

APPLICATION OF THE EMPIRICAL DISTRIBUTION  
FUNCTION TO ENGINEERING DECISIONS

By

DAVID RAY CUNNINGHAM

Bachelor of Science  
Oklahoma State University  
Stillwater, Oklahoma  
1957

Master of Science  
University of Idaho  
Moscow, Idaho  
1959

Submitted to the Faculty of the Graduate College  
of the Oklahoma State University  
in partial fulfillment of the requirements  
for the Degree of  
DOCTOR OF PHILOSOPHY  
August, 1969

NOV 5 1969

APPLICATION OF THE EMPIRICAL DISTRIBUTION  
FUNCTION TO ENGINEERING DECISIONS

Thesis Approved:

*Arthur M. Breipohl*

Thesis Adviser

*Paul A. McCullum*

*J. Lloyd Folber*

*Bennett Basore*

*D. D. Durham*

Dean of the Graduate College

729900

## ACKNOWLEDGEMENTS

I wish to express my appreciation to my thesis adviser, Professor A. M. Breiphol, for his assistance during this research. I also wish to express my appreciation to the other members of my committee and to those professors in the Departments of Electrical Engineering and Mathematics whose courses prepared the background for this work.

The financial support from the National Science Foundation is gratefully acknowledged.

I further wish to express my appreciation to my wife and children for their patience and encouragement.

## TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION. . . . .	1
1.1 Statement of the Problem. . . . .	1
1.2 Existing Solutions. . . . .	1
1.3 Present Contribution. . . . .	2
II. BAYES' EMPIRICAL DISTRIBUTION FUNCTION. . . . .	4
2.1 Introduction. . . . .	4
2.2 Including Prior Information . . . . .	4
2.3 Bayes' Empirical Distribution Function. . . . .	5
III. COMPARISON OF BAYES' EMPIRICAL DISTRIBUTION AND PARAMETRIC BAYESIAN ESTIMATION . . . . .	10
3.1 Introduction. . . . .	10
3.2 Estimation of a Parameter . . . . .	10
3.3 Estimation of a Distribution Function . . . . .	15
3.4 Integral Expected Square Error. . . . .	16
3.5 Discussion. . . . .	23
IV. APPLICATION TO ADAPTIVE BINARY SIGNAL DETECTION . . . . .	24
4.1 Introduction. . . . .	24
4.2 Bayes' Decision Rules . . . . .	24
4.3 Detection With Unknown Noise. . . . .	26
V. BAYES' EMPIRICAL DISTRIBUTION FUNCTION FOR A RANDOM VECTOR. .	30
5.1 Introduction. . . . .	30
5.2 B.E.F. for a Random Vector. . . . .	30
VI. SMOOTHING BAYES' EMPIRICAL DISTRIBUTION FUNCTION. . . . .	34
6.1 Introduction. . . . .	34
6.2 Smoothing B.E.F.. . . . .	34
6.3 Specht's Approximation. . . . .	35
VII. SUMMARY AND CONCLUSIONS . . . . .	38
7.1 Summary . . . . .	38
7.2 Conclusions . . . . .	38
7.3 Recommendations for Further Study . . . . .	39

Chapter	Page
BIBLIOGRAPHY . . . . .	40
APPENDIX A - DERIVATION OF BAYES' EMPIRICAL DISTRIBUTION FUNCTION. .	42
A.1 Introduction . . . . .	42
A.2 The Prior Distribution . . . . .	42
A.3 The Posterior Distribution . . . . .	43
A.4 Restrictions on the Prior Distribution . . . . .	44
A.5 Binomial Sampling Not Required . . . . .	50
APPENDIX B - THEOREMS CONCERNING BAYES' EMPIRICAL DISTRIBUTION FUNCTION. . . . .	52
B.1 Introduction . . . . .	52
B.2 Convergence of B.E.F.. . . . .	52
B.3 Errors of B.E.F. Estimate of $F(x)$ . . . . .	54
B.4 Properties of B.E.F. for Estimation. . . . .	61
B.5 Moments of B.E.F.. . . . .	67
APPENDIX C - CONVENTIONAL BAYESIAN PARAMETER ESTIMATION. . . . .	69
C.1 Introduction . . . . .	69
C.2 Binomial Distribution With Unknown Probability . . . . .	70
C.3 Normal Distribution With Unknown Mean. . . . .	71
C.4 Normal Distribution With Unknown Variance. . . . .	72
C.5 Normal Distribution With Unknown Mean and Variance . . . . .	74
C.6 Rayleigh Distribution With Unknown Parameter . . . . .	76

## LIST OF TABLES

Table	Page
I. Parameter Estimation . . . . .	11

## LIST OF FIGURES

Figure	Page
1. Integral Expected Square Error With Zero Prior Weight . . . . .	18
2. Integral Expected Square Error With Correct Prior . . . . .	19
3. Integral Expected Square Error With Incorrect Prior . . . . .	20

## CHAPTER I

### INTRODUCTION

1.1 Statement of the Problem. Engineering decisions must constantly be made. Is a signal present? Which of  $n$  alternative systems should be produced? Which pattern is present? These are but three of the many decisions which occur in engineering work.

In recent years there has been an increased effort to quantify the factors entering into such decisions. As these factors may be known only approximately and indeed may depend on chance outcomes, the factors are treated as random variables and the decisions are made according to some rule such as minimizing the expected cost. Unfortunately the probability distributions which govern these random variables are often unknown and must be estimated from experience, sample data or both. For these reasons, this field of investigation has become known as "statistical decision theory". A most significant problem in statistical decision theory is the estimation of the probability distribution functions involved. This estimation problem is investigated in this dissertation.

1.2 Existing Solutions. If no prior knowledge is included, the distribution may be estimated from the empirical distribution function (11) or a method suggested by Rosenblatt (13) and investigated by Parzen (10) may be used. Although the empirical approach to statistical decision problems has been investigated (12), the selection of a sample size remains a problem. The advantages of the empirical distribution function

are its simplicity and the fact that it converges with probability one to the true distribution function.

Economics or time limitations frequently dictate the use of a sample size smaller than one might desire. This makes the use of any prior information about the parent distribution very desirable. Although the wisdom of using such prior information is debated (14), its use often appears to be an engineering necessity. For example when determining the number of systems to be tested, should any be tested? This question can only be answered from prior experience. This interesting problem has been investigated by Howard (3,4) with impressive results.

The form of the density function is often assumed and the parameters estimated from sample data. The usual method of including prior knowledge about the density parameters is to select prior estimates of the parameters, treat the parameters as random variables with an assumed prior distribution and modify the distribution of the parameters using Bayes' theorem as sample data becomes available (15). This approach does have two shortcomings. First, if the likelihood function is of the wrong form, the learned distribution may be in error by a considerable amount and will never converge to the true distribution, regardless of the sample size. Second, if more than one parameter is to be learned, the mathematics may become so formidable in appearance that the less mathematically inclined engineer may be discouraged. A significant advantage of a Bayesian testing procedure is that it provides a logical method for selecting sample sizes.

1.3 Present Contribution. In this paper a method is developed, using Bayes' theorem, whereby prior knowledge may be included in the empirical distribution function. A number of theorems are developed

concerning estimates of this form, and the results of this technique are compared to the more conventional Bayesian parametric approach.

Extension of this technique to the nonparametric method of Parzen is suggested, and Specht's approximation (16) is proposed as a possible means of reducing data storage.

The proposed method has the advantages of simplicity and convergence associated with the empirical distribution function and yet retains the advantage of including prior information provided by Bayesian parameter estimation. Thus the selection of a sample size is facilitated.

## CHAPTER II

### BAYES' EMPIRICAL DISTRIBUTION FUNCTION

2.1 Introduction. Given a sequence of independent identically distributed random variables  $\{X_1, X_2, \dots, X_n\}$  with a common cumulative distribution function

$$F(x_i) = P\{X_i \leq x_i\} \quad , \quad (2.1.1)$$

the empirical distribution function (E.D.F.) is defined by (11)

$$F_n(x) = \frac{1}{n} (\text{no. of observations } \leq x \text{ among } X_1, \dots, X_n) \quad . \quad (2.1.2)$$

For a given value of  $x$ ,  $nF_n(x)$  is a binomial random variable. Thus

$$E\{F_n(x)\} = F(x) \quad (2.1.3)$$

and

$$\text{Var}\{F_n(x)\} = \frac{1}{n} F(x)[1 - F(x)] \quad . \quad (2.1.4)$$

2.2 Including Prior Information. The question of interest is "Is it possible to include prior knowledge in an E.D.F.?" The answer is obviously yes. Assume the prior information has a weight equivalent to a sample of size  $w_0$ . The prior distribution function would then be of the form

$$F_{w_0}(x) = \left(\frac{1}{w_0}\right) (\text{no. of observations assumed } \leq x \text{ among an assumed } w_0 \text{ observations}). \quad (2.2.1)$$

That is, a prior E.D.F. could simply be assumed. After an actual sample of size  $n$  the posterior distribution function would have the form

$$F_w(x) = \left( \frac{1}{w_0+n} \right) (\text{no. of observations assumed } \leq x \text{ among } w_0 \text{ assumed observations} + \text{the number of observations } \leq x \text{ among } n \text{ observations } X_1, X_2, \dots, X_n). \quad (2.2.2)$$

On rewriting the above equation becomes

$$F_w(x) = \frac{w_0}{w_0+n} F_{w_0}(x) + \frac{n}{w_0+n} F_n(x), \quad (2.2.3)$$

which is simply a weighted sum of the prior E.D.F. and the sample E.D.F.

2.3 Bayes' Empirical Distribution Function. The restriction that the prior distribution function must be an E.D.F. may be removed in a manner consistent with Bayes' theorem. Assume that the probability distribution function  $F(x)$  of the population is a sample function from a stochastic process with a beta first order density function. The process of learning  $F(x)$  is presented in Appendix A and is approached as outlined below.

Given a random variable  $Q$  having a prior beta density function

$$f_Q(q) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1} (1-q)^{\beta-1} & 0 \leq q \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.3.1)$$

If  $Q$  is the probability that a random variable  $X$  is less than or equal to  $x$ , i.e.,

$$Q = F(x), \quad (2.3.2)$$

where  $\alpha$ ,  $\beta$  and  $q$  are functions of  $x$ , then given  $q$ , the probability that "a" out of  $n$  samples would be less than or equal to  $x$  will be binomial with probability  $q$ . Thus

$$\begin{aligned} P_{A|q,n}(a) &= P\{a \text{ out of } n \text{ samples} \leq x\} \\ &= \binom{n}{a} q^a (1-q)^{n-a} \end{aligned} \quad (2.3.3)$$

On applying Bayes' rule after sampling, the first order density function becomes

$$f_{Q|\mathcal{E}}(q) = \frac{P_{A|q,n}(a)f_Q(q)}{\int_{-\infty}^{\infty} P_{A|q,n}(a)f_Q(u)du} \quad (2.3.4)$$

where  $\mathcal{E}$  is the total experience. It is well known (15) that this posterior density is beta. As shown in Appendix A

$$f_{Q|\mathcal{E}}(q) = \begin{cases} \frac{\Gamma(n + \alpha + \beta)}{\Gamma(a + \alpha)\Gamma(n - a + \beta)} q^{a+\alpha-1} (1-q)^{n-a+\beta-1} \\ 0 \end{cases} \quad \begin{matrix} (2.3.5) \\ \text{otherwise .} \end{matrix}$$

A reasonable choice for the estimate of  $F(x)$  would be the expected value of  $Q$ . Therefore the following definitions are chosen:

$$\hat{F}_X(x) \triangleq E_Q\{Q\} \quad (2.3.6)$$

and

$$\hat{F}_{X|\mathcal{E}}(x) \triangleq E_Q\{Q|\mathcal{E}\} \quad , \quad (2.3.7)$$

where  $E_Q\{\cdot\}$  is the expected value with respect to  $Q$ . In order to be the prior density function  $E_Q\{Q\}$  must be a density function and hence a

nondecreasing function of  $x$ . In order to assure that  $\hat{F}_{X|\mathcal{E}}(x)$  is also a nondecreasing function of  $x$ , it is sufficient to assume that the sum of  $\alpha$  and  $\beta$  is a constant,

$$w_0 = \alpha + \beta \quad .$$

It can easily be shown that

$$\begin{aligned} F_X(x) &= \frac{\alpha(x)}{\alpha(x) + \beta(x)} \\ &= \frac{\alpha(x)}{w_0} \end{aligned} \quad (2.3.8)$$

$$\begin{aligned} F_X(x) &= \frac{a(x) + \alpha(x)}{w_0 + n} \\ &= \frac{w_0}{w_0 + n} \hat{F}_X(x) + \frac{n}{w_0 + n} F_n(x) \quad , \end{aligned} \quad (2.3.9)$$

where  $F_n(x)$  is the empirical distribution function as defined by Equation 2.1.2. Equation 2.3.9 is of exactly the same form as 2.2.3. Hence  $w_0$  as defined in this section may be considered as the equivalent sample size for the prior information. It is important to note, however, that the only restriction on  $w_0$  is

$$w_0 > 0 \quad , \quad (2.3.10)$$

and the only restriction on  $\hat{F}_X(x)$  is that it is a distribution function. In order to make the notation of Equations 2.2.3 and 2.3.9 the same, the following notation will be used:

$$F_{w_0}(x) \triangleq \hat{F}_X(x) \quad (2.3.11)$$

and

$$F_w(x) \triangleq \hat{F}_{X|\mathcal{G}}(x) \quad . \quad (2.3.12)$$

The equation for learning  $F(x)$  is thus

$$F_w(x) = \frac{w_0}{w_0+n} F_{w_0}(x) + \frac{n}{w_0+n} F_n(x) \quad , \quad (2.3.13)$$

where

$F_{w_0}(x)$  is the prior distribution function,

$w_0$  is the equivalent sample size weight of  $F_{w_0}(x)$ ,

$n$  is the sample size, and

$F_n(x)$  is the empirical distribution function.

For convenience  $F_w(x)$  will be termed the Bayes' empirical distribution function (B.E.F.).

A reasonable estimate of the error to be expected in the estimate  $F_w(x)$  of  $F(x)$  is the variance of  $Q$ . As shown in Appendix A,

$$\text{Var}\{Q|\mathcal{G}\} = \frac{1}{w_0+n+1} F_w(x)[1 - F_w(x)] \quad . \quad (2.3.14)$$

In view of Equation 2.1.4 this result is reasonable.

The restriction that an integer number of units must be tested may be easily removed as described in Appendix A. Defining the sample distribution function as

$$F_s(x) \triangleq \frac{1}{s} \text{ (observed quantity of sample } \leq x \text{ out of a sample size } s), \quad (2.3.15)$$

Equation 2.3.13 becomes

$$F_w(x) = \frac{w_0}{w_0+s} F_{w_0}(s) + \frac{s}{w_0+s} F_s(x) \quad , \quad (2.3.16)$$

where

$$0 < s \leq 1. \quad (2.3.17)$$

Thus the general form of the B.E.F. is represented by Equation 2.3.16.

This is simply a weighted average of the prior distribution function and the sample distribution function.

## CHAPTER III

### COMPARISON OF BAYES' EMPIRICAL DISTRIBUTION AND PARAMETRIC BAYESIAN ESTIMATION

3.1 Introduction. The relative merits of using B.E.F. instead of conventional parametric estimation can only be determined when a specific problem is defined. However it is possible to consider certain aspects of the two approaches for a few familiar cases.

3.2 Estimation of a Parameter. A common problem arising in both communications and reliability engineering is the estimation of one or more parameters of a distribution function with an assumed form. The form of the distribution function is assumed from prior experience and often some prior estimate of the unknown parameter is also available. Bayes' rule is used to alter this estimate as data becomes available. Several examples of this type are included in Appendix C.

An alternate method of arriving at an estimate of a parameter is to calculate the parameter from the B.E.F. estimate  $F_w(x)$  of  $F(x)$ .

For example if a random variable  $x$  is assumed to have a normal distribution with known variance  $N^2$  and unknown mean  $\mu$ , the mean may be treated as a random variable and the mean estimated from prior knowledge and data. The conventional Bayesian estimate for  $\mu$  may be found in Equation C.3.3. The estimate given by B.E.F. may be calculated from Theorem B.5.1. The two are listed in Table I for comparison.

Similarly the mean of a normal random variable may be known and the

TABLE I  
PARAMETER ESTIMATES

Normal Distribution With Unknown Mean	
Conventional Bayesian Estimate	B.E.F. Estimate
$\hat{\mu}_n = \frac{\frac{N^2}{\sigma_0^2} \mu_0 + n \left( \frac{1}{n} \sum_{i=1}^n X_i \right)}{\frac{N^2}{\sigma_0^2} + n} \quad (3.2.1)$	$\hat{\mu}_w = \frac{w_0 \mu_0 + n \left( \frac{1}{n} \sum_{i=1}^n X_i \right)}{w_0 + n} \quad (3.2.2)$

$\hat{\mu}_n$  = the estimate of the mean

$N^2$  = the assumed variance

$\mu_0$  = the prior estimate of the mean

$\sigma_0^2$  = the prior variance of the mean

$\{X_1, X_2, \dots, X_n\}$  = the data set

$n$  = the number of data samples

$\hat{\mu}_w$  = the estimate of the mean

$\mu_0$  = the mean given by the prior distribution

$w_0$  = the weight of the prior distribution

$\{X_1, X_2, \dots, X_n\}$  = the data set

$n$  = the number of data samples

TABLE I (Continued)

Normal Distribution With Unknown Variance	
Conventional Bayesian Estimate	B.E.F. Estimate
$\hat{\sigma}_n^2 = \frac{v_0 \varphi_0 + n \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right]}{v_0 + n} \quad (3.2.3)$	$\hat{\sigma}_w^2 = \frac{w_0 \sigma_0^2 + n \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right]}{w_0 + n} \quad (3.2.4)$
<p><math>\hat{\sigma}_n^2</math> = the estimate of the variance</p> <p><math>\varphi_0</math> = the prior estimate of the variance</p> <p><math>v_0</math> = the weight of the prior variance</p> <p><math>\mu</math> = the assumed true mean</p> <p><math>\{X_1, X_2, \dots, X_n\}</math> = the data set</p> <p><math>n</math> = the number of data samples</p>	<p><math>\hat{\sigma}_w^2</math> = the estimate of the variance</p> <p><math>\sigma_0^2</math> = the variance given by the prior distribution</p> <p><math>w_0</math> = the weight of the prior distribution</p> <p><math>\mu</math> = the assumed true mean</p> <p><math>\{X_1, X_2, \dots, X_n\}</math> = the data set</p> <p><math>n</math> = the number of data samples</p>

TABLE I (Continued)

Normal Distribution With Unknown Mean and Variance	
Conventional Bayesian Estimate	B.E.F. Estimate
$\hat{\mu}_n = \frac{w_0 \mu_0 + n \left( \frac{1}{n} \sum_{i=1}^n X_i \right)}{w_0 + n} \quad (3.2.5)$	$\hat{\mu}_w = \frac{w_0 \mu_0 + n \left( \frac{1}{n} \sum_{i=1}^n X_i \right)}{w_0 + n} \quad (3.2.7)$
$\hat{\sigma}_n^2 = \frac{v_0 \varphi_0 + n \left( \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}_n^2 \right) + w_0 (\mu_0^2 - \hat{\mu}_n^2)}{v_0 + n} \quad (3.2.6)$	$\hat{\sigma}_w^2 = \frac{w_0 \sigma_0^2 + n \left( \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}_w^2 \right) + w_0 (\mu_0^2 - \hat{\mu}_w^2)}{w_0 + n} \quad (3.2.8)$

$\hat{\mu}_n$  = the estimate of the mean

$\mu_0$  = the prior estimate of the mean

$w_0$  = the prior weight of the mean

$\{X_1, X_2, \dots, X_n\}$  = the data set

$n$  = the number of data samples

$\varphi_0$  = the prior estimate of the variance

$v_0$  = the prior weight of the variance

$\hat{\mu}_w$  = the estimate of the mean

$\mu_0$  = the mean given by the prior distribution

$w_0$  = the weight of the prior distribution

$\{X_1, X_2, \dots, X_n\}$  = the data set

$n$  = the number of data samples

$\sigma_0^2$  = the variance given by the prior distribution

TABLE I (Continued)

Rayleigh Distribution	
Conventional Bayesian Estimate	B.E.F. Estimate
$\hat{\alpha}_n^2 = \frac{1}{2} \frac{a_0 \frac{b_0}{a_0} + n \left( \frac{1}{n} \sum_{i=1}^n X_i^2 \right)}{a_0 + n} \quad (3.2.9)$	$\hat{\alpha}_w^2 = \frac{1}{2} \frac{w_0 (2\alpha_0^2) + n \left( \frac{1}{n} \sum_{i=1}^n X_i^2 \right)}{w_0 + n} \quad (3.2.10)$
$\hat{\alpha}_n^2$ = the estimate of $\alpha^2 = \frac{1}{2} E\{X^2\}$	$\hat{\alpha}_w^2$ = the estimate of $\alpha^2 = \frac{1}{2} E\{X^2\}$
$\frac{b_0}{a_0}$ = the prior estimate of $2\alpha^2$	$\alpha_0^2$ = the value of $\alpha^2$ given by the prior distribution
$a_0$ = the prior weight of $\frac{b_0}{a_0}$	$w_0$ = the weight of the prior distribution
$\{X_1, X_2, \dots, X_n\}$ = the data set	$\{X_1, X_2, \dots, X_n\}$ = the data set
$n$ = the number of data samples	$n$ = the number of data samples

variance must be estimated, or both the mean and variance are to be estimated. Another familiar example from communications is the estimation of the parameter of a Rayleigh distribution. Estimates of these parameters derived by the conventional Bayesian approach are found in Appendix C. Similar estimates of these parameters given by B.E.F. may be developed using Theorems B.5.1, B.5.2, and B.5.3. These results are included in Table I for comparison.

It is readily apparent that the estimates given by the two approaches are very similar. In fact it appears that in many cases B.E.F. offers a more rational method for Bayesian parameter estimation than does the more conventional method given in Appendix C of choosing a prior distribution for the unknown parameter. The most difficult problem may be the selection of the prior weight  $w_0$ . A weakness in using B.E.F. for parameter estimation can be seen by comparing Equations 3.2.6 and 3.2.8. The B.E.F. estimate does not permit independent weighting of the mean and variance.

3.3 Estimation of a Distribution Function. For the examples of the previous section parameter estimates obtained from B.E.F. are essentially the same as those obtained by the conventional Bayesian parametric method. This does not imply that the B.E.F. estimate of the distribution function is the same as that given by the parametric method. The conventional method yields a distribution function which is a member of a predetermined family with the estimated parameters. The B.E.F. estimate is a weighted average of the prior estimate of the distribution function and the empirical distribution function

$$F_w(x) = \frac{w_0}{w_0+n} F_{w_0}(x) + \frac{n}{w_0+n} F_n(x) \quad (3.3.1)$$

as given by Equation A.4.38.

If the unknown distribution function is truly from the assumed family, it is to be expected that the parametric method would, on the average, yield a better estimate of the distribution function than B.E.F.

3.4 Integral Expected Square Error. A measure of the error of an estimate  $\hat{F}(x)$  of the distribution function  $F(x)$  is

$$I = \int_{-\infty}^{\infty} E\{[\hat{F}(x) - F(x)]^2\}dx \quad (3.4.1)$$

The integral expected square error  $I$  for the B.E.F. estimate  $\hat{F}(x)$  of  $F(x)$  can be calculated from the equation

$$I = \left(\frac{w_0}{w_0+n}\right)^2 \int_{-\infty}^{\infty} [F_{w_0}(x) - F(x)]^2 dx + \left(\frac{n}{w_0+n}\right)^2 \int_{-\infty}^{\infty} \frac{1}{n} F(x)[1-F(x)]dx \quad (3.4.2)$$

developed in Theorem B.3.2. For  $F(x)$  a normal random variable with variance  $\sigma^2$  the integral

$$\int_{-\infty}^{\infty} F(x)[1-F(x)]dx \cong 0.564\sigma \pm 0.004\sigma \quad (3.4.3)$$

was evaluated using numerical integration. Given  $F_{w_0}(x)$  is a normal distribution with mean one and variance two and  $F(x)$  is normal with zero mean and variance one, the integral

$$\int_{-\infty}^{\infty} [F_{w_0}(x) - F(x)]^2 dx \cong 0.244 \pm 0.004 \quad (3.4.4)$$

was evaluated numerically. Thus Equations 3.4.2, 3.4.3 and 3.4.4 yield the integral expected square error for the B.E.F. estimates of the

distribution function  $F(x)$  shown in the upper curves of Figures 1, 2 and 3.

The integral expected square error  $I$  for parametric estimates with unknown mean and variance are plotted on the same graphs.  $\hat{F}(x)$  is normal with mean and variance given by the appropriate equation in Table I. For the parametric cases, the prior weight given on the graph is the value of the term in the conventional Bayesian estimate of Table I that corresponds to  $w_0$  in the B.E.F. estimate of the parameter. In the interest of keeping computer time short, values were calculated only for ten, thirty and one hundred samples. Using

$$\int_{-\infty}^{\infty} E\{[\hat{F}(x)-F(x)]^2\}dx = E\left\{\int_{-\infty}^{\infty} [\hat{F}(x)-F(x)]^2 dx\right\} \quad (3.4.5)$$

the expected value was estimated by taking the average of one hundred integrals. The numerical methods used limit the accuracy of these estimates to approximately ten per cent. The lower bound for the unknown mean case in Figure 3 was calculated for  $\hat{F}(x)$  normal with zero mean and variance two. This is the lower bound for this example because the variance of the estimate is always two and the minimum error occurs when the estimate of the mean is the true value. Thus

$$I = \int_{-\infty}^{\infty} [\hat{F}(x)-F(x)]^2 dx \cong 0.0199 \pm 0.004 \quad . \quad (3.4.6)$$

If the correct form of the distribution is chosen and the prior density of the parameter does not exclude the true value, the figures indicate the anticipated result, i.e., the conventional parametric method provides a better result than given by B.E.F. Figure 3 indicates,

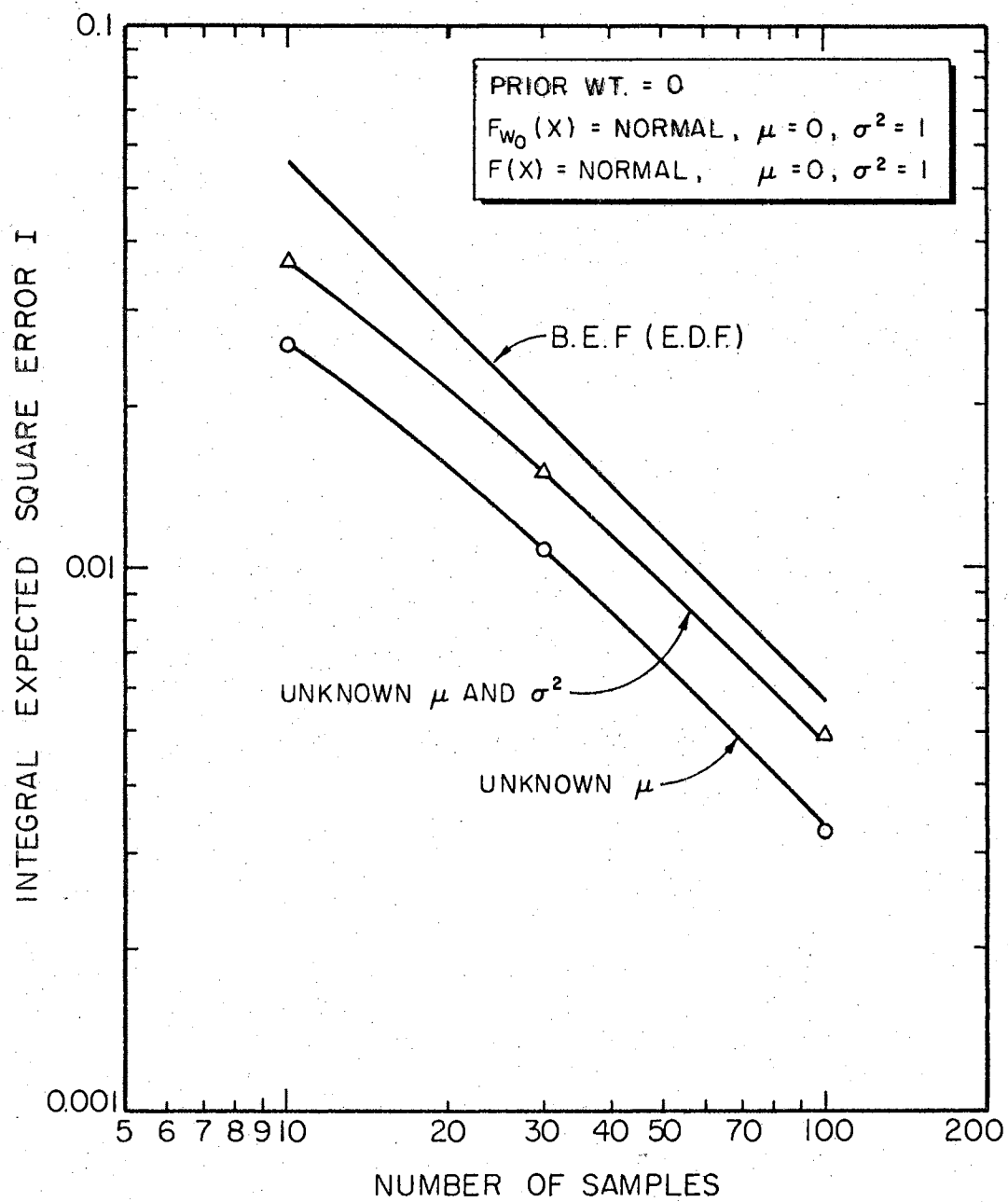


Figure 1. Integral Expected Square Error With Zero Prior Weight

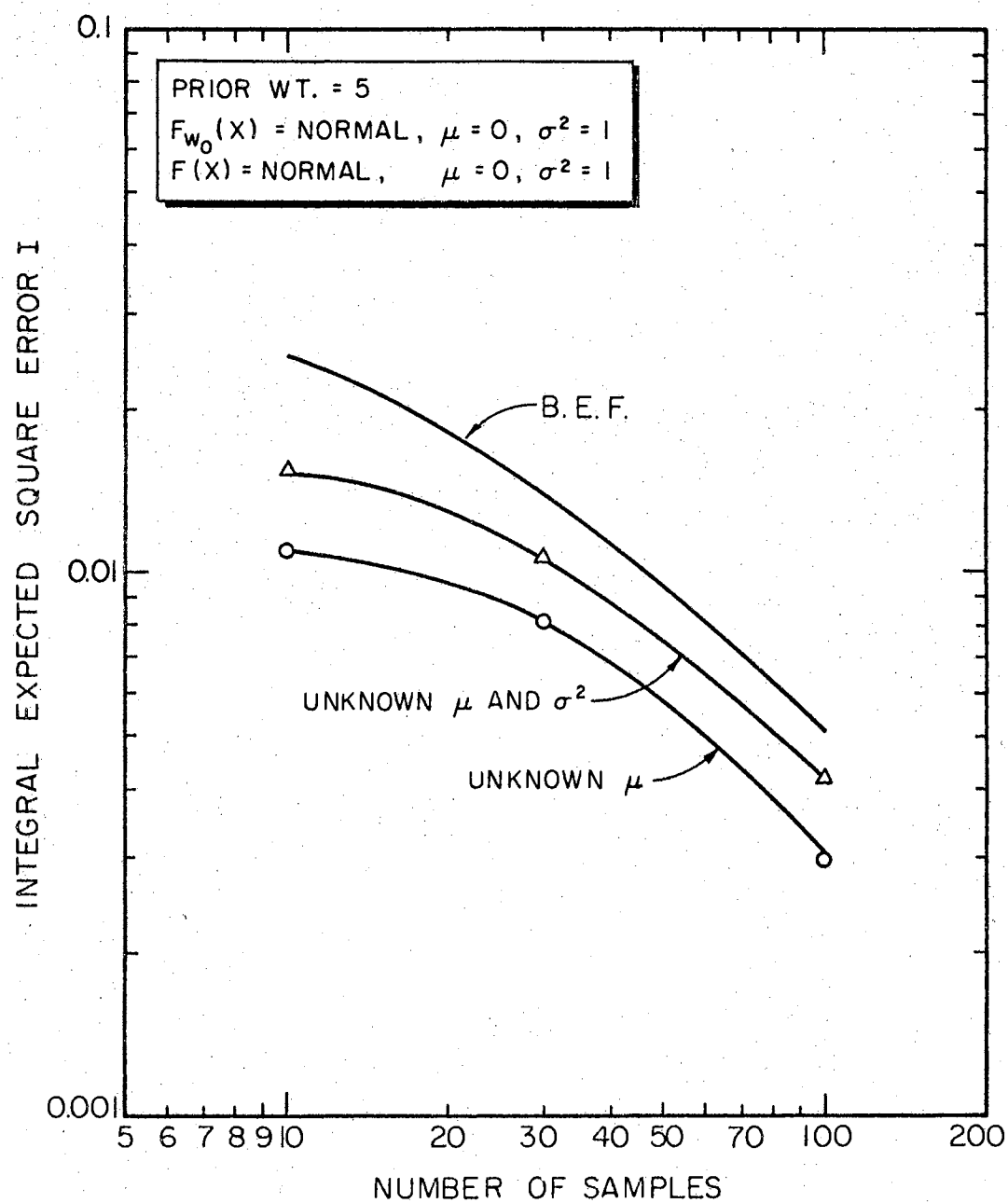


Figure 2. Integral Expected Square Error With Correct Prior

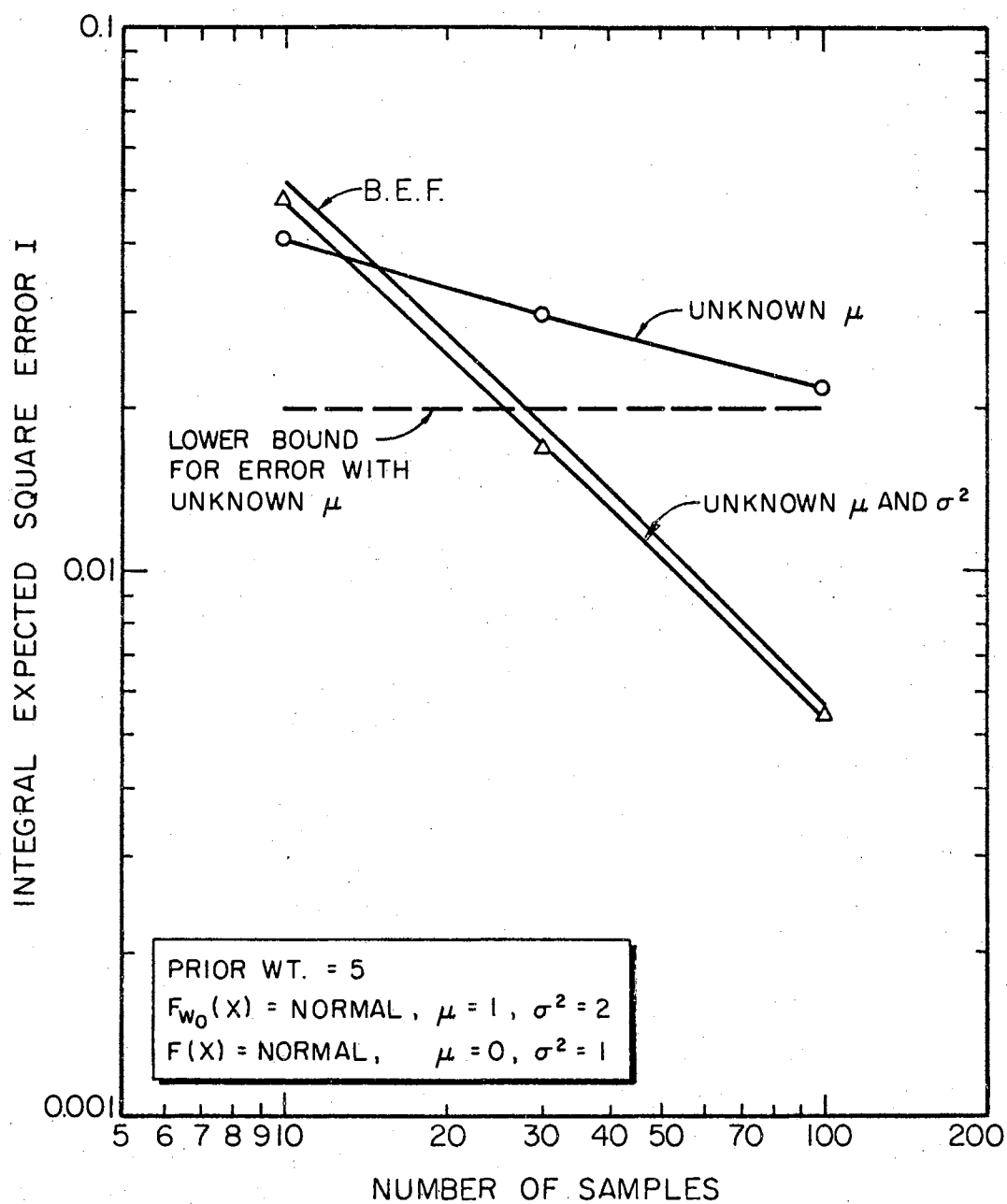


Figure 3. Integral Expected Square Error With Incorrect Prior

however, that for only a relatively small error in the form (in this example the wrong variance) of the distribution function, the B.E.F. estimate of the distribution function may be superior after a small number of samples.

For a Rayleigh distribution

$$F_x(x) = \begin{cases} 1 - \exp\left(-\frac{x^2}{2\alpha^2}\right) & x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.4.7)$$

Thus

$$\begin{aligned} \int_{-\infty}^{\infty} F(x)[1-F(x)]dx &= \frac{\pi}{2} (\sqrt{2} - 1)\alpha \\ &\approx 0.367\alpha \end{aligned} \quad (3.4.8)$$

If the prior estimate  $F_{w_0}(x)$  is Rayleigh with parameter  $\alpha_0$  and  $F(x)$  is Rayleigh with parameter  $\alpha$ , then

$$\int_{-\infty}^{\infty} [F_{w_0}(x) - F(x)]^2 dx = \sqrt{\pi} \left[ \frac{\alpha_0 + \alpha}{2} - \frac{\alpha_0 \alpha \sqrt{2}}{\sqrt{\alpha_0^2 + \alpha^2}} \right] \quad (3.4.9)$$

Using Equations 3.4.8 and 3.4.9, Equation 3.4.2 becomes

$$I = \left(\frac{w_0}{w_0+n}\right)^2 \sqrt{\pi} \left[ \frac{\alpha_0 + \alpha}{2} - \frac{\alpha_0 \alpha \sqrt{2}}{\sqrt{\alpha_0^2 + \alpha^2}} \right] + \left(\frac{n}{w_0+n}\right)^2 \cdot \frac{1}{n} \cdot \frac{\pi}{2} (\sqrt{2} - 1)\alpha \quad (3.4.10)$$

Interchanging the order of expectation and integration as in Equation 3.4.5, Equation 3.4.9 indicates that for the parametric estimate of the Rayleigh distribution

$$I = E\left\{\sqrt{\pi} \left[ \frac{\alpha_n + \alpha}{2} - \frac{\alpha_n \alpha \sqrt{2}}{\sqrt{\alpha_n^2 + \alpha^2}} \right]\right\}, \quad (3.4.11)$$

where  $\alpha$  is the true value of the parameter and  $\alpha_n$  is the estimate of the parameter given in Table I.

An approximate value for Equation 3.4.11 was obtained in the following manner. Ten independent samples from a Rayleigh distribution with parameter  $\alpha = 1$  were generated by the computer. A value for  $\alpha_n$  was calculated from Equation 3.2.9 with the prior weight set equal to zero. This calculation was repeated 100 times and the average of the values was taken as the estimate of  $I$  in Equation 3.4.11. The value of 0.016 obtained was less than the value 0.0367 obtained from Equation 3.4.10 with  $w_0 = 0$ ,  $n = 10$  and  $\alpha = 1$ . Thus, as would be expected, the expected value of the integral square error of the parametric estimate is less than the error for the nonparametric B.E.F. estimate.

An interesting observation can be made. For  $X$  Rayleigh distributed

$$\begin{aligned} \text{Var}\{X\} &= \alpha^2 \left(2 - \frac{\pi}{2}\right) \\ &= \sigma_x^2 \end{aligned} \quad (3.4.12)$$

Thus

$$\begin{aligned} \int_{-\infty}^{\infty} F(x)[1-F(x)]dx &= \frac{\pi}{2} (\sqrt{2} - 1) \frac{\sigma_x}{\sqrt{2 - \frac{\pi}{2}}} \\ &= 0.565\sigma_x \end{aligned}, \quad (3.4.13)$$

which is essentially the same as Equation 5.3.3 for  $F(x)$  normal. The significance, if any, of this observation is not known.

### 3.5 Discussion. The examples presented indicate three points.

The estimation of a parameter by B.E.F. can be expected to yield a result very similar to that obtained by the conventional Bayesian parametric method. Second, as would be expected, if the form of the true distribution function is known, the parametric approach is to be preferred to the nonparametric B.E.F. Third, a relatively small error in the form of the assumed distribution function can make the nonparametric method superior in accuracy (as measured by the expected integral square error) after a relatively small number of samples.

As discussed by Spraggins (15), in many cases conventional Bayesian estimation of parameters leads to a recursive estimation scheme which requires only the storage of the prior estimate and the present datum point. These results are believed to hold for B.E.F. parameter estimation also, but this remains to be proved. Estimation of the distribution function by B.E.F. of course requires the storage of all data. In many cases this objection may be overcome by limiting the resolution of the data, e.g., use a histogram instead of the empirical distribution function. A further possibility is suggested by the work of Specht (16) as discussed in Chapter VI.

## CHAPTER IV

### APPLICATION TO ADAPTIVE BINARY SIGNAL DETECTION

4.1 Introduction. This chapter considers the detection of a binary signal in the presence of additive noise where the probabilities of the signal having the value zero or one are known, but the distribution of the noise is unknown.

Such a detection problem can exist in radar. The decision is made by comparing the received signal and a threshold to decide whether a target is present or not. The threshold is set to minimize the effect of an incorrect decision. Samples of the noise may be obtained when it is known that no target is present. In addition some idea about the noise distribution usually exists, e.g., the noise was measured at a previous time. If the noise distribution changes rather slowly, the old estimate of the distribution serves as a prior estimate of the noise at the present. Given the form of the noise distribution, the Bayesian parametric method is applicable. With the form of the noise distribution unknown, B.E.F. offers a method for combining prior knowledge and current data. Some aspects of the errors involved in applying B.E.F. are discussed.

4.2 Bayes' Decision Rules. Let the signal  $S$  be a binary random variable taking on the values zero or one with probabilities

$$P\{1\} = p \qquad (4.2.1)$$

and

$$P\{0\} = 1 - p \quad . \quad (4.2.2)$$

In the presence of independent additive noise  $X$ , the received signal  $Z$  is defined by

$$Z = S + X \quad , \quad (4.2.3)$$

It follows that

$$F_{Z|S=0}(z) = F_X(z) \quad (4.2.4)$$

and

$$F_{Z|S=1}(z) = F_X(z-1) \quad . \quad (4.2.5)$$

Define  $D_1=1$  as the decision that  $S=1$  and  $D_0=0$  as the decision that  $S=0$ . The following deterministic decision rules are used.

$$Z \geq y \Rightarrow D_1 \quad (4.2.6)$$

and

$$Z < y \Rightarrow D_0 \quad . \quad (4.2.7)$$

The cost  $C_{D=i|S=j}$  of making a decision  $D$  depends on the value of  $S$ . The reasonable assumption that  $C_{i|i} < C_{i|j}$  for  $i \neq j$  is usually made.

Given the distribution function of  $X$ , the average risk  $R_y$  associated with the threshold  $y$  is

$$\begin{aligned} R_y(F_X, p) = & pC_{1|1} \int_y^\infty dF_{Z|S=1} + pC_{0|1} \int_{-\infty}^y dF_{Z|S=1} \\ & + (1-p)C_{1|0} \int_y^\infty dF_{Z|S=0} + (1-p)C_{0|0} \int_{-\infty}^y dF_{Z|S=0} \quad . \quad (4.2.8) \end{aligned}$$

Using Equations 4.2.4 and 4.2.5, Equation 4.2.8 becomes

$$R_y(F_X, p) = pC_{1|1} + p(C_{0|1} - C_{1|1})F_X(y-1) + (1-p)C_{1|0} + (1-p)(C_{0|0} - C_{1|0})F_X(y) \quad (4.2.9)$$

Given  $p$  and  $F_X$ , the value of  $y_0$  of  $y$  that minimizes our expected risk  $R_y$  may be found by the familiar methods of calculus. The case to be considered is for  $p$  known and  $F_X$  unknown.

4.3 Detection With Unknown Noise. If the distribution function  $F_X$  of the noise is known, the risk  $R_y$  may be minimized as stated. In practice  $F_X$  is usually unknown and some estimate  $\hat{F}_X$  of  $F_X$  must be used. Define

$$\hat{R}_Y = R_y(\hat{F}_X, p) = pC_{1|1} + p(C_{0|1} - C_{1|1})\hat{F}_X(y-1) + (1-p)C_{1|0} + (1-p)(C_{0|0} - C_{1|0})\hat{F}_X(y) \quad (4.3.1)$$

The value  $\hat{y}_0$  of  $y$  which minimizes  $\hat{R}_Y$  is used as the estimate of the value  $y_0$  which minimizes  $R_y$ .

Some interesting properties of  $\hat{R}_Y$  may be derived for  $\hat{F}_X$  the B.E.F. estimate of  $F_X$ .

Theorem 4.3.1. If  $p$  and  $C_{i|j}$ ,  $i, j=0,1$  are finite constants and  $\hat{F}_X = F_w$ ,  $R_Y$  converges uniformly in  $y$  to  $R_y$  with probability one for  $-\infty < y < \infty$  as  $n \rightarrow \infty$ , i.e.,

$$P\{\lim_{n \rightarrow \infty} [\sup_x |\hat{R}_Y - R_y|] = 0\} = 1 \quad (4.3.2)$$

Proof. From the definitions of  $R_y$  and  $\hat{R}_Y$

$$\hat{R}_Y - R_y = p(C_{0|1} - C_{1|1})[F_w(y-1) - F_X(y-1)] + (1-p)(C_{0|0} - C_{1|0})[F_w(y) - F_X(y)] \quad (4.3.3)$$

The desired result follows immediately from Theorem B.2.2 which states that

$$P\{\lim_{n \rightarrow \infty} [\sup_x |F_w(x) - F_X(x)|] = 0\} = 1 \quad .$$

Theorem 4.3.2. If  $\hat{F}_X = F_w$ , then

$$E\{\hat{R}_Y\} = \frac{w_0}{w_0+n} R_y(F_{w_0}, p) + \frac{n}{w_0+n} R_y(F_X, p) \quad (4.3.4)$$

where  $F_{w_0}$  is the prior estimate of  $F_X$ . If  $F_{w_0}$  is an unbiased estimate of  $F_X$ , then

$$E\{\hat{R}_Y\} = R_y \quad . \quad (4.3.5)$$

Proof. From the definitions of  $\hat{R}_Y$  and  $R_y$

$$\begin{aligned} E\{\hat{R}_Y\} &= E\{pC_{1|1} + p(C_{0|1} - C_{1|1}) \left[ \frac{w_0}{w_0+n} F_{w_0}(y-1) + \frac{n}{w_0+n} F_n(y-1) \right] \\ &\quad + (1-p)C_{1|0} + (1-p)(C_{0|0} - C_{1|0}) \left[ \frac{w_0}{w_0+n} F_{w_0}(y) + \frac{n}{w_0+n} F_n(y) \right]\} \\ &= \frac{w_0}{w_0+n} \{pC_{1|1} + p(C_{0|1} - C_{1|1}) F_{w_0}(y-1) + (1-p)C_{1|1} + (1-p)(C_{0|0} - C_{1|0}) F_{w_0}(y)\} \\ &\quad + \frac{n}{w_0+n} \{pC_{1|1} + p(C_{0|1} - C_{1|1}) F_X(y-1) + (1-p)C_{1|1} + (1-p)(C_{0|0} - C_{1|0}) F_X(y)\} \\ &= \frac{w_0}{w_0+n} R_y(F_{w_0}, p) + \frac{n}{w_0+n} R_y(F_X, p) \quad . \end{aligned} \quad (4.3.6)$$

The second desired result follows directly from the definition of an unbiased estimate and the above step.

In a decision theory problem such as this the most significant measure of the error of the estimate  $\hat{F}_X$  of  $F_X$  is

$$E\{R_{y_0} - R_{Y_0}\} \quad , \quad (4.3.7)$$

where  $Y_0$  is the random variable that minimizes  $\hat{R}_y$  and  $y_0$  is the value of  $y$  which minimizes  $R_y$ . Equation 4.3.7 must be evaluated numerically for most applications. A less meaningful though more convenient measure of the error of the estimate is

$$E\{(R_{y_0} - \hat{R}_{y_0})^2\} \quad . \quad (4.3.8)$$

If  $F_X$  is known,  $y_0$  can be calculated. Thus for  $\hat{F}_X = F_w$ , given any prior distribution  $F_{w_0}$  and prior weight  $w_0$  Equation 4.3.8 can be evaluated from the following theorem.

Theorem 4.3.3. If  $\hat{F}_X = F_w$ , then

$$\begin{aligned} E\{(R_y - \hat{R}_y)^2\} = & A^2 \left\{ \left( \frac{w_0}{w_0+n} \right)^2 [F_{w_0}(y-1) - F_X(y-1)]^2 + \left( \frac{n}{w_0+n} \right)^2 \frac{1}{n} F_X(y-1) [1 - F_X(y-1)] \right\} \\ & + B^2 \left\{ \left( \frac{w_0}{w_0+n} \right)^2 [F_{w_0}(y) - F_X(y)]^2 + \left( \frac{n}{w_0+n} \right)^2 \frac{1}{n} F_X(y) [1 - F_X(y)] \right\} \\ & - 2AB \left\{ \left( \frac{w_0}{w_0+n} \right)^2 [F_{w_0}(y-1) - F_X(y-1)] [F_{w_0}(y) - F_X(y)] \right. \\ & \left. + \left( \frac{n}{w_0+n} \right)^2 \frac{1}{n} F_X(y-1) [1 - F_X(y)] \right\} \quad , \quad (4.3.9) \end{aligned}$$

where

$$A = p(C_{0|1} - C_{1|1}) \quad (4.3.10)$$

and

$$B = - (1 - p)(C_{0|0} - C_{1|0}) \quad . \quad (4.3.11)$$

Proof. From the definitions of  $R_y$ ,  $\hat{R}_y$  and  $F_w = \hat{F}_X$

$$\begin{aligned}
(R_y - \hat{R}_y)^2 &= A^2 [F_X(y-1) - F_W(y-1)]^2 + B^2 [F_X(y) - F_W(y)]^2 \\
&- 2AB[F_X(y-1) - F_W(y-1)][F_X(y) - F_W(y)] \quad . \quad (4.3.12)
\end{aligned}$$

$$\begin{aligned}
E\{(R_y - \hat{R}_y)^2\} &= A^2 E\{[F_X(y-1) - F_W(y-1)]^2\} + B^2 E\{[F_X(y) - F_W(y)]^2\} \\
&- 2AB E\{[F_X(y-1) - F_W(y-1)][F_X(y) - F_W(y)]\} \quad . \quad (4.3.13)
\end{aligned}$$

On applying Theorems B.3.1 and B.3.4 the previous equation becomes

$$\begin{aligned}
E\{(R_y - \hat{R}_y)^2\} &= \\
&A^2 \left\{ \left( \frac{w_0}{w_0+n} \right)^2 [F_{w_0}(y-1) - F_X(y-1)]^2 + \left( \frac{n}{w_0+n} \right)^2 \frac{1}{n} F_X(y-1)[1 - F_X(y-1)] \right\} \\
&+ B^2 \left\{ \left( \frac{w_0}{w_0+n} \right)^2 [F_{w_0}(y) - F_X(y)]^2 + \left( \frac{n}{w_0+n} \right)^2 \frac{1}{n} F_X(y)[1 - F_X(y)] \right\} \\
&- 2AB \left\{ \left( \frac{w_0}{w_0+n} \right)^2 [F_{w_0}(y-1) - F_X(y-1)][F_{w_0}(y) - F_X(y)] \right. \\
&\left. + \left( \frac{n}{w_0+n} \right)^2 \frac{1}{n} F_X(y-1)[1 - F_X(y)] \right\} \quad . \quad (4.3.14)
\end{aligned}$$

If  $F_X$  is normal with zero mean and variance  $\sigma^2$ ,  $y_0$  may easily be found (1)

$$y_0 = \frac{1}{2} + \sigma^2 \ln\left(\frac{B}{A}\right) \quad . \quad (4.3.15)$$

Thus for this particular case Equation 4.3.8 can be easily evaluated with the aid of Theorem 4.3.3.

Although in a real problem  $F_X$  would be unknown, a general idea of  $F_X$  would exist. Therefore the theorems presented here are of value in obtaining a feel for the errors involved in a practical binary detection problem and are especially valuable in that the influence of the prior distribution is explicitly shown.

## CHAPTER V

### BAYES' EMPIRICAL DISTRIBUTION FUNCTION FOR A RANDOM VECTOR

5.1 Introduction. An estimate of the distribution function of a random vector must often be obtained from prior information and sample data. Keehn (5) for example has developed a Bayesian method for estimating the mean and covariance matrices for an n-variate normal distribution. Some thoughts are presented in the following section about the extension of B.E.F. to k-dimensional random vectors.

5.2 B.E.F. for a Random Vector. Consider the k-dimensional random vector

$$\underline{X} = (X_1, X_2, \dots, X_k) \quad , \quad (5.2.1)$$

and a sample vector

$$\underline{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,k}) \quad (5.2.2)$$

of  $\underline{X}$ . Guided by the one dimensional development of Appendix A, let

$$Q = F_{\underline{X}}(\underline{x}) \quad , \quad (5.2.3)$$

and assume that Q has a first order beta density

$$f_Q(q) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1} (1-q)^{\beta-1} & 0 \leq q \leq 1 \\ 0 & \text{otherwise,} \end{cases} \quad (5.2.4)$$

where  $q$ ,  $\alpha$  and  $\beta$  are real valued functions of the random vector  $\underline{X}$  such that

$$0 \leq q(\underline{x}) \leq 1 \quad (5.2.5)$$

$$\alpha(\underline{x}) > 0 \quad (5.2.6)$$

and

$$\beta(\underline{x}) > 0 \quad (5.2.7)$$

As in Appendix A, let

$$w_0 = \alpha(\underline{x}) + \beta(\underline{x}) \quad (5.2.8)$$

where  $w_0$  is not a function of  $\underline{x}$ , i.e., a constant. Thus from the properties of a beta density,

$$\begin{aligned} \mu_Q &= E\{Q\} \\ &= \frac{\alpha(\underline{x})}{w_0} \end{aligned} \quad (5.2.9)$$

and

$$\begin{aligned} \sigma_Q^2 &= \text{Var}\{Q\} \\ &= \frac{\alpha(\underline{x})\beta(\underline{x})}{(w_0+1)w_0^2}, \end{aligned} \quad (5.2.10)$$

where the dependence of  $\mu_Q$  and  $\sigma_Q^2$  on  $\underline{x}$  is understood.

Given  $n$  independent samples of  $\underline{x}$

$$\{\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n\}, \quad (5.2.11)$$

applying Bayes' theorem as in Section A.3 yields posterior values

$$\mu_Q' = \frac{a(\underline{x}) + \alpha(\underline{x})}{w_0 + n} \quad (5.2.12)$$

and

$$\sigma_Q^{2'} = \frac{[a(\underline{x}) - \alpha(\underline{x})][n - a(\underline{x}) + \beta(\underline{x})]}{[w_0 + n + 1]w_0^2} \quad (5.2.13)$$

where

$$a(\underline{x}) = (\text{the number of trials out of } n \text{ samples such that } x_{i,1} \leq x_1, x_{i,2} \leq x_2, \dots, x_{i,k} \leq x_k) \quad (5.2.14)$$

With the exception of the argument being  $\underline{x}$  instead of  $x$ , Equations 5.2.9, 5.2.10, 5.2.12 and 5.2.13 correspond to Equations A.3.4, A.3.5, A.3.6 and A.3.7 respectively.

Using  $\mu_Q$  as the prior estimate  $F_{w_0}(\underline{x})$  of  $F(\underline{x})$  and  $\mu_Q'$  as the posterior estimate  $F_w(\underline{x})$  of  $F(\underline{x})$ , Equations 5.2.9 and 5.2.12 yield

$$F_w = \frac{w_0}{w_0 + n} \cdot \frac{\alpha(\underline{x})}{w_0} + \frac{n}{w_0 + n} \cdot \frac{a(\underline{x})}{n} \quad (5.2.15)$$

From Equation 5.2.14 it can be seen that  $a(\underline{x})/n$  is the empirical distribution function for the  $k$ -dimensional random vector  $\underline{X}$ , i.e.,

$$F_n(\underline{x}) = \frac{a(\underline{x})}{n} \quad (5.2.16)$$

Thus Equation 5.2.15 becomes

$$F_w(\underline{x}) = \frac{w_0}{w_0 + n} F_{w_0}(\underline{x}) + \frac{n}{w_0 + n} F_n(\underline{x}) \quad (5.2.17)$$

Hence B.E.F. for a random vector is of the same form as B.E.F. for a random variable.

As Equation 5.2.17 is the weighted average of two distribution

functions,  $F_w(\underline{x})$  can readily be shown to be a distribution function also. Convergence and error theorems similar to those of Appendix B remain to be proved for  $\underline{X}$  a random vector.

## CHAPTER VI

### SMOOTHING BAYES' EMPIRICAL DISTRIBUTION FUNCTION

6.1 Introduction. As was shown in Theorem B.4.1, the B.E.F. estimate  $F_w(x)$  of  $F(x)$  is an unbiased estimate of  $F(x)$  if the prior distribution function  $F_{w_0}(x)$  is an unbiased estimate of  $F(x)$ . Regardless of the prior,  $F_w(x)$  is an asymptotically unbiased estimate of  $F(x)$  for finite  $w_0$  as shown in Theorem B.4.2. In spite of this B.E.F. may not be a desirable estimate in some applications. If the distribution function  $F(x)$  is known to be continuous, the discontinuous nature of  $F_w(x)$  may be disconcerting if not an actual problem. Therefore some form of smoothing may be desirable.

6.2 Smoothing B.E.F. Parzen (10) has developed a method for estimating the density function  $f(x)$  from  $n$  independent samples  $\{X_1, X_2, \dots, X_n\}$ . This estimate  $g_n(x)$  of  $f(x)$ , where

$$F(x) = \int_{-\infty}^x f(u) du \quad ,$$

is of the form

$$\begin{aligned} g_n(x) &= \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) dF_n(y) \\ &= \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right) \quad . \end{aligned}$$

$F_n(x)$  is the familiar empirical distribution function and  $K(y)$  is a weighting function satisfying certain conditions. From an engineering

point of view this is equivalent to time domain filtering where  $K(x)$  is the filter and  $x$  is analogous to time. From another point of view it is a weighted average of  $n$  density functions<sup>1</sup>  $\frac{1}{h} K(\frac{x-X_j}{h})$ , where  $X_j$  determines the shift of  $K$  with respect to the origin and  $h$  determines the spread about  $X_j$  of  $K$ . Thus a smoothed estimate of  $F(x)$  is

$$G_n(x) = \int_{-\infty}^x g_n(y) dy \quad .$$

A natural application of this to B.E.F. would be to use

$$G_w(x) = \frac{w_0}{w_0+n} F_{w_0}(x) + \frac{n}{w_0+n} G_n(x)$$

for the estimate of  $F(x)$ . Properties of this estimate need to be investigated. Because of the nature of  $G_w(x)$ , it is to be expected that  $G_w(x)$  would have properties similar to  $G_n(x)$ . The properties of  $G_n(x)$  have been investigated by several authors (7,8,10,13,17,18).

6.3 Specht's Approximation. A major difficulty in applying the E.D.F. to engineering problems is that all data must be kept in storage. This difficulty carries over to B.E.F. Specht (16) has developed a series approximation for Parzen's method that requires a fixed storage capacity. It would appear that this approach might be used to simultaneously reduce data storage requirements and provide smoothing of B.E.F.

Specht's approximation chooses a weighting function  $K(x)$  of the form

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) \quad . \quad (6.3.1)$$

---

<sup>1</sup> $K(x)$  is not necessarily a density function.

Thus

$$g_n(x) = \int_{-\infty}^x \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{(x-y)^2}{2\sigma^2}\right] dF_n(y) \quad , \quad (6.3.2)$$

or

$$g_n(x) = \frac{1}{\sigma \sqrt{2\pi}} \frac{1}{n} \sum_{i=1}^n \exp\left(-\frac{(x-X_i)^2}{2\sigma^2}\right) \quad . \quad (6.3.3)$$

Writing

$$\exp\left[-\frac{(x-X_i)^2}{2\sigma^2}\right] = \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(\frac{xx_i}{\sigma^2}\right) \exp\left(-\frac{X_i^2}{2\sigma^2}\right) \quad (6.3.4)$$

and expanding  $\exp(xx_i/\sigma^2)$  in a Taylor's series, Equation 6.3.3 becomes

$$g_n(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \sum_{r=0}^{\infty} C_{r,n} x^r \quad (6.3.5)$$

where

$$C_{r,n} = \frac{1}{r! \sigma^{2r}} \frac{1}{n} \sum_{i=1}^n X_i^r \exp\left(-\frac{X_i^2}{2\sigma^2}\right) \quad . \quad (6.3.6)$$

Noting that

$$C_{r,n+1} = \frac{n}{n+1} C_{r,n} + \frac{1}{r! \sigma^{2r}} \cdot \frac{1}{n+1} X_{n+1}^r \exp\left(-\frac{X_{n+1}^2}{2\sigma^2}\right) \quad , \quad (6.3.7)$$

it can be seen that a recursive relation exists for  $C_{r,n}$ . Hence for a fixed number of terms  $M$  in the Taylor's series approximation

$$g_n(x) \approx \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \sum_{r=0}^M C_{r,n} x^r \quad . \quad (6.3.8)$$

Thus the approximation requires the storage of a fixed number of terms regardless of the sample size. A number of properties of this

approximation are investigated by Specht and should be of value in applying it to B.E.F.

## CHAPTER VII

### SUMMARY AND CONCLUSIONS

7.1 Summary. Taking a specific weighted average of the prior estimate of a distribution function and the empirical distribution function as the posterior estimate of a distribution function was shown to be consistent with Bayes' theorem. This result, referred to as Bayes' empirical distribution function (B.E.F.), was compared to conventional parametric estimation using theorems developed in Appendix B. An application of B.E.F. to the detection of a binary signal in unknown noise was given. A method for extending B.E.F. to the estimation of a distribution function for finite dimensional random vectors was outlined. The adaptation of B.E.F. to Parzen's method was suggested for obtaining continuous estimates of a distribution function.

7.2 Conclusions. B.E.F. offers a simple, logical method for combining prior knowledge and independent sample data to estimate a distribution function. B.E.F. was shown to converge to the true distribution function with probability one regardless of the prior distribution. Given the true form of the distribution and a prior density for a parameter which does not exclude the true value, conventional Bayesian parametric estimation is superior to B.E.F. If, however, the assumed form of the distribution function is incorrect or the assumed prior excludes the true value of a parameter, B.E.F. may yield a superior estimate after a relatively small number of samples. Thus B.E.F. allows the use of prior

information as does conventional Bayesian parametric estimation, but B.E.F. will converge (with probability one) to the true distribution whereas conventional Bayesian estimation may not. B.E.F. may also be used for parametric Bayesian estimation with results very similar to the conventional Bayesian technique. For most applications it appears that B.E.F. can replace the conventional Bayesian technique either for non-parametric or parametric estimation.

7.3 Recommendations for Further Study. While additional investigation of the application of B.E.F. to specific engineering problems is of interest, more general investigations should prove more valuable. Convergence and error theorems need to be proved for the extension to random vectors outlined in Chapter V. The smoothing suggested in Chapter VI needs further study to determine if the improvement in the estimate, if any, justifies the added complexity. Extension of B.E.F. to dependent samples should be a fruitful field for further investigation.

## BIBLIOGRAPHY

1. Harman, W. W. Principles of the Statistical Theory of Communication. New York: McGraw Hill, 1963.
2. Hogg, Robert V. and Allen F. Craig. Introduction to Mathematical Statistics. New York: The Macmillan Company, 1965.
3. Howard, Ronald A. "Information Value Theory." IEEE Transactions on Systems Science and Cybernetics. Vol. SSC-2 (August, 1966).
4. Howard, Ronald A. "Bayesian Decision Models for System Engineering." IEEE Transactions on Systems Science and Cybernetics. Vol. SSC-1 (November, 1965).
5. Keehn, Daniel G. "A Note on Learning for Gaussian Properties." IEEE Transactions on Information Theory. Vol. IT-11 (January, 1965).
6. Loeve, Michel. Probability Theory. New York: D. Van Nostrand Company, Inc., 1960.
7. Murthy, V. K. "Estimation of Probability Density." Annals of Mathematical Statistics. Vol. 36 (1965).
8. Nadaraya, E. A. "On Non-Parametric Estimates of Density Functions and Regression Curves." Theory of Probability and Its Applications. Vol. 10 (1965).
9. Papoulis, Athanasios. Probability, Random Variables, and Stochastic Processes. New York: McGraw Hill, 1965.
10. Parzen, Emanuel. "On Estimation of a Probability Density Function and Mode." Annals of Mathematical Statistics. Vol. 33 (1962).
11. Rao, C. Radhakrishna. Linear Statistical Inference and Its Applications. New York: John Wiley & Sons, 1965.
12. Robbins, H. "The Empirical Bayes Approach to Statistical Decision Problems." Annals of Mathematical Statistics. Vol. 35 (1964).
13. Rosenblatt, Murray. "Remarks on Some Nonparametric Estimates of a Density Function." Annals of Mathematical Statistics. Vol. 27 (1956).

14. Savage, Leonard J. "Bayesian Statistics." In Recent Developments in Information and Decision Processes. Ed. Robert E. Machol and Paul Gray. New York: Macmillan Company, 1963.
15. Spragins, John. "A Note on the Iterative Application of Bayes' Rule." IEEE Transactions on Information Theory. Vol. IT-11 (October, 1965).
16. Specht, Donald E. "Series Estimation of a Probability Density Function." (To be published in Technometrics.)
17. Watson, G. S. and M. R. Leadbetter. "On the Estimation of the Probability Density, I." Annals of Mathematical Statistics. Vol. 34 (1963).
18. Woodroffe, Michael. "On the Maximum Deviation of the Sample Density." Annals of Mathematical Statistics. Vol. 38 (1967).

## APPENDIX A

### DERIVATION OF BAYES' EMPIRICAL DISTRIBUTION FUNCTION

A.1 Introduction. Assume that the distribution function of a population is to be learned from a set of independent samples. Further assume that some prior knowledge of the distribution function is available. This knowledge is in the form of a prior distribution function and the degree of confidence in the prior distribution function. Bayes' theorem is used to develop a method of combining the prior distribution with the sample data to obtain a posterior distribution.

A.2 The Prior Distribution. Assume that the probability distribution function  $F(x)$  of the population is a sample function from a stochastic process with a beta first order density function. Also assume that the prior knowledge about  $F(x)$  is in the form of the parameters of the beta function. If  $Q$  is the probability that a sample  $X$  from the population is less than or equal to  $x$ , i.e.,

$$\begin{aligned} Q &= P(X \leq x) \\ &= F(x), \end{aligned} \quad (\text{A.2.1})$$

then  $Q$  has the prior beta density function

$$f_Q(q) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1} (1-q)^{\beta-1} & 0 \leq q \leq 1 \\ & \alpha, \beta > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A.2.2})$$

where

$\alpha$ ,  $\beta$  and  $q$  are functions of  $x$ .

Then given  $q$ , the probability that  $a$  out of  $n$  samples would be less than or equal to  $x$  will be binomial with probability  $q$ . Thus

$$\begin{aligned} P_{A|q,n}(a) &= P\{a \text{ out of } n \text{ samples } \leq x\} \\ &= \binom{n}{a} q^a (1-q)^{n-a}, \quad a=1,2,\dots,n \end{aligned} \quad (A.2.3)$$

A.3 The Posterior Distribution. After sampling the population, Bayes' theorem may be used to obtain the posterior density function for  $Q$ . Bayes' theorem gives

$$f_{Q|\mathcal{E}}(q) = \frac{P_{A|q,n}(a)f_Q(q)}{\int_{-\infty}^{\infty} P_{A|q,n}(a)f_Q(u)du}, \quad (A.3.1)$$

after substitution of Equations A.2.2 and A.2.3, Equation A.3.1 becomes

$$f_Q(q) = \begin{cases} \frac{\binom{n}{a} q^a (1-q)^{n-a} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1} (1-q)^{\beta-1}}{\int_0^1 \binom{n}{a} u^a (1-u)^{n-a} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} (1-u)^{\beta-1} du} & 0 \leq q \leq 1 \\ 0 & \text{otherwise,} \end{cases} \quad (A.3.2)$$

and on integration

$$f_{Q|\mathcal{E}}(q) = \begin{cases} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(a+\alpha)\Gamma(n-a+\beta)} q^{a+\alpha-1} (1-q)^{n-a+\beta-1} & 0 \leq q \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (A.3.3)$$

The posterior density function is again beta and the well known (15) fact that a beta prior distribution and a binomial sampling law will yield a beta distribution on the iterative application of Bayes' rule may be readily established by induction.

It can easily be shown that for the beta function

$$\begin{aligned}\mu_Q &= E\{Q\} \\ &= \frac{\alpha}{\alpha + \beta}\end{aligned}\tag{A.3.4}$$

$$\begin{aligned}\sigma_Q^2 &= \text{Var}\{Q\} \\ &= \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}\end{aligned}\tag{A.3.5}$$

and similarly

$$\begin{aligned}\mu_Q' &= E\{Q|\xi\} \\ &= \frac{a+\alpha}{\alpha+\beta+n}\end{aligned}\tag{A.3.6}$$

$$\begin{aligned}\sigma_Q^{2'} &= \text{Var}\{Q|\xi\} \\ &= \frac{(a+\alpha)(n-a+\beta)}{(\alpha+\beta+n+1)(\alpha+\beta+n)^2}\end{aligned}\tag{A.3.7}$$

A.4 Restrictions on the Prior Distribution. The model assumes that the prior knowledge of the distribution function  $F(x)$  is in the form of the parameters  $\alpha$  and  $\beta$  of the beta density. It is reasonable to assume that the prior distribution function is best estimated by the expected value of  $Q$ . Thus

$$\begin{aligned}\hat{F}_X(x) &\triangleq \mu_Q \\ &= \frac{\alpha(x)}{\alpha(x) + \beta(x)}\end{aligned}\tag{A.4.1}$$

and similarly

$$\begin{aligned}\hat{F}_{X|\xi}(x) &\triangleq \mu_Q' \\ &= \frac{a(x) + \alpha(x)}{\alpha(x) + \beta(x) + n} \quad .\end{aligned}\tag{A.4.2}$$

This implies that  $\mu_Q$  must be a distribution function and hence that

$$\mu_Q \text{ is continuous from the right,} \tag{A.4.3}$$

$$\mu_Q(-\infty) = 0, \tag{A.4.4}$$

$$\mu_Q(+\infty) = 1, \text{ and} \tag{A.4.5}$$

$$x > x_0 \Rightarrow \mu_Q(x) - \mu_Q(x_0) \geq 0 \quad . \tag{A.4.6}$$

$\mu_Q'$  is a distribution function if and only if it satisfies the same four conditions. If  $\mu_Q$  satisfies the first condition, Equation A.4.2 implies  $\mu_Q'$  also satisfies the same condition. The second and third conditions are satisfied by  $\mu_Q'$  if they are satisfied by  $\mu_Q$ . It remains to be shown what restrictions on the prior assumptions are sufficient to assure that  $\mu_Q'$  is an increasing function of  $x$ .

As  $\alpha$  and  $\beta$  are uniquely determined when  $\mu_Q$  and  $\sigma_Q^2$  are selected, the necessary restriction is developed in terms of restrictions on  $\sigma_Q^2$  given  $\mu_Q$ . From Equations A.3.4 and A.3.5

$$\mu_Q = \frac{\alpha}{\alpha + \beta}$$

and

$$\sigma_Q^2 = \frac{1}{\alpha + \beta + 1} \mu_Q (1 - \mu_Q) \quad . \tag{A.4.7}$$

As the beta distribution is only defined for  $0 < \alpha < \infty$  and  $0 < \beta < \infty$ ,

$$0 < \sigma_Q^2 < \mu_Q(1-\mu_Q) \quad . \quad (A.4.8)$$

Let

$$x > x_0 \quad ,$$

then

$$\mu_Q(x) - \mu_Q(x_0) \geq 0 \quad ,$$

and

$$\mu'_Q(x) - \mu'_Q(x_0) \geq 0 \quad (A.4.9)$$

must also be true. For simplicity the notation will be simplified as follows:

$$\mu_Q(x) = \mu \quad (A.4.10)$$

$$\mu_Q(x_0) = \mu_0 \quad (A.4.11)$$

$$\mu'_Q(x) = \mu' \quad (A.4.12)$$

$$\mu'_Q(x_0) = \mu'_0 \quad (A.4.13)$$

$$\Delta\mu' = \mu' - \mu'_0 \quad , \quad (A.4.14)$$

and similarly for  $\sigma_Q^2(x)$ .

Let  $C$  be a function of  $x$  such that

$$0 < C < 1 \quad (A.4.15)$$

and

$$\sigma^2 = C\mu(1-\mu) \quad . \quad (A.4.16)$$

Similarly

$$\sigma_0^2 = C_0\mu_0(1-\mu_0) \quad . \quad (A.4.17)$$

From Equations A.3.4, A.3.5, A.3.6, A.3.7 and A.4.14

$$\Delta\mu' = \frac{(a-\mu)C + \mu}{(n-1)C + 1} - \frac{(a-\mu_0)C_0 + \mu_0}{(n-1)C_0 + 1} \quad , \quad (A.4.18)$$

and

$$\Delta\mu' \geq 0 \quad . \quad (A.4.19)$$

On rewriting Equations A.4.18 and A.4.19 become

$$\frac{[(a-\mu)C+\mu][(n-1)C_0+1] - [(a_0-\mu_0)C_0+\mu_0][(n-1)C+1]}{[(n-1)C+1][(n-1)C_0+1]} \geq 0 \quad . \quad (A.4.20)$$

From the definitions:  $n > 0$ ,  $0 < C < 1$ , and  $0 < C_0 < 1$ . Therefore the denominator of Equation A.4.20 is greater than zero. Thus

$$[(a-\mu)C+\mu][(n-1)C_0+1] - [(a_0-\mu_0)C_0+\mu_0][(n-1)C+1] \geq 0 \quad . \quad (A.4.21)$$

On expanding and rearranging Equation A.4.21 becomes

$$\begin{aligned} & [nC_0C(a-a_0)] + [(1-C_0)(1-C)(\mu-\mu_0)] + [nC_0(1-C)\mu - nC(1-C_0)\mu_0] \\ & + [C(1-C_0)a - C_0(1-C_0)a_0] \geq 0 \quad . \quad (A.4.22) \end{aligned}$$

Now  $a$  and  $\mu$  are nondecreasing functions of  $x$ ,  $n \geq 0$  and  $0 < C < 1$ .

Therefore the first two terms in the above expression are nonnegative.

If

$$C = C_0 = K \quad , \quad (A.4.23)$$

that is,  $C$  is a constant, the third and fourth terms become

$$[nK(1-K)(\mu - \mu_0)] + [K(1-K)(a - a_0)] \quad , \quad (A.4.24)$$

which is also nonnegative.

From the above it can be concluded that in order to assure that  $\mu^i$  is a nondecreasing function of  $x$ , it is sufficient to choose the prior such that  $\mu_q$  is nondecreasing and

$$\sigma_q^2 = K\mu_q(1-\mu_q) \quad , \quad (A.4.25)$$

where  $K$  is a constant. That this will remain true for repeated sampling may be easily proved by induction.

The posterior mean and variance of  $Q$  become

$$\mu^i = \frac{a + \mu \frac{1-K}{K}}{\frac{1-K}{K} + n} \quad (A.4.26)$$

$$\sigma^{2^i} = \frac{1}{\frac{1-K}{K} + n + 1} \mu^i (1 - \mu^i) \quad (A.4.27)$$

$$= K^i \mu^i (1 - \mu^i) \quad (A.4.28)$$

where

$$K^i = \frac{1}{\frac{1-K}{K} + n + 1} \quad . \quad (A.4.29)$$

If

$$w_0 \triangleq \frac{1-K}{K} \quad , \quad (A.4.30)$$

then

$$0 < w_0 \quad (\text{A.4.31})$$

and Equation A.4.2 becomes

$$\hat{F}_{X|\mathcal{E}}(x) = \frac{a + w_0 \hat{F}_X}{w_0 + n} \quad (\text{A.4.32})$$

$$= \frac{a}{w_0 + n} + \frac{w_0}{w_0 + n} \hat{F}_X(x) \quad (\text{A.4.33})$$

$$= \frac{w_0}{w_0 + n} \hat{F}_X(x) + \frac{n}{w_0 + n} F_n(x) \quad , \quad (\text{A.4.34})$$

where

$$F_n(x) \triangleq \frac{1}{n} \text{ (no. of observations } \leq x \text{ among } X_1, \dots, X_n) \quad (\text{A.4.35})$$

is the empirical distribution function.  $w_0$  may be considered an equivalent sample size weight on the prior distribution. For convenience

$$F_{w_0}(x) \triangleq \hat{F}_X(x) \quad (\text{A.4.36})$$

$$F_w(x) \triangleq \hat{F}_{X|\mathcal{E}}(x) \quad . \quad (\text{A.4.37})$$

Thus from Equations A.4.34, A.4.35, A.4.36 and A.4.37

$$F_w(x) = \frac{w_0}{w_0 + n} F_{w_0}(x) + \frac{n}{w_0 + n} F_n(x) \quad . \quad (\text{A.4.38})$$

Similarly Equation A.4.27 becomes

$$\sigma^{2'} = \frac{1}{w_0 + n + 1} F_w(x) [1 - F_w(x)] \quad (\text{A.4.39})$$

The above equation is of course a measure of the expected value of the squared error in the estimate  $F_w(x)$  of  $F(x)$ .

A.5 Binomial Sampling Not Required. In the previous sections it was assumed that the sampling was binomial, i.e., an integer number of units,  $a$ , out of an integer number of units tested satisfied the less than or equal criteria. It may be desirable to remove this restriction. Consider the case of estimating the distribution of the diameter  $X$  of wire manufactured by a machine. Assume that a sample of length  $s = 100$  feet is chosen. For some diameter  $X = x_0$ , conceivably a length of  $\sqrt{2}$  feet of the sample could be less than or equal to  $x$ . Thus the sample distribution for  $X = x_0$  would be  $F_s(x_0) = \sqrt{2}/100$ . If instead of the assumption A.2.3

$$P_{A|q,n}(a) = \binom{n}{a} q^a (1-q)^{n-a} \quad a=1,2,\dots,n, \quad (A.5.1)$$

it is assumed that

$$f_{a|q,s}(a|q,s) = \begin{cases} Bq^a(1-q)^{s-a} & 0 < a < s \\ 0 & \text{otherwise,} \end{cases} \quad (A.5.2)$$

where

$$B = \begin{cases} \frac{\ln q - \ln(1-q)}{q^s - (1-q)^s} & q \neq \frac{1}{2}, 0 < q < 1 \\ \frac{1}{s} 2^s & q = \frac{1}{2} \end{cases} \quad (A.5.3)$$

Then the restriction of binomial sampling is removed. It is an easy matter to show that all of the derived formulas remain unchanged in form and Equation A.4.38 becomes

$$F_w(x) = \frac{w_0}{w_0+s} F_{w_0}(x) + \frac{s}{w_0+s} F_s(x) \quad (\text{A.5.4})$$

where

$$0 < s \quad (\text{A.5.5})$$

$$0 < w_0 \quad (\text{A.5.6})$$

and

$$F_w(w_0) \triangleq \frac{1}{s} \text{ (observed quantity of sample } \leq x \text{ out of a sample size } s). \quad (\text{A.5.7})$$

## APPENDIX B

### THEOREMS CONCERNING BAYES' EMPIRICAL DISTRIBUTION FUNCTION

B.1 Introduction. Bayes' empirical distribution function (B.E.F.) is closely related to the conventional empirical distribution function (E.D.F.). Therefore it is to be expected that the properties of the two will be similar. Several theorems are developed that are useful not only in applying B.E.F. to engineering problems, but are also useful in comparing the relative merits of choosing the B.E.F. approach rather than the E.D.F. or a conventional Bayesian parametric method.

B.2 Convergence of B.E.F. The proof of Theorems B.2.1 and B.2.2 are a direct consequence of the relation between B.E.F. and the E.D.F.

Theorem B.2.1.  $F_w(x)$  converges uniformly in  $x$  to  $F_n(x)$  for  $-\infty < x < +\infty$ , i.e.,  $\lim_{n \rightarrow \infty} [\sup_x |F_w(x) - F_n(x)|] = 0$ .

Proof. Given  $\epsilon > 0$ ,  $0 < w_0 < \infty$ , select  $N$  such that  $N > w_0/\epsilon$ .  $\forall n > N$ ,  $-\infty < x < \infty$ ,

$$\begin{aligned} |F_w(x) - F_n(x)| &= \left| \frac{w_0}{w_0+n} F_{w_0}(x) + \frac{n}{w_0+n} F_n(x) - F_n(x) \right| \\ &= \left| \frac{w_0}{w_0+n} [F_{w_0}(x) - F_n(x)] \right| \\ &\leq \left| \frac{w_0}{w_0+n} \right| \end{aligned}$$

$$\begin{aligned}
&< \frac{w_0}{w_0 + \frac{w_0}{\epsilon}} \\
&= \frac{\epsilon}{1 + \epsilon} \\
&< \epsilon,
\end{aligned}$$

from the definitions of  $F_n(x)$ ,  $F_{w_0}(x)$ , and  $F_w(x)$  and the first statement. Thus  $F_w(x)$  converges uniformly in  $x$  to  $F_n(x)$ , by the previous step and the definitions of uniform convergence.

Theorem B.2.2.  $F_w(x)$  converges uniformly in  $x$  to  $F(x)$  with probability one for  $-\infty < x < \infty$ , i.e.

$$P\{\lim_{n \rightarrow \infty} [\sup_x |F_w(x) - F(x)|] = 0\} = 1.$$

Proof. It was shown by Glivenko (6) that

$$P\{\lim_{n \rightarrow \infty} [\sup_x |F_n(x) - F(x)|] = 0\} = 1.$$

Given  $\epsilon > 0$ , let  $\epsilon' = \epsilon/2$ . There exists an  $N'$  such that  $n > N' \Rightarrow$

$$P\{[\sup_x |F_n(x) - F(x)|] < \epsilon'\} = 1,$$

by the first step and the definition of a limit. There exists an  $N''$  such that  $n > N'' \Rightarrow$

$$\sup_x |F_w(x) - F_n(x)| < \epsilon',$$

by Theorem B.2.1 and the definition of uniform convergence. Then  $n > N = \max(N', N'')$  and  $\sup_x |F_n(x) - F(x)| < \epsilon' \Rightarrow$

$$\sup_x |F_n(x) - F(x)| + \sup_x |F_w(x) - F_n(x)| < \epsilon' + \epsilon' = \epsilon,$$

from the previous two steps. But

$$\begin{aligned}
 \sup_x |F_w(x) - F(x)| &= \sup_x |F_w(x) - F_n(x) + F_n(x) - F(x)| \\
 &\leq \sup_x [|F_n(x) - F(x)| + |F_w(x) - F_n(x)|] \\
 &\leq \sup_x |F_n(x) - F(x)| + \sup_x |F_w(x) - F_n(x)|,
 \end{aligned}$$

because from the triangle inequality,

$$\sup_{\substack{a \in A \\ b \in B}} [a + b] \leq \sup_{a \in A} [a] + \sup_{b \in B} [b]$$

and the above step. Hence  $n > N$  and  $\sup_x |F_n(x) - F(x)| < \epsilon' \Rightarrow$

$$\sup_x |F_w(x) - F(x)| < \epsilon,$$

by the two previous statements. Thus

$$P\{\sup_x |F_w(x) - F(x)| < \epsilon\} = 1$$

for  $n > N$ , by the second and the above statement. Hence

$$P\{\lim_{n \rightarrow \infty} [\sup_x |F_w(x) - F(x)|] = 0\} = 1,$$

by the second and the above statements and the definition of a limit.

B.3 Errors of B.E.F. Estimate of  $F(x)$ . The theorems in this section give an indication of the errors to be expected in estimating the distribution function. These theorems are of use in some decision problems.

Theorem B.3.1. The expected value of squared error of the B.E.F. estimate of  $F(x)$  in terms of the prior distribution  $F_{w_0}(x)$  and the true distribution  $F(x)$  is

$$E\{[F_w(x) - F(x)]^2\} = \left(\frac{w_0}{w_0+n}\right)^2 [F_{w_0}(x) - F(x)]^2 + \left(\frac{n}{w_0+n}\right)^2 \frac{1}{n} F(x)[1-F(x)] ,$$

that is, the expected squared error is a weighted average of the squared error of the prior estimate and the expected value of the squared error of the E.D.F. estimate.

Proof.

$$\begin{aligned} E\{[F_w(x) - F(x)]^2\} &= E\left\{\left[\frac{w_0}{w_0+n} F_{w_0}(x) + \frac{n}{w_0+n} F_n(x) - F(x)\right]^2\right\} \\ &= E\left\{\left(\frac{w_0}{w_0+n}\right)^2 F_{w_0}^2(x) + \left(\frac{n}{w_0+n}\right)^2 F_n^2(x) + F^2(x) + 2 \frac{w_0 n}{(w_0+n)^2} F_{w_0}(x) F_n(x) \right. \\ &\quad \left. - 2 \frac{w_0}{w_0+n} F_{w_0}(x) F(x) - 2 \frac{n}{w_0+n} F_n(x) F(x)\right\} \\ &= \left(\frac{w_0}{w_0+n}\right)^2 F_{w_0}^2(x) + \left(\frac{n}{w_0+n}\right)^2 \left\{\frac{1}{n} F(x)[1-F(x)] + F^2(x)\right\} \\ &\quad + F^2(x) + 2 \frac{w_0 n}{(w_0+n)^2} F_{w_0}(x) F(x) - 2 \frac{w_0}{w_0+n} F_{w_0}(x) F(x) \\ &\quad - 2 \frac{n}{w_0+n} F^2(x) , \end{aligned}$$

from the definitions of  $F_w(x)$ ,  $F_{w_0}(x)$  and  $F_n(x)$ . The desired result is obtained by rearranging the result of the above step.

Lemma B.3.1. Given a random variable  $x$  with probability distribution function  $F(x)$  and finite mean  $\mu$  and variance  $\sigma^2$ , then

$$\lim_{x \rightarrow -\infty} \{x F(x)[1-F(x)]\} = 0$$

and

$$\lim_{x \rightarrow +\infty} \{x F(x)[1-F(x)]\} = 0 .$$

Proof.

$$\sigma^2 + \mu^2 = \int_{-\infty}^{\infty} x^2 dF(x) = \lim_{y \rightarrow \infty} \int_{-y}^y x^2 dF(x) \text{ for } y > 0, \quad ,$$

from the definitions of  $\mu$ ,  $\sigma^2$  and the integral over an infinite interval.

$$\int_{-\infty}^y x^2 dF(x) \leq \sigma^2 + \mu^2, \quad ,$$

because  $x^2 > 0$  and  $F(x)$  is monotone nondecreasing.

The total variation of  $F(x)$  over the interval  $[-\infty, -y]$  is less than or equal to  $(\sigma^2 + \mu^2)/y^2$  because  $F(x)$  is monotone nondecreasing, and  $x^2 \geq y^2$  for  $x \in [-\infty, -y]$ . Thus  $F(-y) \leq (\sigma^2 + \mu^2)/y^2$ , from above step and  $F(-\infty)=0$ .

$$\lim_{x \rightarrow -\infty} |xF(x)[1-F(x)]| \leq \lim_{x \rightarrow -\infty} |xF(x)| \leq \lim_{x \rightarrow -\infty} \left| x \frac{\sigma^2 + \mu^2}{x^2} \right| = 0, \quad ,$$

because  $0 \leq F(x) \leq 1$  and the previous step. The proof that

$$\lim_{x \rightarrow -\infty} \{xF(x)[1-F(x)]\} = 0$$

is similar.

Theorem B.3.2. The integral of the expected value of the squared error of the B.E.F. estimate  $F_w(x)$  of  $F(x)$  in terms of the prior distribution function  $F_{w_0}(x)$  and the true distribution function  $F(x)$  is

$$\begin{aligned} I &= \int_{-\infty}^{\infty} E\{[F_w(x) - F(x)]^2\} dx \\ &= \left(\frac{w_0}{w_0+n}\right)^2 \int_{-\infty}^{\infty} [F_{w_0}(x) - F(x)]^2 dx + \left(\frac{n}{w_0+n}\right)^2 \int_{-\infty}^{\infty} \frac{1}{n} F(x)[1-F(x)] dx. \end{aligned}$$

Further

$$I \leq \left( \frac{w_0}{w_0+n} \right)^2 \int_{-\infty}^{\infty} [F_{w_0}(x) - F(x)]^2 dx + \left( \frac{n}{w_0+n} \right)^2 \frac{1}{n} \int_{-\infty}^{\infty} |x - \mu| dF(x)$$

where  $\mu = E\{x\}$ .

A less stringent though occasionally more convenient upper bound is

$$I \leq \left( \frac{w_0}{w_0+n} \right)^2 \int_{-\infty}^{\infty} [F_{w_0}(x) - F(x)]^2 dx + \left( \frac{n}{w_0+n} \right)^2 \frac{1}{n} \left( 1 + \int_{-\infty}^{\infty} (x-\mu)^2 dF(x) \right)$$

Proof. The first statement follows directly from Theorem B.3.1.

Let  $\mu = E\{x\}$  and  $Z = X - \mu$ , then  $G(z) = F(z + \mu)$  and

$$\int_{-\infty}^{\infty} F(x)[1-F(x)]dx = \int_{-\infty}^{\infty} G(z)[1-G(z)]dz$$

From the above statement and the definition of an integral over an infinite interval it is sufficient to consider

$$\begin{aligned} \lim_{y \rightarrow \infty} \int_{-y}^y G(z)[1-G(z)]dz &= \lim_{y \rightarrow \infty} \left\{ zG(z)[1-G(z)] \Big|_{-y}^y - \int_{-y}^y z dG(z) + \int_{-y}^y z dG^2(z) \right\} \\ &= \lim_{y \rightarrow \infty} \left\{ zG(z)[1-G(z)] \Big|_{-y}^y + \int_{-y}^y z dG^2(z) \right\} \\ &= \lim_{y \rightarrow \infty} \left\{ \int_{-y}^y z dG^2(z) \right\} \end{aligned}$$

The reduction is accomplished using integration by parts,  $E\{Z\} = 0$  and Lemma B.3.1. Now

$$\begin{aligned}
\lim_{y \rightarrow \infty} \int_{-y}^y G(z)[1-G(z)]dz &= \lim_{y \rightarrow \infty} \left\{ \int_{-y}^y z dG^2(z) \right\} \\
&\leq \lim_{y \rightarrow \infty} \left\{ \int_0^y z dG^2(z) \right\} \\
&\leq \lim_{y \rightarrow \infty} \left\{ \int_0^y 2z dG(z) \right\} \\
&= \lim_{y \rightarrow \infty} \left\{ \int_{-y}^y |z| dG(z) \right\} \\
&= \int_{-\infty}^{\infty} |z| dG(z) \\
&= \int_{-\infty}^{\infty} |x-\mu| dF(x) \quad ,
\end{aligned}$$

where these steps are justified by the previous statement,  $G(z)$  nondecreasing,

$$G^2(Z+\Delta) - G^2(Z) = [G(Z+\Delta) - G(Z)][G(Z+\Delta) + G(Z)] \leq 2[G(Z+\Delta) - G(Z)]$$

and the definition of  $Z$ . The final inequality follows easily from

$$\begin{aligned}
\int_{-\infty}^{\infty} |x-\mu| dF(x) &= \int_{-\infty}^{\mu-1} |x-\mu| dF(x) + \int_{\mu-1}^{\mu+1} |x-\mu| dF(x) + \int_{\mu+1}^{\infty} |x-\mu| dF(x) \\
&\leq \int_{-\infty}^{\mu+1} (x-\mu)^2 dF(x) + \int_{\mu-1}^{\mu+1} |x-\mu| dF(x) + \int_{\mu+1}^{\infty} (x-\mu)^2 dF(x) \\
&\leq \int_{-\infty}^{\infty} (x-\mu)^2 dF(x) + \int_{\mu-1}^{\mu+1} |x-\mu| dF(x) \\
&\leq \int_{-\infty}^{\infty} (x-\mu)^2 dF(x) + 1 \quad .
\end{aligned}$$

Although Theorems B.3.1 and B.3.2 are of value, the true distribution function  $F(x)$  would not be available in practice. Hence the following theorem is of interest.

Theorem B.3.3. The estimate of the expected squared error of the B.E.F. estimate  $F_w(x)$  of  $F(x)$  in terms of  $F_{w_0}(x)$  and  $F_n(x)$  is

$$\begin{aligned} & \frac{1}{w_0+n+1} F_w(x)[1-F_w(x)] \\ &= \frac{1}{w_0+n+1} \left\{ \left( \frac{w_0}{w_0+n} \right)^2 F_{w_0}(x)[1-F_{w_0}(x)] + \left( \frac{n}{w_0+n} \right)^2 F_n(x)[1-F_n(x)] \right. \\ & \quad \left. + \frac{w_0 n}{(w_0+n)^2} \{ F_{w_0}(x)[1-F_n(x)] + F_n(x)[1-F_{w_0}(x)] \} \right\} . \end{aligned}$$

Proof. The first expression results from the assumed first order beta density and is the variance given by this density as noted in Appendix A, Equation A.4.39. The second expression follows from the definition of  $F_w(x)$ .

The following lemma is of use in predicting errors in decision problems.

Lemma B.3.2. Given  $a > 0$ ,

$$E\{F_n(x-a)F_n(x)\} = \frac{1}{n} F(x-a)[1-F(x)] + F(x-a)F(x)$$

and

$$E\{[F_n(x-a)-F(x-a)][F_n(x)-F(x)]\} = \frac{1}{n} F(x-a)[1-F(x)] .$$

Proof. Let

$$U = nF_n(x-a) ,$$

$$V = n[F_n(x) - F_n(x-a)] \quad ,$$

$$p_1 = P\{X \leq x - a\}$$

$$= F(x)$$

and

$$p_2 = P\{x - a < X \leq x\}$$

$$= F(x) - F(x-a) \quad ,$$

then  $U$  and  $V$  are jointly trinomial, i.e.

$$f_{U,V}(u,v) = \begin{cases} \frac{n!}{u!v!(n-u-v)!} p_1^u p_2^v (1-p_1-p_2)^{n-u-v} & \begin{array}{l} u \text{ and } v \text{ nonnegative} \\ \text{integers and} \\ u + v \leq n \end{array} \\ 0 & \text{otherwise} \end{cases} \quad ,$$

and the properties of this density function are well known (2).

Thus

$$\begin{aligned} E\{UV\} &= E_U\{E_V\{UV|U\}\} \\ &= E_U\{UE_V\{V|U\}\} \\ &= E_U\left\{U(n-U) \frac{p_2}{1-p_1}\right\} \\ &= \frac{p_2}{1-p_1} [n^2 p_1 - n p_1 (1-p_1) - n^2 p_1^2] \\ &= p_1 p_2 n(n-1) \quad . \end{aligned}$$

Hence

$$\begin{aligned}
n^2 E\{F_n(x-a)F_n(x)\} &= E\{U(U+V)\} \\
&= E\{U^2+UV\} \\
&= np_1(1-p_1)+n^2 p_1^2+p_1 p_2 n(n-1) \\
&= nF(x-a)[1-F(x-a)]+n^2 F^2(x-a)+n(n-1)F(x-a)[F(x)-F(x-a)] \\
&= nF(x-a)-nF(x-a)F(x)+n^2 F(x-a)F(x) \\
&= nF(x-a)[1-F(x)]+n^2 F(x-a)F(x) .
\end{aligned}$$

On dividing the above equation by  $n^2$  the first half of the lemma is proved. The proof of the second half follows from this and the definitions.

B.4 Properties of B.E.F. for Estimation. The following theorems are especially useful when the B.E.F. is used for parameter estimation and for comparing the B.E.F. estimate with a parametric Bayesian estimate.

Theorem B.4.1. If the prior estimate  $F_{w_0}(x)$  is an unbiased estimate of  $F(x)$ , then  $F_w(x)$  is an unbiased estimate of  $F(x)$ .

Proof.

$$\begin{aligned}
E\{F_w(x)\} &= E\left\{\frac{w_0}{w_0+n} F_{w_0}(x) + \frac{n}{w_0+n} F_n(x)\right\} \\
&= \frac{w_0}{w_0+n} F(x) + \frac{n}{w_0+n} F(x) \\
&= F(x) .
\end{aligned}$$

Thus by the definition of an unbiased estimate,  $F_w(x)$  is an unbiased

estimate.

Theorem B.4.2. Regardless of the prior estimate  $F_{w_0}(x)$ ,  $F_w(x)$  is an asymptotically unbiased estimate of  $F_w(x)$  as  $n$  approaches infinity.

Proof.

$$\begin{aligned}\lim_{n \rightarrow \infty} E\{F_w(x)\} &= \lim_{n \rightarrow \infty} E\left\{\frac{w_0}{w_0+n} F_{w_0}(x) + \frac{n}{w_0+n} F_n(x)\right\} \\ &= \lim_{n \rightarrow \infty} \left\{\frac{w_0}{w_0+n} F_{w_0}(x)\right\} + \lim_{n \rightarrow \infty} \left\{\frac{n}{w_0+n} F(x)\right\} \\ &= F(x)\end{aligned}$$

Hence by the definition  $F_w(x)$  is an asymptotically unbiased estimate of  $F(x)$ .

Theorem B.4.3. Let  $\theta$  be any parameter of  $F(x)$  such that

$$\theta = \int_a^b u(x) dF(x)$$

where  $-\infty \leq a < b \leq +\infty$  and  $u(x)$  is a measurable function of  $x$  such that  $u(x)$  does not depend on an unknown parameter, i.e.,  $u(x)$  is a statistic, then if  $F_w(x)$  is an unbiased estimate of  $F(x)$  such that

$$\hat{\theta}_w = \int_a^b u(x) dF_w(x)$$

exists, then  $\hat{\theta}_w$  is an unbiased estimate of  $\theta$ .

Proof. By Fubini's theorem

$$\begin{aligned}
E\{\hat{\theta}_w\} &= E\left\{\int_a^b u(x) dF_w(x)\right\} \\
&= \int_a^b u(x) dE\{F_w(x)\} \\
&= \int_a^b u(x) dF(x) \\
&= \theta .
\end{aligned}$$

Thus by definition  $\hat{\theta}_w$  is an unbiased estimate of  $\theta$ .

It is important to note that Theorem B.4.3 does not state that  $\hat{\theta}_w$  converges to  $\theta$  as  $n$  approaches infinity. However the following theorem can be proved.

Theorem B.4.4. Given  $\theta$  and  $\hat{\theta}_w$  as defined in Theorem B.4.3 where  $u(x)$  is Reimann integrable with respect to  $F(x)$ , then

$$P\left\{\lim_{n \rightarrow \infty} |\hat{\theta}_w - \theta| = 0\right\} = 1 ,$$

i.e.,  $\hat{\theta}_w$  converges to  $\theta$  with probability 1 as  $n$  approaches infinity.

Proof. Let  $G$  be the set of outcomes such that  $F_w(x)$  converges uniformly to  $F(x)$  in  $x$ . Then

$$\hat{\theta}_w = \int_a^b u(x) dF_w(x) = u(b)F_w(b) - F_w(a)u(a) - \int_a^b F_w(x) du(x) .$$

$F_w(x)$  converges uniformly in  $x$  to  $F(x)$  implies  $\int_a^b F_w(x) du(x)$  converges uniformly to  $\int_a^b F(x) du(x)$ . Thus  $\int_a^b u(x) dF_w(x)$  converges to  $\theta = \int_a^b u(x) dF(x)$ . For set  $G$ ,  $P\{G\} = 1$ , thus

$$P\left\{\lim_{n \rightarrow \infty} |\hat{\theta}_w - \theta| = 0\right\} = 1 .$$

If the restriction that  $u(x)$  is Reimann integrable with respect to  $F(x)$  is replaced with the restriction  $u(x)$  is bounded and continuous, the proof follows from the Helly-Bray theorem (11).

Lemma B.4.1. Let  $\theta$  be defined by

$$\theta = \int_a^b u(x) dF(x)$$

as described in Theorem B.4.4. Then

$$E\left\{\left[\int_a^b u(x) dF_n(x)\right]^2\right\} = \left(1 - \frac{1}{n}\right)\left[\int_a^b u(x) dF(x)\right]^2 + \frac{1}{n} \int_a^b u^2(x) dF(x)$$

and

$$E\left\{\int_{-\infty}^{\infty} u(x) dF_n(x)\right\}^2 = \left(1 - \frac{1}{n}\right)E\{u(x)\}^2 + \frac{1}{n} E\{u^2(x)\}.$$

Proof. Let  $F(x_1) \times F(x_2)$  be the product distribution over the product space  $X_1 \times X_2$ .

$$\begin{aligned} E\left\{\left[\int_a^b u(x) dF_n(x)\right]^2\right\} &= E\left\{\left[\int_a^b u(x_1) dF_n(x_1)\right]\left[\int_a^b u(x_2) dF_n(x_2)\right]\right\} \\ &= E\left\{\int_a^b \int_a^b u(x_1) u(x_2) dF_n(x_1) \times F_n(x_2)\right\} \\ &= \int_a^b \int_a^b u(x_1) u(x_2) dE\{F_n(x_1) F_n(x_2)\} \\ &= \int_a^b \int_a^b u(x_1) u(x_2) d\left[\left(1 - \frac{1}{n}\right)F(x_1) \times F(x_2) + \frac{1}{n} F(x_1) \times U(x_2 - x_1) \right. \\ &\quad \left. + \frac{1}{n} F(x_2) \times U(x_1 - x_2)\right] \end{aligned}$$

$$= (1 - \frac{1}{n}) \int_a^b \int_a^b u(x_1)u(x_2)dF(x_1)dF(x_2) + \frac{1}{n} \int_a^b \int_a^b u(x_1)u(x_2)d[F(x_1) \\ \times U(x_2 - x_1) + F(x_2) \times U(x_1 - x_2)]$$

where  $U(\cdot)$  is the unit step function. This follows from Lemma B.3.2.

Evaluation of the integrals yields the desired results.

Lemma B.4.2.

$$E\left\{\left[\int_a^b u(x)dF_n(x) - \int_a^b u(x)dF(x)\right]^2\right\} = \frac{1}{n} \left\{\int_a^b u^2(x)dF(x) - \left[\int_a^b u(x)dF(x)\right]^2\right\}.$$

Proof.

$$E\left\{\left[\int_a^b u(x)dF_n(x) - \int_a^b u(x)dF(x)\right]^2\right\} \\ = E\left\{\left[\int_a^b u(x)dF_n(x)\right]^2 - 2\int_a^b u(x)dF_n(x) \int_a^b u(x)dF(x) + \left[\int_a^b u(x)dF(x)\right]^2\right\} \\ = E\left\{\left[\int_a^b u(x)dF_n(x)\right]^2\right\} - 2\left[\int_a^b u(x)dF(x)\right]^2 + \left[\int_a^b u(x)dF(x)\right]^2,$$

The result then follows directly from Lemma B.4.1.

Theorem B.4.5. The expected value of the squared error of the estimate

$$\theta_w = \int_a^b u(x)dF_w(x)$$

of the parameter

$$\theta = \int_a^b u(x) dF(x)$$

is given by

$$E\{(\theta_w - \theta)^2\} = \left[ \left( \frac{w_0}{w_0+n} \theta_0 + \frac{n}{w_0+n} \theta \right) - \theta \right]^2 + \frac{n}{(w_0+n)^2} \left[ \int_a^b u^2(x) dF(x) - \theta^2 \right]$$

where

$$\theta_0 = \int_a^b u(x) dF_{w_0}(x)$$

In particular when

$$\theta = E\{u(x)\} \quad ,$$

$$E\{\theta_w - \theta\}^2 = \left[ \left( \frac{w_0}{w_0+n} \theta_0 + \frac{n}{w_0+n} \theta \right) - \theta \right]^2 + \frac{n}{(w_0+n)^2} [E\{u^2(x)\} - \theta^2] \quad .$$

Proof.

$$\begin{aligned} E\{(\theta_w - \theta)^2\} &= E\{\theta_w^2 - 2\theta\theta_w + \theta^2\} \\ &= E\{\theta_w^2\} - 2\theta E\{\theta_w\} + \theta^2 \\ &= E\left\{\left(\frac{w_0}{w_0+n}\right)^2 \theta_0^2 + 2 \frac{w_0 n}{(w_0+n)^2} \theta_0 \theta + \left(\frac{n}{w_0+n}\right)^2 \theta^2\right\} - 2\theta E\{\theta_w\} + \theta^2 \\ &= \left(\frac{w_0}{w_0+n}\right)^2 \theta_0^2 + 2 \frac{w_0 n}{(w_0+n)^2} \theta_0 \theta + \left(\frac{n}{w_0+n}\right)^2 \left[ \left(1 - \frac{1}{n}\right) \theta^2 + \frac{1}{n} \int_a^b u^2(x) dF(x) \right] \\ &\quad - 2\theta \left[ \frac{w_0}{w_0+n} \theta_0 + \frac{n}{w_0+n} \theta \right] + \theta^2 \quad . \end{aligned}$$

The above follows from the definitions and Lemma B.4.1. This may be rewritten as

$$\begin{aligned} E\{(\theta_w - \theta)^2\} &= \left(\frac{w_0}{w_0+n} \theta_0 + \frac{n}{w_0+n} \theta\right)^2 - 2\theta \left[\frac{w_0}{w_0+n} \theta_0 + \frac{n}{w_0+n}\right] \\ &\quad + \theta^2 + \frac{n}{(w_0+n)^2} \left[\int_a^b u^2(x) dF(x) - \theta^2\right] \\ &= \left[\left(\frac{w_0}{w_0+n} \theta_0 + \frac{n}{w_0+n} \theta\right) - \theta\right]^2 + \frac{n}{(w_0+n)^2} \left[\int_a^b u^2(x) dF(x) - \theta^2\right]. \end{aligned}$$

### B.5 Moments of B.E.F.

Theorem B.5.1. The estimate of the  $k^{\text{th}}$  moment of the random variable  $x$  as given by B.E.F. is

$$\int_{-\infty}^{\infty} x^k dF_w(x) = \frac{w_0}{w_0+n} \int_{-\infty}^{\infty} x^k dF_{w_0}(x) + \frac{n}{w_0+n} \int_{-\infty}^{\infty} x^k dF_n(x).$$

The expected value of the squared error of this estimate is given by Theorem B.4.5.

Proof. The proof follows immediately from the definitions and Theorem B.4.5.

Theorem B.5.2. If the true mean  $\mu$  is known, the estimate of the variance given by B.E.F. is

$$\sigma_w^2 = \int_{-\infty}^{\infty} (x - \mu)^2 dF_w(x) = \frac{w_0}{w_0+n} \sigma_0^2 + \frac{n}{w_0+n} \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right],$$

where  $\sigma_0^2$  is the estimate of the variance given by  $F_{w_0}(x)$ . The expected value of the squared error of this estimate is given by Theorem B.4.5.

Proof. The proof of the above follows directly from the definitions and the fact that  $\mu$  is a known constant.

Theorem B.5.3. If the mean  $\mu$  is unknown, and

$$\mu_w = \int_{-\infty}^{\infty} x F_w(x) \quad ,$$

then the estimate of the variance given by B.E.F. is

$$\sigma_w^2 = \int_{-\infty}^{\infty} (x - \mu_w)^2 dF_w(x) = \frac{w_0}{w_0+n} \sigma_0^2 + \frac{n}{w_0+n} \left( \frac{1}{n} \sum_{i=1}^n x_i^2 - \mu_w^2 \right) + \frac{w_0}{w_0+n} (\mu_0^2 - \mu_w^2) \quad .$$

Proof.

$$\begin{aligned} \int_{-\infty}^{\infty} (x - \mu_w)^2 dF_w(x) &= \frac{w_0}{w_0+n} \int_{-\infty}^{\infty} (x - \mu_w)^2 dF_{w_0}(x) + \frac{n}{w_0+n} \int_{-\infty}^{\infty} (x - \mu_w)^2 dF_n(x) \\ &= \frac{w_0}{w_0+n} [\sigma_0^2 + \mu_0^2 - 2\mu_0\mu_w + \mu_w^2] + \frac{n}{w_0+n} \left[ \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\mu_w \frac{1}{n} \sum_{i=1}^n x_i + \mu_w^2 \right] \end{aligned}$$

Rearranging the above equation yields the desired result.

## APPENDIX C

### CONVENTIONAL BAYESIAN PARAMETER ESTIMATION

C.1 Introduction. The inclusion of prior information in the estimate of a distribution function is very desirable in many applications. This is especially true in the decision theory approach to problem solution. For example in selecting the number of units to be tested from a lot of resistors, should the number to be tested be greater than zero?

In the conventional method for including prior information in the estimate of a distribution function, the distribution function is assumed to come from a family of distribution functions characterized by certain parameters. One or more parameters of this distribution are treated as random variables with an assumed prior density function. As sample data becomes available the prior density is modified by Bayes' rule to obtain a posterior density for the parameters. The conditional mean or mode of a parameter is usually used as a point estimate of the unknown parameter.

In the examples presented the prior probability densities for the parameters to be estimated were selected so that application of Bayes' rule yielded a posterior density of the same family. In such a case the prior is said to be "reproducing". Spragins (15) has shown that given  $n$  independent observations  $\{X_1, X_2, \dots, X_n\}$  characterized by the joint density  $f_{X_1, X_2, \dots, X_n | \theta}(x_1, x_2, \dots, x_n | \theta)$ , then a reproducing prior density  $f_\theta(\theta)$  exists if and only if the observations admit a sufficient statistic expressible as a vector of fixed dimension (regardless of  $n$ ). If

a reproducing prior does not exist, the computations will grow without bound as the data set increases in number.

C.2 Binomial Distribution With Unknown Probability. Consider a random variable  $Y$  which takes on the value 1 with probability  $q$  and 0 with probability  $1-q$ . Assume  $n$  independent values of  $Y$  are available and let

$$X = \sum_{i=1}^n Y_i, \quad (C.2.1)$$

then  $X$  will be binomially distributed with probability mass function.

$$P_X(x) = \begin{cases} \binom{n}{x} q^x (1-q)^{n-x} & x=0,1,\dots,n \\ 0 & \text{otherwise} \end{cases} \quad (C.2.2)$$

Further assume that  $Q$  is a random variable with a beta distribution, i.e.

$$f_Q(q) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1} (1-q)^{\beta-1} & 0 < q < 1 \\ 0 & \text{otherwise} \end{cases} \quad (C.2.3)$$

Given  $n$  independent samples of  $Y$ , Bayes' theorem

$$f_{Q|X}(q) = \frac{P_{X|Q}(x|q,n)f_Q(q)}{\int_{-\infty}^{\infty} P_{X|Q}(x|u,n)f_Q(u)du} \quad (C.2.4)$$

yields

$$f_{Q|X}(q) = \begin{cases} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} q^{x+\alpha-1} (1-q)^{n-x+\beta-1} & \\ 0 & \text{otherwise} \end{cases} \quad (C.2.5)$$

If the mean of  $Q$  is used as the point estimate of  $Q$ , then the prior estimate is

$$\begin{aligned} \hat{Q} &= E\{Q\} \\ &= \frac{\alpha}{\alpha+\beta} \end{aligned} \quad (C.2.6)$$

and the posterior estimate is

$$\begin{aligned} \hat{Q}' &= E\{Q|X\} \\ &= \frac{\alpha+x}{\alpha+\beta+n} \\ &= \frac{\alpha+\beta}{\alpha+\beta+n} \left( \frac{\alpha}{\alpha+\beta} \right) + \frac{n}{\alpha+\beta+n} \left( \frac{x}{n} \right) \end{aligned} \quad (C.2.7)$$

The latter equation shows that the posterior estimate of  $q$  is simply a weighted average of the prior estimate and the sample value of  $q$ . Thus  $\alpha+\beta$  is the equivalent prior sample size.

C.3 Normal Distribution With Unknown Mean. Consider a normally distributed random variable  $X$  with known variance  $N^2$  and unknown mean  $Y$ , i.e.

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi N^2}} \exp \left[ -\frac{(x-y)^2}{2N^2} \right] \quad (C.3.1)$$

The mean  $Y$  is to be estimated from prior knowledge and sample data.

Assume  $Y$  is normally distributed with prior mean  $\mu_0$  and variance  $\sigma_0^2$ . Given a data set  $\{X_1, X_2, \dots, X_n\}$ , it can easily be shown that on iterative application of Bayes' rule

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|u)f_Y(u)du}, \quad (C.3.2)$$

$f_{Y|X_1, \dots, X_n}$  will be normal with conditional mean

$$\mu = \frac{\frac{N^2}{\sigma_0^2} \mu_0 + n \left( \frac{1}{n} \sum_{i=1}^n X_i \right)}{\frac{N^2}{\sigma_0^2} + n} \quad (C.3.3)$$

and variance

$$\sigma_n^2 = \frac{\frac{N^2}{\sigma_0^2} \sigma_0^2}{\frac{N^2}{\sigma_0^2} + n} \quad (C.3.4)$$

Taking the conditional mean of  $Y$  as a point estimate of  $Y$ ,  $X$  would be estimated to be normally distributed with mean given by Equation C.3.3 and variance  $\sigma_n^2$ .  $\sigma_n^2$  is a measure of the variance of the estimate of the mean of  $X$ .

C.4 Normal Distribution With Unknown Variance. Consider a normally distributed random variable  $X$  with known mean  $\mu$  and unknown variance  $\sigma^2$ . Let

$$z = \frac{1}{\sigma^2}, \quad ,$$

then the probability density function for  $X$  is

$$f_{X|Z}(x|z) = \sqrt{\frac{z}{2\pi}} \exp\left[-\frac{1}{2} z(x-\mu)^2\right] \quad (C.4.1)$$

Assume  $Z$  is a random variable with a Wishart distribution, i.e.

$$f_Z(z) = \begin{cases} \frac{1}{\Gamma(\frac{v-1}{2})} \left(\frac{v\varphi}{2}\right)^{\frac{v-1}{2}} (z)^{\frac{v-3}{2}} \exp\left(-\frac{1}{2} v \varphi z\right) & z > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (C.4.2)$$

where  $\Gamma(\cdot)$  is the gamma function and the parameters  $v > 3$  and  $\varphi > 0$ .

Given  $v = v_0$ ,  $\varphi = \varphi_0$  and a sample value of  $X$ , Bayes' rule

$$f_{Z|X}(z|x) = \frac{f_{X|Z}(x|z)f_Z(z)}{\int_{-\infty}^{\infty} f_{X|Z}(x|u)f_Z(u)du} \quad (C.4.3)$$

yields the result that  $f_{Z|X}$  is Wishart with parameters

$$v_1 = v_0 + 1 \quad (C.4.4)$$

and

$$\varphi_1 = \frac{v_0}{v_0+1} \varphi_0 + \frac{1}{v_0+1} (x-\mu)^2 \quad (C.4.5)$$

Using induction it can easily be shown that given a sample set

$\{X_1, X_2, \dots, X_n\}$ ,  $f_{Z|X_1, \dots, X_n}$  is Wishart with parameters

$$v_n = v_0 + n \quad (C.4.6)$$

and

$$\varphi_n = \frac{v_0}{v_0+n} \varphi_0 + \frac{n}{v_0+n} \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right] \quad . \quad (C.4.7)$$

Although

$$\frac{1}{E\{Z\}} = \frac{v}{v-1} \varphi \quad , \quad (C.4.8)$$

it is more convenient to consider  $\varphi_n$  as the point estimate of the variance. In this case  $v_0$  becomes an equivalent prior sample size while  $\varphi_0$  is the prior estimate of the variance. Thus  $\varphi_n$  is an estimate of the variance and is a weighted average of the prior variance and the variance obtained from the sample data using the known mean.

C.5 Normal Distribution With Unknown Mean and Variance. Consider a normal random variable with unknown mean  $Y$  and unknown variance  $\sigma^2$ . As in the previous section define

$$z = \frac{1}{\sigma^2} \quad . \quad (C.5.1)$$

Then

$$f_{X|Y,Z}(x|y,z) = \sqrt{\frac{z}{2\pi}} \exp\left[-\frac{z(x-y)^2}{2}\right] \quad . \quad (C.5.2)$$

The mean and variance are to be estimated from prior knowledge and sample data.

Treating  $Y$  and  $Z$  as random variables, Keehn (5) has shown that  $f_{Y,Z}(y,z)$  reproduces itself with respect to  $f_{X|Y,Z}(x|y,z)$ , if  $f_{Y,Z}(y,z)$  is a composite Gaussian-Wishart density function. Thus the mean  $Y$  and the reciprocal variance are assumed to have a density of the form

$$f_{Y,Z}(y,z) = \begin{cases} \left(\frac{wz}{2\pi}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2} wz(y-\mu)^2\right] \left[\Gamma\left(\frac{v-1}{2}\right)\right]^{-1} \left(\frac{v\varphi}{2}\right)^{\frac{v-1}{2}} (z)^{\frac{v-3}{2}} \exp\left(-\frac{1}{2} v\varphi z\right) & z > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (C.5.3)$$

where  $\Gamma(\cdot)$  is the gamma function and  $\varphi > 0$ ,  $w > 0$  and  $v > 3$ . The parameters  $w$ ,  $\mu$ ,  $v$  and  $\varphi$  are given the following interpretation by Keehn.  $\mu$  is the estimate of the mean of  $X$  while  $w$  reflects the confidence in this estimate.  $\varphi$  is the estimate of  $\sigma^2$  and  $v$  is the confidence that the estimate  $\varphi$  is the true value of  $\sigma^2$ .

It can be shown that given prior values  $w_0$ ,  $\mu_0$ ,  $v_0$  and  $\varphi_0$ , a sample set  $\{X_1, X_2, \dots, X_n\}$  on application of Bayes' rule yields posterior values:

$$w_n = w_0 + n, \quad (C.5.4)$$

$$\mu_n = \frac{w_0}{w_0+n} \mu_0 + \frac{n}{w_0+n} \left(\frac{1}{n} \sum_{i=1}^n X_i\right), \quad (C.5.5)$$

$$v_n = v_0 + n, \quad (C.5.6)$$

and

$$\varphi_n = \frac{v_0}{v_0+n} \varphi_0 + \frac{n}{v_0+n} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \mu_n^2\right) + \frac{w_0}{v_0+n} (\mu_0^2 - \mu_n^2). \quad (C.5.7)$$

The estimate of the density  $f_X(x)$  is then normal with mean  $\mu_n$  given by Equation C.5.5 and variance  $\varphi_n$  given by Equation C.5.7. Examination of these equations indicates that  $w_0$  is an equivalent sample size of the prior estimate of the mean  $\mu_0$ , while  $v_0$  can be considered an equivalent sample size of the prior variance estimate  $\varphi_0$ .

C.6 Rayleigh Distribution With Unknown Parameter. Let  $X$  be a random variable with a Rayleigh distribution, i.e.

$$f_X(x) = \begin{cases} \frac{1}{\alpha^2} x \exp\left(-\frac{x^2}{2\alpha^2}\right) & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{C.6.1})$$

Let

$$C = \frac{1}{2\alpha^2}, \quad (\text{C.6.2})$$

and assume that the parameter  $C$  is a random variable with a gamma distribution. Then

$$f_C(c) = \begin{cases} \frac{1}{\Gamma(a)} b^a c^{a-1} \exp(-bc) & 0 < c \\ 0 & \text{otherwise} \end{cases} \quad (\text{C.6.3})$$

where  $0 < a$  and  $0 < b$ . Bayes' rule

$$f_{C|X}(c|x) = \frac{f_{X|C}(x|c)f_C(c)}{\int_{-\infty}^{\infty} f_{X|C}(x|u)f_C(u)du} \quad (\text{C.6.4})$$

yields

$$f_{C|X}(c|x) = \frac{1}{\Gamma(a_0+1)} (x^2+b_0)^{a_0+1} c^{a_0} \exp[-c(x^2+b_0)] \quad 0 < c, \quad (\text{C.6.5})$$

where  $a_0$  and  $b_0$  are the prior values of the parameters  $a$  and  $b$ . Given a

sample set  $\{X_1, X_2, \dots, X_n\}$ , it can be shown that repeated application of Bayes' rule yields a gamma distribution with posterior values of the parameters

$$a_n = a_0 + n \quad (\text{C.6.6})$$

and

$$b_n = b_0 + \sum_{i=1}^n X_i^2 \quad (\text{C.6.7})$$

Using the conditional mean of  $C$  as the point estimate of  $C$  gives

$$E\{C|X_1, \dots, X_n\} = \frac{a_0 + n}{\sum_{i=1}^n X_i^2 + b_0} \quad (\text{C.6.8})$$

For a Rayleigh distribution it is well known (9) that

$$E\{X\} = \frac{1}{2} \sqrt{\frac{\pi}{C}} = \alpha \sqrt{\frac{\pi}{2}} \quad (\text{C.6.9})$$

and

$$E\{X^2\} = \frac{1}{C} = 2\alpha^2 \quad (\text{C.6.10})$$

Thus the point estimate of the mean square value of  $X$  is

$$E\{X^2\} = \frac{a_0}{a_0 + n} \frac{b_0}{a_0} + \frac{n}{a_0 + n} \left( \frac{1}{n} \sum_{i=1}^n X_i^2 \right) \quad (\text{C.6.11})$$

Therefore  $b_0/a_0$  can be considered as the prior estimate  $2\alpha_0^2$  of the mean square value of  $X$  while  $a_0$  is an equivalent sample size for the prior estimate.

VITA 2

David Ray Cunningham

Candidate for the Degree of

Doctor of Philosophy

Thesis: APPLICATION OF THE EMPIRICAL DISTRIBUTION FUNCTION TO  
ENGINEERING DECISIONS

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in El Reno, Oklahoma, November 15, 1934, the son of Mr. and Mrs. Ray Cunningham; married with two children.

Education: Graduated from El Reno High School, El Reno, Oklahoma, in May, 1953; attended El Reno Junior College from September, 1953, to June, 1954; received the Bachelor of Science degree from Oklahoma State University in May, 1957, and the Master of Science degree from the University of Idaho in May, 1959, with majors in Electrical Engineering; completed the requirements for the Doctor of Philosophy degree at Oklahoma State University in August, 1969.

Professional Experience: Engineering Training Program, General Electric Company, June, 1957, to September, 1958; research and development of advanced recording techniques, circuitry and instrumentation, General Electric Company, July, 1959, to September, 1964, and June, 1965, to September, 1966; part-time teaching, Oklahoma State University, September, 1964, to June, 1965, and September, 1966, to June, 1967.