noninfor-
mative prior. Thus, this chapter focuses on demonstrating
the insensitivity of the posterior distribution to minor
changes in the prior by showing that the previously se-
lected transformation, $g(p)$, has a likelihood that is
approximately data translated. That is, the likelihood
for $g(p)$ is nearly independent of the data y except for
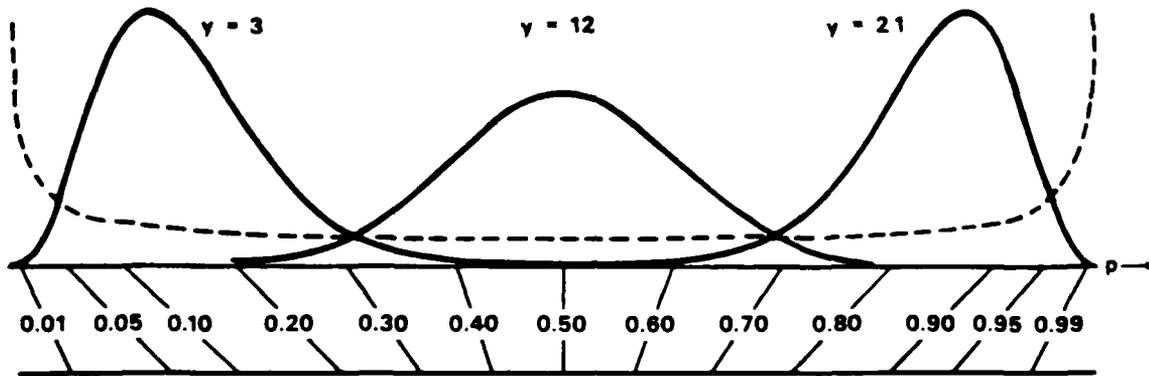its location.

3.1 NONINFORMATIVE PRIOR

As mentioned in the previous chapter, and documented
in Reference 5, if a noninformative prior is locally
uniform in $g(p)$, then the corresponding noninformative
prior for p is locally proportional to dg/dp, assuming the
transformation is one to one. The transformation selected
here, and used in the derivation of the Bayesian posterior
probability distribution, is:

$$g(p) = \sin^{-1}\sqrt{p}.$$

Suppose, for illustration, that there are $n=24$
trials. Then figure (a) in Exhibit 3-1 shows the stan-
dardized likelihood for $y=3$, $y=12$, and $y=21$ observed fail-
ures. The standardized likelihoods have different shapes,
depending on the number of observed failures. Figure (b)
in Exhibit 3-1 is the corresponding diagram obtained by
plotting the standardized likelihood in the transformed
metric, $g(p)$. Because the standardized likelihood curves
are defined only up to a constant, there is no y-axis
indicated, and the curves show only proportionality.
Although in terms of $g(p)$, the likelihood curves are not
exactly identical in shape and spread, they are nearly
so. Thus, in this metric, the likelihood curve is nearly
data translated, and a locally uniform prior distribution

EXHIBIT 3-1
Standardized Likelihoods (Solid Curves)
and Noninformative Priors (Broken Curves)

a) The binomial mean p
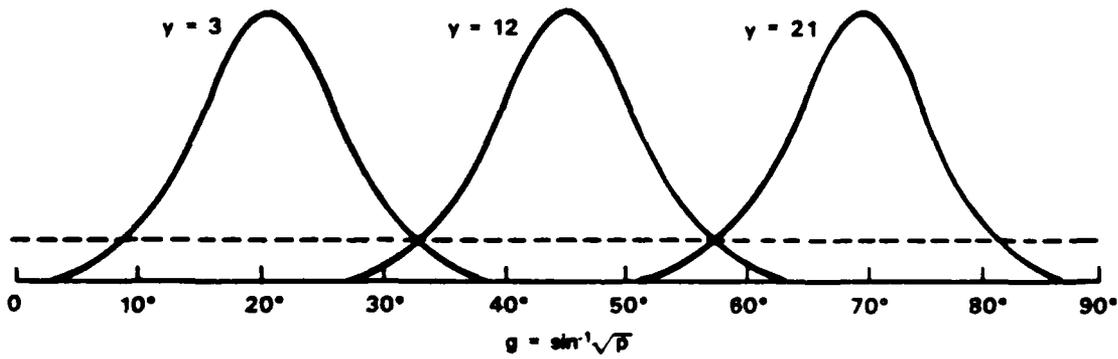
b) The transformed mean g = $\sin^{-1}\sqrt{p}$

(the dotted line in figure (b)) is nearly noninformative.
This in turn implies that the corresponding nearly nonin-
formative prior for p (the dotted line in figure (a)) is
proportional to dq/dp, as described in the previous
chapter:

$$f(p) \propto dq/dp \propto [p(1-p)]^{-1/2}.$$

3.2 SENSITIVITY

The insensitivity of the posterior distribution to
minor changes in the prior has been demonstrated logically
by the following steps:

- If a noninformative prior is locally uniform in a
  transformed metric, then the corresponding non-
  informative prior for the untransformed parameter
  is locally proportional to the derivative of the
  transformation with respect to the parameter,
  assuming the transformation is one to one.

- For the binomial mean, p, the transformation
  $\sin^{-1}\sqrt{p}$ yields a standardized likelihood that
  is approximately data translated; that is, its
  shape is independent of its location.

- A locally uniform prior for the transformed
  metric is noninformative, because, when it is
  multiplied by the standardized likelihood of the
  transformed metric, the resultant posterior
  distribution changes uniformly over its whole
  range.

- Therefore, the derivative of $\sin^{-1}\sqrt{p}$ yields the
  noninformative prior for the original param-
  eter, p.

The insensitivity of the posterior distribution to the
selected prior is thus inferred deductively.

4.0   SAMPLE SIZE

4.0 SAMPLE SIZE

This chapter investigates the sensitivity of the out-
put Bayesian failure probability distribution to vari-
ations in the input sample size parameter assuming a non-
informative prior distribution. For this investigation,
sample size is set at the discrete values of 10, 20, 50,
100, and 200 observations. For each sample size, the
output failure probability density functions and cumu-
lative distribution functions are calculated corresponding
to discrete percentages of observed failures of 50, 20,
and 10 percent. Also, for each discrete sample size, the
output failure probability cumulative distribution func-
tion is studied for the cases of no observed failures and
exactly on observed failure. After investigating the
effect of sample size on the posterior distributions,
conclusions are drawn regarding the implications of the
sensitivities for test planning and analyses.

4.1 SENSITIVITY

Exhibit 4-1 shows the Bayesian posterior failure prob-
ability density functions corresponding to discrete sample
sizes. The number of failures used for each sample size
is 50 percent of the sample size. Thus, for a sample size
equal to 10, the number of failures is 5; for a sample
size equal to 20, the number of failures is 10. The
resultant plot shows how the probability density functions
for the larger sample sizes are more tightly grouped about
the 50 percent failure probability point, while the lower
sample sizes are more widely distributed over the range of
possible failure probabilities. This narrow concentration
for high sample sizes and wide spread for low sample sizes
shows that, for a fixed percentage of observed failures,
larger sample sizes produce a greater likelihood of the
failure probability being located near a single point.
Put another way, in the case of 50 percent failure prob-
ability, an analyst would be more confident that the
actual failure probability is near 0.50 with a sample size
of 200 than with a sample size of 10.

Exhibits 4-2 and 4-3 show similar results for the
cases of 20 percent and 10 percent observed failures,
respectively. For the case of 20 percent observed fail-
ures in Exhibit 4-2, the larger sample sizes produce post-
erior failure probability density functions that are not
only grouped more tightly around a single probability
point (0.20), but are also less skewed than for smaller

EXHIBIT 4-1

Probability Density Functions:
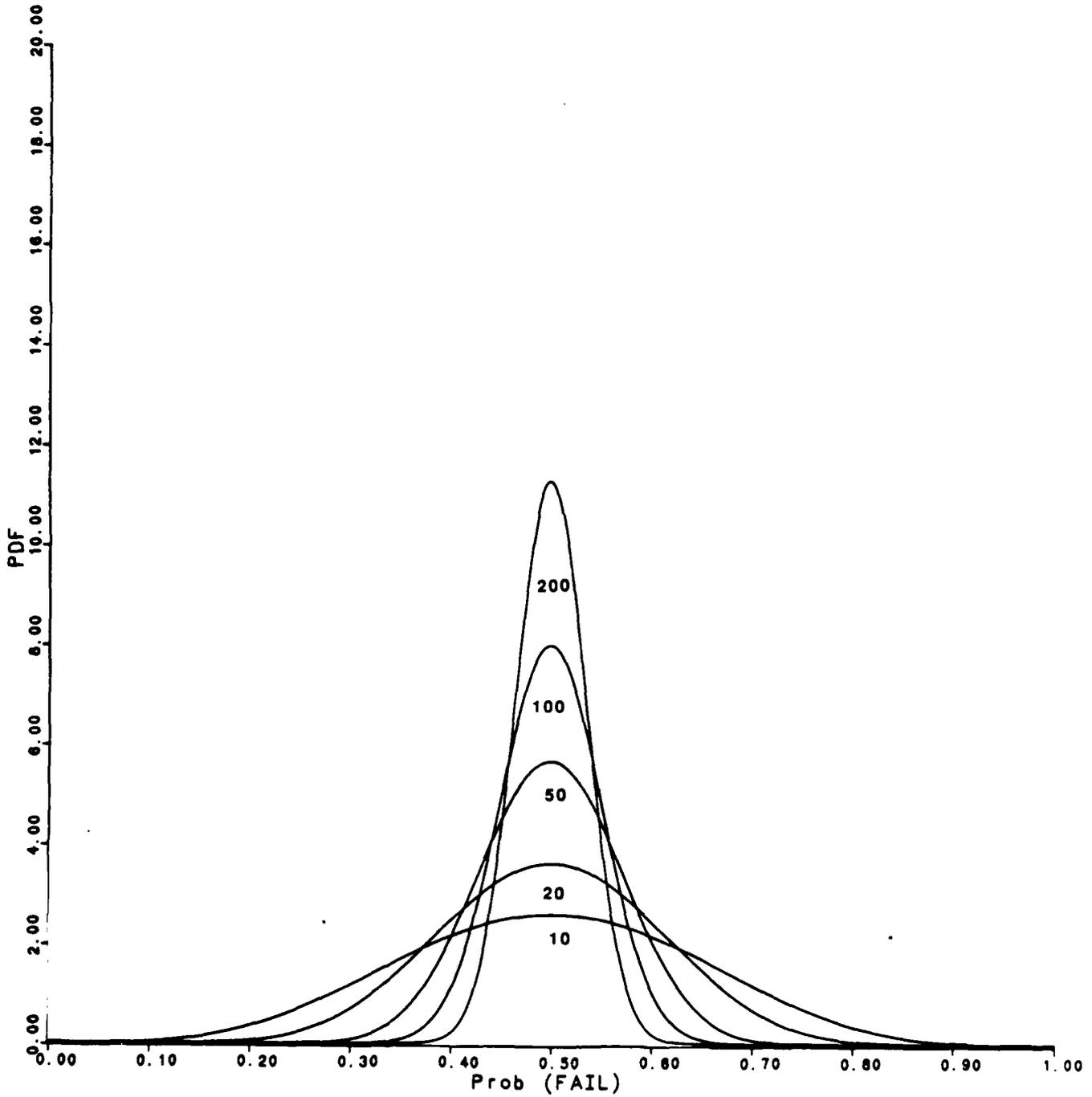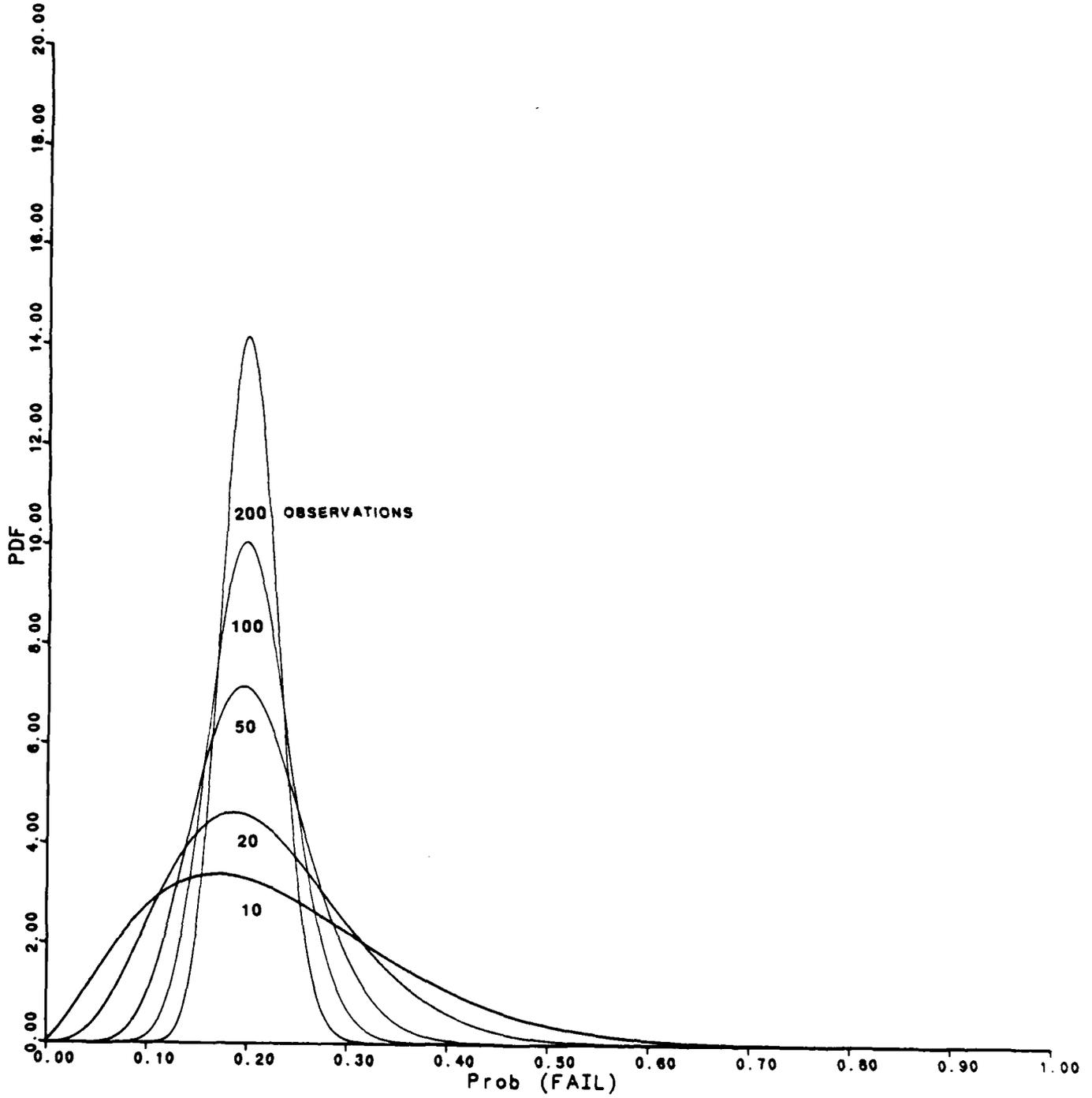50 Percent Observed Failures

4-2

EXHIBIT 4-2
Probability Density Functions:
20 Percent Observed Failures

sample sizes. This skewness for smaller sample sizes is even more pronounced for the case of 10 percent observed failures in Exhibit 4-3. Another observation from Exhibits 4-2 and 4-3 resulting from the skewness phenomenon is that, for small percentages of observed failures, the mode of the probability density functions becomes increasingly less than the mean of the distribution as the sample size decreases.

Exhibit 4-4 shows the cumulative distribution functions corresponding to the probability density functions in Exhibit 4-1 for 50 percent observed failures and the different discrete sample sizes. The curves show that, for smaller sample sizes, there is a greater estimated probability that the actual failure probability is much less than or much greater than the observed failure rate (0.50). For example, using a sample size of 20, there is an estimated probability of 0.25 (y-axis) that the actual failure probability is less than 0.40 (x-axis), while for a sample size of 200 the estimated probability is nearly zero. Likewise, for a sample size of 200, the estimated probability is nearly 1.00 that the actual failure probability is less than 0.60, while for a sample size of 10, the estimated probability is only 0.75. This means that for sample size of 10, there is still an estimated 27 percent probability that the actual failure probability is greater than 0.60 even though there were only 50 percent observed failures. For a large sample size of 200, such a probability approaches zero.

Exhibits 4-5 and 4-6 show the cumulative distribution functions corresponding to the cases of 20 percent and 10 percent observed failures, respectively, for the different discrete sample sizes. These cumulative distribution functions correspond to the probability density functions in Exhibits 4-2 and 4-3, respectively. Exhibit 4-5 shows, for example, that with a small sample size of 10 with 20 percent observed failures (2 observed failures), there is an estimated probability of only 0.90 (y-axis) that the failure probability is less than 0.40 (x-axis). Put another way, there is an estimated 10 percent probability that the actual failure probability is greater than 0.40 even though the fraction of observed failures is only 0.20. For sample sizes above 50, the estimated probability is effectively zero that the actual probability is greater than 0.40.

In the case of 10 percent failures (Exhibit 4-6), the implication of the skewed probability density functions for small sample sizes becomes apparent. This implication

EXHIBIT 4-3
Probability Density Functions:
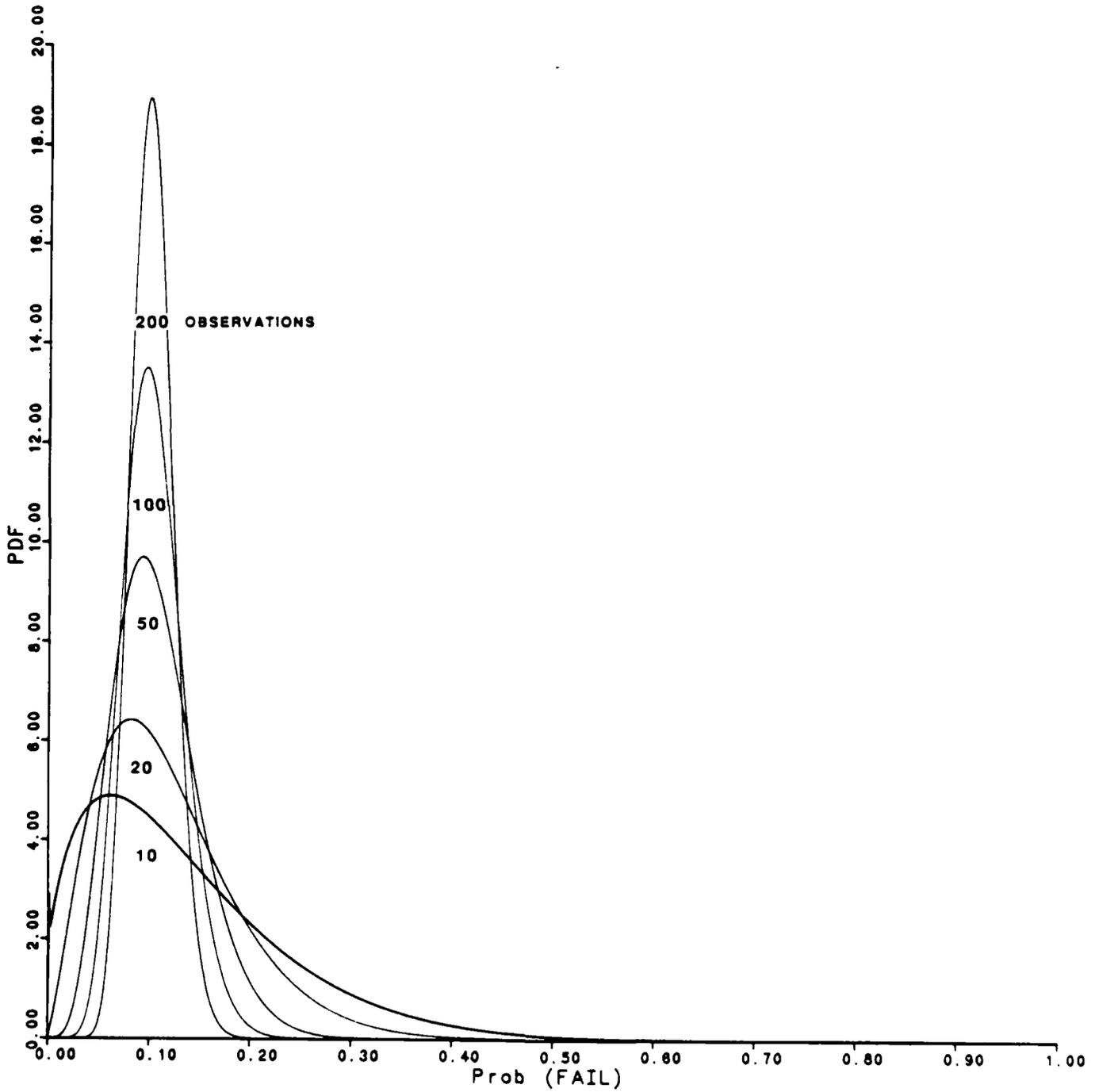## 10 Percent Observed Failures

EXHIBIT 4-4
Cumulative Distribution Functions:
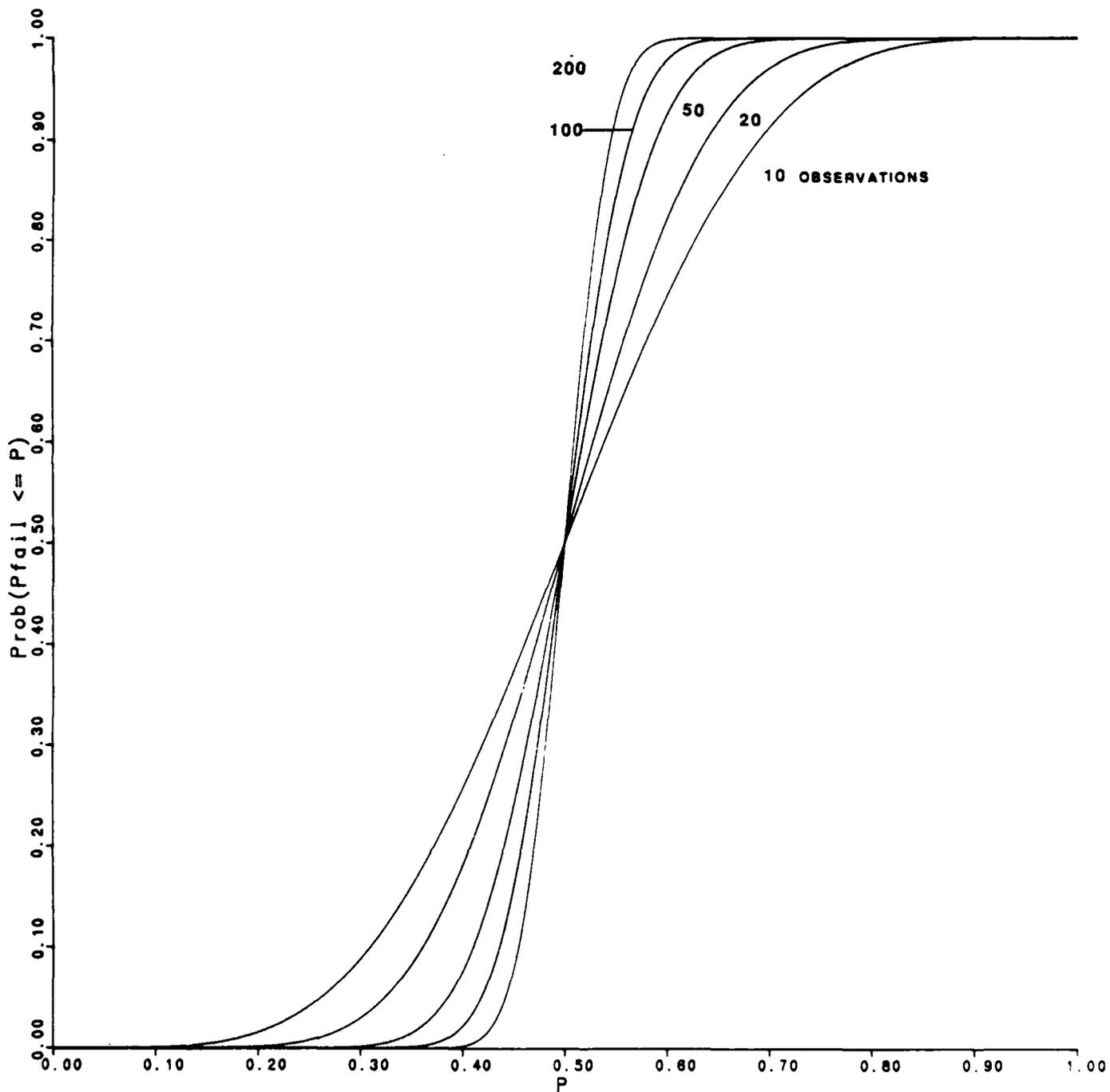50 Percent Observed Failures

EXHIBIT 4-5
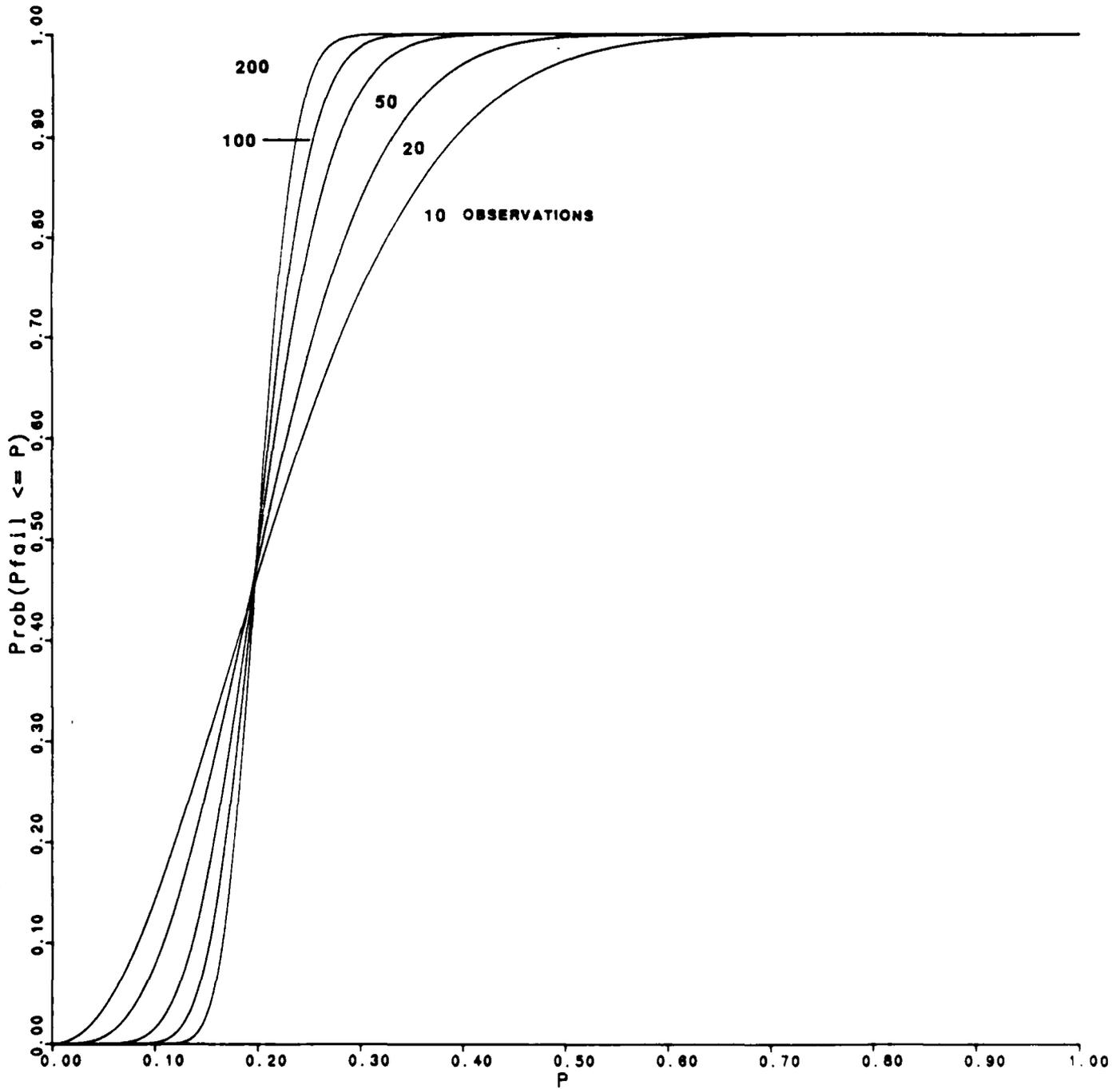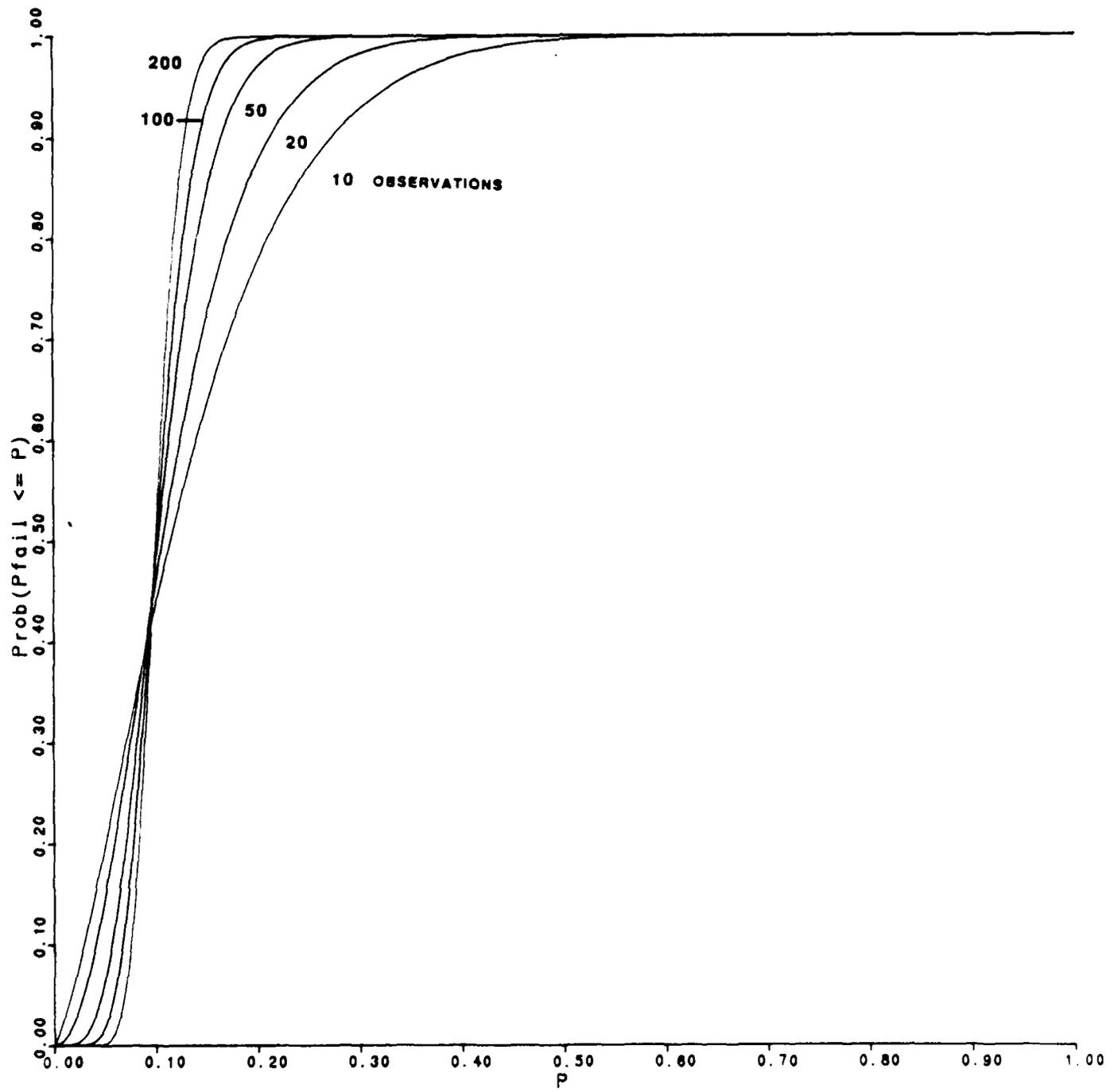Cumulative Distribution Functions:
20 Percent Observed Failures

EXHIBIT 4-6
Cumulative Distribution Functions:
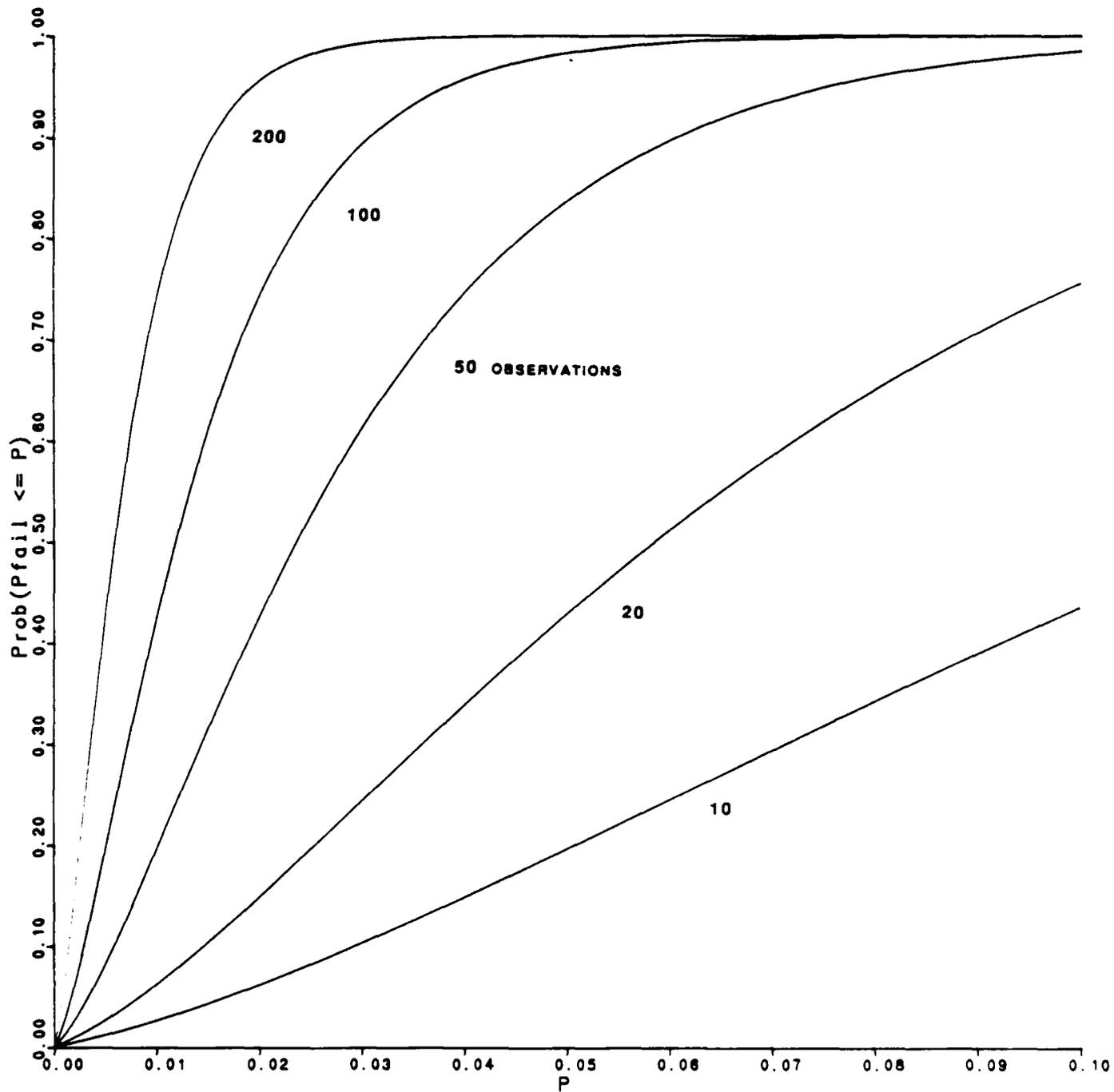10 Percent Observed Failures

may be demonstrated by using the cumulative distribution
functions in Exhibit 4-6 to construct confidence intervals
for failure probability. For example, a 90 percent,
2-sided confidence interval for failure probability may be
constructed by identifying the probability points (x-axis)
corresponding to the 0.05 and 0.95 cumulative prob-
abilities (y-axis). With a sample size of 10, this ap-
proach would yield a 90 percent, 2-sided confidence inter-
val of approximately [0.02, 0.32]. The center of this
2-sided confidence interval is at approximately 0.17,
which is greater than the 0.10 observed failure rate.
This indicates that the actual failure probability is
estimated as more likely to be greater than the observed
0.10 rate for small sample sizes. A 2-sided, 90-percent
confidence interval constructed similarly for the case of
a sample size of 200 is approximately [0.07, 0.13]. This
confidence interval is centered approximately at 0.10,
which is the observed failure rate, and it has a half
width of approximately 0.03. Thus, there is high confi-
dence with large sample sizes that the actual failure
probability is near the observed failure rate.

In addition to investigating the sensitivity of the
output Bayesian failure probability distributions to
sample size with fixed percentages of observed failures,
it is possible to investigate the effect of sample size in
the case where there is only one observed failure. For a
sample size of 10, this case corresponds to the 10 percent
observed failure rate addressed earlier. Exhibit 4-7
shows the cumulative distribution functions for the dif-
ferent discrete sample sizes and exactly one failure. It
should be noted that the curves in this exhibit are plot-
ted only over the x-axis probability range from 0.00 to
0.10. This is done to allow greater resolution on the
curves corresponding to high sample sizes, in the range
from 0.00 to 0.10 (y-axis). The curves show, for example,
that with a sample size of 10, there is a 0.44 (y-axis)
probability that the actual failure probability is less
than 0.10. For a sample size of 20, the probability that
the actual failure probability is less than 0.10 rises to
approximately 0.75. A sample size of 50 yields a 98 per-
cent probability that the actual failure probability is
less than 0.10, while sample sizes above 100 effectively
yield a 100 percent probability that the failure prob-
ability is less than 0.10.

A similar analysis of the curves can be performed for
the 0.01 probability point (x-axis). A sample size of 10
yields only about a 0.03 probability (y-axis) that the
actual failure probability is less than 0.01. With a

4-9

EXHIBIT 4-7
Cumulative Distribution Functions:
One Observed Failure

sample size of 200, the probability rises to 0.75 that the actual failure probability is less than 0.01.

Two-sided, 90 percent confidence intervals for failure probability can also be constructed from the cumulative distribution functions in Exhibit 4-7 for sample sizes of 50, 100, and 200. Such confidence intervals cannot be constructed from this exhibit for sample sizes of 10 to 20 because the 95 percent (y-axis) cumulative probability points are not plotted beyond the 0.10 x-axis range. A 90 percent, 2-sided confidence interval for the actual failure probability with a sample size of 100 is [0.002, 0.038]. This confidence interval is not centered on 0.01, which is the observed failure rate.

Exhibit 4-8 shows how the Bayesian posterior failure probability cumulative distribution functions are sensitive to different sample sizes for the case where no failures occur. It should be noted that the x-axis is only plotted over the range from 0.00 to 0.01. The curves show, for example, that with 34 percent confidence (y-axis), the actual failure probability is less than 0.01 (x-axis) if no failures occur in a sample of size 10. With 84 percent confidence, the actual failure probability is less than 0.01 with a sample size of 100, but is less than 0.001 with a confidence of only 0.34. These curves can be used to select test sample sizes required to achieve a given level of statistical confidence that the actual failure probabiity of the equipment under test is arbitrarily close to zero in the event of no failures. With a sample size of 200, there is a 95 percent confidence that the failure probability is less than 0.01.

4.2 CONCLUSIONS

The plotted curves and specific examples of the preceding section show how the output Bayesian failure probability density functions and cumulative distribution functions are sensitive to different sample sizes. The sample size comparisons at different fixed, observed failure percentages and the case of one observed failure show how the sensitivity of the output failure probability distributions to sample size changes as a function of the observed failure rate. The sensitivity to this observed failure rate is studied in the next chapter.

The cumulative distribution function plots show the diminishing returns achieved as sample size increases. Thus, as sample size increases, the increase in confidence about the actual value of the failure probability becomes

EXHIBIT 4-8
Cumulative Distribution Functions:
No Observed Failures
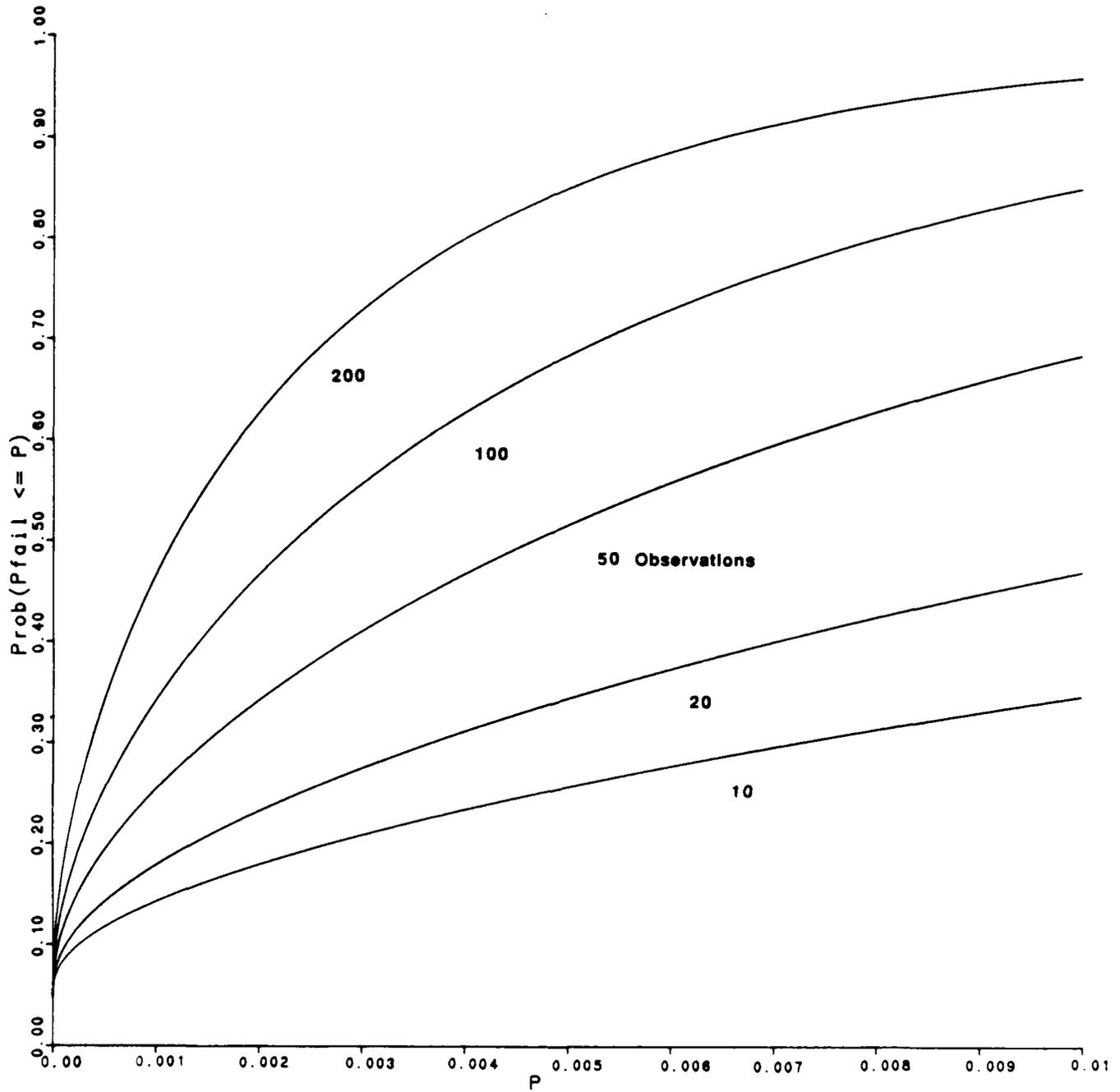
succeedingly smaller per sample. Therefore, these curves may help bound the sample sizes deemed necessary to achieve a desired level of statistical confidence in a particular test circumstance.

The cumulative distribution function curves also show that the greatest increase in confidence per unit sample increases most rapidly when the number of observed failures is nearly half of the total sample. Conversely, for a relatively small (near zero) or large (near 100) percent of observed failures, the confidence increase per unit sample is smaller. This is illustrated by the fact that, in Exhibit 4-4, the curves are more "spread out" laterally than they are in Exhibit 4-6. To estimate confidently the actual failure probability, relatively large sample sizes are required if the equipment fails nearly half the time. Smaller sample sizes are sufficient to achieve the same confidence if the equipment fails either rarely or almost always.

5.0 NUMBER OF FAILURES

5.0   NUMBER OF FAILURES

This chapter presents findings on the sensitivity of
the output Bayesian failure probability distributions to
variations in the number of observed failures in the input
data.   The investigation shows how the location and shape
of the Bayesian posterior probability density function and
cumulative distribution function change in response to
changing percentages of observed failures for different
sample sizes.

5.1   SENSITIVITY

In this analysis, the number of observed failures is
fixed at 10, 20 and 50 percent of the sample size for sam-
ple sizes of 10, 20, 50, 100, and 200.   The case of
exactly one observed failure and no failures is also in-
vestigated for the sample sizes noted.   For selected fixed
sample sizes, the probability density functions are plot-
ted to demonstrate graphically the change in location and
shape as the number of observed failures changes.   The
cumulative distribution functions are plotted for each
fixed sample size to draw numerical examples of prob-
abilistic statements that can be made regarding sensitiv-
ity to the number of observed failures.

Exhibit 5-1 shows the output Bayesian posterior prob-
ability density functions corresponding to 1, 2, and 5
observed failures with a sample size af 10.   These numbers
of observed failures correspond to observed failure rates
of 10, 20, and 50 percent of the sample size.   The curves
show that as the number of observed failures decreases,
the central location of the probability density function
decreases correspondingly, while the shape of the function
narrows.   Skewness, however, increases with the decreasing
number of observed failures.   Also, the 10 percent and 20
percent curves show that the mode of the probability dens-
ity function is less than the observed failure rate, de-
creasing as the number of observed failures decreases.
For example, the mode of the curve for 20 percent observed
failures is at approximately 0.17 (x-axis), while the mode
of the curve for 10 percent is at approximately 0.05.
Although the curves do not show it directly, the means of
the distributions may be expected to more closely match
the observed failure rates.

For comparison, Exhibit 5-2 shows the probability
density functions for different numbers of observed fail-
ures with a sample size of 20.   The numbers of observed

EXHIBIT 5-1
Probability Density Functions:
Sample Size Of 10
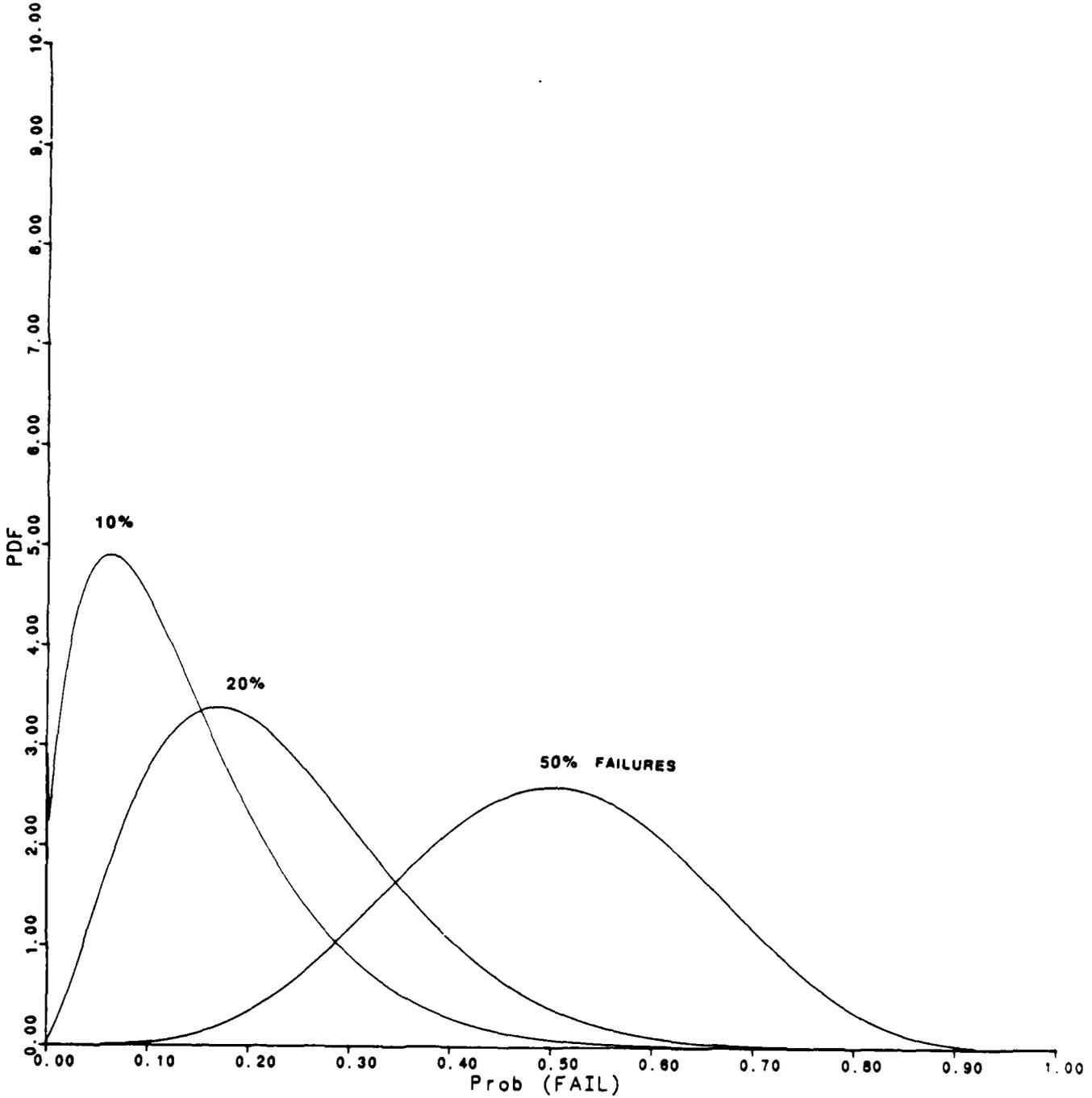
failures are 2, 4, and 10, corresponding to 10, 20, and 50 percent observed failure rates. Also, the probability density function from exactly one observed failure is plotted. Comparison of Exhibit 5-2 with Exhibit 5-1 shows that the basic sensitivities of function location and shape remain the same, regardless of sample size. Sample size does, however, have an effect on sensitivity to percentage of observed failures. The 20 sample size case shows that the modes of distributions are nearer to the percentage of observed failures than they are with the 10 sample size case. Also, the shape of the density functions changes less radically, as the number of observed failures decreases, for the larger sample size case. The single observed failure case in Exhibit 5-1 can be compared to the single observed failure case in Exhibit 5-2, but can also be compared to the 10 percent observed failure rate in Exhibit 5-2.

Exhibit 5-3 shows the probability density functions for different percentages of observed failure with a sample size of 50. For graphical representation purposes, the scale of the y-axis has been increased to accommodate the high function value corresponding to the single observed failure case. Thus, Exhibit 5-3 should not be compared directly to Exhibit 5-2 to contrast the 50 sample size case with the 20 sample size case. Instead, the 20 sample size case is plotted again in Exhibit 5-4 using the same y-axis scale as the 50 sample size case. The two different scales maintain resolution for the curves corresponding to the lower sample sizes.

Comparison between the 20 and 50 sample size cases shows how the sensitivity of the probability density functions to percentage of observed failures changes in response to different sample sizes. With larger sample sizes, the shapes of the probability density functions changes less radically as the percentage of observed failures decreases, and their modes and locations of central tendancy are nearer to the percentage of observed failures.

While the probability density functions show how the location and shape of the output Bayesian posterior failure probability distributions are sensitive to changes in the number or percentage of observed failures, they cannot be used directly to demonstrate the sensitivity of estimated failure probabilities. Thus, cumulative distribution functions corresponding to percent observed failures of 10, 20, and 50 percent are examined for sample sizes of 10, 20, 50, 100, and 200.

EXHIBIT 5-2
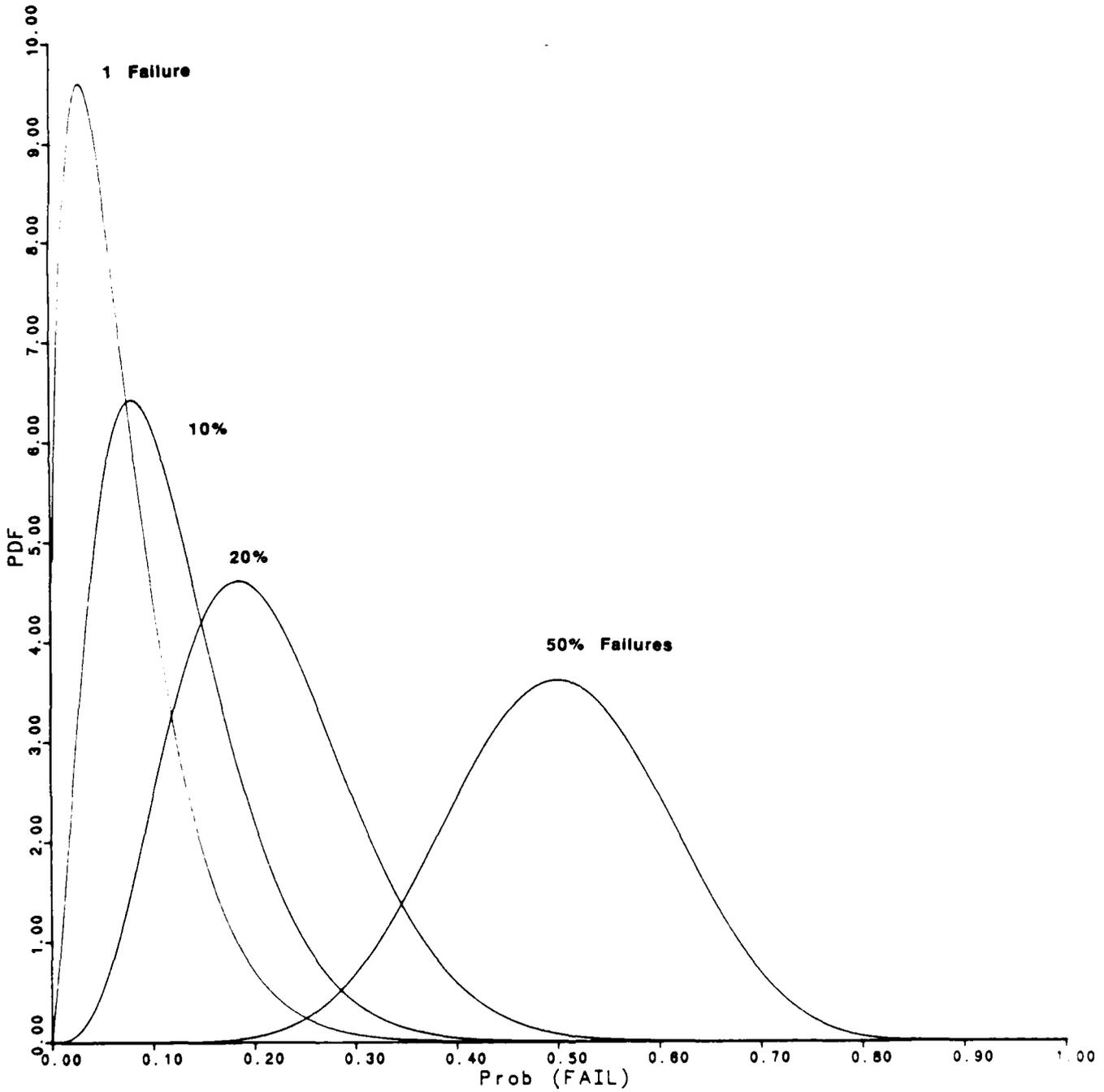## Probability Density Functions
Sample Size Of 20

5-4

EXHIBIT 5-3
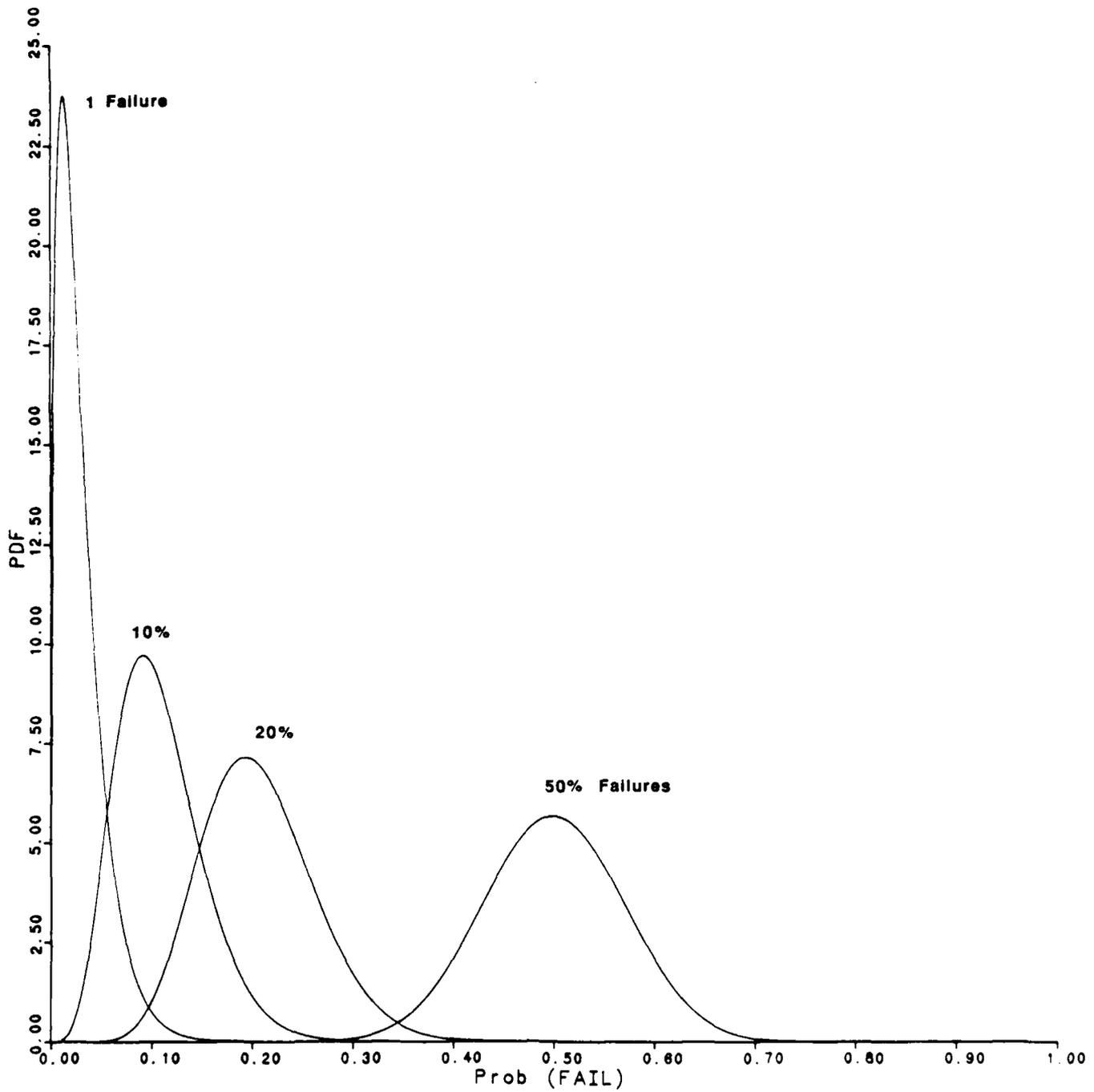Probability Density Functions:
Sample Size Of 50

EXHIBIT 5-4
Probability Density Functions:
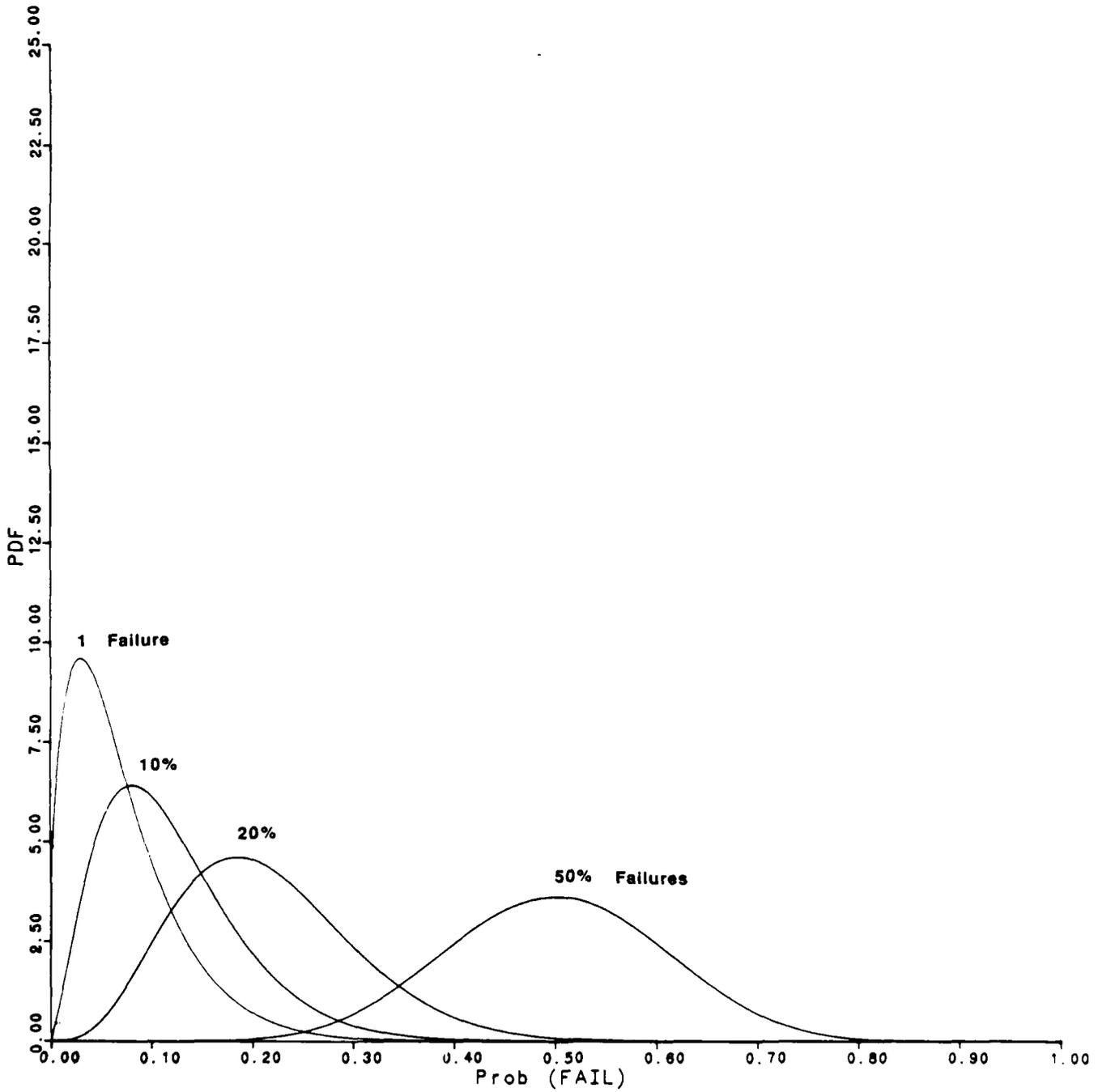### Sample Size Of 20

Exhibit 5-5 shows the cumulative distribution functions corresponding to 1, 2, and 5 observed failures with a sample size of 10. For example, the estimated probability is 0.45 (y-axis) that the actual failure probability is less than 0.10 (x-axis) when the observed failure rate is 10 percent. The estimated probability that the actual failure probability is less than 0.20 when the observed failure rate is 20 percent rises to approximately 0.48. With a 50 percent observed failure rate, the estimated probability that the actual failure probability is less than 50 percent is 0.50.

Exhibits 5-6, 5-7, 5-8, and 5-9 show the cumulative distribution functions at observed failure rates of 10, 20, and 50 percent for sample sizes of 20, 50, 100, and 200 respectively. These exhibits indicate how sensitivity of the output Bayesian failure probability distributions to percentage of observed failures changes in response to increasing sample size. For instance, the example just given for the sample size of 10, showing how the estimated probability that the actual failure probability is less than the observed failure rate decreases as the observed failure rate decreases, yields quite different results when the sample size is 200. Exhibit 5-9 shows that, with a sample size of 200, the estimated probability that the actual failure probability is less than the observed failure rate is 0.50 regardless of the observed failure rate. The cumulative distribution functions in Exhibit 5-9 corresponding to a sample size of 200 have almost identical shape, and are translated only with respect to location depending on the observed failure rate. The cumulative distribution functions in Exhibit 5-5 corresponding to a sample size of 10 are translated not only with respect to location, they also have different shapes.

5.2 CONCLUSIONS

The number of observed failures affects the location of the Bayesian posterior failure probability distribution as expected. That is, the central tendancy of the distribution lies near the observed failure rate. The shape of the distribution varies in response to different numbers of observed failures -- this variability is especially pronounced when the percent of observed failures is near zero or 100 -- but the shape variability with respect to number of observed failures diminishes as the sample size increases.
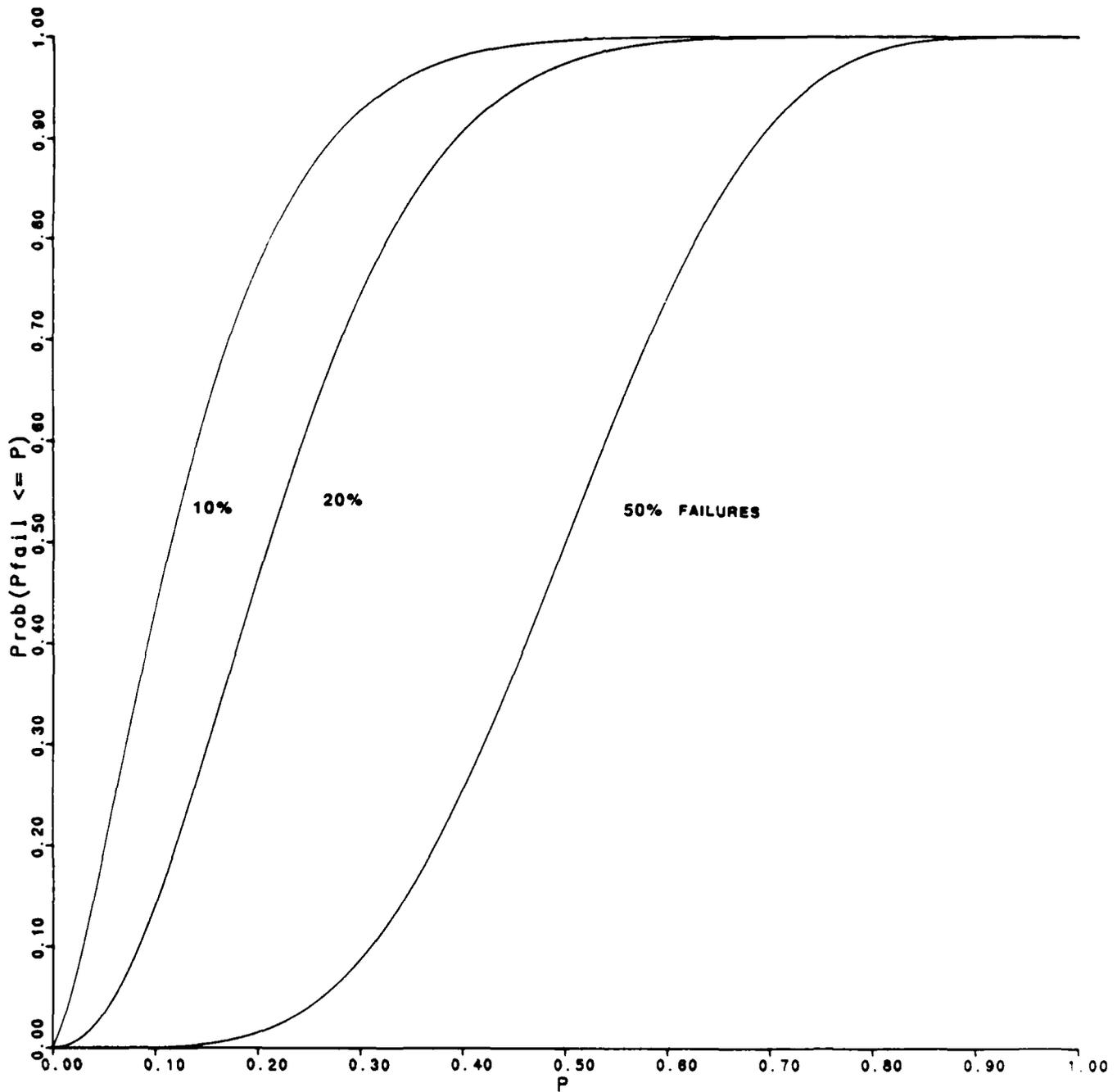
EXHIBIT 5-5
Cumulative Distribution Functions:
Sample Size Of 10

EXHIBIT 5-6
Cumulative Distribution Functions:
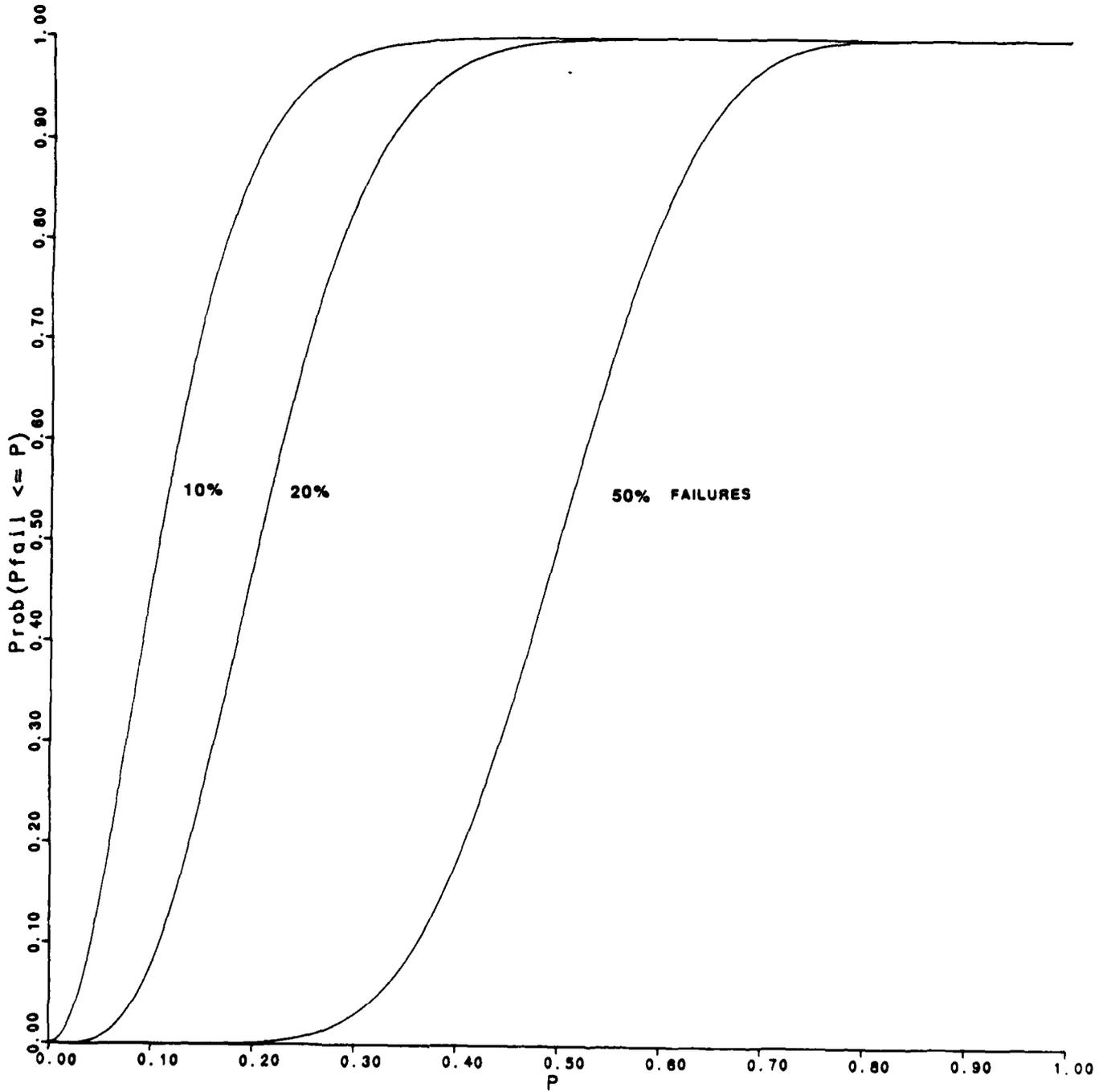Sample Size Of 20

EXHIBIT 5-7
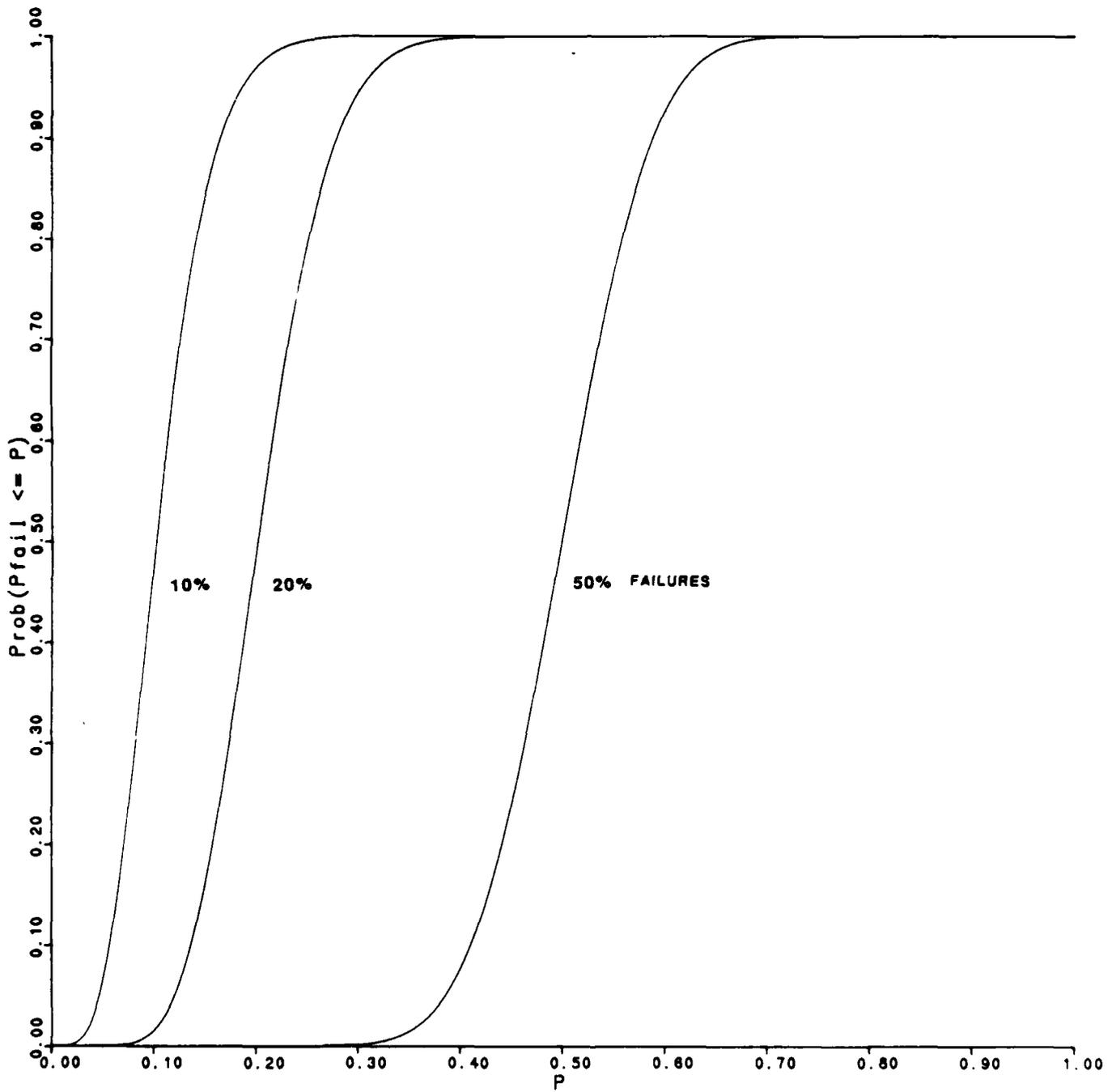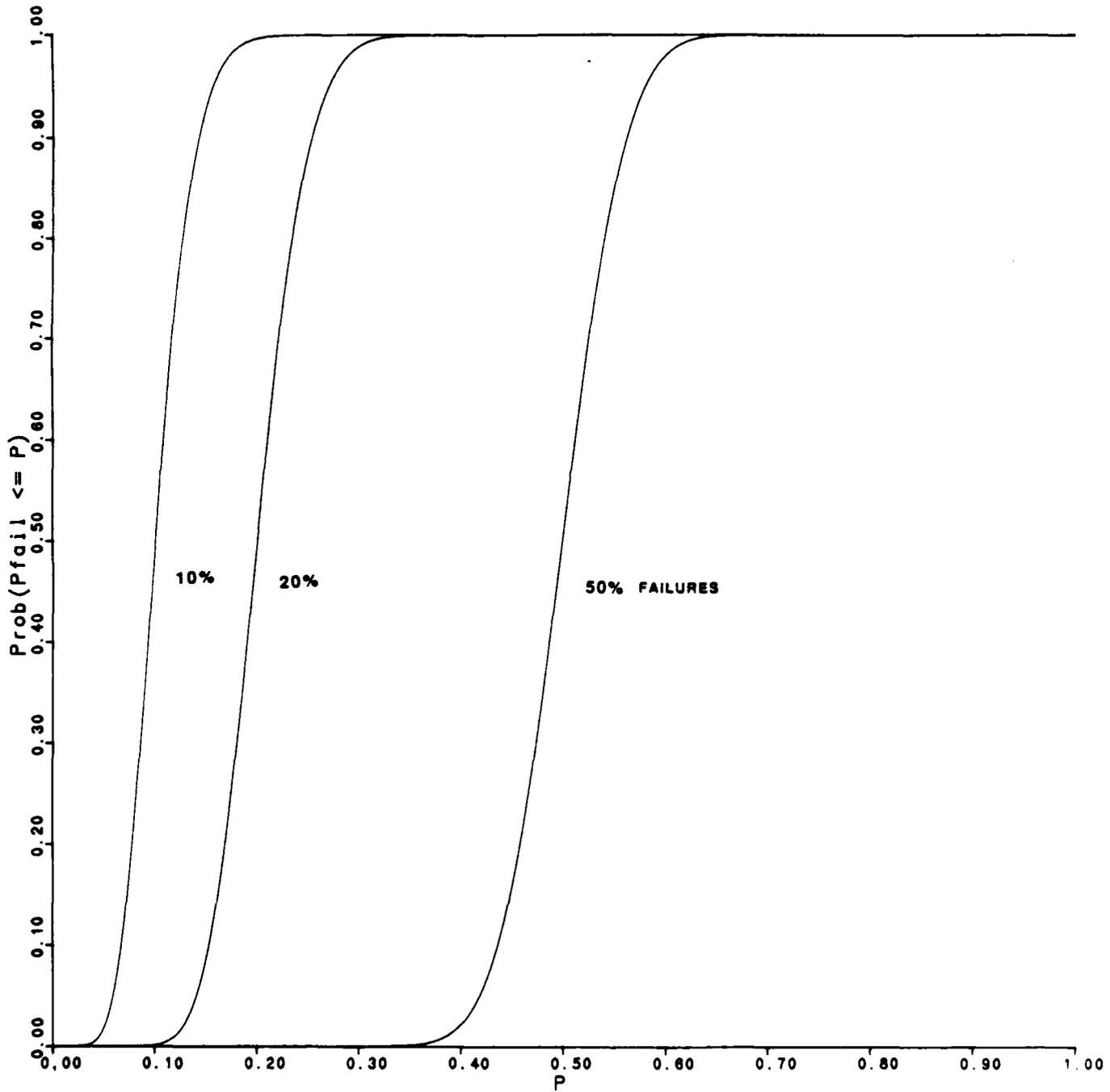Cumulative Distribution Functions:
Sample Size Of 50

EXHIBIT 5-8
Cumulative Distribution Functions:
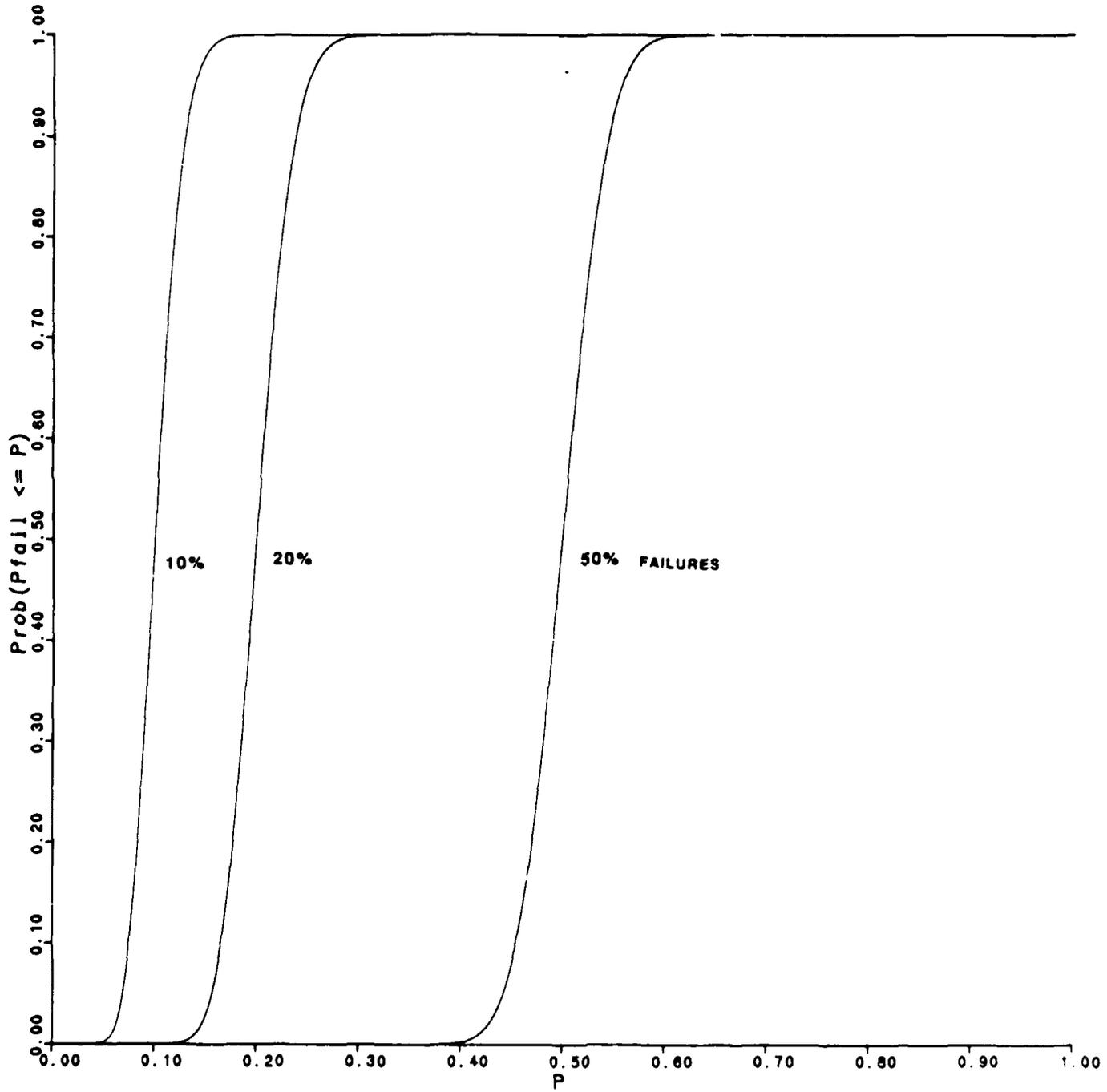## Sample Size Of 100

EXHIBIT 5-9
Cumulative Distribution Functions:
Sample Size Of 200

6.0   SUMMARY

6.0 SUMMARY

This report presents a sensitivity study of a statis-
tical model used by the OMNCS to estimate HEMP-induced
failure probability distributions of telecommunications
switches and transmission facilities. The model uses
Bayesian statistical theory to interpret test data from
EMP test programs conducted by the OMNCS and other organ-
izations. The outputs of the model support nuclear weap-
ons effects analyses of major NSEP telecommunications
networks, and assist EMP test planning. This sensitivity
study explores the meaning and limitations of the model's
output by quantifying its response to variations in its
inputs.

The Bayesian approach used to characterize HEMP-
induced failure probabilities of telecommunications equip-
ment combines observed test data with assumed prior know-
ledge to form posterior estimates of failure performance.
The test data are comprised of the total number of test
observations (sample size) and the number of observed
failures for a particular type of equipment at a given
HEMP stress level or stress level range. The assumed
prior knowledge is embodied in a prior probability distri-
bution of the parameter being estimated (failure prob-
ability). The application of Bayes' Theorem allows the
data and prior distribution to produce a posterior prob-
ability distribution for failure probability. The poster-
ior distribution is proportional to the product of the
failure probability likelihood and the prior distri-
bution. This output distribution is reflected in a cumu-
lative distribution function and its associated prob-
ability density function.

The prior distribution used in this application of the
Bayesian methodology is a noninformative prior distribu-
tion. That is, no prior knowledge is assumed relative to
the data regarding the value of the parameter being esti-
mated. The insensitivity of the output posterior failure
probability distribution to variations in the selected
prior is demonstrated in this study using a theoretical
derivation. This insensitivity is demonstrated by finding
a transformation of the parameter being estimated for
which a locally uniform prior is noninformative. Follow-
ing a theorem, the corresponding noninformative prior of
the untransformed parameter is locally proportional to the
derivative of the transformation with respect to the para-
meter. Because an inverse sine transformation produces a

likelihood curve that is nearly data translated, the insensitivity of the resultant posterior distribution follows on theoretical grounds.

The test data sample size affects the output failure probability distribution primarily through the variance of the distributions. Small sample sizes produce large posterior distribution variances; large sample sizes produce small variances. The resulting relationship between sample size and estimation precision is quantified graphically using cumulative distribution plots. These plots can be used to construct confidence interval estimates of failure probability. For example, if no failures are observed during testing, a sample size of 150 provides a 90 percent confidence that the actual failure probability is less than 0.01.

The number of failures observed during testing affects primarily the location of the posterior failure probability distributions. The central tendency of the distribution lies near the observed failure rate, as shown graphically in cumulative distribution function and probability density function plots. The shape of the distribution varies in response to different numbers of observed failures. This variability is especially pronounced when the percent of observed failures is near zero or 100. However, this shape variability with respect to number of observed failures diminishes as the sample size increases.

The results of this sensitivity study can be used to support subsequent sensitivity studies of the network connectivity analysis models that use the Bayesian posterior failure distributions as inputs. Also, the sample size sensitivity curves can be used to guide HEMP test planning for telecommunications equipment.

APPENDIX A

BAYESIAN MODEL COMPUTER CODE

```
c*************************************************************************
c
c  This Program is a Bayesian Failure Probability Model.
c  The program calculates the cumulative distribution for the Bernoulli
c  parameter ( Mean Failure Probability ) with Noninformative Prior
c  Distribution.
c
c*************************************************************************
c
c
c
c  Variable Declaration and definition:
c
c  -------------------------------------------
c

        CHARACTER   Ans*1,              ! User Interactive Character string
       *            Eqtested*25,        ! Description for Equipment
       *            Stresslvl*10,       ! Level of Stress used for this data
       *            IObuf*10            ! An I/O Buffer used with system Functions


        REAL*8      A,B,                ! Incomplete Beta Function Interval
       *            W1,W2,              ! Intermediate Function values
       *            Wnum,Wden,          !   "           "         "
       *            Term,               !   "           "         "
       *            CDF,                ! Storage Variables for Cumulative Density
       *            Previous            ! Function and Posterior Distribution
       *            PDF,                ! Function.
       *            Percent,            ! Interval value. Interval is between 0-1
       *            ZA(7),              ! Array used in CDF Polynomial Calculation
       *            Zsave,              ! Storage values for the Polynomial Result
       *            Z,                  !   "
       *            Pz,                 !   "
       *            Sum,                !   "
       *            Coeff(7),           ! Coefficeint array
       *            Resolution          ! Requested resolution between 0-1


        INTEGER*4   Ssize,              ! Sample Size
       *            Failures,           ! Number of Failures
       *            Index,              ! Loop Upper Bound variable
       *            Ans_1,              ! System funcion I/O conversion variables
       *            Eqtested_1,         !   "
       *            Stresslvl_1,        !   "
       *            IObuf_1,            !   "
       *            LIB$Get_input,      ! System I/O Functions
       *            LIB$Get_symbol,     !   "
       *            LIB$Set_symbol,     !   "
       *            OTS$CVT_TI_L,       !   "
       *            OTS$CVT_T_F,        !   "
       *            OTS$CVT_L_TI,       !   "
       *            LIB$STOP            !   "


c
c
c  Initialize Coeff, first ZA, and strings  variables.
c
```

```fortran
c  ---------------------------------------------------------
c  Coeff(1) is never used.
c
        Coeff(2) = 0.049867347D0
        Coeff(3) = 0.0211410061D0
        Coeff(4) = 0.0032776263D0
        Coeff(5) = 3.80036D-05
        Coeff(6) = 4.88906D-05
        Coeff(7) = 5.383D-06

        ZA(1) = 1.0D0

        Ans(1:) = ' '
        Eqtested(1:) = ' '
        Stresslvl(1:) = ' '
        Not_Done =.TRUE.

c
c  Open Ouput file for results
c
        OPEN(UNIT=7,Status='NEW')

c
c  Loop over all input data
c
        DO WHILE (Not_Done)


c  Initialize Cumulative Density Function, Previous retainer, and Counter

        CDF = 0.0D0
        Previous = 0.0D0

c
c  Prompt Operator for program continuation
c
        ISTAT = LIB$Get_input(Ans,'Is there data to be processed (Y/N) ',
     *                           Ans_1)

        IF (Ans .EQ. 'Y')  THEN

c
c  Retrieve necessary information for a run.
c
            ISTAT = LIB$Get_input(Eqtested,'Equipment Tested ',Eqtested_1)
            IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))

            ISTAT = LIB$Get_input(Stresslvl,'Stress Level  ',Stresslvl_1)
            IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))

            ISTAT = LIB$Get_input(IObuf,'Sample Size ',IObuf_1)
            IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))

            ISTAT = OTS$CVT_TI_L(IObuf(1:IObuf_1),Ssize)
            IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))
```

A-2

```fortran
                  ISTAT = LIB$Get_input(IObuf,'Failures ',IObuf_l)
                  IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))

                  ISTAT = OTS$CVT_TI_L(IObuf(1:IObuf_l),Failures)
                  IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))

            ELSE
c                 Run is complete

                  Not_Done = .FALSE.

            ENDIF

            IF (Not_Done) THEN

c
c                 Write header information to the file for this iteration
c
                  WRITE(07,104) Eqtested
                  WRITE(07,103) Stresslvl
                  WRITE(07,102) Ssize
                  WRITE(07,101) Failures
                  WRITE(07,100)
                  WRITE(07,99)

104               FORMAT(' Equipment Tested: ',A25)
103               FORMAT(' EMP Stress Level: ',A10)
102               FORMAT(' Sample Size:      ',I6)
101               FORMAT(' Failures:         ',I6)
100               FORMAT(' Prob(fail)          CDF              PDF')
99                FORMAT('-----------------------------------------------')


c
c     Setup A and B for this run, the incomplete Beta Function Interval.
c

                  A = Failures + 0.5D0
                  B = 0.5D0 + (Ssize - Failures)

c
c     For percentage values from 1% to 100% calculate CDF and PDF
c     Increments are input from keyboard.
c

                  ISTAT = LIB$Get_symbol('RESOLUTION',IObuf,Iobuf_l)
                  IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))

                  ISTAT = OTS$CVT_TI_L(IObuf(1:Iobuf_l),Index)
                  IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))
                  RESOLUTION  = INDEX
c
c     Loop over resolution desired converting to an interval between 0-1
c

                  DO  10 J = 0, Index

                        Percent = J / Resolution
c
```

```fortran
c     Has the cumulative distribution reached 1?
c     If so there is no need to calculate any more.
c

                        IF ( CDF .LT. 1.0D0 ) THEN

c
c     Test the validity of the Numerical Approximation
c     If the test is true we force the PDF to 0.
c     Implementation of the second approximation was not appropriate
c

                            Test = ( A + B - 1.0D0) * ( 1.0D0 - Percent)

                            IF ( Test .LT. 0.8D0 ) THEN

                                Previous = 1.0D0
                                CDF = 1.0D0
                                PDF = CDF - Previous
                                Previous = CDF

                            ELSE

c
c     Calculate the probability point for Standard Normal Approximation
c

                                W1 = (B * Percent)**0.3333333D0
                                W2 = (A * (1.0D0 - Percent))**0.333333D0
                                Wnum = 3.0D0*((W1*(1.0D0-(1.0D0/(9.0D0*B))))
     *                                 -(W2*(1.0D0-(1.0D0/(9.0D0*A)))))

                                Wden = DSQRT( ((W1*W1)/B) + ((W2*W2)/A) )

c
c               Calculate the standard normal CDF
c

                                Z = Wnum / Wden
                                Zsave = 0.0D0

                                IF ( Z .LE. 0.0D0 ) THEN

                                    Zsave = Z
                                    Z = -Z
                                ENDIF

                                Sum = 1.0D0
                                DO 20 I = 2,7

                                    ZA(I) = ZA(I-1) * Z
                                    Term = ZA(I) * Coeff(I)
                                    Sum = Sum + Term

20                              CONTINUE

                                PZ = 1.0D0 - (0.5D0 * (Sum**(-16D0)))

                                IF ( Zsave .NE. 0.0D0 ) THEN
```

A-4

```
                                     Z = Zsave
                                     PZ = 1.0D0 - PZ

                            ENDIF

c
c       Assign Calculated value to the Storage value
c
                                     CDF = PZ

c       Calculate the PDF from the new and previous CDF

                                     PDF = (CDF - Previous) * Resolution
                                     Previous = CDf

                            ENDIF

                    ELSE

                            PDF = CDF - Previous
                            Previous = CDF

                    ENDIF


c
c   Write Results of this iteraion to the output file.
c
                    WRITE(07,98) Percent,CDF,PDF
98                  FORMAT(T3,F6.4,T16,F10.7,T32,F15.9)


c
c   Next percentage
c
10              CONTINUE

        ENDIF                           ! For Not_done Checking

        END DO                          ! For Controlling Program Execution

        CLOSE(07)

c
c   System Variable updates used in control program
c
        ISTAT = OTS$CVT_L_TI(Count,IObuf)
        IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))

        ISTAT = LIB$Set_symbol('COUNT',IObuf(1:1))
        IF (.NOT. ISTAT) CALL LIB$STOP(%VAL(ISTAT))

        WRITE(06,*) 'End Of Program'
        END
```

REFERENCES

1) Estimated Electromagnetic Pulse Responses of Selected
    Telecommunications Assets (U), National Communications
    System, February 1985, Secret.

2) Probabilistic System-Level Effects of EMP: Worst Case
    Analysis (U), National Communications System,
    April 1985, Secret.

3) Probabilistic System-Level Effects of EMP: Present
    System Analysis (U), National Communications System,
    September 1985, Secret.

4) Probabilistic System-Level Effects of EMP: Nets Net-
    work Analysis (U), National Communications System,
    February 1986, Secret.

5) Bayesian Inference in Statistical Analysis, G. Box and
    G. Tiao; Addison Wesley Publishing Company; Reading,
    Massachusetts, 1973.

6) Handbook of Mathematical Functions, edited by
    M. Abramowitz and I. A. Stegun; National Bureau of
    Standards.

END

9-87

DTIC