1969-12

A Comparison of Various Non-Parametric Discriminating Procedures When the Populations Are Bivariate Exponentials

Lieberman, Jay Edward

Monterey, California. Naval Postgraduate School

[redacted]

United States
Naval Postgraduate School

THESIS

A COMPARISON OF VARIOUS
NON-PARAMETRIC DISCRIMINATING PROCEDURES
WHEN THE POPULATIONS ARE BIVARIATE EXPONENTIALS

by

Jay Edward Lieberman

December 1969

November 7, 2017

SUBJECT: Change in distribution statement for *A Comparison of Various Non-Parametric Discriminating Procedures When the Populations Are Bivariate Exponentials* – December 1969.

1. Reference: Lieberman, Jay Edward. *A Comparison of Various Non-Parametric Discriminating Procedures When the Populations Are Bivariate Exponentials*. Monterey, CA: Naval Postgraduate School, December 1969. UNCLASSIFIED [Approved for public release, distribution is unlimited. Internally Distributed Report]

2. Upon consultation with NPS faculty, the School has determined that this thesis may distributed externally with no limitations effective November 7, 2017.

University Librarian
Naval Postgraduate School

THIS PAGE INTENTIONALLY LEFT BLANK

[redacted]

[redacted]

A Comparison of Various Non-Parametric Discriminating Procedures When the Populations Are Bivariate Exponentials

by

Jay Edward Lieberman
Lieutenant (junior grade), United States Navy
B.S.I.E., University of Michigan, 1968

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 1969

Author [redacted]

Approved by [redacted]
Thesis Advisor

[redacted]
Chairman, Department of Operations Analysis

[redacted]
Academic Dean

ABSTRACT

A comparison of the error probabilities for various discriminating rules is performed in the two population case when nothing is known of the populations other than they are bivariate negative exponential. In most cases, the absolute difference between the error probabilities for each function was very small. However, the Euclidean distance function consistently performed as well as, and sometimes superior to any of the others studied in this thesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

ACKNOWLEDGMENTS

I wish to thank Dr. J. R. Borsting for suggesting the topic of this thesis and his guidance during its preparation. I am very grateful to Dr. Rex Schudde for his constant aid and assistance in the computer programming. I wish also to thank Professor Ralph Disney, University of Michigan, for the encouragement to seek this degree.

I. INTRODUCTION

The discrimination problem arises when one has several measurements on an individual and it is desired to classify him as belonging to one of a number of categories on the basis of these observations. A decision rule is sought that assigns him to a population in some optimum fashion.

For purposes of this paper, it is assumed that the individual (a random variable Z) is known to be distributed according to one or two multivariate distributions F or G. The approach to the classification problem is dependent on the amount of information known about F and G.

(1) The distributions are completely known. In this case we are to determine which population the individual belongs to based on the new observation. We assume that the distributions for each population are completely known. This problem is solved by the Neyman Pearson Lemma.

(2) F and G are known except for the values of one or more parameters. This case has received little attention except when the assumption of normality is made. An estimation of certain parameters is required and then either a linear discriminant function or a likelihood ratio test is made to classify the observation. When the number of sample points present approaches infinity, these methods are optimal. However, when the sample size is finite, almost nothing is known with respect to an optimal assignment procedure.

(3) F and G are completely unknown. This problem is one of a nonparametric classification of the individual to one of the two populations.

Nothing is assumed about F or G other than their existence. The problem is to assign Z to one of these populations based only on the observations.

The discriminating function used in this thesis will be the rule of the nearest neighbor. This states that the random variable Z will be assigned to population F or G depending upon whichever has the closest sample point. If in fact the distribution of Z is known, we can measure the effectiveness of the non-parametric discriminating rule in terms of the probability of an erroneous assignment. In this thesis a simulation is used to determine error probabilities for various distance functions. The resulting error probabilities are then compared.

II. INDEPENDENT BIVARIATE EXPONENTIALS

If the populations F and G are distributed as independent, bivariate negative exponentials, their densities are as follows:

$$F = \begin{cases} f_{X_1Y_1}(x_1, y_1) = \lambda_1 \lambda_2 e^{-(\lambda_1 x_1 + \lambda_2 y_1)} & x_1, y_1, \lambda_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$G = \begin{cases} f_{X_2Y_2}(x_2, y_2) = \mu_1 \mu_2 e^{-(\mu_1 x_2 + \mu_2 y_2)} & x_2, y_2, \mu_i \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

If independence is assumed between X_1 and Y_1 and between X_2 and Y_2, then

$$F_{X_1}(x_1) = (1 - e^{-\lambda_1 x_1})$$

and $F_{Y_1}(y_1) = (1 - e^{-\mu_1 y_1})$.

The probability integral transformation asserts that F_{X_1} and F_{Y_1} are distributed uniformly on the interval between zero and one. Hence, values of the point (x_1, y_1) can be simulated as,

$$x_1 = -\frac{\ln(RN)}{\lambda_1}$$

$$y_1 = -\frac{\ln(RN)}{\lambda_2} \quad .$$

RN is a uniformly distributed random variable on the interval (0,1).

In the rule of the nearest neighbor, one can use any one of a number of distance functions. It is possible to consider many different geometrical shapes; circles, squares, rectangles, triangles etc. Time constraints have limited the scope of this investigation to circles and symmetrical (about point Z) squares and rectangles.

Distance, D_E, as defined by the circle, is the Euclidean distance from point Z to the sample points present from F and G

$$D_E = \sqrt{(Z_X - X_i)^2 + (Z_Y - Y_i)^2} \quad \forall i \ .$$

Measurement by the square distance function, D_S, is obtained by constructing a square centered at point Z whose sides are parallel to the axes. This distance can be expressed as

$$D_S = \text{MAX}\left\{ \left|\left(X_i - Z_X\right)\right| , \left|\left(Y_i - Z_Y\right)\right| \right\} \quad \forall i \ .$$

The rectangular distance function, D_R, is similar to the square distance function in that it is a rectangle centered at point Z with sides parallel to the axes. The ratio of the vertical sides to the horizontal sides is defined as "A". This distance can be expressed as

$$D_R = \text{MAX}\left\{ A\left| X_i - Z_X \right| , \left| Y_i - Z_Y \right| \right\} \quad \forall i \ .$$

In this investigation A was chosen to equal 2.0.

When all these distances are determined, the minimum is found for each function and Z is assigned to the population to which the nearest point belongs. Since we do know the distribution of Z, we can ascertain if a correct decision was made. If this procedure is performed many times we will be able to estimate the error probabilities for each distance function and then determine if one is superior.

The two types of error, P_1 and P_2 are defined as follows:

P_1 = Pr (Assignment of Z to population 2 | it came from population 1)
P_2 = Pr (Assignment of Z to population 1 | it came from population 2).

Rydzweski [4] began an investigation into this problem, comparing the square distance function with the Euclidean distance function. Excessive computer time to perform numerical integration limited his investigation to only a few parameter values for the square distance function. Simulation was chosen to expand this work because it consumes less computer time than does numerical integration.

In a simulation, it is essential to determine the number of replications to be performed. It was assumed that if the parameters of both distributions were equal, any discriminating function would be indifferent as to which distribution the assignment was made. This is comparable to a Bernoulli trial with p=q=0.5. As a check on the validity of the simulation the program was run several times at 5,000; 10,000; and 20,000 replications. Each time the pseudo uniform (0,1) random number generator (URN) was started at a different position. The following data were obtained for P_1 using the Euclidean discriminating function,

Replications	95% Confidence Interval	Width
5,000	$.4956 \leq \bar{X} \leq .5078$	.0122
10,000	$.4939 \leq \bar{X} \leq .5031$	.0100
20,000	$.4972 \leq \bar{X} \leq .5058$	.0086.

These confidence limits fell well within those of a normal approximation to the binomial at the 0.05 level. It was felt that 10,000 replications gave sufficient accuracy for this thesis.

As a further check on the validity of the simulation, the results are compared with those of Rydzewski's in Table I.

	$\lambda_1 = \lambda_2 = 10$		$\lambda_1 = \lambda_2 = 30$		$\lambda_1 = \lambda_2 = 50$	
	$\mu_1 = \mu_2 = 10$		$\mu_1 = \mu_2 = 10$		$\mu_1 = \mu_2 = 10$	
Sample Size	Sim.	Intg.	Sim.	Intg.	Sim.	Intg.
1	.3402	.3434	.2398	.2387	.1336	.1318
4	.3794	.3820	.2886	.2842	.1722	.1733
8	.4010	.3939	.2972	.2994	.1828	.1862
20	.4000	.4020	.3046	.3012	.1962	.1969

TABLE I

A comparison between simulated error probabilities and those obtained through numerical integration

Lockett [3] has shown that P_1 and P_2 in the univariate case depend on λ and μ only through $C = \lambda/\mu$. Because of independence, this can be extended to the bivariate case, let $C = \lambda_1/\mu_1$, and $D = \lambda_2/\mu_2$. The following tables contain P_1 and P_2 for a wide range of parameters. The arguments in the tables are C, D, and the sample size (N). Values for C are listed horizontally and values of D are listed vertically. If C and D are both less than one, P_1 is found in table A-1 and P_2 is found in table A-2. If both C and D are greater than one, these ratios should be inverted to form C' and D'. P_1 can be located in table A-2 and P_2 in table A-1. If $C \leq 1$ and $D > 1$, P_1 is listed in table B-1 and P_2 in table B-2. If $C \geq 1$ but $D < 1$, these ratios should again be inverted to form C' and D'. P_1 can be located in table B-2 and P_2 in table B-1.

As an example of this procedure, assume $\lambda_1 = \lambda_2 = 20$ and $\mu_1 = \mu_2 = 10$ and the desired sample size is 1. Inverting both ratios, $C' = \frac{1}{2}$ and $D' = \frac{1}{2}$. From table A-2-1 we find $P_1 = 0.341$ and from table A-1-1, $P_2 = 0.550$. Rydzewski also evaluated P_1 and P_2 for these parameters (see Table I) and the simulation concurs favorably.

D \ C	.1	.2	.3	.5	.7	.9	1.0	
	.439	.455	.452	.448	.441	.449	.439	EUC
.1	.421	.442	.440	.443	.438	.447	.438	SQ
	.424	.452	.456	.454	.445	.453	.439	REC
	.458	.502	.513	.511	.501	.419	.493	EUC
.2	.443	.482	.499	.506	.498	.487	.491	SQ
	.440	.484	.508	.511	.508	.489	.492	REC
	.457	.503	.527	.534	.528	.526	.514	EUC
.3	.447	.487	.514	.528	.522	.522	.511	SQ
	.435	.483	.511	.533	.525	.524	.511	REC
	.452	.515	.552	.553	.555	.531	.538	EUC
.5	.447	.508	.538	.550	.549	.526	.536	SQ
	.441	.499	.537	.542	.548	.520	.532	REC
	.438	.506	.534	.552	.547	.533	.531	EUC
.7	.433	.502	.528	.546	.543	.530	.533	SQ
	.432	.495	.522	.539	.539	.524	.524	REC
	.441	.489	.532	.540	.529	.521	.507	EUC
.9	.438	.488	.527	.537	.527	.522	.507	SQ
	.436	.484	.525	.540	.531	.523	.503	REC
	.438	.484	.522	.527	.527	.511	.498	EUC
1.0	.437	.482	.523	.528	.531	.507	.494	SQ
	.436	.481	.521	.532	.524	.512	.493	REC

TABLE A-1

N = 1

D \ C	.1	.2	.3	.5	.7	.9	1.0	
	.044	.072	.099	.122	.137	.145	.142	EUC
.1	.047	.076	.102	.124	.140	.148	.151	SQ
	.053	.090	.116	.146	.166	.170	.178	REC
	.077	.129	.167	.208	.225	.243	.246	EUC
.2	.080	.131	.172	.211	.231	.247	.248	SQ
	.089	.142	.186	.236	.257	.280	.281	REC
	.095	.170	.204	.265	.293	.318	.311	EUC
.3	.096	.175	.212	.269	.301	.321	.315	SQ
	.101	.179	.226	.289	.322	.350	.345	REC
	.127	.205	.268	.337	.385	.397	.415	EUC
.5	.130	.211	.276	.341	.386	.402	.416	SQ
	.131	.215	.275	.350	.385	.417	.436	REC
	.129	.226	.293	.380	.435	.449	.462	EUC
.7	.134	.232	.302	.380	.438	.451	.464	SQ
	.129	.230	.295	.384	.441	.464	.475	REC
	.146	.239	.314	.401	.458	.477	.489	EUC
.9	.149	.243	.319	.402	.462	.476	.492	SQ
	.141	.236	.306	.397	.457	.479	.491	REC
	.151	.248	.318	.415	.461	.491	.501	EUC
1.0	.150	.252	.322	.419	.460	.488	.497	SQ
	.144	.242	.320	.411	.453	.487	.505	REC

TABLE A-2

N = 1

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.237	.277	.293	.294	.306	.297	.295	EUC
.1	.230	.272	.291	.293	.303	.298	.296	SQ
	.235	.280	.303	.308	.319	.309	.305	REC
	.274	.335	.361	.381	.389	.388	.384	EUC
.2	.271	.329	.358	.382	.388	.390	.387	SQ
	.268	.329	.367	.397	.403	.403	.397	REC
	.290	.361	.409	.443	.454	.439	.435	EUC
.3	.287	.356	.403	.438	.451	.436	.435	SQ
	.281	.360	.400	.452	.464	.458	.441	REC
	.299	.385	.441	.489	.496	.488	.487	EUC
.5	.297	.383	.437	.485	.498	.490	.487	SQ
	.291	.380	.433	.487	.498	.493	.498	REC
	.294	.387	.441	.496	.514	.504	.505	EUC
.7	.291	.386	.440	.483	.511	.503	.506	SQ
	.284	.377	.437	.492	.513	.504	.501	REC
	.295	.387	.437	.495	.511	.508	.505	EUC
.9	.296	.387	.434	.496	.508	.505	.507	SQ
	.290	.380	.431	.488	.511	.505	.502	REC
	.296	.368	.434	.495	.504	.506	.497	EUC
1.0	.293	.369	.433	.492	.505	.505	.499	SQ
	.290	.367	.430	.490	.504	.505	.499	REC

TABLE A-1

N = 3

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.063	.097	.126	.166	.177	.179	.186	EUC
.1	.065	.101	.129	.169	.178	.181	.186	SQ
	.070	.108	.137	.174	.193	.190	.194	REC
	.102	.164	.204	.248	.265	.279	.288	EUC
.2	.105	.167	.209	.252	.270	.283	.291	SQ
	.107	.169	.218	.261	.274	.296	.299	REC
	.130	.202	.257	.301	.336	.249	.355	EUC
.3	.134	.202	.254	.303	.342	.354	.362	SQ
	.134	.207	.258	.316	.356	.360	.367	REC
	.158	.251	.303	.385	.409	.433	.433	EUC
.5	.161	.251	.307	.390	.410	.431	.430	SQ
	.163	.252	.307	.388	.423	.436	.436	REC
	.173	.273	.335	.410	.439	.463	.467	EUC
.7	.174	.272	.335	.409	.445	.465	.467	SQ
	.173	.269	.340	.410	.443	.465	.472	REC
	.191	.286	.348	.428	.462	.485	.495	EUC
.9	.192	.289	.353	.432	.465	.481	.499	SQ
	.187	.284	.350	.438	.464	.488	.499	REC
	.175	.296	.355	.432	.475	.506	.499	EUC
1.0	.178	.301	.353	.432	.470	.503	.502	SQ
	.174	.294	.349	.434	.471	.505	.506	REC

TABLE A-2

N = 3

D \ C	.1	.2	.3	.5	.7	.9	1.0	
	.182	.225	.248	.269	.269	.263	.261	EUC
.1	.181	.224	.247	.267	.269	.263	.263	SQ
	.185	.229	.258	.279	.281	.275	.275	REC
	.221	.288	.316	.345	.364	.359	.352	EUC
.2	.217	.283	.320	.346	.364	.358	.357	SQ
	.223	.291	.329	.353	.379	.370	.369	REC
	.236	.327	.366	.404	.426	.425	.410	EUC
.3	.234	.324	.360	.403	.426	.424	.410	SQ
	.234	.326	.373	.413	.435	.434	.434	REC
	.266	.356	.407	.463	.475	.487	.478	EUC
.5	.265	.354	.403	.461	.477	.487	.479	SQ
	.255	.350	.403	.466	.477	.489	.482	REC
	.263	.356	.414	.486	.496	.508	.512	EUC
.7	.263	.355	.411	.484	.494	.504	.507	SQ
	.256	.351	.410	.482	.491	.506	.509	REC
	.260	.359	.425	.479	.507	.511	.500	EUC
.9	.260	.361	.424	.478	.506	.513	.500	SQ
	.252	.355	.419	.475	.504	.508	.498	REC
	.259	.352	.402	.474	.496	.491	.499	EUC
1.0	.260	.359	.406	.474	.498	.493	.501	SQ
	.254	.348	.399	.471	.496	.495	.500	REC

Table A-1

N = 5

C D	.1	.2	.3	.5	.7	.9	1.0	
	.070	.114	.144	.175	.188	.198	.198	EUC
.1	.072	.117	.146	.176	.198	.200	.197	SQ
	.073	.124	.151	.182	.195	.205	.207	REC
	.112	.175	.215	.266	.282	.302	.298	EUC
.2	.115	.179	.220	.268	.284	.304	.297	SQ
	.120	.183	.223	.269	.288	.307	.305	REC
	.137	.222	.263	.328	.341	.369	.369	EUC
.3	.140	.224	.261	.328	.344	.370	.368	SQ
	.142	.229	.260	.332	.347	.377	.369	REC
	.174	.264	.324	.389	.419	.439	.435	EUC
.5	.178	.265	.325	.391	.420	.437	.436	SQ
	.175	.262	.325	.389	.423	.442	.437	REC
	.185	.281	.349	.420	.457	.480	.476	EUC
.7	.187	.288	.355	.418	.458	.481	.479	SQ
	.186	.281	.350	.424	.461	.487	.479	REC
	.190	.297	.359	.432	.476	.496	.495	EUC
.9	.193	.300	.362	.434	.474	.493	.497	SQ
	.192	.300	.358	.431	.473	.492	.494	REC
	.201	.300	.359	.436	.479	.491	.498	EUC
1.0	.201	.302	.365	.434	.477	.493	.500	SQ
	.203	.305	.365	.435	.478	.493	.499	REC

TABLE A-2

N = 5

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.149	.194	.210	.243	.243	.244	.238	EUC
.1	.149	.190	.208	.245	.242	.245	.238	SQ
	.149	.194	.214	.252	.257	.256	.248	REC
	.183	.253	.285	.338	.350	.339	.340	EUC
.2	.184	.251	.285	.340	.347	.340	.343	SQ
	.184	.251	.291	.344	.353	.350	.352	REC
	.213	.294	.339	.386	.399	.401	.406	EUC
.3	.213	.293	.334	.387	.398	.404	.404	SQ
	.211	.294	.338	.394	.410	.412	.412	REC
	.230	.330	.386	.450	.478	.476	.482	EUC
.5	.230	.332	.384	.448	.475	.478	.484	SQ
	.226	.321	.380	.452	.476	.480	.486	REC
	.240	.345	.404	.472	.499	.502	.506	EUC
.7	.239	.342	.402	.469	.497	.504	.504	SQ
	.231	.337	.395	.465	.499	.503	.504	REC
	.237	.336	.407	.476	.504	.512	.517	EUC
.9	.234	.338	.407	.477	.504	.512	.516	SQ
	.233	.327	.402	.469	.501	.509	.512	REC
	.236	.340	.398	.470	.507	.506	.500	EUC
1.0	.236	.337	.401	.470	.507	.510	.501	SQ
	.231	.336	.394	.467	.505	.502	.500	REC

TABLE A-1

N = 10

D \ C	.1	.2	.3	.5	.7	.9	1.0	
	.076	.129	.158	.184	.193	.210	.219	EUC
.1	.070	.128	.160	.183	.194	.210	.219	SQ
	.081	.131	.161	.188	.196	.216	.219	REC
	.122	.189	.224	.277	.298	.317	.306	EUC
.2	.124	.189	.227	.279	.298	.322	.210	SQ
	.127	.190	.230	.282	.303	.316	.314	REC
	.150	.230	.278	.334	.355	.370	.377	EUC
.3	.151	.230	.280	.335	.353	.374	.378	SQ
	.154	.235	.280	.333	.358	.370	.380	REC
	.181	.277	.336	.402	.432	.448	.446	EUC
.5	.182	.277	.341	.400	.434	.449	.445	SQ
	.183	.281	.336	.404	.430	.448	.443	REC
	.200	.293	.363	.431	.470	.485	.486	EUC
.7	.200	.293	.363	.430	.469	.486	.485	SQ
	.202	.299	.360	.426	.462	.481	.480	REC
	.209	.309	.378	.450	.489	.497	.496	EUC
.9	.213	.309	.379	.454	.487	.499	.496	SQ
	.213	.312	.373	.453	.487	.498	.492	REC
	.212	.309	.381	.442	.482	.514	.500	EUC
1.0	.213	.310	.378	.449	.484	.511	.499	SQ
	.216	.310	.374	.442	.481	.504	.498	REC

TABLE A-2

N = 10

D \ C	.1	.2	.3	.5	.7	.9	1.0	
	.122	.166	.195	.221	.229	.235	.228	EUC
.1	.121	.166	.193	.221	.230	.239	.229	SQ
	.124	.168	.199	.227	.237	.243	.235	REC
	.167	.228	.264	.314	.325	.331	.330	EUC
.2	.167	.225	.266	.316	.325	.332	.328	SQ
	.164	.229	.270	.320	.332	.342	.336	REC
	.193	.265	.315	.365	.389	.388	.394	EUC
.3	.192	.266	.314	.367	.385	.394	.394	SQ
	.190	.262	.315	.372	.394	.399	.402	REC
	.217	.299	.371	.431	.462	.467	.458	EUC
.5	.218	.300	.370	.435	.464	.466	.456	SQ
	.217	.301	.369	.430	.462	.468	.466	REC
	.230	.319	.392	.451	.483	.497	.491	EUC
.7	.229	.320	.389	.453	.481	.500	.491	SQ
	.224	.320	.384	.447	.484	.499	.499	REC
	.229	.326	.399	.452	.486	.505	.502	EUC
.9	.230	.326	.397	.454	.492	.505	.501	SQ
	.226	.322	.391	.451	.492	.501	.498	REC
	.228	.324	.398	.470	.482	.510	.511	EUC
1.0	.227	.323	.398	.472	.487	.509	.512	SQ
	.226	.323	.393	.470	.478	.511	.508	REC

TABLE A-1

N = 20

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.090	.134	.159	.195	.211	.215	.216	EUC
.1	.091	.136	.163	.197	.215	.220	.215	SQ
	.093	.134	.165	.195	.215	.216	.215	REC
	.128	.198	.204	.281	.309	.318	.317	EUC
.2	.130	.197	.242	.279	.312	.318	.316	SQ
	.129	.195	.243	.286	.313	.314	.313	REC
	.159	.233	.289	.343	.372	.380	.378	EUC
.3	.162	.236	.289	.345	.373	.386	.377	SQ
	.160	.235	.292	.340	.374	.383	.379	REC
	.194	.286	.341	.403	.440	.452	.460	EUC
.5	.193	.285	.345	.401	.440	.455	.457	SQ
	.194	.286	.339	.401	.439	.451	.457	REC
	.206	.307	.362	.432	.472	.485	.481	EUC
.7	.206	.308	.366	.435	.471	.486	.482	SQ
	.206	.303	.362	.435	.467	.487	.483	REC
	.215	.308	.374	.444	.482	.448	.496	EUC
.9	.216	.310	.381	.444	.487	.499	.496	SQ
	.213	.311	.384	.443	.484	.498	.503	REC
	.218	.313	.376	.452	.489	.498	.502	EUC
1.0	.219	.314	.381	.447	.491	.500	.502	SQ
	.218	.314	.382	.448	.486	.490	.509	REC

TABLE A-2

N = 20

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.427	.455	.479	.483	.458	.436	.425	EUC
2.	.433	.464	.499	.493	.467	.477	.434	SQ
	.436	.476	.482	.517	.491	.472	.457	REC
	.419	.448	.462	.446	.431	.397	.391	EUC
3.	.427	.461	.479	.461	.449	.412	.406	SQ
	.434	.475	.505	.497	.488	.458	.448	REC
	.412	.430	.432	.434	.391	.365	.344	EUC
5.	.420	.446	.456	.448	.412	.388	.368	SQ
	.427	.466	.486	.443	.470	.451	.424	REC
	.414	.435	.445	.402	.390	.338	.322	EUC
7.	.425	.458	.469	.431	.417	.365	.349	SQ
	.434	.479	.502	.487	.474	.432	.415	REC
	.410	.427	.432	.405	.368	.329	.309	EUC
9.	.423	.447	.457	.437	.397	.358	.340	SQ
	.430	.471	.494	.486	.462	.435	.413	REC
	.402	.431	.437	.402	.362	.322	.309	EUC
10.	.414	.450	.461	.433	.395	.351	.366	SQ
	.423	.474	.498	.486	.467	.429	.412	REC
20.	.406	.426	.430	.380	.345	.307	.290	EUC
	.417	.448	.458	.425	.382	.341	.323	SQ
	.426	.474	.497	.483	.453	.423	.406	REC

TABLE B-1

N = 1

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.154	.260	.344	.426	.488	.523	.532	EUC
2.	.154	.252	.346	.425	.488	.522	.531	SQ
	.146	.245	.327	.411	.477	.512	.519	REC
	.149	.253	.331	.422	.473	.505	.514	EUC
3.	.151	.252	.331	.424	.468	.502	.510	SQ
	.143	.240	.317	.407	.465	.497	.500	REC
	.141	.250	.316	.407	.462	.488	.502	EUC
5.	.140	.249	.312	.400	.456	.478	.491	SQ
	.133	.240	.307	.396	.462	.484	.492	REC
	.151	.244	.317	.408	.466	.474	.495	EUC
7.	.150	.243	.314	.398	.456	.464	.481	SQ
	.144	.241	.308	.399	.453	.473	.491	REC
	.142	.234	.306	.392	.443	.468	.486	EUC
9.	.140	.232	.294	.383	.431	.460	.474	SQ
	.138	.231	.300	.387	.442	.474	.489	REC
	.140	.243	.306	.396	.439	.469	.472	EUC
10.	.138	.237	.300	.392	.425	.461	.462	SQ
	.136	.237	.297	.402	.441	.472	.483	REC
	.141	.233	.301	.384	.433	.452	.458	EUC
20.	.134	.227	.291	.369	.417	.437	.444	SQ
	.136	.228	.299	.383	.436	.465	.471	REC

TABLE B-2

N = 1

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.269	.353	.383	.436	.452	.448	.434	EUC
2.	.271	.356	.396	.443	.455	.451	.438	SQ
	.276	.361	.405	.450	.463	.460	.488	REC
	.264	.325	.366	.396	.407	.393	.387	EUC
3.	.270	.336	.322	.403	.414	.396	.391	SQ
	.275	.352	.392	.425	.437	.421	.416	REC
	.252	.311	.327	.340	.339	.334	.320	EUC
5.	.262	.324	.343	.357	.353	.349	.331	SQ
	.273	.348	.366	.393	.395	.390	.370	REC
	.252	.288	.306	.326	.316	.299	.289	EUC
7.	.262	.302	.323	.348	.333	.317	.304	SQ
	.275	.331	.364	.394	.390	.367	.358	REC
	.251	.297	.303	.308	.295	.280	.272	EUC
9.	.259	.312	.321	.329	.315	.297	.288	SQ
	.273	.341	.359	.385	.379	.352	.347	REC
	.238	.292	.305	.310	.297	.275	.266	EUC
10.	.250	.306	.323	.331	.320	.293	.285	SQ
	.263	.337	.367	.388	.383	.355	.345	REC
	.240	.256	.280	.277	.260	.244	.231	EUC
20.	.252	.298	.306	.307	.287	.267	.256	SQ
	.268	.332	.357	.374	.358	.340	.337	REC

TABLE B-1

N = 3

D \ C	.1	.2	.3	.5	.7	.9	1.0	
	.197	.298	.359	.437	.483	.499	.489	EUC
2.	.197	.296	.361	.439	.479	.499	.491	SQ
	.193	.297	.358	.439	.480	.497	.494	REC
	.187	.281	.342	.415	.460	.465	.474	EUC
3.	.187	.277	.344	.411	.454	.462	.472	SQ
	.185	.282	.350	.419	.465	.480	.487	REC
	.182	.271	.323	.385	.408	.431	.431	EUC
5.	.180	.264	.320	.382	.403	.433	.427	SQ
	.184	.227	.336	.401	.431	.451	.455	REC
	.162	.263	.304	.368	.389	.412	.406	EUC
7.	.160	.257	.299	.362	.384	.402	.400	SQ
	.168	.272	.315	.386	.415	.438	.434	REC
	.164	.244	.301	.352	.380	.385	.395	EUC
9.	.161	.238	.294	.346	.372	.380	.390	SQ
	.170	.261	.216	.384	.407	.423	.426	REC
10.	.168	.252	.300	.355	.380	.384	.386	EUC
	.162	.248	.294	.347	.371	.376	.375	SQ
	.173	.264	.212	.326	.402	.421	.420	REC
	.156	.234	.272	.322	.349	.355	.355	EUC
20.	.150	.231	.266	.311	.337	.346	.347	SQ
	.163	.250	.294	.352	.381	.392	.393	REC

TABLE B-2

N = 3

D \ C	.1	.2	.3	.5	.7	.9	1.0	
	.237	.309	.380	.423	.441	.440	.432	EUC
2.	.238	.311	.382	.429	.441	.442	.438	SQ
	.241	.315	.389	.432	.445	.453	.443	REC
	.227	.293	.334	.378	.392	.384	.381	EUC
3.	.231	.299	.339	.382	.394	.386	.381	SQ
	.240	.307	.358	.401	.411	.400	.399	REC
	.216	.268	.291	.322	.326	.319	.314	EUC
5.	.221	.278	.300	.333	.333	.328	.323	SQ
	.233	.304	.329	.367	.371	.362	.348	REC
	.197	.256	.275	.286	.283	.286	.279	EUC
7.	.208	.267	.288	.302	.307	.298	.291	SQ
	.224	.298	.328	.347	.353	.336	.328	REC
	.204	.245	.261	.275	.271	.259	.259	EUC
9.	.214	.258	.277	.292	.285	.274	.271	SQ
	.231	.289	.322	.345	.337	.323	.314	REC
	.200	.245	.253	.270	.258	.253	.243	EUC
10.	.210	.258	.271	.287	.278	.267	.260	SQ
	.228	.292	.311	.343	.388	.325	.311	REC
	.187	.233	.237	.230	.229	.208	.202	EUC
20.	.200	.251	.260	.256	.252	.230	.223	SQ
	.220	.287	.313	.323	.318	.294	.286	REC

TABLE B-1

N = 5

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.201	.298	.369	.431	.464	.482	.484	EUC
2.	.202	.297	.367	.429	.462	.485	.483	SQ
	.201	.299	.380	.436	.463	.492	.490	REC
	.193	.292	.348	.412	.442	.446	.456	EUC
3.	.194	.290	.349	.405	.442	.444	.453	SQ
	.197	.300	.360	.422	.458	.460	.471	REC
	.183	.267	.319	.360	.391	.393	.404	EUC
5.	.180	.265	.319	.358	.387	.392	.396	SQ
	.188	.281	.334	.382	.419	.424	.424	REC
	.174	.265	.296	.338	.362	.364	.373	EUC
7.	.172	.259	.267	.335	.356	.364	.366	SQ
	.178	.279	.313	.364	.392	.405	.404	REC
	.167	.252	.291	.336	.356	.352	.356	EUC
9.	.165	.245	.288	.328	.352	.346	.356	SQ
	.175	.264	.315	.365	.384	.385	.394	REC
	.164	.243	.285	.327	.339	.348	.350	EUC
10.	.161	.234	.279	.317	.340	.346	.349	SQ
	.173	.257	.309	.353	.377	.394	.386	REC
	.160	.237	.259	.290	.307	.308	.314	EUC
20.	.156	.232	.256	.284	.301	.303	.305	SQ
	.167	.258	.283	.327	.352	.360	.358	REC

TABLE B-2

N = 5

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.214	.300	.352	.424	.448	.450	.456	EUC
2.	.214	.304	.357	.426	.448	.453	.458	SQ
	.200	.302	.357	.422	.443	.453	.453	REC
	.200	.276	.321	.369	.392	.387	.389	EUC
3.	.203	.279	.323	.371	.393	.390	.390	SQ
	.211	.291	.329	.377	.402	.393	.396	REC
	.181	.243	.269	.309	.322	.312	.323	EUC
5.	.186	.249	.276	.313	.326	.315	.324	SQ
	.200	.268	.298	.333	.350	.335	.345	REC
	.172	.215	.254	.271	.275	.276	.274	EUC
7.	.178	.240	.258	.277	.281	.281	.280	SQ
	.196	.249	.289	.308	.309	.308	.302	REC
	.167	.207	.227	.247	.252	.242	.249	EUC
9.	.174	.218	.235	.257	.262	.249	.255	SQ
	.193	.249	.272	.293	.294	.284	.287	REC
	.164	.203	.220	.238	.240	.231	.238	EUC
10.	.171	.214	.230	.248	.253	.238	.245	SQ
	.192	.246	.269	.288	.292	.275	.280	REC
	.152	.178	.188	.192	.187	.177	.178	EUC
20.	.164	.194	.203	.208	.206	.191	.189	SQ
	.188	.237	.253	.265	.261	.242	.243	REC

TABLE B-1

N = 10

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.206	.300	.364	.430	.459	.475	.467	EUC
2.	.205	.299	.362	.429	.463	.474	.467	SQ
	.213	.304	.377	.440	.472	.481	.477	REC
	.200	.289	.342	.391	.421	.439	.432	EUC
3.	.197	.288	.340	.384	.420	.438	.431	SQ
	.201	.295	.352	.400	.438	.448	.447	REC
	.187	.263	.303	.346	.367	.383	.376	EUC
5.	.182	.363	.304	.343	.367	.379	.376	SQ
	.189	.273	.320	.366	.394	.401	.400	REC
	.174	.241	.282	.323	.344	.342	.335	EUC
7.	.172	.240	.280	.321	.344	.343	.337	SQ
	.182	.259	.302	.357	.372	.371	.366	REC
	.173	.236	.271	.302	.319	.322	.315	EUC
9.	.168	.233	.270	.298	.315	.320	.314	SQ
	.179	.252	.294	.330	.353	.352	.351	REC
	.170	.232	.266	.294	.309	.312	.306	EUC
10.	.164	.228	.263	.290	.306	.311	.303	SQ
	.177	.248	.288	.324	.347	.345	.344	REC
	.156	.212	.241	.261	.270	.275	.264	EUC
20.	.153	.209	.236	.254	.270	.275	.264	SQ
	.167	.231	.267	.296	.313	.311	.307	REC

TABLE B-2

N = 10

C / D	.1	.2	.3	.5	.7	.9	1.0	
	.206	.299	.344	.398	.449	.457	.443	EUC
2.	.204	.298	.343	.402	.449	.456	.443	SQ
	.210	.296	.341	.401	.447	.448	.451	REC
	.192	.260	.315	.361	.384	.392	.399	EUC
3.	.194	.259	.316	.362	.385	.394	.399	SQ
	.197	.266	.318	.372	.388	.389	.401	REC
	.171	.219	.263	.299	.308	.316	.319	EUC
5.	.173	.222	.265	.302	.313	.321	.320	SQ
	.183	.236	.273	.313	.321	.324	.327	REC
	.156	.195	.229	.253	.265	.264	.269	EUC
7.	.160	.200	.234	.261	.269	.268	.271	SQ
	.176	.218	.251	.278	.284	.278	.282	REC
	.148	.178	.206	.224	.231	.234	.235	EUC
9.	.153	.181	.211	.230	.236	.234	.241	SQ
	.170	.207	.237	.253	.257	.256	.252	REC
	.145	.172	.197	.214	.218	.220	.225	EUC
10.	.150	.176	.202	.219	.222	.224	.289	SQ
	.168	.203	.231	.245	.245	.246	.241	REC
	.129	.139	.152	.153	.153	.156	.156	EUC
20.	.135	.149	.161	.163	.163	.163	.163	SQ
	.161	.186	.206	.204	.202	.196	.192	REC

TABLE B-1

N = 20

D \ C	.1	.2	.3	.5	.7	.9	1.0	
	.206	.297	.360	.425	.443	.467	.472	EUC
2.	.202	.296	.362	.424	.442	.474	.469	SQ
	.211	.304	.369	.432	.454	.471	.473	REC
	.205	.276	.324	.373	.407	.411	.418	EUC
3.	.204	.277	.322	.372	.406	.413	.420	SQ
	.210	.286	.333	.383	.417	.424	.425	REC
	.190	.251	.285	.319	.340	.356	.353	EUC
5.	.188	.248	.285	.320	.342	.355	.356	SQ
	.195	.263	.300	.336	.363	.366	.376	REC
	.178	.229	.257	.287	.305	.314	.314	EUC
7.	.175	.228	.258	.286	.305	.313	.314	SQ
	.184	.245	.282	.308	.333	.335	.335	REC
	.169	.215	.240	.264	.283	.288	.285	EUC
9.	.166	.212	.240	.263	.282	.288	.283	SQ
	.177	.234	.267	.289	.313	.312	.314	REC
	.164	.209	.234	.253	.272	.279	.275	EUC
10.	.162	.207	.234	.255	.271	.278	.273	SQ
	.174	.229	.261	.282	.305	.304	.306	REC
	.149	.182	.205	.218	.223	.231	.224	EUC
20.	.147	.181	.202	.216	.222	.230	.222	SQ
	.163	.209	.232	.248	.262	.264	.263	REC

TABLE B-2

N = 20

III. DEPENDENT BIVARIATE EXPONENTIALS

Any dependent bivariate distribution, $F_{X,Y}$, can be simulated by means of the marginal, F_X, and conditional, $F_{X|Y}$, distributions. The first step is to solve for one value through either marginal distribution and the probability integral transformation. Then use this as the argument in the corresponding conditional distribution and use the probability integral transformation again. This procedure will yield a point from the distribution $F_{X,Y}$.

Gumble [2] states that given two probability functions F_X and F_Y, a bivariate distribution can be constructed by means of the equation

$$F_{X,Y}(x,y) = F_X(x)\, F_Y(y) \left[1 + \alpha \left\{ 1 - F_X(x) \right\} \left\{ 1 - F_Y(y) \right\} \right]$$

$$-1 \le \alpha \le 1 .$$

If the marginal distributions are exponential, the equation becomes

$$F_{X,Y}(x,y) = \begin{cases} (1-e^{-x})\,(1-e^{-y})\,[1 + \alpha\, e^{-x-y}] & x,\ y \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

The marginal distribution of Y is clearly $(1-e^{-y})$. The conditional distribution of X given Y can be found by integrating the given conditional density function $f_{X|Y}$ between zero and x,

$$F_{X,Y}(x|y) = \int_0^x \left\{ e^{-x}(1 + \alpha - 2\,\alpha e^{-y}) - 2\,\alpha e^{-2x}\,(1 - 2\,e^{-y}) \right\} dx .$$

Setting the above equal to a random number (RN) we obtain,

$$(\alpha - 2\,\alpha\, e^{-y})e^{-2x} - (1 + \alpha - 2\,\alpha e^{-y})\, e^{-x} - RN = 0 .$$

Solving by the quadratic formula,

$$e^{-x} = \frac{(1 + \alpha - 2\,\alpha e^{-y}) \pm \sqrt{(1 + \alpha - 2\,\alpha e^{-y})^2 - 4\,(\alpha - 2\,\alpha e^{-y})\,RN}}{2\,(\alpha - 2\,\alpha e^{-y})} \, .$$

The value of the random variable X can be determined by taking the natural logarithm of both sides. The negative root is chosen so that the expression will lie on the interval (0,1) and hence X will be real and greater than or equal to zero.

The correlation between X and Y in this distribution is given by the expression $\rho = \frac{\alpha}{4}$. Hence we have the highest positive correlation between variables when $\alpha = +1$ and the greatest negative correlation when $\alpha = -1$. The correlation, however, varies only within the narrow limits $-0.25 \le \rho \le +0.25$.

This distribution was simulated to determine error probabilities as in the independent case but with the addition of a fourth distance function. The rectangular distance function was rotated 45 degrees clockwise around point Z. This was accomplished by a rotation of the coordinate axes and a corresponding adjustment to the data points. The values for P_1 and P_2 are listed in tables D-1 and D-2. The arguments for entering the tables are alpha from the first distribution and beta from the second.

β \ α		-1.0	-.50	0.0	+.50	+1.0
-1.0	EUC	.505	.509	.496	.494	.482
	SQ	.503	.505	.498	.496	.482
	REC	.502	.505	.497	.491	.478
	ROT REC	.497	.504	.492	.483	.466
- .5	EUC	.496	.496	.498	.490	.483
	SQ	.498	.499	.497	.491	.479
	REC	.497	.504	.505	.489	.476
	ROT REC	.499	.494	.492	.482	.475
0	EUC	.493	.499	.493	.505	.488
	SQ	.494	.500	.494	.503	.489
	REC	.493	.500	.497	.504	.487
	ROT REC	.499	.503	.498	.495	.487
+ .5	EUC	.486	.496	.505	.505	.496
	SQ	.488	.496	.502	.506	.497
	REC	.491	.494	.501	.502	.501
	ROT REC	.503	.503	.505	.499	.491
+1.0	EUC	.498	.499	.497	.498	.494
	SQ	.502	.499	.497	.499	.496
	REC	.504	.495	.504	.500	.494
	ROT REC	.509	.512	.504	.504	.496

TABLE D-1

N = 2

P_1 given alpha and beta for dependent bivariate exponentials

β \ α		-1.0	-.50	0.0	+.50	+1.0
-1.0	EUC	.504	.500	.497	.488	.491
	SQ	.505	.501	.500	.491	.485
	REC	.507	.506	.499	.488	.491
	ROT REC	.504	.507	.500	.501	.503
- .5	EUC	.499	.500	.504	.495	.501
	SQ	.502	.501	.505	.501	.502
	REC	.500	.493	.505	.499	.498
	ROT REC	.499	.499	.507	.504	.507
0	EUC	.496	.496	.506	.500	.503
	SQ	.494	.496	.509	.500	.503
	REC	.498	.496	.506	.506	.510
	ROT REC	.482	.493	.502	.502	.508
+.5	EUC	.486	.490	.502	.503	.505
	SQ	.488	.490	.500	.502	.505
	REC	.495	.493	.500	.503	.510
	ROT REC	.476	.486	.502	.506	.509
+1.0	EUC	.481	.495	.489	.508	.505
	SQ	.479	.494	.489	.503	.509
	REC	.483	.494	.484	.500	.503
	ROT REC	.466	.483	.486	.491	.504

TABLE D-2

N = 2

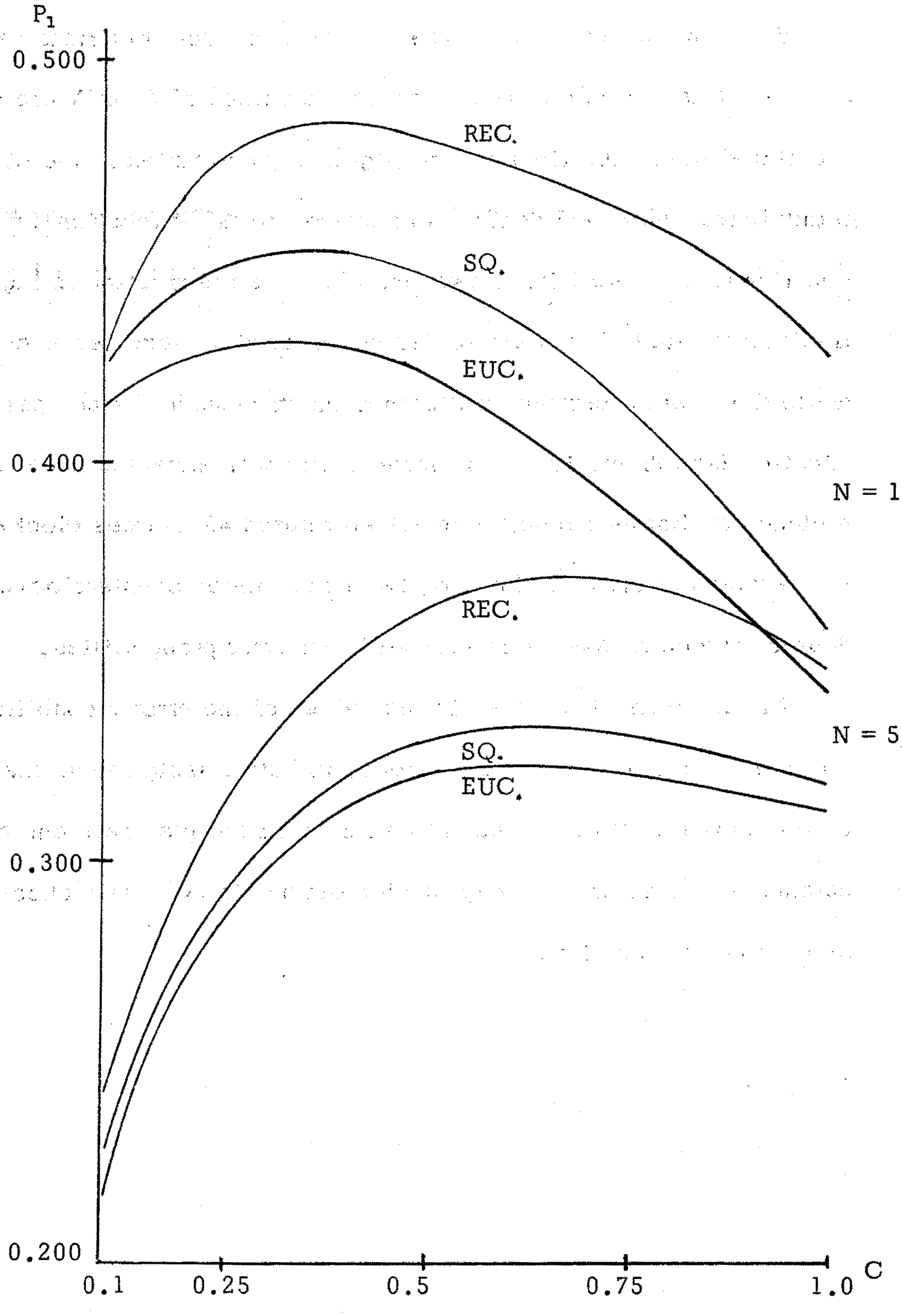
P_2 given alpha and beta for dependent bivariate exponentials

IV. CONCLUSIONS

The goal of this thesis was to compare different distance functions in the rule of the nearest neighbor for bivariate negative exponential populations. When C (λ_1/μ_1) and D (λ_2/μ_2) are less than one, there is virtually no difference between the Euclidean, square, and rectangular discriminating functions. This generalization remains true even when the sample size becomes small. The variation that occurs is in the third decimal place and is not significant. This implies that when C and D are small, the actual discriminating function used may be unimportant and one can safely employ Euclidean distance for assignment purposes. However, when D is allowed to increase, the choice of discriminating functions becomes important. Figures I and II clearly show that for small sample sizes, the Euclidean distance function is superior for the cases considered to either the square or rectangular discriminating functions respectively. When the sample increases, there is no discernible difference between the Euclidean and square distance functions and the absolute differences between them and the rectangular distance function decreases (see Figure II). This reduction in the discrimination power between the functions is in agreement with Fix and Hodges [1] who prove under fairly general regularity conditions that there is no difference in discriminating functions when the limit of the number of sample points approaches infinity.

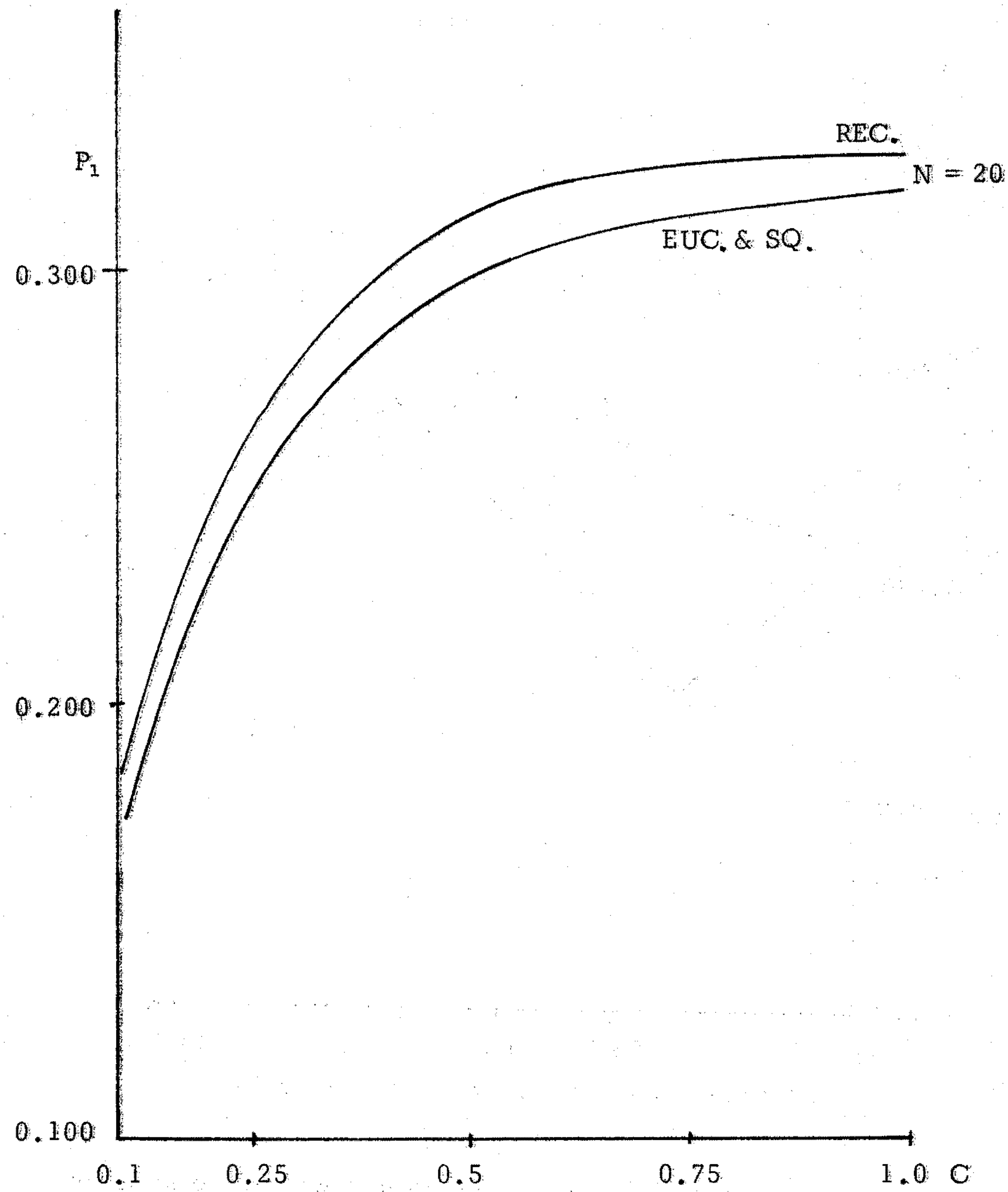
When the sample size is small in the dependent bivariate case, there are some interesting observations. If the values of X and Y are positively correlated in one distribution and negatively correlated in the other, the rotated rectangular (45 degrees clockwise) distance function seems to be superior (see Figure III). However, when one distribution is highly negatively correlated, Euclidean distance appeared to make better assignments while the rotated rectangular distance function made the poorest classification (see Figure IV). This latter result was expected because the rectangular distance function remained rotated 45 degrees clockwise. If the rectangular distance function had been rotated counterclockwise 45 degrees it would have again yielded lower error probabilities.

The absolute differences in the values of the error probabilities may not be as important as the consistency in better assignments that one function gives. It can be seen that for small sample sizes and certain parameter ratios, one can improve his diagnosis by wisely choosing his discriminating function.



Comparison of distance functions by probability of erroneous assignment with fixed D = 5.

Figure I



Comparison of distance functions by probability of erroneous assignment with fixed D = 5.

Figure II

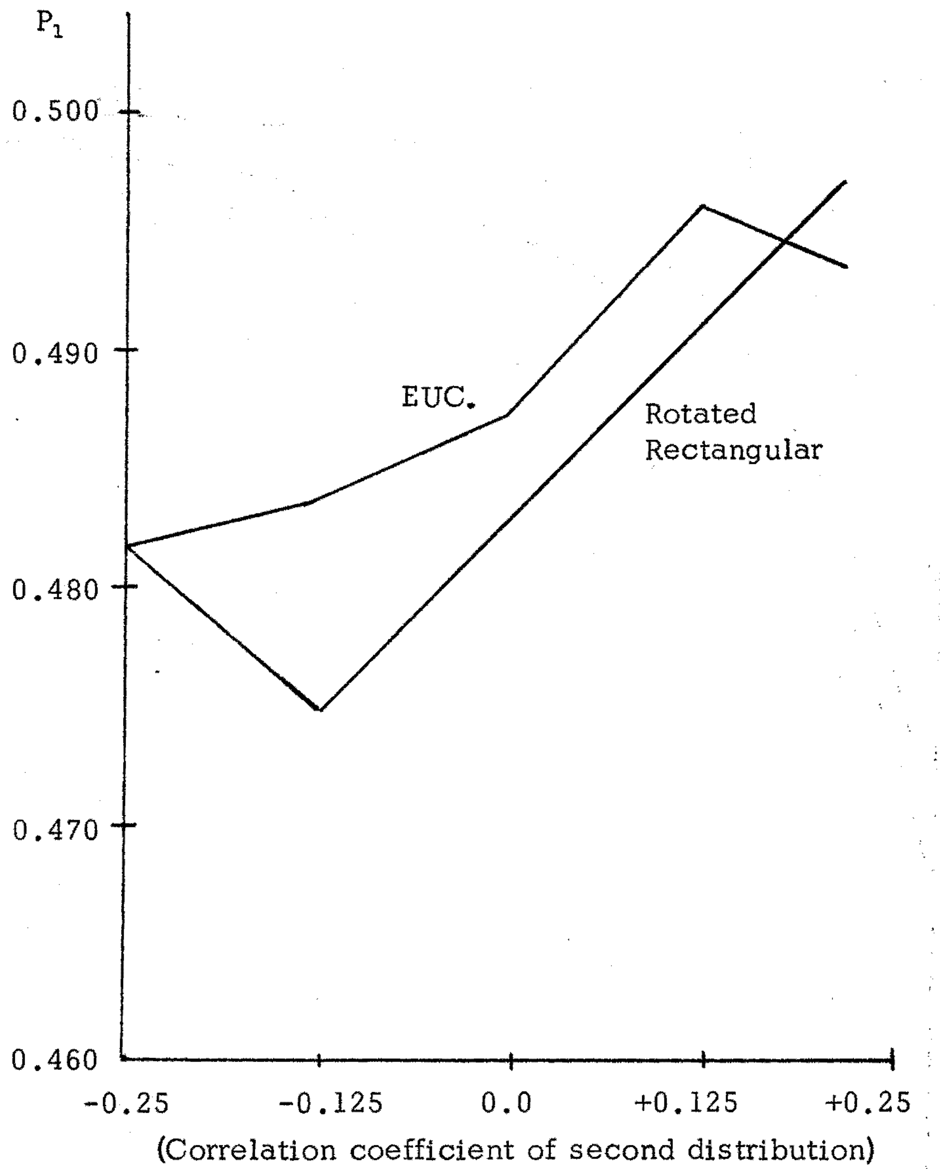


Figure III

Probability of erroneous assignment with correlation of first distribution set at + 0.25 and sample size = 2

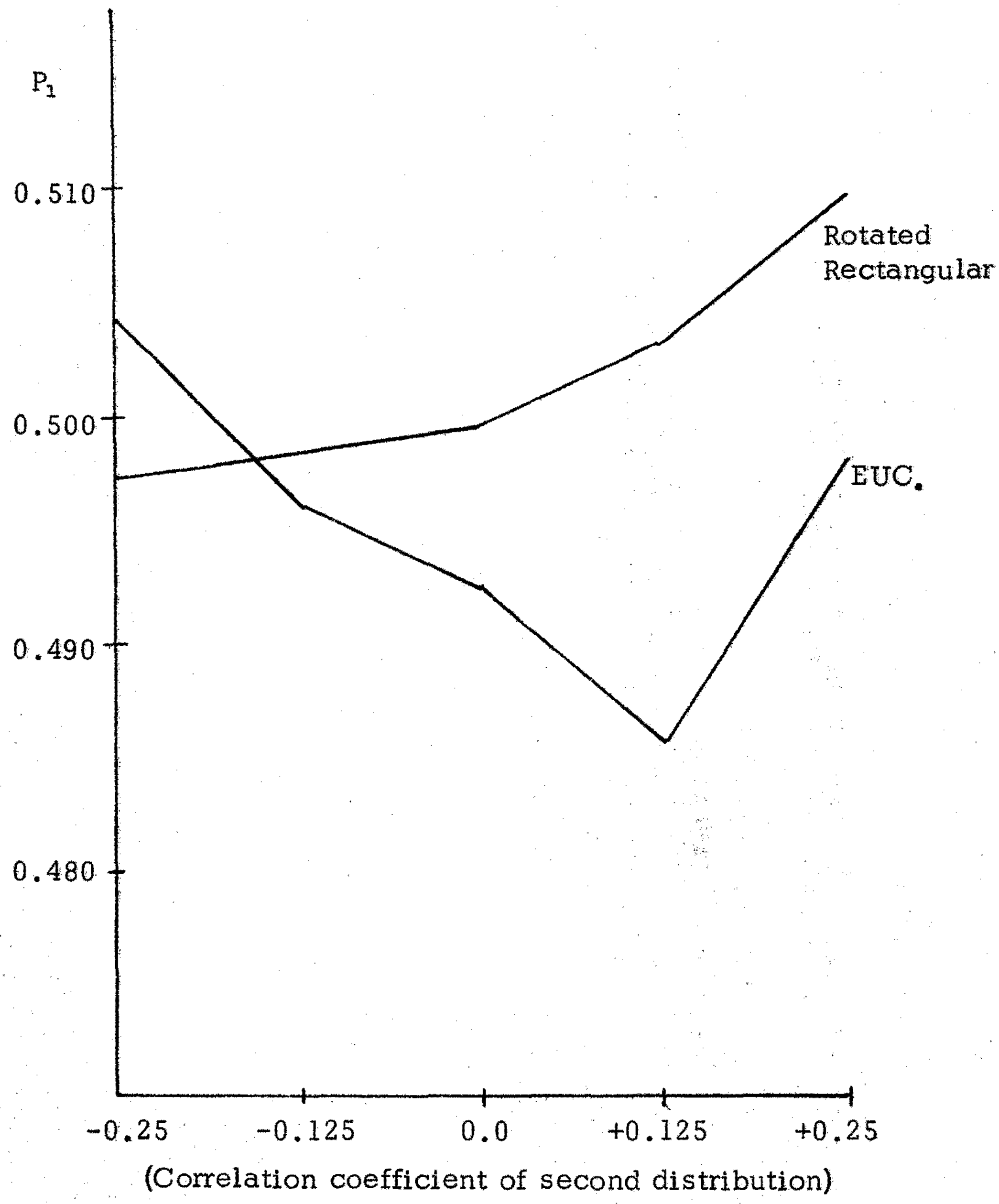


Figure IV

Probability of erroneous assignment with correlation of first distribution set at - 0.25 and sample size = 2

```
C                    APPENDIX I
C
C
C
C     DISCRIMINATORY ANALYSIS SIMULATION: JAY E. LIEBERMAN
C     INDEPENDENT BI-VARIATE NEGATIVE EXPONENTIALLY DISTRIBUTED
C     POPULATIONS
      DATA A/2.0/
      INTEGER PE1,PE2,PS1,PS2,PR1,PR2
      REAL LAM1,LAM2,MU1,MU2
C     NU IS THE NUMBER OF DATA CARDS
C     'N' IS THE NUMBER OF SAMPLE POINTS TO BE GENERATED FROM EACH DISTN
C     'NREPS' IS THE NUMBER OF REPLICATIONS DONE IN THE SIMULATION
C     LAMBDAS AND MUS ARE THE PARAMETERS OF THE BI-VARIATE EXPONENTIALS
      DUMMY=URN(0)
      READ(5,830)NU
      DO 1800 II=1,NU
      READ(5,840)N,NREPS,LAM1,MU1,LAM2,MU2
      ZNREPS=NREPS
      PE1=0
      PE2=0
      PS1=0
      PS2=0
      PR1=0
      PR2=0
      DO 480 NR=1,NREPS
      T1E=10.E50
      T2E=10.E50
      T1S=10.E50
      T2S=10.E50
      T1R=10.E50
      T2R=10.E50
      D1E=10.E50
      D2E=10.E50
      D1S=10.E50
      D2S=10.E50
      D1R=10.E50
      D2R=10.E50
      ZX1=-ALOG(URN(1))/LAM1
      ZY1=-ALOG(URN(1))/LAM2
      ZX2=-ALOG(URN(1))/MU1
      ZY2=-ALOG(URN(1))/MU2
      DO 100 J=1,N
      X1=-ALOG(URN(1))/LAM1
      Y1=-ALOG(URN(1))/LAM2
      X2=-ALOG(URN(1))/MU1
      Y2=-ALOG(URN(1))/MU2
C     COMPUTE COMMON FACTORS FOR DISTANCE FORMULAS
      XX1=ABS(ZX1-X1)
```

```
      XX2=ABS(ZX2-X2)
      YY1=ABS(ZY1-Y1)
      YY2=ABS(ZY2-Y2)
      AX1=ABS(ZX1-X2)
      AX2=ABS(ZX2-X1)
      AY1=ABS(ZY1-Y2)
      AY2=ABS(ZY2-Y1)
C     MEASUREMENT BY THE EUCLIDEAN DISTANCE FUNCTION
      T1E=AMIN1(XX1**2+YY1**2,T1E)
      T2E=AMIN1(XX2**2+YY2**2,T2E)
      D1E=AMIN1(AX1**2+AY1**2,D1E)
      D2E=AMIN1(AX2**2+AY2**2,D2E)
C     MEASUREMENT BY THE SQUARE DISTANCE FUNCTION
      T1S=AMIN1(AMAX1(XX1,YY1),T1S)
      T2S=AMIN1(AMAX1(XX2,YY2),T2S)
      D1S=AMIN1(AMAX1(AX1,AY1),D1S)
      D2S=AMIN1(AMAX1(AX2,AY2),D2S)
C     MEASUREMENT BY THE RECTANGULAR DISTANCE FUNCTION
      T1R=AMIN1(AMAX1(A*XX1,YY1),T1R)
      T2R=AMIN1(AMAX1(A*XX2,YY2),T2R)
      D1R=AMIN1(AMAX1(A*AX1,AY1),D1R)
      D2R=AMIN1(AMAX1(A*AX2,AY2),D2R)
  100 CONTINUE
      IF(T1E.GT.D1E) PE1=PE1+1
      IF(T1S.GT.D1S) PS1=PS1+1
      IF(T1R.GT.D1R) PR1=PR1+1
      IF(T2E.GT.D2E) PE2=PE2+1
      IF(T2S.GT.D2S) PS2=PS2+1
      IF(T2R.GT.D2R) PR2=PR2+1
  480 CONTINUE
      ZPE1=PE1
      ZPE2=PE2
      ZPS1=PS1
      ZPS2=PS2
      ZPR1=PR1
      ZPR2=PR2
      P1ERRE=ZPE1/ZNREPS
      P2ERRE=ZPE2/ZNREPS
      P1ERRS=ZPS1/ZNREPS
      P2ERRS=ZPS2/ZNREPS
      P1ERRR=ZPR1/ZNREPS
      P2ERRR=ZPR2/ZNREPS
      WRITE(6,700) P1ERRE
      WRITE(6,740) P2ERRE
      WRITE(6,710) LAM1,LAM2,MU1,MU2
      WRITE(6,720) NREPS
      WRITE(6,730) N
      WRITE(6,760) P1ERRS
```

```
      WRITE(6,780) P2ERRS
      WRITE(6,710) LAM1,LAM2,MU1,MU2
      WRITE(6,720) NREPS
      WRITE(6,730) N
      WRITE(6,800) P1ERRR
      WRITE(6,820) P2ERRR
      WRITE(6,710) LAM1,LAM2,MU1,MU2
      WRITE(6,720) NREPS
      WRITE(6,730) N
 1800 CONTINUE
  700 FORMAT(1H1,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'2ND DISTRIBUTION, GIVEN THAT IT CAME FROM THE 1ST,'/,22X
     2'(CALLED TYPE 1 ERROR) USING THE EUCLIDEAN DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  710 FORMAT(1H0,21X,'THE PARAMETERS OF THE BI-VARIATE EXPONENTIAL WERE
     1AS FOLLOWS:'/22X,'FROM DISTRIBUTION 1: LAM(1)='F5.2,' LAM(2)='
     2F5.2,/22X,'FROM DISTRIBUTION 2:  MU(1)='F5.2,'  MU(2)='F5.2)
  720 FORMAT(1H0,21X,'THE NUMBER OF REPLICATIONS WAS 'I5)
  730 FORMAT(1H0,21X,'THE SAMPLE SIZE WAS 'I2)
  740 FORMAT(1H0,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'1ST DISTRIBUTION, GIVEN THAT IT CAME FROM THE 2ND,'/,22X
     2'(CALLED TYPE 2 ERROR) USING THE EUCLIDEAN DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  760 FORMAT(1H1,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'2ND DISTRIBUTION, GIVEN THAT IT CAME FROM THE 1ST,'/,22X
     2'(CALLED TYPE 1 ERROR) USING THE SQUARE DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  780 FORMAT(1H0,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'1ST DISTRIBUTION, GIVEN THAT IT CAME FROM THE 2ND,'/,22X
     2'(CALLED TYPE 2 ERROR) USING THE SQUARE DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  800 FORMAT(1H1,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'2ND DISTRIBUTION, GIVEN THAT IT CAME FROM THE 1ST,'/,22X
     2'(CALLED TYPE 1 ERROR) USING A RECTANGULAR DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  820 FORMAT(1H0,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'1ST DISTRIBUTION, GIVEN THAT IT CAME FROM THE 2ND,'/,22X
     2'(CALLED TYPE 2 ERROR) USING A RECTANGULAR DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  830 FORMAT(I2)
  840 FORMAT(I2,3XI5,5X,F5.2,5X,F5.2,5X,F5.2,5X,F5.2)
      STOP
      END
```

```
C      DISCRIMINATORY ANALYSIS SIMULATION: JAY E. LIEBERMAN
C      DEPENDENT BI-VARIATE NEGATIVE EXPONENTIALLY DISTRIBUTED POPULATIONS
       DATA A/2.0/
       INTEGER PE1,PE2,PS1,PS2,PR1,PR2,PRR1,PRR2
C      'NU' IS THE NUMBER OF DATA CARDS
C      'N' IS THE NUMBER OF SAMPLE POINTS TO BE GENERATED FROM EACH DISTN
C      'NREPS' IS THE NUMBER OF REPLICATIONS DONE IN THE SIMULATION
C      ALPHA & BETA ARE THE PARAMETER OF THE DEPENDENT DISTNS
       DUMMY=URN(0)
       READ(5,830)NU
       DO 1800 II=1,NU
       READ(5,840)N,NREPS,ALPHA,BETA
       ZNREPS=NREPS
       PE1=0
       PE2=0
       PS1=0
       PS2=0
       PR1=0
       PR2=0
       PRR1=0
       PRR2=0
       DO 480 NR=1,NREPS
       T1E=10.E50
       T2E=10.E50
       T1S=10.E50
       T2S=10.E50
       T1R=10.E50
       T2R=10.E50
       T1RR=10.E50
       T2RR=10.E50
       D1E=10.E50
       D2E=10.E50
       D1S=10.E50
       D2S=10.E50
       D1R=10.E50
       D2R=10.E50
       D1RR=10.E50
       D2RR=10.E50
       U1=URN(1)
       U2=URN(1)
       ZC1=ALPHA*(1.-U1-U1)
       ZC2=BETA*(1.-U2-U2)
       ZC1P=1.+ZC1
       ZC2P=1.+ZC2
       ZY1=-ALOG(U1)
       ZY2=-ALOG(U2)
       IF(ZC1.EQ.0.0)GO TO 70
       ZZ1=ZC1P**2-4.*ZC1*URN(1)
```

```
      ZX1=-ALOG((ZC1P-SQRT(ZZ1))/(ZC1+ZC1))
      GO TO 71
   70 ZX1=-ALOG(URN(1))
   71 IF(ZC2.EQ.0.0)GO TO 72
      ZZ2=ZC2P**2-4.*ZC2*URN(1)
      ZX2=-ALOG((ZC2P-SQRT(ZZ2))/(ZC2+ZC2))
      GO TO 73
   72 ZX2=-ALOG(URN(1))
   73 DO 100 J=1,N
      U1=URN(1)
      U2=URN(1)
      Y1=-ALOG(U1)
      Y2=-ALOG(U2)
      C1=ALPHA*(1.-U1-U1)
      C2=BETA*(1.-U2-U2)
      C1P=1.+C1
      C2P=1.+C2
      IF(C1.EQ.0.0)GO TO 80
      V1=C1P**2-4.*C1*URN(1)
      X1=-ALOG((C1P-SQRT(V1))/(C1+C1))
      GO TO 81
   80 X1=-ALOG(URN(1))
   81 IF(C2.EQ.0.0)GO TO 82
      V2=C2P**2-4.*C2*URN(1)
      X2=-ALOG((C2P-SQRT(V2))/(C2+C2))
      GO TO 83
   82 X2=-ALOG(URN(1))
C     COMPUTE COMMON FACTORS FOR DISTANCE FORMULAS
   83 XX1=ABS(ZX1-X1)
      XX2=ABS(ZX2-X2)
      YY1=ABS(ZY1-Y1)
      YY2=ABS(ZY2-Y2)
      AX1=ABS(ZX1-X2)
      AX2=ABS(ZX2-X1)
      AY1=ABS(ZY1-Y2)
      AY2=ABS(ZY2-Y1)
C     MEASUREMENT BY THE EUCLIDEAN DISTANCE FUNCTION
      T1E=AMIN1(XX1**2+YY1**2,T1E)
      T2E=AMIN1(XX2**2+YY2**2,T2E)
      D1E=AMIN1(AX1**2+AY1**2,D1E)
      D2E=AMIN1(AX2**2+AY2**2,D2E)
C     MEASUREMENT BY THE SQUARE DISTANCE FUNCTION
      T1S=AMIN1(AMAX1(XX1,YY1),T1S)
      T2S=AMIN1(AMAX1(XX2,YY2),T2S)
      D1S=AMIN1(AMAX1(AX1,AY1),D1S)
      D2S=AMIN1(AMAX1(AX2,AY2),D2S)
C     MEASUREMENT BY THE RECTANGULAR DISTANCE FUNCTION
      T1R=AMIN1(AMAX1(A*XX1,YY1),T1R)
```

```
      T2R=AMIN1(AMAX1(A*XX2,YY2),T2R)
      D1R=AMIN1(AMAX1(A*AX1,AY1),D1R)
      D2R=AMIN1(AMAX1(A*AX2,AY2),D2R)
C     MEASMT BY ROTATED RECTANGULAR DISTANCE FUNCTION
C     A COMMON FACTOR OF 2/SQRT(2) HAS BEEN LEFT OFF THE FOLLOWING
C     EIGHT FACTORS
      ZX1P=ZX1-ZY1
      ZX2P=ZX2-ZY2
      ZY1P=ZX1+ZY1
      ZY2P=ZX2+ZY2
      X1P=X1-Y1
      Y1P=X1+Y1
      X2P=X2-Y2
      Y2P=X2+Y2
      T1RR=AMIN1(AMAX1(A*ABS(ZX1P-X1P),ABS(ZY1P-Y1P)),T1RR)
      T2RR=AMIN1(AMAX1(A*ABS(ZX2P-X2P),ABS(ZY2P-Y2P)),T2RR)
      D1RR=AMIN1(AMAX1(A*ABS(ZX1P-X2P),ABS(ZY1P-Y2P)),D1RR)
      D2RR=AMIN1(AMAX1(A*ABS(ZX2P-X1P),ABS(ZY2P-Y1P)),D2RR)
  100 CONTINUE
      IF(T1E.GT.D1E) PE1=PE1+1
      IF(T1S.GT.D1S) PS1=PS1+1
      IF(T1R.GT.D1R) PR1=PR1+1
      IF(T1RR.GT.D1RR) PRR1=PRR1+1
      IF(T2E.GT.D2E) PE2=PE2+1
      IF(T2S.GT.D2S) PS2=PS2+1
      IF(T2R.GT.D2R) PR2=PR2+1
      IF(T2RR.GT.D2RR) PRR2=PRR2+1
  480 CONTINUE
      ZPE1=PE1
      ZPE2=PE2
      ZPS1=PS1
      ZPS2=PS2
      ZPR1=PR1
      ZPR2=PR2
      ZPRR1=PRR1
      ZPRR2=PRR2
      P1ERRE=ZPE1/ZNREPS
      P2ERRE=ZPE2/ZNREPS
      P1ERRS=ZPS1/ZNREPS
      P2ERRS=ZPS2/ZNREPS
      P1ERRR=ZPR1/ZNREPS
      P2ERRR=ZPR2/ZNREPS
      P1RRR=ZPRR1/ZNREPS
      P2RRR=ZPRR2/ZNREPS
      WRITE(6,700) P1ERRE
      WRITE(6,740) P2ERRE
      WRITE(6,710) ALPHA,BETA
      WRITE(6,720) NREPS
```

```
      WRITE(6,730) N
      WRITE(6,760) P1ERRS
      WRITE(6,780) P2ERRS
      WRITE(6,710) ALPHA,BETA
      WRITE(6,720) NREPS
      WRITE(6,730) N
      WRITE(6,800) P1ERRR
      WRITE(6,820) P2ERRR
      WRITE(6,710) ALPHA,BETA
      WRITE(6,720) NREPS
      WRITE(6,730) N
      WRITE(6,850) P1RRR
      WRITE(6,860) P2RRR
      WRITE(6,710) ALPHA,BETA
      WRITE(6,720) NREPS
      WRITE(6,730) N
 1800 CONTINUE
  700 FORMAT(1H1,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'2ND DISTRIBUTION, GIVEN THAT IT CAME FROM THE 1ST,'/,22X
     2'(CALLED TYPE 1 ERROR) USING THE EUCLIDEAN DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  710 FORMAT(1H0,21X,'THE PARAMETER OF THE DEPENDENT BI-VARIATE'
     1/,22X'EXPONENTIALS WAS AS FOLLOWS:'/22X,
     2'FROM DISTRIBUTION 1: ALPHA='F5.2,/22X,
     3'FROM DISTRIBUTION 2:  BETA='F5.2)
  720 FORMAT(1H0,21X,'THE NUMBER OF REPLICATIONS WAS 'I5)
  730 FORMAT(1H0,21X,'THE SAMPLE SIZE WAS 'I2)
  740 FORMAT(1H0,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'1ST DISTRIBUTION, GIVEN THAT IT CAME FROM THE 2ND,'/,22X
     2'(CALLED TYPE 2 ERROR) USING THE EUCLIDEAN DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  760 FORMAT(1H1,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'2ND DISTRIBUTION, GIVEN THAT IT CAME FROM THE 1ST,'/,22X
     2'(CALLED TYPE 1 ERROR) USING THE SQUARE DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  780 FORMAT(1H0,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'1ST DISTRIBUTION, GIVEN THAT IT CAME FROM THE 2ND,'/,22X
     2'(CALLED TYPE 2 ERROR) USING THE SQUARE DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  800 FORMAT(1H1,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'2ND DISTRIBUTION, GIVEN THAT IT CAME FROM THE 1ST,'/,22X
     2'(CALLED TYPE 1 ERROR) USING A RECTANGULAR DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  820 FORMAT(1H0,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
     1'1ST DISTRIBUTION, GIVEN THAT IT CAME FROM THE 2ND,'/,22X
     2'(CALLED TYPE 2 ERROR) USING A RECTANGULAR DISTANCE FUNCTION'/,22X
     3'*NEAREST NEIGHBOR RULE* IS:'//,35X,F7.5)
  830 FORMAT(I2)
```

```
840 FORMAT(I2,3XI5,5X,F5.2,5X,F5.2)
850 FORMAT(1H1,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
   1'2ND DISTRIBUTION, GIVEN THAT IT CAME FROM THE 1ST,'/,22X
   2'(CALLED TYPE 1 ERROR) USING A RECTANGULAR DISTANCE FUNCTION'/,22X
   3'WHICH HAS BEEN ROTATED 45 DEGREES CLOCKWISE IS:'//,35X,F7.5)
860 FORMAT(1H0,21X,'THE PROBABILITY OF ASSIGNING POINT Z TO THE'/,22X
   1'1ST DISTRIBUTION, GIVEN THAT IT CAME FROM THE 2ND,'/,22X
   2'(CALLED TYPE 2 ERROR) USING A RECTANGULAR DISTANCE FUNCTION'/,22X
   3'WHICH HAS BEEN ROTATED 45 DEGREES CLOCKWISE IS:'//,35X,F7.5)
    STOP
    END
```

APPENDIX II

If the error probabilities cannot be found for the desired parameters, the simulation can be run for any values of λ and $\mu > 0$. The program requires as input:

A) On data card #1 in format I2, the number of data cards to be processed (not including itself).

B) For the remaining data cards:

1) In columns 1 and 2 in integer format is the number of sample points desired (N),

2) In columns 6-10 in integer format is the number of replications to be performed (NREPS),

3) In columns 16-20, 26-30, 36-40, 46-50 are the parameters of the distributions λ_1, λ_2, μ_1, and μ_2 respectively (format F5.2).

The program will print out P_1 and P_2 for each distance function and the respective input data.

The simulation for dependent exponentials can also be run for any values of alpha and beta between -1 and +1. The input is similar except that on the data cards alpha is in columns 16-20 and beta is in columns 26-30. The form of the output is similar to that received in the independent case.

BIBLIOGRAPHY

1. Fix, Evelyn and Hodges, J. L., Jr., Non-Parametric Discrimination: Small Sample Performances, Report prepared under contract between between University of California and School of Aviation Medicine, Randolph Field, Texas, February 1951.

2. Gumbel, E. J., "Bivariate Exponential Distributions," Journal of the American Statistical Association, v. 55, p. 698-707, December 1960.

3. Lockett, Joseph Louis III, A Comparison of Some Parametric and Non-Parametric Discrimination Procedures in Negative Exponential Populations, Master's Thesis at United States Naval Postgraduate School, Monterey, California, 1968.

4. Rydzewski, James C., Non-Parametric Classification Error Probabilities for Two Dimensional Exponential Populations, Master's Thesis at United States Naval Postgraduate School, Monterey, California, 1966.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library, Code 0212 U. S. Naval Postgraduate School Monterey, California 93940	2
3.	Dr. J. R. Borsting, Code 55Bg Department of Operations Analysis U. S. Naval Postgraduate School Monterey, California 93940	10
4.	LT(j.g.) Jay E. Lieberman 219 E. Dorrance Street Kingston, Pennsylvania 18704	2
5.	Mr. D. F. Ancona III U. S. Army Tank Automotive Command Warren, Michigan 48093	1

[redacted]

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Naval Postgraduate School Monterey, California 93940	Unclassified
	2b. GROUP

3. REPORT TITLE

A Comparison of Various Non-Parametric Discriminating Procedures When the Populations Are Bivariate Exponentials

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Master's Thesis; December 1969

5. AUTHOR(S) (First name, middle initial, last name)

Jay Edward Lieberman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1969	55	4

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

A comparison of the error probabilities for various discriminating rules is performed in the two population case when nothing is known of the populations other than they are bivariate negative exponential. In most cases, the absolute difference between the error probabilities for each function was very small. However, the Euclidean distance function consistently performed as well as, and sometimes superior to any of the others studied in this thesis.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

Security Classification

A-31408

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Non-Parametric Discriminating Procedures						
Bivariate Negative Exponential Populations						