ESTIMATING STATISTICALLY SIGNIFICANT DIFFERENCES
BETWEEN A PAIR OF BETA DISTRIBUTIONS

by
KRISHNAN LAKSHMINARAYAN

B. Tech (Mechanical Engineering)
Indian Institute of Technology, Madras, 1974

_____________________________________
A MASTER'S REPORT

submitted in partial fulfillment of the
requirements of the degree

MASTER OF SCIENCE

Department of Industrial Engineering
KANSAS STATE UNIVERSITY
Manhattan, Kansas
1980

Approved by:

Major Professor
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iii</td>
</tr>
<tr>
<td>1.0 INTRODUCTION.</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Objective.</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Literature Search.</td>
<td>1</td>
</tr>
<tr>
<td>1.3 The Problem.</td>
<td>2</td>
</tr>
<tr>
<td>1.4 Analysis</td>
<td>4</td>
</tr>
<tr>
<td>1.4.1 Measures of Separation.</td>
<td>5</td>
</tr>
<tr>
<td>2.0 THE $\chi^2$ TEST STATISTIC</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Construction of the Test</td>
<td>8</td>
</tr>
<tr>
<td>3.0 DETERMINATION OF THE INDEX OF NON-CONGRUITY</td>
<td>13</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>3.2 Procedures for Determining the Points of Intersection.</td>
<td>17</td>
</tr>
<tr>
<td>3.2.1 Evaluation of $\delta$ for Case (i) (one intersection point).</td>
<td>17</td>
</tr>
<tr>
<td>3.2.1.1 Evaluation of $\delta_1$</td>
<td>18</td>
</tr>
<tr>
<td>3.2.1.2 Evaluation of $\delta_2$</td>
<td>18</td>
</tr>
<tr>
<td>3.2.2 Evaluation of $\delta$ for Two Intersection Points (case (ii))</td>
<td>19</td>
</tr>
<tr>
<td>3.3 Random Sampling from a Beta Distribution</td>
<td>20</td>
</tr>
<tr>
<td>4.0 RESULTS</td>
<td>24</td>
</tr>
<tr>
<td>4.1 Conclusions</td>
<td>26</td>
</tr>
<tr>
<td>5.0 ACKNOWLEDGEMENTS</td>
<td>36</td>
</tr>
<tr>
<td>6.0 REFERENCES</td>
<td>37</td>
</tr>
<tr>
<td>APPENDIX I: Computer Programs</td>
<td>38</td>
</tr>
</tbody>
</table>
THIS BOOK CONTAINS NUMEROUS PAGES WITH DIAGRAMS THAT ARE CROOKED COMPARED TO THE REST OF THE INFORMATION ON THE PAGE. THIS IS AS RECEIVED FROM CUSTOMER.
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Graphs of beta densities for various values of the parameters $a$ and $b$.</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>$\chi^2$ distribution showing right hand tail area for a given $\chi^2$ value.</td>
<td>12</td>
</tr>
<tr>
<td>3.1</td>
<td>Graphs of beta $(a_1,b_1)$ and beta $(a_2,b_2)$ showing $\delta$ (one intersection point).</td>
<td>14</td>
</tr>
<tr>
<td>3.2</td>
<td>Graph of $g(x)$ against $x$. The intersection points are represented by $x_0$, $x_1$, and $x_2$.</td>
<td>15</td>
</tr>
<tr>
<td>3.3</td>
<td>Shape of the cumulative distribution function for beta $(a,b)$.</td>
<td>21</td>
</tr>
<tr>
<td>4.1</td>
<td>Relationship of $\eta_1$ and $\delta$.</td>
<td>28</td>
</tr>
<tr>
<td>4.2</td>
<td>Relationship of $x_{ps}^2$ and $\delta$.</td>
<td>29</td>
</tr>
<tr>
<td>4.3</td>
<td>Relationship of $x_{ps}^2$ and $\eta_1$.</td>
<td>30</td>
</tr>
<tr>
<td>4.4</td>
<td>Relationship of $x_{ps}^2$ and $\eta_2$.</td>
<td>31</td>
</tr>
<tr>
<td>4.5</td>
<td>Relationship of $\delta$ and $\eta_2$.</td>
<td>32</td>
</tr>
<tr>
<td>4.6</td>
<td>Relationship of $(x_{ps}^2-x_s^2)$ and $\eta_2$.</td>
<td>33</td>
</tr>
<tr>
<td>4.7</td>
<td>Relationship of $x_{ps}^2$ and $x_s^2$.</td>
<td>34</td>
</tr>
<tr>
<td>4.8</td>
<td>Relationship of $x_s^2$ and $\delta$.</td>
<td>35</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Probability elements for model beta (4,4) and alternative beta (3,3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Results using a beta (4,4) model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.0 INTRODUCTION

1.1 Objective

The beta distribution has been used frequently to fit sample data (1). It has been used, for example, as a prior distribution for a binomial proportion ((1) Chapter 8). For a given set of data, different methods can be used to estimate the beta parameters, usually yielding different estimates. In order to distinguish between a pair of beta distributions, it may be desirable to invent an index, or indices, that reveal the extent of these differences.

The objective of this investigation was to develop a measure (or measures) of separation between a pair of beta distributions, and to correlate these measures with some standard test of significance, such as the $\chi^2$ goodness-of-fit test. An important sub-objective of this research was to develop a computer code that can be used to obtain these correlations for a wide variety of beta distributions.

1.2 Literature Search

Several coefficients have been suggested in statistical literature to reflect the fact that some probability distributions are "closer together" than others, and consequently that it may be "easier to distinguish" between the distributions of one pair than between those of another (2). Such coefficients have been variously called measures of distance between two distributions (3), measures
of separation (4), measures of discriminatory information (5,6) and measures of variation distance (7). These coefficients have the common property of increasing as the two distributions involved "move apart".

Rao (4) states the need for a distance function between two populations, and describes mathematical concepts involving the construction of such a function. Mahalanobis (9) defined a generalized distance to be used in comparing normal populations. Karl Pearson proposed a measure of racial likeness (10) which has been used by anthropologists for analyzing skeletal remains. He defined a coefficient of racial likeness (C.R.L) which is a measure of the distance between two populations. Adhikari and Joshi (3) give a complete summary of the various coefficients of distance developed by several investigators.

1.3 The Problem

The concept of significance has been developed for analyzing sample data and for making inferences about the populations being sampled. Various tests have been developed for comparing groups of sample data with each other or for comparing a sample with a population (or model). The question to be asked when comparing two populations is: How different is different?

The type of distribution considered here is the beta family described by the density function

\[ f(x|a,b) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} \quad (0 \leq x \leq 1) \quad (1.1) \]

where

\[ B(a,b) = \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)} \quad (1.2) \]
and the gamma function is defined as

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x} \, dx \quad [0 < a < \infty].$$

(1.3)

For convenience of terminology, such a density function (or a random variable thus distributed) is described as beta \((a, b)\).

In general, by varying the parameter \(a\) and \(b\) of the beta distribution, a variety of curve shapes are obtained (Figure 1.1). For this reason, the beta distribution is often suitable in modeling work for random variables over the interval \((0, 1)\).

Density \(A\) can be called significantly different from density \(B\) if a random sample from \(A\) does not pass a \(\chi^2\) goodness-of-fit test, when tested against \(B\) (or vice versa). The significance level of an observed random sample (or perhaps the average of many) can be noted and compared with some index which depends only on the parameters of the curves. However, the construction of random samples presents some difficulties:

(i) Beta random samples are expensive to generate, when concentrated over a small range in \((0, 1)\).

(ii) True random samples would require many points to ascertain the relationship between any two indices of non-equality, because there would be the random fluctuation effect which would obscure the basic relationship.

For this reason, emphasis is placed in this study on "pseudo-samples" where the "observed" frequency for a class is taken to be the same as the "expected" frequency for that class for the particular distribution under consideration. The chi-squared value obtained from testing density \(A\) against density \(B\) under these conditions is referred to as a "pseudo-chi-squared" statistic, and has the following properties:
(a) It is zero when density A and density B are the same

(b) It has no random component and is purely a measure of lack of fit, since the "sample" is a perfect replica of its generating density function.

The objective, as explained in more detail below, is to

i) Compute a pseudo-chi-square test statistic for a sample from density A tested against density B. Here B is defined as the "model" and A is defined as the "alternative".

ii) Compute an "index of non-congruity", denoted by $\delta$, and relate this to the pseudo chi-square value.

iii) Compute certain other indices that are easily determined from the parameters of both distributions.

iv) Generate a random sample from A and test against B using the chi-square goodness-of-fit test.

v) Lastly, study the variation of the chi-square value obtained in iv) with the indices obtained in iii).

1.4 Analysis

Theoretically, two distributions are considered equal only if their density functions are identical. But in practice minor variations in the parameters may be of no consequence because two "nearby" densities may yield samples that are indistinguishable from each other. For example, it may be impossible to detect any difference in a random sample from a beta (3,3) density and one from a beta (3.2, 3.5) density. If it were possible to find the sampling distribution of the statistics $\hat{a}$ and $\hat{b}$ (the estimates of $a$ and $b$), then the two distributions could be considered indistinguishable if it were possible, using the computed $\hat{a}_i, \hat{b}_i$ ($i = 1,2$), to accept the null hypothesis:
\[ H_0: \ a_1 = a_2 \]
\[ b_1 = b_2 \]

against the alternative:
\[ H_1: \ a_1 \neq a_2 \]
\[ b_1 \neq b_2 \]

In the absence of knowledge of the distribution of \( \hat{a} \) and \( \hat{b} \), other techniques must be employed.

An alternative approach would be to first construct "expected" frequencies for each class interval for each distribution. These are called "perfect" samples. One distribution is then referred to as the "model", and can be tested against the "alternative", using a chi-squared goodness of fit. The chi-square value thus obtained will be a "pseudo" chi-square value whenever we construct "perfect" samples from each distribution. If a genuine random sample were drawn from the "alternative" and compared with the "model" using the chi-squared test statistic, the chi-square value thus obtained would be the usual chi-squared test statistic. Discussion of the computation of the chi-squared value is given in Chapter II.

1.4.1 Measures of Separation

Let beta \( (a_1, b_1) \) and beta \( (a_2, b_2) \) be the beta distributions under consideration (Figures 3.1 (i) and (ii)). The "index of non-congruity", is defined as

\[ \delta = \int_0^1 |f(x|a_1, b_1) - f(x|a_2, b_2)| \, dx \quad (1.4) \]
where

$$0 \leq \delta \leq 2.$$  

If the two curves are identical, then $\delta = 0$. However, if the two curves are very different, $\delta$ may be near the maximum possible value of 2.

Various other indices of non-equality are also computed in this study. These are defined as

$$\eta_1 = |\mu_1 - \mu_2| + |\sigma_1 - \sigma_2|$$  \hspace{1cm} (1.5)

$$\eta_2 = \left| \frac{\mu_1}{\sigma_1} - \frac{\mu_2}{\sigma_2} \right|$$  \hspace{1cm} (1.6)

where $\mu_i$ and $\sigma_i$ are, respectively, the mean and standard deviation of the $i^{th}$ distribution, $i = 1,2$.

Results have been obtained using beta $(4,4)$ as the model and beta $(3,b_2)$ as the alternative, where $b_2$ has been varied in the interval [2,4] in steps of 0.1. The pseudo sample size used is 100. Relationships have been obtained between $\eta_1$ and $\delta$, $\delta$ and $\chi^2$, $\eta_1$ and $\chi^2$, and $\eta_2$ and $\chi^2$. Lastly, random samples have been drawn from the alternative beta and the $\chi^2$ value thus obtained ($\chi^2$) has been plotted against $\eta_1$ and $\eta_2$. 
Fig. 1.1. Graphs of beta densities for various values of the parameters $a$ and $b$. (From (11)).
2.0 THE $x^2$ TEXT STATISTIC

2.1 Construction of the Test

Let $f(x|a_1, b_1)$ represent the density function for the model and $f(x|a_2, b_2)$ represent the density function for the alternative distribution. The first step in computing the chi-squared value is to obtain a frequency distribution for both the model and the alternative.

The ordinate of the density function at a point does not represent the probability at that point. For this reason, a finite number of classes are constructed over the interval $[0,1]$. The probability element for each class interval is normally computed using the equation

$$F = \int_{x_1}^{x_2} f(x|a,b) \, dx \quad x_1, x_2 \in [0,1] \quad (2.1)$$

where $x_1$ and $x_2$ are the lower and upper limits, respectively, of the class interval. Since it is not possible in general to analytically evaluate the integral in Eq. (2.1), an approximation is made by evaluating the ordinate at the midpoint of the class interval and multiplying by the class width. This probability element has the physical meaning of relative frequency, since it denotes the probability of an event occurring in that class interval. In the case of the alternative curve, these probability elements are "pseudo" relative frequencies, since they are obtained from an exact model and not from some random process occurring in nature.

In the computer program in Appendix I the probability element is designated as PRBEL. Using the subscript 1 for the model and 2 for the alternative, let $N$ be the number of class intervals (cells) chosen. Then the pseudo chi-square statistic is defined for pseudo-sample size $M$ as
\[ x_{ps}^2 = M \sum_{i=1}^{N} \frac{[(PRBEL)_{1i} - (PRBEL)_{2i}]^2}{(PRBEL)_{1i}} \]  

(2.2)

with degrees of freedom \(DF = N - 1\).

The following example illustrates the construction of the chi-square test statistic.

**EXAMPLE** Let beta (4,4) be the model and beta (3,3) be the alternative.

A sample size \(M = 50\) is assumed, and \(N = 10\). Thus

\[ f(x|4,4) = \frac{1}{B(4,4)} x^3(1-x)^3 \quad [0 \leq x \leq 1] \]  

(2.3)

\[ f(x|3,3) = \frac{1}{B(3,3)} x^2(1-x)^2 \quad [0 \leq x \leq 1] \]  

(2.4)

The "observed" frequencies, designated OF, and the "expected" frequencies, designated EF, were obtained using the program developed for this investigation (Appendix 1). These results are illustrated in Table 2.1, which also illustrates the probability elements for each class interval. If a sample of \(M\) observations is assumed, then each probability element must be multiplied by \(M\) to obtain the "frequency count" for that particular cell. In the computer program, OF and EF were designated as \(PRBEL(L,I,M),\) (for the \(L^{th}\) curve, \(I^{th}\) class, sample size \(M\)). Thus

\[ PRBE(L,I,M) = PRBEL(L,I) \times M \]  

(2.5)

where

\[ L = 1, \text{ (model) } I = 1, \ldots, N \]

\[ = 2 \text{ (alternative) } \]

The OF and EF values in Table 2.2 correspond to the PRBE values computed above.
In order to compute the pseudo-chi square test statistic a classical (though perhaps overly strict) requirement is that the expected frequency count in each interval be at least 5. This is ensured by grouping adjacent intervals whenever necessary. This reduces the effective number of intervals. In the above problem classes 1, 2 and 3 are grouped together to form one interval and intervals 8, 9, and 10 are grouped together to form another, the others remaining unaltered. The effective number of intervals is then reduced to 10-6+2 = 6. The pseudo-chi-square statistic then becomes

\[ \chi^2_{ps} = M \sum_{i=1}^{6} \frac{(PRBEl(I,I) - PRBEl(2,I))^2}{PRBEl(1,I)} \]  \hspace{1cm} (2.6) 

and carrying out the computations yields:

\[ \chi^2_{ps} = 1.474 \text{ with df = 6-1 = 5.} \]

As explained below, the next step is to calculate the significance level \( \hat{\alpha} \) of the test.

The usual \( \chi^2 \) table gives the critical point for a given \( \alpha \). But it is desired to know what significance level, called \( \hat{\alpha} \), is associated with an observed \( \chi^2 \) value. In the present work a function subprogram CADTR (12) is employed for this purpose. The input parameters of the subroutine are the known chi-square value and the degrees of freedom, both of which are treated as real continuous parameters (abbreviated as \( \chi^2 \) and DF, respectively, for convenience).

The resultant \( \hat{\alpha} \) value calculated by subprogram CADTR is \( \hat{\alpha} = 0.9161 \), for \( \chi^2_{ps} = 1.474 \text{ with df = 5.} \)
<table>
<thead>
<tr>
<th>CLASS MARK</th>
<th>(PRBEL)_{2i}</th>
<th>(PRBEL)_{1i}</th>
</tr>
</thead>
<tbody>
<tr>
<td>.05</td>
<td>.0068</td>
<td>.0015</td>
</tr>
<tr>
<td>.15</td>
<td>.0488</td>
<td>.0290</td>
</tr>
<tr>
<td>.25</td>
<td>.1055</td>
<td>.0923</td>
</tr>
<tr>
<td>.35</td>
<td>.1553</td>
<td>.1648</td>
</tr>
<tr>
<td>.45</td>
<td>.1838</td>
<td>.2123</td>
</tr>
<tr>
<td>.55</td>
<td>.1838</td>
<td>.2123</td>
</tr>
<tr>
<td>.65</td>
<td>.1553</td>
<td>.1648</td>
</tr>
<tr>
<td>.75</td>
<td>.1055</td>
<td>.0923</td>
</tr>
<tr>
<td>.85</td>
<td>.0488</td>
<td>.0290</td>
</tr>
<tr>
<td>.95</td>
<td>.0068</td>
<td>.0015</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

TABLE 2.1  Probability elements for model beta (4,4) and alternative beta (3,3).
Fig. 2.1. $\chi^2$ distribution showing right hand tail area for a given $\chi^2$ value.
3.0 Determination of the Index of Non-Congruity

3.1 Introduction

The index of non-congruity, $s$, is calculated in the following manner. As before, let $f(x|a_1,b_1)$ and $f(x|a_2,b_2)$ be two beta distributions described in the manner of Figure 3.1.

It can be proved that there will be at least one point of intersection for the two distributions but never more than two. Let the points of intersection be denoted by

$$x = x_i, \quad i = 1, 2 \quad (3.1)$$

At each such $x$

$$f(x|a_1,b_1) = f(x|a_2,b_2) \quad (3.2)$$

Thus,

$$\frac{1}{B(a_1,b_1)} x^{a_1-1} (1-x)^{b_1-1} = \frac{1}{B(a_2,b_2)} x^{a_2-1} (1-x)^{b_2-1}$$

hence

$$x^{a_1-a_2} (1-x)^{b_1-b_2} = \frac{B(a_1,b_1)}{B(a_2,b_2)}$$

$$= \frac{\Gamma(a_1) \Gamma(b_1)}{\Gamma(a_1+b_1)} \cdot \frac{\Gamma(a_2+b_2)}{\Gamma(a_2) \Gamma(b_2)} = C \quad (3.3)$$

The right hand side of equation (3.3) is easily evaluated by subroutine GMMA, documented in [12], which evaluates the gamma function for real arguments.

In order to determine the points of intersection, it is necessary to determine the zeros of the equation
Fig. 3.1(i). Graphs of $\text{beta}(a_1, b_1)$ and $\text{beta}(a_2, b_2)$ showing $\delta$ (two intersection points).

Fig. 3.1(ii). Graphs of $\text{beta}(a_1, b_1)$ and $\text{beta}(a_2, b_2)$ showing $\delta$ (one intersection point).
Fig. 3.2. Graph of $g(x)$ against $x$. 
\[ g(x) = x^{a_1-a_2}(1-x)^{b_1-b_2} - C = 0 \]  \hspace{1cm} (3.4)

The critical point is determined by differentiating Eq. (3.4) and setting the result to 0 at \( x = x_0 \), the critical point. Thus,

\[
\frac{dg(x)}{dx} \bigg|_{x=x_0} = \frac{a_1-a_2}{x_0} (1-x_0)^{b_1-b_2} + \frac{b_1-b_2}{(b_1-b_2)(1-x_0)^{b_1-b_2}} = 0
\]

hence,

\[
x_0^{a_1-a_2-1} (1-x_0)^{b_1-b_2-1} \left[ (a_1-a_2)(1-x_0) - (b_1-b_2)x_0 \right] = 0
\]

\[
(a_1-a_2)(1-x_0) - (b_1-b_2)x_0 = 0
\]

\[
x_0 = \frac{a_1-a_2}{(a_1-a_2) + (b_1-b_2)} \hspace{1cm} (3.5)
\]

Equation (3.6) reveals that there is only one critical point over \([0,1]\), given by \( x_0 \). Furthermore

\[
\begin{align*}
g(x)_{x=0} &= -C \\
g(x)_{x=1} &= -C
\end{align*}
\]

Using (3.6) and (3.7) one can describe the general shape of \( g(x) \) over \( x \) as shown by curve 1 and curve 2 in Figure 3.2.

Curve 1, Figure (3.2), illustrates the case where two intersection points at \( x_1 \) and \( x_2 \) were obtained. Curve 2 in the same figure illustrates the case where one intersection point was obtained. Referring to curve 1, \( x_0 \) is the maximum point and \( g(x_0) \), from equation (3.5), is positive. Referring to curve 2, \( x'_0 \) is the maximum point and \( g(x'_0) \) is zero. In
the latter case, the point of maximum is also the point of intersection of the two curves.

3.2 Procedure for Determining the Points of Intersection

A first step in determining the intersection points $x_1$ and $x_2$ is to establish whether the function $g(x)$ is of the type represented by curve 1 or by curve 2, referring to Figure 3.1. To do this, first evaluate

$$x_0 = \frac{a_1-a_2}{(a_1-a_2) + (b_1-b_2)} \quad (3.7)$$

and then evaluate $g(x_0)$ from equation (3.5). The value of $g(x_0)$ determines the number of intersection points. Two cases are possible:

(i) $g(x_0) = 0$ ⇒ one root at $x_0$

(ii) $g(x_0) > 0$ ⇒ two roots, $x_1$ and $x_2$

3.2.1 Evaluation of $\delta$ for case (i) (one intersection point)

The intersection point $x_0$ is determined by equation (3.6). The "index of non-congruity" $\delta$ then becomes

$$= \int_0^{x_0} |f(x|a_1,b_1) - f(x|a_2,b_2)| \, dx + \int_{x_0}^1 |f(x|a_1,b_1) - f(x|a_2,b_2)| \, dx$$

$$= \delta_1 + \delta_2 \quad (3.8)$$

where

$$\delta_1 = \int_0^{x_0} |f(x|a_1,b_1) - f(x|a_2,b_2)| \, dx \quad (3.9)$$

$$\delta_2 = \int_0^1 |f(x|a_1,b_1) - f(x|a_2,b_2)| \, dx \quad (3.10)$$
3.2.1.1 Evaluation of $\delta_1$

Since $f(x|a,b)$ is a positive function, Eq. (3.9) can also be written as:

$$\delta_1 = \left| \int_0^{x_0} f(x|a_1,b_1) \, dx - \int_0^{x_0} f(x|a_2,b_2) \, dx \right|$$

Each of the integrals is evaluated separately. However, neither of them can be evaluated analytically. Some form of series expansion method must be used, and subroutine BDTR [12] does this. The subroutine, documented in [12], computes $I_x(a,b)$ defined by:

$$I_x(a,b) = \int_0^x f(y|a,b) \, dy. \quad (3.11)$$

Then

$$\delta_1 = |I_{x_0}(a_1,b_1) - I_{x_0}(a_2,b_2)| \quad (3.12)$$

3.2.1.2 Evaluation of $\delta_2$

Recall the formula for $\delta_2$ (Eq. (3.10)) as

$$\delta_2 = \int_{x_0}^1 |f(x|a_1,b_1) - f(x|a_2,b_2)| \, dx$$

$$= \left| \int_{x_0}^1 f(x|a_1,b_1) \, dx - \int_{x_0}^1 f(x|a_2,b_2) \, dx \right| \quad (3.13)$$

where $x_0$ is the intersection point. Each of the above two integrals can be evaluated separately. Using the relationship:

$$I_x(a,b) = \int_0^x f(y|a,b) \, dy = 1 - \int_x^1 f(y|a,b) \, dy \quad (3.14)$$

each of the integrals in equation (3.12) is expressed as:
\[
\int_{x_0}^{1} f(x, a_i, b_i) \, dx = 1 - \int_{0}^{x_0} f(x, a_i, b_i) \, dx \\
= 1 - I_{x_0} (a_i, b_i) \quad i = 1, 2 \tag{3.15}
\]

hence

\[
\delta_2 = \left| 1 - I_{x_0} (a_1, b_1) - (1 - I_{x_0} (a_2, b_2)) \right| \\
= \left| I_{x_0} (a_2, b_2) - I_{x_0} (a_1, b_1) \right| \tag{3.16}
\]

Equation (3.16) for \( \delta_2 \) is identical to equation (3.12) for \( \delta_1 \). Thus

\[
\delta_1 = \delta_2 \tag{3.17}
\]

3.2.2 Evaluation of \( \delta \) for two intersection points (case (ii))

Let \( x_1 \) and \( x_2 \) be the intersection points shown in Figure 1.2(i).

The index of non-congruity \( \delta \) is given by

\[
\delta = \delta_1 + \delta_2 + \delta_3
\]

where

\[
\delta_1 = \int_{0}^{x_1} \left| f(x, a_1, b_1) - f(x, a_2, b_2) \right| \, dx \tag{3.18}
\]

\[
\delta_2 = \int_{x_1}^{x_2} \left| f(x, a_1, b_1) - f(x, a_2, b_2) \right| \, dx \tag{3.19}
\]

\[
\delta_3 = \int_{x_2}^{1} \left| f(x, a_1, b_1) - f(x, a_2, b_2) \right| \, dx \tag{3.20}
\]

Before computing each of the above integrals, it is necessary to evaluate the points of intersection, \( x_1 \) and \( x_2 \), using equation (3.5) which is reproduced below:

\[
g(x) = x^{a_1-a_2} (1-x)^{b_1-b_2} + C = 0 \tag{3.5}
\]
Knowing that \( g(x_0) > 0 \) (recall Figure 3.1, curve 1) the roots of the nonlinear equation \( g(x) \) must be obtained using a search technique. Mueller's iterative scheme, documented in [12], pages 217-218, is a convenient scheme to use, and has been incorporated into the computer program in Appendix 1.

3.3 Random Sampling from a Beta Distribution

Several methods can be used for generating random beta variates [13]. The method used here is a table look-up procedure with linear interpolation. Briefly, this method involves setting up a table of values of the cumulative distribution of beta \((a,b)\) in the range. A random number is then interpolated directly into the cumulative tabled value to yield the desired beta variate. A discussion of the procedure used follows:

Let \( f(x|a,b) \) defined by equations (1.1) and (1.2) be the given beta density function. The first step in generating a random sample is to set up a table of cumulative distribution function (c.d.f.) values. Subroutine BDTR, documented in [1], pages 78-80, can be employed for this purpose. The shape of the cumulative distribution function

\[
f(x) = \int_0^x f(t|a,b) dt
\]

has the form depicted in Figure (3.3).

The next step is to generate a random number in the interval \([0,1]\) and set this value equal to \( F(x) \). Using linear interpolation (if necessary) the corresponding beta variate \( x \) can then be obtained.
Fig. 3.3. Shape of cumulative distribution function for beta \((a,b)\).
Referring to Figure 3.3, \( R_1 \) is a uniform random number and the corresponding beta variate is \( x_1 \).

The random number generator used in this investigation was obtained from the random number package "Super Duper" [14]. The function UNI(0) generates a uniform random variable in the half open interval \([0,1]\).

The cumulative table is constructed with the beta variate \( x \) ranging from 0.0 (0.005) 1.0. For example, let \( \text{RANUM}(I) \) be the \( i^{th} \) random number generated, and let \( \text{PROB}(K) \) and \( \text{PROB}(K-1) \) be the tabulated cumulative probability values with associated values of the beta variate \( X(K) \) and \( X(K-1) \), respectively, such that

\[
\text{PROB}(K-1) < \text{RANUM}(I) < \text{PROB}(K)
\]

and

\[
x(K-1) < \text{DEV}(I) < X(K)
\]

where \( \text{DEV}(I) \) is the desired variate.

\[
\text{DEV}(I) = \frac{x(K-1)[\text{PROB}(K) - \text{RANUM}(I)] + x(K)[\text{RANUM}(I) - \text{PROB}(K-1)]}{[x(K) - x(K-1)]}
\]

which is a linear interpolation method.

The next step is to group these random beta deviates into a frequency distribution.

This grouping is done as before for the case of the pseudo-random samples (Chapter 2). To compare these random frequencies (which may correspond to a real life situation) to the model (which is an assumed beta distribution) the \( \chi^2 \) goodness-of-fit test is used (as was done for the case of the pseudo-random samples). The \( \chi^2 \) value obtained in this manner is designated \( \chi^2_S \).
Comparisons were made between δ (discussed in this chapter), η₁ and η₂ (presented in Chapter 1) and the χ² values obtained between a pair of beta distributions. These comparisons will be discussed in the next chapter.
4.0 RESULTS

Results are presented in this chapter for a few special cases of beta distributions. A beta (4,4) was used as the model, and beta (3, b_2) were used as alternative distributions, where b_2 was varied in the range (2,4) in steps of 0.1. Eight sets of relationships between the various measures of separation were studied, and each is briefly described below:

i) $\eta_1$ versus $\delta$ The results indicate that the index $\eta_2$ defines certain ranges over which $\eta_1$ and $\delta$ are related. Referring to Figure 4.1, we see that an increase in $\delta$ (associated with a pair of curves that are increasingly different) corresponds to an increase in $\eta_1$, as might be expected.

ii) $\delta$ versus $\chi^2_{ps}$ Again here the index $\eta_2$ plays an auxiliary discriminating role in the functional relationship. Two distinct ranges of $\eta_2$ over which the relationship of $\delta$ against $\chi^2_{ps}$ is valid have been established, as shown in Figure 4.2. As might be expected, an increase in $\delta$ corresponds to an increase in $\chi^2_{ps}$.

iii) $\eta_1$ versus $\chi^2_{ps}$ Referring to Figure 4.3, we see that over two ranges of $\eta_2$, $\eta_1$ increases as $\chi^2_{ps}$ increases.

iv) $\eta_2$ versus $\chi^2_{ps}$ Figure 4.4 depicts this relationship. A minimum value of $\chi^2_{ps}$ is obtained at $\eta_2 = 0.354$. This corresponds to a $\chi^2_{ps}$ value of 2.947, indicating that beta (4,4) and beta (3,3) are the least different from each other among those tested.
v) \( \eta_2 \) versus \( \delta \) Figure 4.5 shows the relationship between \( \eta_2 \) and \( \delta \). A parabolic type relationship is again obtained, with a minimum at \( \eta_2 = 0.379 \). The value of \( \delta \) at this point is 0.144. Comparing Figures 4.4 and 4.5, it appears that \( \eta_2 \) is a better index to work with, since the relationships obtained have been independent of \( \eta_1 \) or any other index.

vi) \( \eta_2 \) versus (\( x_{ps}^2 - x_s^2 \)) Figure 4.6 shows the relationship. It was expected that in general \( x_{ps}^2 < x_s^2 \), and a fixed quantity \( R \) (i.e., a random component) would have to be added to the \( x_{ps}^2 \) value to obtain a more "realistic" approximation to a random sample value. Further, it was expected that \( R \) would be approximately constant from distribution to distribution and would always be positive. Figure 4.6, however, reveals something quite different. We see that \( x_s^2 \) fluctuates around \( x_{ps}^2 \), indicating that in the limit, i.e., as \( x_2 \) increases indefinitely, \( x_s^2 \) would approach the \( x_{ps}^2 \) value.

vii) \( x_{ps}^2 \) versus \( x_s^2 \) This relationship is shown in Fig. (4.7). A least squares linear fit has been obtained for the data.

\[
x_s^2 = 1.518 x_{ps}^2 - 2.949
\]

The correlation coefficient for this case is

\[
r^2 = 0.944.
\]

viii) \( x_s^2 \) versus \( \delta \) Figure (4.8) depicts this relationship. A power curve of the form

\[
y = ax^b, \quad x,y > 0
\]

has been fitted to the data. This fit is given by

\[
x_s^2 = 298.35 \delta^{2.4}
\]

and the correlation coefficient is

\[
r^2 = 0.91
\]
4.1 Conclusions

The results have been obtained for a few particular cases of beta distributions. Certain limited conclusions can be drawn from these cases, which may not apply for other model and alternative distributions. These conclusions are presented briefly below:

(i) For $\delta > .27$ (see Figures (4.2) and (4.8)), a $\chi^2_s$ test would show significance at the $\alpha = .05$ level of significance. This would apply for $a_2 = 3$ and $b_2 > 3.7$ or $b_2 < 2.5$. For $\alpha = .01$, then for $\delta > .26$ a $\chi^2_s$ test would show significance. This would apply for $a_2 = 3$ and $b_2 > 3.8$ or $b_2 < 2.4$.

(ii) The ranges of $n_2$ over which the $\chi^2_{ps}$ test would show significance are illustrated in Figure 4.4. At $\alpha = .05$ and .01, and an approximate range $n_2 > .5$ or $n_2 < .2$ a $\chi^2_{ps}$ test would show significance.

The results obtained from the investigations for these special cases have been disappointing to the investigator in that the measures of separation developed are shown to be only moderately good, and are probably not promising enough to warrant further investigation.
<table>
<thead>
<tr>
<th>( a_2 )</th>
<th>( b_2 )</th>
<th>( \chi^2_{ps} )</th>
<th>df</th>
<th>( \hat{\alpha} )</th>
<th>( \chi^2_s )</th>
<th>df</th>
<th>( \hat{\alpha} )</th>
<th>( \chi^2_{ps}-\chi^2_s )</th>
<th>( n_1 )</th>
<th>( n_2 )</th>
<th>( \delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>3.982</td>
<td>5</td>
<td>.5521</td>
<td>-3.982</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3.9</td>
<td>17.11</td>
<td>5</td>
<td>.0043</td>
<td>16.91</td>
<td>5</td>
<td>.0047</td>
<td>0.20</td>
<td>.075</td>
<td>.535</td>
<td>.308</td>
</tr>
<tr>
<td>3</td>
<td>3.8</td>
<td>14.27</td>
<td>5</td>
<td>.0140</td>
<td>13.82</td>
<td>5</td>
<td>.0168</td>
<td>0.45</td>
<td>.070</td>
<td>.518</td>
<td>.280</td>
</tr>
<tr>
<td>3</td>
<td>3.7</td>
<td>11.68</td>
<td>5</td>
<td>.0394</td>
<td>10.53</td>
<td>5</td>
<td>.0616</td>
<td>1.15</td>
<td>.065</td>
<td>.501</td>
<td>.252</td>
</tr>
<tr>
<td>3</td>
<td>3.6</td>
<td>9.802</td>
<td>5</td>
<td>.1330</td>
<td>7.99</td>
<td>5</td>
<td>.1568</td>
<td>1.81</td>
<td>.059</td>
<td>.483</td>
<td>.224</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>7.389</td>
<td>5</td>
<td>.1933</td>
<td>5.381</td>
<td>5</td>
<td>.3712</td>
<td>2.01</td>
<td>.054</td>
<td>.465</td>
<td>.197</td>
</tr>
<tr>
<td>3</td>
<td>3.4</td>
<td>5.725</td>
<td>5</td>
<td>.3339</td>
<td>4.264</td>
<td>5</td>
<td>.5121</td>
<td>1.46</td>
<td>.048</td>
<td>.445</td>
<td>.174</td>
</tr>
<tr>
<td>3</td>
<td>3.3</td>
<td>4.419</td>
<td>5</td>
<td>.4908</td>
<td>4.499</td>
<td>5</td>
<td>.4800</td>
<td>-0.08</td>
<td>.042</td>
<td>.424</td>
<td>.156</td>
</tr>
<tr>
<td>3</td>
<td>3.2</td>
<td>3.499</td>
<td>5</td>
<td>.6235</td>
<td>5.229</td>
<td>5</td>
<td>.3885</td>
<td>-1.73</td>
<td>.036</td>
<td>.402</td>
<td>.146</td>
</tr>
<tr>
<td>3</td>
<td>3.1</td>
<td>2.997</td>
<td>5</td>
<td>.7004</td>
<td>3.371</td>
<td>5</td>
<td>.6430</td>
<td>-0.37</td>
<td>.029</td>
<td>.379</td>
<td>.144</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>2.947</td>
<td>5</td>
<td>.7081</td>
<td>3.784</td>
<td>5</td>
<td>.5809</td>
<td>-0.84</td>
<td>.022</td>
<td>.354</td>
<td>.149</td>
</tr>
<tr>
<td>3</td>
<td>2.9</td>
<td>3.384</td>
<td>5</td>
<td>.6409</td>
<td>3.602</td>
<td>5</td>
<td>.6080</td>
<td>-0.22</td>
<td>.032</td>
<td>.328</td>
<td>.161</td>
</tr>
<tr>
<td>3</td>
<td>2.8</td>
<td>4.344</td>
<td>5</td>
<td>.5010</td>
<td>1.976</td>
<td>5</td>
<td>.8524</td>
<td>2.37</td>
<td>.042</td>
<td>.301</td>
<td>.179</td>
</tr>
<tr>
<td>3</td>
<td>2.7</td>
<td>5.862</td>
<td>5</td>
<td>.3199</td>
<td>4.270</td>
<td>5</td>
<td>.5112</td>
<td>1.59</td>
<td>.053</td>
<td>.272</td>
<td>.203</td>
</tr>
<tr>
<td>3</td>
<td>2.6</td>
<td>7.972</td>
<td>5</td>
<td>.1578</td>
<td>8.630</td>
<td>5</td>
<td>.1248</td>
<td>-0.658</td>
<td>.063</td>
<td>.240</td>
<td>.232</td>
</tr>
<tr>
<td>3</td>
<td>2.5</td>
<td>10.70</td>
<td>5</td>
<td>.0576</td>
<td>13.79</td>
<td>5</td>
<td>.0170</td>
<td>-3.09</td>
<td>.074</td>
<td>.207</td>
<td>.266</td>
</tr>
<tr>
<td>3</td>
<td>2.4</td>
<td>14.08</td>
<td>5</td>
<td>.0151</td>
<td>17.36</td>
<td>5</td>
<td>.0038</td>
<td>-3.28</td>
<td>.085</td>
<td>.172</td>
<td>.303</td>
</tr>
<tr>
<td>3</td>
<td>2.3</td>
<td>18.11</td>
<td>5</td>
<td>.0028</td>
<td>27.62</td>
<td>5</td>
<td>4.33x10^{-4}</td>
<td>-9.51</td>
<td>.097</td>
<td>.133</td>
<td>.343</td>
</tr>
<tr>
<td>3</td>
<td>2.2</td>
<td>22.80</td>
<td>5</td>
<td>.369x10^{-3}</td>
<td>35.83</td>
<td>5</td>
<td>1.07x10^{-6}</td>
<td>-13.03</td>
<td>.109</td>
<td>.092</td>
<td>.386</td>
</tr>
<tr>
<td>3</td>
<td>2.1</td>
<td>28.03</td>
<td>5</td>
<td>.341x10^{-4}</td>
<td>43.11</td>
<td>5</td>
<td>1.79x10^{-7}</td>
<td>-14.98</td>
<td>.121</td>
<td>.048</td>
<td>.431</td>
</tr>
<tr>
<td>3</td>
<td>2.0</td>
<td>34.05</td>
<td>5</td>
<td>.241x10^{-4}</td>
<td>50.15</td>
<td>5</td>
<td>5.96x10^{-8}</td>
<td>-16.10</td>
<td>.133</td>
<td>0.000</td>
<td>.479</td>
</tr>
</tbody>
</table>

**TABLE 4.1** Results using a beta (4,4) model.
Fig. 4.1. Relationship of $\eta_1$ and $\varepsilon$. 
Fig. 4.2. Relationship of $x_{ps}^2$ and $s$. 
Fig. 4.3. Relationship of $\chi^2_{ps}$ and $\eta_1$. 
Fig. 4.4. Relationship of $\chi^2_{ps}$ and $\eta_2$. 
Fig. 4.5. Relationship of $s$ and $\eta_2$. 
Fig. 4.6. Relationship of \((x_{ps} - x_s)^2\) and \(\eta_2\).
Fig. 4.7. Relationship of $\chi^2_{ps}$ and $\chi^2_s$. 

regression line
Fig. 4.8. Relationship of \( \chi_s^2 \) and \( \delta \).
5.0 ACKNOWLEDGEMENTS

The author wishes to express his sincerest gratitude to Dr. Doris Grosh for her patient and considerate guidance throughout the duration of this investigation, and in the preparation of this report. Thanks are also extended to Shelly Kemnitz for her excellent and prompt typing of the manuscript. Gratitude is extended to the Industrial Engineering Department for their financial support of this investigation.
6.0 REFERENCES


APPENDIX I

This appendix contains the computer program used to compute \( \eta_1, \eta_2, \delta_1 \) and \( \chi^2 \) for a pair of beta distributions. Also included is an explanation of the symbols used in the program.
SYMBOLS USED IN PROGRAM

ALPHA(L), BETA(L)  input parameters, a and b, of the beta distributions. L=1 (model), L=2 (alternative).

CC, CD CE  symbols used in the calculation of

\[ B(a,b) = \frac{r(a) P(b)}{r(a+b)} \]

PMU(I)  mean of \( I^{th} \) distribution, I=1 (model), I=2 (alternative).

SIG(I)  std. deviation \( I^{th} \) distribution, I=1 (model), I=2 (alternative).

CON  \( B(a,b) \)

DELTA  index of non-congruity

DEV(L,I)  midpoint of class intervals for the \( L^{th} \) distribution, \( I^{th} \) class

DENS(L,I)  ordinates at the DEV(L,I) points.

PRBEL(L,I)  probability element for the \( L^{th} \) distribution, \( I^{th} \) class.

PRBE(L,I,M)  frequency at the \( I^{th} \) class for sample size M.

PRB(I,J,M)  "corrected" frequency taking into consideration the requirement that \( E_{I,J} \geq 5 \) for the \( J^{th} \) class and sample size M.

CHSQ(M)  the Chi-square statistic for sample size M.

J-1  degrees of freedom after computing the "effective" number of classes to satisfy the requirement that \( E_{I,J} \geq 5 \).

PROBAB(M)  right hand tail area of suitably chosen chi-square distribution for sample size M.
DIMENSION PROB(2), PRO(2), CC(2), CD(2), CF(2), A(2), PMU(2), SIG(2), XPROBA(2,2), XDEV(2,10), PRBEL(2,10), DENS(2,10), PRBE(2,10,500), XPRBI(2,10,500), CHSQ(500), PROBAI(500)
COMMUN ALPHA(2), BETA(2), CON
EXTERNAL FCT

DO 2 I=1,2
PROB(I)=0.0
PRO(I)=0.0
CC(I)=0.0
CD(I)=0.0
CF(I)=0.0
A(I)=0.0
PMU(I)=0.0
2 SIG(I)=0.0
DO 3 J=1,2
4 PKUBAI(J,K)=0.0
5 DO 9 L=1,10
6 DEV(J,L)=0.0
7 PRBEL(J,L)=0.0
8 DENS(J,L)=0.0
9 CONTINUE
10 CONTINUE
11 DO 7 N=1,500
12 CHSQ(N)=0.0
13 CONTINUE

READ IN VALUES OF THE PARAMETERS ALPHA AND BETA OF THE TWO DISTRIBUTIONS
THE CLASS WIDTH DINGA, THE NUMBER OF CLASS INTERVALS I NUM, THE MAXIMUM SAMPLE
SIZE MMX, AND THE STEP SIZE FOR COMPUTATIONS N
1 FORMAT (5F10.4,E310)

DO 102 I=1,2
XX=ALPHA(I)
CALL GAMMA (XX, XX, XER)
CC(I)=XX
XX=META(1)
CALL GAMMA (XX, GX, IER)
CJ(1)=GX
XX=ALPHA(1)+BETA(1)
CALL GAMMA (XX, GX, IER)

102 CV(1)=GX
CLN=1/(CL1*CG(1))/(CG(2)*CD(2))*CE(2)/CE(1)
WRITE (18, 103) CON

103 FORMAT ('1', 'THE VALUE OF THE CONSTANT IS', 2X, F9.4)

C
Determine the number of intersection points

C
AA=ALPHA(1)-ALPHA(2)
BB=BETA(1)-BETA(2)
IF (AA.GT.0.AND.ABS(BB).GT.0) GO TO 21
IF (AB.SQRT(AA).GT.0.AND.BB.EQ.0) GO TO 22
IF (AA.LT.0.AND.BB.GT.0) GO TO 123
IF (AA.GT.0.AND.BB.LT.0) GO TO 124
IF (AA.GT.0.AND.BB.GT.0) GO TO 125

C
Determine the intersection points R11 and R12

C
121 X0=1-CON**(1/BB))
G1 TO 104
122 X0 = CON**(1/AA)
GO TO 104
123 AAA=0.01
B=1.0
EPS=0.01
IEND=30
CALL RTMI (X.F,FCT, B, AAA, EPS, IEND, IER)
R(1)=X
WRITE (6, 107) R(1)
GO TO 115
124 AAA=0.0
B=0.99
EPS=0.01
IEND=30
CALL RTMI (X.F,FCT, B, AAA, EPS, IEND, IER)
R(1)=X
WRITE (6, 107) R(1)
GO TO 115
125 X0=AA/(AAA+BB)
FX=1*X0**2*(ALPHA(1)-ALPHA(2))**2*(1-X0)**2*(BETA(1)-BETA(2))**2*CON
IF (FXO) 104, 104, 105
104 WRITE (6, 107) X0
107 FORMAT ('1', 'THERE IS ONE INTERSECTION POINT', '1', 'THIS POINT IS X=
X*, F7.4)
R(1)=X0
GO TO 115
105 AAA=0.0
B=0.0
EPS=0.01
IEND=20
CALL RTMI (X.F,FCT, A, AAA, EPS, IEND, IER)
R(1)=X
CALL RTMI (X.F,FCT, B, ABB, EPS, IEND, IER)
R(2)=X
WRITE (6, 108) R(1), R(2)
C COMPUTE THE PSEUDO CHI-SQUARE FOR SAMPLES OF SIZE M

WRITE (6,48)
48 FORMAT('1',*T30,'SAMPLE SIZE',*8X,'CHI-SQUARE',*8X,'DEGREES OF FREEDOM
XM',*8X,'ALPHA/T30.111111.11111111.11',*8X,'10(1,-1),8X,18(1,-1),8X,5(1,-1))
DO 130 N=50,MAX,N
131 L=1.2
DO 132 I=1,N
132 PRBE(I,1,M)=PRBE(I,1,1)*M
CONTINUE
J=0
YY=PRBE(2,1,1)
Y=PRBE(1,1,1)
CHSU(M)=0.0
I=1
52 IF(YY.GE.5.0) GO TO 53
YY=YY+PRBE(1,1,1)
YY=YY+PRBE(2,1,1,1)
I=I+1
53 IF(I.EQ.INUM) GO TO 54
GO TO 52
54 I=I+1
IF(I.EQ.INUM) GO TO 54
J=J+1
PRB(1,J,M)=YY
PRB(2,J,M)=YY
YY=PRBE(1,1,1,1,1)
YY=YY+PRBE(2,1,1,1,1)
GO TO 52
56 IF(PRBE(1,1,1,1,1,1_GE.5.0) GO TO 55
J=J+1
PRB(1,J,M)=YY+PRBE(1,1,1,1,1,1)
PRB(2,J,M)=YY+PRBE(2,1,1,1,1,1)
GO TO 56
55 IF(PKB(1,1,1,1,1,1_GE.5.0) GO TO 75
J=J+1
PRB(1,J,M)=YY+PRBE(1,1,1,1,1,1)
PRB(2,J,M)=YY+PRBE(2,1,1,1,1,1)
GO TO 56
75 J=J+1
PRB(1,J,M)=PRBE(1,1,1,1,1)
PRB(2,J,M)=PRBE(2,1,1,1,1)
J=J+1
PRB(1,J,M)=PRBE(1,1,1,1,1)
PRB(2,J,M)=YY+PRBE(2,1,1,1,1)
DO 133 K=1,J
133 CHSU(M)=CHSU(M)+((PRB(2,K,M)-PRB(1,K,M))*1#21/PRB(1,K,M))
DEF J=1.
CHSU=CHSU(M)
108 FORMAT('THERE ARE TWO INTERSECTION POINTS'/0', 'THESE ARE XI')
X1=7.4, X2=1.4
GO TO 120
115 XXX=K(1)
C
C DETERMINE THE INDEX OF NON-CONGRUITY, DELTA
C
DO 110 I=1,2
ALPH=ALPHA(I)
BET=BETA(I)
CALL MOBETA (XXX,ALPH,BET,P,IER)
P=PK(1)
110 PKU(I)=1-P
DEL1=ABS(PKU(I))
DEL2=ABS(PKU(2))
DELTA=DEL1*DEL2
GO TO 300
120 DO 118 J=1,2
XXX=RI(J)
DJ=RI(J)
ALPH=ALPHA(J)
BET=BETA(J)
CALL MOBETA (XXX,ALPH,BET,P,IER)
118 PKU(I,J)=P
DEL1=PKU(I,1)-PKU(1,2)
DEL2=PKU(2,1)-PKU(2,2)
DEL3=PKU(2,1)-PKU(2,2)
DEL4=1-PKU(2,2)
DELTA=ABS(DEL1)+ABS(DEL2)+ABS(DEL3-DEL2)+ABS(DEL5-DEL4)
300 WRITE (6,101) DELTA
101 FORMAT(999,99,99999,99)
C
C DETERMINE PROBABILITY INTERVALS FOR EACH CLASS INTERVAL FOR EACH DISTRIBUTION
C
DO 300 L=1,2
DEV(L,1)=0.0
PKBL(L,1)=0.0
DO 10 I=2,1000
DEV(L,1)=DEV(L,1)+DINCR
10 CONTINUE
I=2
20 XXX=DEV(L,1)
BET=BETA(I)
ALPH=ALPHA(I)
CALL MOBETA (XXX,ALPH,BET,P,IER)
PKBL(L,1)=P-PKBL(L,1-1)
IF (IER.EQ.9) GO TO 30
WRITE (6,23), L, IER
22 FORMAT (1HO,'ERROR IN ITERATION',149, 'BOTH ERROR CODE 15', 149.
1', ' END OF PROGRAM.
')
WRITE (6,23)
23 FORMAT (1HO)
GO TO 200
30 IF (I.EQ.1000) GO TO 301
I=1
GO TO 20
301 CONTINUE
WRITE (6,42) ALPHA(I),BETA(I),ALPHA(2),BETA(2)
Determine the right hand tail area (PROBAB1) of a suitably chosen chi-squared distribution. N denotes the sample size.

\[
\text{PROBAB1} = \Phi(\text{CHIOUTH} - \text{CHISQ}(\text{N})
\]

Determine mean and standard deviation (PMU and SIG respectively) of the two distributions.

\[
\text{DU} = 134 \text{ L} = 1,2 \\
\text{PMU(L)} = \text{ALPHA(L)} + \text{BETA(L)} \\
\text{SIG(L)} = \sqrt{\text{ALPHA(L)} \times \text{BETA(L)}}/((\text{ALPHA(L)} + \text{BETA(L)})^{0.5}) \\
\text{XETA(L)} = 1.11) \\
\text{WRITE}(6,57) \text{M}, \text{LHSQUM1}, \text{OFPROBAB1} \\
\text{IF}(\text{PROBAB1} .LT. 0.0) \text{GO TO 200} \\
57 \text{FORMAT('0.1,13.1,13x,E13.4,13x,F5.2,13x,E13.4')} \\
130 \text{CONTINUE}
\]

Compute the values of the parameters ETA(1) and ETA(2).

\[
\text{DU} = \text{SIG(1)} - \text{SIG(2)} \\
\text{E1} = \text{ABS(1)} + \text{ABS(UDU)} \\
\text{E2} = (\text{PMU(L)}/\text{SIG(1)} - \text{PMU(L)}/\text{SIG(2)}) \\
\text{DUDD} = \text{UDU}
\]

Print results.

\[
\text{WRITE}(6,98) \\
98 \text{FORMAT('1,1,T15,1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1',1'
ESTIMATING STATISTICALLY SIGNIFICANT DIFFERENCES
BETWEEN A PAIR OF BETA DISTRIBUTIONS

by

KRISHNAN LAKSHMINARAYAN

B. Tech. (Mechanical Engineering)
Indian Institute of Technology, Madras, 1974

AN ABSTRACT OF A MASTER'S REPORT
submitted in partial fulfillment of the
requirements of the degree

MASTER OF SCIENCE
Department of Industrial Engineering
KANSAS STATE UNIVERSITY
Manhattan, Kansas

1980
ABSTRACT

Computational procedures have been developed for estimating measures of separation between a pair of beta distributions. These measures were variously defined as:

(a) the amount of non-congruity of the two density curves

(b) the measure of separation of the two means and standard deviations and

(c) the discrepancy of the \(\mu/\sigma\) ratios.

They have been compared with each other and with a \(\chi^2\) goodness-of-fit test for a specific group of model and alternative distributions. The results obtained were not conclusive, i.e., the measures of separation developed were shown to be only moderately good for the particular cases studied, and are probably not promising enough to warrant further investigations.