CROSBY, David S., 1938-
THE STRUCTURE OF CERTAIN TWO-DIMENSIONAL
PROBABILITY MEASURES.

University of Arizona, Ph.D., 1967
Mathematics

University Microfilms, Inc., Ann Arbor, Michigan

THE STRUCTURE OF CERTAIN TWO-DIMENSIONAL PROBABILITY MEASURES

by

David S. Crosby

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MATHEMATICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 6 7

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

I hereby recommend that this dissertation prepared under my direction by David S. Crosby

entitled "The Structure of Certain Two-Dimensional Probability Measures"

be accepted as fulfilling the dissertation requirement of the degree of Doctor of Philosophy

D. E. Myers — Dissertation Director — 10-5-66 Date

After inspection of the dissertation, the following members of the Final Examination Committee concur in its approval and recommend its acceptance:*

[signature]	10-5-66
[signature]	10-5-66
M. S. Cheema	10-5-66
P. K. Bhattacharya	10-5-66
D. E. Myers	10-5-66

*This approval and acceptance is contingent on the candidate's adequate performance and defense of this dissertation at the final oral examination. The inclusion of this sheet bound into the library copy of the dissertation is evidence of satisfactory performance at the final examination.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: David S. Crosby

ACKNOWLEDGMENTS

I wish to thank Dr. Donald E. Myers for his interest and encouragement during the preparation of this work.

I also wish to thank Dr. P. K. Bhattacharya for suggesting improvements in several of the proofs.

TABLE OF CONTENTS

ABSTRACT

It is found that a large class of two-dimensional probability measures are absolutely continuous with respect to the measure defined by the product of the marginal measures. These include the most common two-dimensional probability measures of mathematical statistics. By considering the measure as an integral with respect to the product measure, new proofs are given for known results.

Necessary and sufficient conditions are given on the conditional expectation for a two-dimensional measure to belong to an especially simple class of two-dimensional measures with given marginals. In certain special cases the marginals of members of this class are shown to be independent if and only if the distribution function of the sum of two marginals is equal to the convolution of the distribution functions of the marginals. If two random variables have non-zero covariance and if their two-dimensional Lebesgue-Stieltjes measure is of this class then the measure is shown to be completely determined by the two regression curves.

If the marginal probability measures are finite discrete then the convex set of two-dimensional measures with these marginals is shown to have a finite number of extreme points. It is shown that such measures can be associated with convex polyhedrons in Euclidean space. In the special case where the marginals are identical and uniform all the extreme points are obtained. Using known results about distribution functions a method is given for obtaining two extreme points for any such convex set of two-dimensional measures.

CHAPTER 1
INTRODUCTION

In this paper we examine the structure of certain two-dimensional probability measures. Specifically, those which are absolutely continuous with respect to the measure defined by the product of the marginals are considered.

In recent years certain special types of such two-dimensional probability measures have been studied. Lancaster and Kagleson have considered measures such that the Radon-Nikodym derivative of the measure with respect to the product of marginals is square integrable. See [13], [14], [15], [16]. Farlie has considered a class of such measures which have a relatively simple Radon-Nikodym derivative in order to evaluate certain measures of association. See [3], [4], [5]. Silvery uses the Radon-Nikodym derivative to obtain a measure of association. See [24].

This dissertation is divided into four parts. The first of these (Chapter 2) gives some standard results about two-dimensional probability measures. The second of these (Chapter 3) is mainly expository in nature. No claim is made for depth or originality of the results of that chapter. The theorems from Theorem 3.16 on are special cases of more general theorems. It should be noted, however, that the proofs are original and that the methods used should be useful at least in a pedogogical way.

The original parts of this dissertation are contained in parts 3 and 4 (Chapter 4 and 5). Chapter 4 deals with a special class of two-dimensional measures where the Radon-Nikodym derivative has an especially simple form. The main result of this chapter is Theorem 4.2 which gives necessary and sufficient conditions on the conditional expectation for a measure to belong to this class. With the exception of Lemma 4.1 and Theorem 4.13 all the results in this chapter are original. Lemma 4.1 is proven by Farlie in a special case, and Theorem 4.13 is suggested by Farlie although it is neither stated completely nor proven.

Chapter 5 contains the most original idea of this dissertation. It treats two-dimensional probability measures with finite discrete marginals as a convex set and establishes a correspondence between such measures and polyhedrons in Euclidean space. The theorems contained in this chapter are believed to be original.

Theorems, corollaries and lemmas are numbered in order according to the chapter in which they occur. Definitions are numbered in the same way but in a separate sequence. Numbers in brackets occurring in the body of the text refer to the list of references at the end.

CHAPTER 2
PRELIMINARIES

In this chapter we will present the notation and some basic definitions and theorems necessary to the development of the theory of bivariate measures which are absolutely continuous with respect to the product of the marginals.

$(W_1, \mathcal{A}, \mu_1)$ and $(W_2, \mathcal{B}, \mu_2)$ will represent two measure spaces where W_1 and W_2 are sets, $\mathcal{A}$ and $\mathcal{B}$ are the σ- algebras on them and μ_1 and μ_2 are the probability measures. Points of W_1 and W_2 will be designated by x and y respectively. $(W_1 \times W_2, \mathcal{A} \times \mathcal{B}, \mu_1 \times \mu_2)$ will denote the product probability space as is constructed in [12].

Random variables will be designated by X and Y, their distribution functions by $F_1(x)$ and $F_2(y)$ respectively, and the associated density functions, if they exist, by $f_1(x)$ and $f_2(y)$. Their joint distribution will be designated by $F(x,y)$ and their joint density function, if it exists, by $f(x,y)$. The expected value of a random variable or a Baire function of a random variable will be designated by E [ ]. The Lebesgue-Stieltjes measure on $(-\infty,\infty) \times (-\infty,\infty)$ determined by $F(x,y)$ will be denoted by μ and the Lebesgue-Stieltjes measures determined by $F_1(x)$ and $F_2(y)$ on $(-\infty,\infty)$ will be denoted by μ_1 and μ_2 respectively.

Let $(W_1 \times W_2, \mathcal{A} \times \mathcal{B}, \mu)$ be a probability measure space.

<u>DEFINITION 2.1</u>. μ_1 is the marginal probability measure of μ on W_1 if $\mu_1(A) = \mu(A \times W_2)$ for all sets A belonging to $\mathcal{A}$. μ_2 is the marginal measure of μ on W_2 if $\mu_2(B) = \mu(W_1 \times B)$ for all sets B belonging to $\mathcal{B}$.

<u>DEFINITION 2.2</u>. If $F(x,y)$ is the joint distribution function of two random variables X and Y then the distribution function of X is given by $F_1(x) = F(x,\infty)$ and is called the marginal distribution function of X. The distribution function of Y is given by $F_2(y) = F(\infty,y)$.

In this dissertation when we speak of a unique function we will actually be referring to an equivalent class of functions. Two functions will be said to be equivalent with respect to some measure if they are equal almost everywhere.

The following theorems are standard and are to be found in almost any text which deals with probability theory in a measure theoretic way. We have chosen to use the form given in [21].

<u>THEOREM 2.1</u>. Let $(W_1 \times W_2, \mathcal{A} \times \mathcal{B}, \mu)$ be a probability measure space. Let μ_1 be the marginal measure of μ on W_1, and μ_2 be the marginal measure of μ on W_2. Then if B is any measurable set in $\mathcal{B}$, there is defined almost everywhere μ_1 a unique function $\mu_2(B|x)$ which is integrable over W_1 with respect to μ_1 and has the property that

$$\mu(A \times B) = \int_A \mu_2(B|x)d\mu_1$$

for every measurable A in $\mathfrak{A}$.

<u>DEFINITION 2.3</u>. The number $\mu_2(B|x)$ is called the conditional probability that y belongs to B under the condition that x takes on its assigned value.

There is a similar theorem and definition which defines a function $\mu_1(A|y)$ on W_2 where the number $\mu_1(A|y)$ is called the conditional probability that x belongs to the set A under the condition that y takes its assigned value.

<u>THEOREM 2.2</u>. Suppose that h(y) is integrable with respect to μ_2, then there is defined for almost all x a unique function $E[h(y)|x]$ which is integrable over W_1 with respect to μ_1 and has the property that

$$\int_{A \times W_2} h(y)\ d\mu = \int_A E[h(y)|x]\ d\mu_1$$

for every set A in $\mathfrak{A}$. If $A = W_1$ then

$$E[h(y)] = \int_{W_1} E[h(y)|x]\ d\mu_1 \ .$$

<u>DEFINITION 2.4</u>. The number $E[h(y)|x]$ is called the conditional expectation of h(y) under the condition that x takes on its assigned value.

The two marginal measures μ_1 and μ_2 will be called independent if μ, the measure on the product space, is that determined by the product of the two marginals.

<u>THEOREM 2.3</u>. Suppose that for almost all x a probability measure $\mu_{2,x}(B)$ can be defined on the measurable sets B of $\mathcal{B}$ to satisfy

$$\mu(A \times B) = \int_A \mu_{2,x}(B)\ d\mu_1$$

for all A in $\mathcal{Q}$. Then if B is any measurable set $\mu_{2,x}(B) = \mu_2(B|x)$ for almost all x, and if h(x,y) is integrable with respect to μ, then h(x,y) is integrable with respect to $\mu_2(\ |x)$ for almost all x and

$$\int_{W_1 \times W_2} h(x,y)\ d\mu = \int_{W_1} \int_{W_2} h(x,y) d\mu_2(\ |x) d\mu_1 \quad .$$

If h(y) is integrable with respect to μ_2 it is integrable with respect to $\mu_2(\ |x)$ and

$$E[h(y)|x] = \int_{W_1} h(y)\ d\mu_2(\ |x)$$

for almost all x.

Moreover, the marginals are independent if and only if $\mu_2(\ |x) = \mu_2$ for almost all x.

THEOREM 2.4. If $W_2 = (-\infty,\infty)$, a probability measure $\mu_2(\ |x)$ can be defined on W_2 for almost all x to satisfy the condition of Theorem 2.3.

For the proofs of the above theorems see [21] pages 65-72.

Since in this paper we are going to be concerned with measures which are absolutely continuous with respect to other measures, we will rely on the Radon-Nikodym theorem. The form of this theorem will be that given in [12].

Since we are going to stipulate that the bivariate measure μ is absolutely continuous with respect to the product measure $\mu_1 \times \mu_2$, and since μ and $\mu_1 \times \mu_2$ are both totally finite we will have a function $1 + R(x,y)$ determined almost everywhere $\mu_1 \times \mu_2$ such that

$$\mu(E) = \int_E [1 + R(x,y)]d\mu_1 \times \mu_2$$

for all measurable sets E belonging to $W_1 \times W_2$.

When we speak of the joint distribution function $F(x,y)$ of X and Y being absolutely continuous with respect to the product of the marginal distribution functions, we will mean that the Lebesgue-Stieltjes measure μ is absolutely continuous with respect to the product measure determined by $\mu_1 \times \mu_2$.

CHAPTER 3
GENERAL RESULTS

In this chapter we are going to present some general properties of the bivariate probability measures which are absolutely continuous with respect to the measure defined by the product of their marginals. We thus assume that there exists a function $1 + R(x,y)$ such that

$$\mu(E) = \int_E \left[1 + R(x,y)\right] d\mu_1 \times \mu_2$$

for all measurable sets E in $W_1 \times W_2$.

In order that μ be a probability measure with the proper marginals there are several properties that $R(x,y)$ must have.

THEOREM 3.1. If μ is a signed measure defined on $(W_1 \times W_2, \mathcal{A} \times \mathcal{B})$ such that

$$\mu(E) = \int_E \left[1 + R(x,y)\right] d\mu_1 \times \mu_2$$

for all sets E in $\mathcal{A} \times \mathcal{B}$, then μ is a probability measure on $(W_1 \times W_2, \mathcal{A} \times \mathcal{B})$ with marginals μ_1 on $(W_1, \mathcal{A})$ and μ_2 on $(W_2, \mathcal{B})$ if and only if

$$R(x,y) \geq -1 \text{ almost everywhere } \mu_1 \times \mu_2,$$

$$\int_{W_2} R(x,y)\, d\mu_2 = 0 \text{ almost everywhere } \mu_1,$$

and

$$\int_{W_1} R(x,y)\, d\mu_1 = 0 \text{ almost everywhere } \mu_2.$$

PROOF. If μ is a probability measure on $\mathcal{A} \times \mathcal{B}$ then $\mu(E) \geq 0$ for all sets in $\mathcal{A} \times \mathcal{B}$. From this it follows that $1 + R(x,y) \geq 0$ almost everywhere $\mu_1 \times \mu_2$ and hence $R(x,y) \geq -1$ almost everywhere $\mu_1 \times \mu_2$.

Suppose that μ is a probability measure with marginal μ_1 on W_1. By definition

$$\mu_1(A) = \mu(A \times \Omega_2)$$

for all A in $\mathcal{A}$, and by hypothesis

$$\mu(A \times \Omega_2) = \int_{A \times W_2} \{1 + R(x,y)\}\, d\mu_1 \times \mu_2.$$

By an application of Fubini's theorem we get the following equality

$$\mu_1(A) = \int_A \left\{ \int_{W_2} (1 + R(x,y)\, d\mu_2) \right\} d\mu_1$$

$$= \int_A \left\{ \int_{W_2} 1\, d\mu_2 + \int_{W_2} R(x,y)\, d\mu_2 \right\} d\mu_1.$$

Since

$$\mu_1(A) = \int_A 1 \cdot d\mu_1 ,$$

we have by the uniqueness assertion of the Radon-Nikodym theorem that

$$\int_{W_2} 1 \cdot d\mu_2 + \int_{W_2} R(x,y)\, d\mu_2 = 1 \text{ almost everywhere } \mu_1 .$$

Now, μ_2 is a probability measure so that

$$\int_{W_2} 1 \cdot d\mu_2 = 1.$$

Hence, it is immediate that

$$\int_{W_2} R(x,y)\, d\mu_2 = 0 \text{ almost everywhere } \mu_1 .$$

Similarly we find that

$$\int_{W_1} R(x,y)\, d\mu_1 = 0 \text{ almost everywhere } \mu_2 .$$

This finishes the necessity part of the proof.

Suppose that $R(x,y) \geq -1$ almost everywhere $\mu_1 \times \mu_2$ and that

$$\int_{W_2} R(x,y)\, d\mu_2 = 0 \text{ almost everywhere } \mu_1$$

and

$$\int_{W_1} R(x,y)\, d\mu_1 = 0 \text{ almost everywhere } \mu_2 .$$

Then μ defined by

$$\mu(E) = \int_E (1 + R(x,y))d\mu_1 \times \mu_2$$

for all E in $\mathcal{Q} \times \mathcal{B}$ will be a probability measure. Because $\mu(E) \geq 0$ for all measurable E and

$$\mu(W_1 \times W_2) = \int_{W_1 \times W_2} (1 + R(x,y))d\mu_1 \times d\mu_2 ,$$

we have by Fubini's theorem

$$\mu(W_1 \times W_2) = \int_{W_1} \left[\int_{W_2} (1 + R(x,y))d\mu_2 \right] d\mu_1$$

$$= \int_{W_1} \left[\int_{W_2} 1\, d\mu_2 + \int_{W_2} R(x,y)\, d\mu_2 \right] d\mu_1$$

$$= \int_{W_1} 1 \cdot d\mu_1 = 1.$$

That μ will have the marginal μ_1 on W_1 is clear from the definition of the marginal and upon replacing W_1 by any A in $\mathfrak{A}$ in the above computation.

A similar proof will show that μ_2 is the marginal measure of μ on W_2.

Q.E.D.

For a simple method of obtaining examples of such functions, R(x,y), see Chapter 4.

The following theorem shows that the bivariate measures we are considering are a special case of those of Theorem 2.3.

<u>THEOREM 3.2</u>. If the bivariate measure μ is absolutely continuous with respect to the product measure $\mu_1 \times \mu_2$ then there exists a measure $\mu_2(\ |x)$ on W_2 for almost all x such that

$$\mu(A \times B) = \int_A \mu_2(B|x)\, d\mu_1 \quad \text{for}$$

every measurable set of the form A x B where A belongs to $\mathfrak{A}$ and B belongs to $\mathfrak{B}$.

PROOF. By hypothesis

$$\mu(A \times B) = \int_{A \times B} [1 + R(x,y)] \, d\mu_1 \times \mu_2 .$$

This identity by Fubini's theorem is equivalent to

$$\mu(A \times B) = \int_A \left[\int_B [1 + R(x,y)] \, d\mu_2 \right] d\mu_1 .$$

A necessary and sufficient condition for a measurable subset E of $W_1 \times W_2$ to have $\mu_1 \times \mu_2$ measure zero is that

$$\mu_2(y|(x,y) \text{ belongs to } E) = 0 \text{ for almost all } x.$$

For this result see [12] page 147. From this result and from the fact that $R(x,y) \geq -1$ almost everywhere $\mu_1 \times \mu_2$, it follows that $R(x,y) \geq -1$ almost everywhere μ_2 for almost all x.

We also have that

$$\int_{W_2} R(x,y) \, d\mu_2 = 0 \text{ almost everywhere } \mu_1 ,$$

and hence

$$\mu_2(\ |x) = \int [1 + R(x,y)]\, d\mu_2$$ is a probability measure on W_2 for almost all x.

Q.E.D.

By the above theorem we are able to apply the results of Theorem 2.3 to the measures we are considering.

COROLLARY 3.3. If μ meets the conditions of Theorem 3.2 the for B in $\mathcal{B}$

$$\mu_2(B|x) = \int_B [1 + R(x,y)]\, d\mu_2 \text{ where}$$

$\mu_2(B|x)$ is the conditional probability of B given x.

PROOF. This result follows immediately from Theorems 2.3 and 3.2.

Q.E.D.

COROLLARY 3.4. If μ meets the conditions of Theorem 3.2 and if $\alpha(x,y)$ is integrable in μ then

$$\int_{W_2} |\alpha(x,y)|[1 + R(x,y)]\, d\mu_2 < \infty \text{ for almost all } x.$$

PROOF. By Theorem 2.3 $\alpha(x,y)$ is integrable with respect to $\mu_2(\ |x)$ for almost all x.

Hence,

$$\int_{W_2} |\alpha(x,y)| d\mu_2(\ |x) < \infty \text{ for almost all } x.$$

Since

$\mu_2(\ |x) = \int [1 + R(x,y)] d\mu_2$, we have a corollary to the Radon-Nikodym theorem that

$$\int_{W_2} |\alpha(x,y)| d\mu_2(\ |x) = \int_{W_2} |\alpha(x,y)|(1 + R(x,y)) d\mu_2$$

for all functions $\alpha(x,y)$ which are integrable with respect to $\mu_2(\ |x)$. From this it follows that

$$\int_{W_2} |\alpha(x,y)|(1 + R(x,y)) d\mu_2 < \infty \text{ for almost all } x.$$

Q.E.D.

COROLLARY 3.5. If μ meets the conditions of Theorem 3.2 and if $h(y)$ is integrable with respect to μ_2, then it is integrable with respect to $\mu_2(\ |x)$ for almost all x and

$$E[h(y)|x] = \int_{W_2} h(y)(1 + R(x,y))d\mu_2 .$$

PROOF. By Theorem 2.3

$$E[h(y)|x] = \int_{W_2} h(y)d\mu_2(\ |x) \text{ exists}$$

for almost all x. By a procedure similar to that used in Corollary 3.4 we have

$$E[h(y)|x] = \int_{W_2} h(y)(1 + R(x,y))d\mu_2 .$$

Q.E.D.

COROLLARY 3.6. If μ meets the conditions of Theorem 2.2 and if $\alpha(x,y)$ is integrable with respect to μ and $\mu_1 \times \mu_2$ then $\alpha(x,y)R(x,y)$ is integrable in $\mu_1 \times \mu_2$.

PROOF. By a corollary to the Radon-Nikodym theorem

$$\int_{W_1 \times W_2} \alpha(x,y)d\mu = \int_{W_1 \times W_2} \alpha(x,y)(1 + R(x,y))d\mu_1 \times \mu_2$$

for all functions $\alpha(x,y)$ which are integrable with respect to μ.

Since $\alpha(x,y)$ is integrable with respect to μ, $\alpha(x,y)(1 + R(x,y))$ is integrable with respect to $\mu_1 \times \mu_2$. Also, $\alpha(x,y)$ is integrable with respect to $\mu_1 \times \mu_2$ by hypothesis.

Since $\alpha(x,y) \cdot R(x,y) = \alpha(x,y)(1 + R(x,y)) - \alpha(x,y)$ and because the difference of two integrable functions is integrable, $\alpha(x,y)R(x,y)$ is integrable with respect to $\mu_1 \times \mu_2$.

Q.E.D.

<u>COROLLARY 3.7</u>. If μ meets the conditions of Theorem 3.2 and if $h(y)$ is integrable with respect to μ_2 then $h(y) \cdot R(x,y)$ is integrable with respect to μ_2 for almost all x.

<u>PROOF</u>. By Theorem 2.3 $h(y)$ is integrable with respect to $\mu_2(\ |x)$ for almost all x.

Since

$$\int_{W_2} h(y)d\mu_2(\ |x) = \int_{W_2} h(y)(1 + R(x,y))d\mu_2$$

$h(y)(1 + R(x,y))$ is integrable with respect to μ_2 for almost all x.

Since $h(y)R(x,y) = h(y)(1 + R(x,y)) - h(y)$ and because the difference of two integrable functions is integrable, $h(y) \cdot R(x,y)$ is integrable with respect to μ_2 for almost all x.

Q.E.D.

It should be noted at this point that there are functions $\alpha(x,y)$ which are integrable with respect to μ and not integrable with respect to $\mu_1 \times \mu_2$ and conversely.

<u>COROLLARY 3.8</u>. If μ meets the conditions of Theorem 3.2, then μ_1 and μ_2 are independent if and only if $R(x,y) = 0$ almost everywhere $\mu_1 \times \mu_2$.

<u>PROOF</u>. This result follows immediately from Theorem 3.2 and Theorem 2.3.

<u>THEOREM 3.9</u>. If μ and μ' are two measures on $(W_1 \times W_2, \mathfrak{A} \times \mathfrak{B})$ with the same marginal μ_1 on W_1 such that for almost all x two probability measures $\mu_2(\ |x)$ and $\mu_2'(\ |x)$ can be defined on the measurable sets B of W_2 satisfying

$$\mu(A \times B) = \int_A \mu_2(B|x)d\mu_1$$

and

$$\mu'(A \times B) = \int_A \mu_2(B|x)d\mu_1$$

for all A belonging to $\mathfrak{A}$. Then $\mu' = \mu$ if and only if $\mu_2(B|x) = \mu_2'(B|x)$ for almost all x and all B in $\mathfrak{B}$.

PROOF. If $\mu_2(B|x) = \mu_2'(B|x)$ for almost every x, then $\mu'(A \times B) = \mu(A \times B)$ for all sets of the form $A \times B$ where A is in $\mathfrak{A}$ and B is in $\mathfrak{B}$.

By the additivity of the measures μ' and μ,

$$\mu'((A_1 \times B_1) \cup (A_2 \times B_2) \cup \ldots \cup (A_n \times B_n))$$

$$= \mu((A_1 \times B_1) \cup (A_2 \times B_2) \cup \ldots \cup (A_n \times B_n))$$

where $\{A_i \times B_i\}$ is a class of disjoint sets of the type given above. Thus the two measures μ and μ' are identical on the algebra of all finite unions of sets of the form $A_i \times B_i$ where A is in $\mathfrak{A}$ and B is in $\mathfrak{B}$. The σ- algebra $\mathfrak{A} \times \mathfrak{B}$ is defined to be the smallest extension of this field. By the uniqueness assertion of the Extension theorem $\mu' = \mu$ on $(W_1 \times W_2, \mathfrak{A} \times \mathfrak{B})$. For the statement and proof of this theorem see [17] page 87. This finishes the proof of sufficiency.

Suppose that $\mu' = \mu$ on $(W_1 \times W_2, \mathfrak{A} \times \mathfrak{B})$. Then for each B

$$\int_A \mu_2(B|x)d\mu_1 = \int_A \mu_2'(B|x)d\mu_1 \text{ for}$$

every A in $\mathfrak{A}$. Hence, using the theorem that if the integral of two functions is the same over all measurable sets the functions are identical almost everywhere, we get that

$$\mu_2(B|x) = \mu'_2(B|x) \text{ for almost all } x.$$

Q.E.D.

The following theorems give sufficient conditions for the two-dimensional measure μ to be absolutely continuous with respect to the measure $\mu_1 \times \mu_2$.

THEOREM 3.10. If the measures μ_1 and μ_2 are independent then μ is absolutely continuous with respect to the product measure.

PROOF. If we put $(1 + R(x,y)) = 1$, this result is obvious.

Q.E.D.

THEOREM 3.11. If one of the marginal measures, say μ_1, is discrete then the two-dimensional measure must be absolutely continuous with respect to the product measure.

PROOF. Since the measure μ_1 is discrete there are at most a countable number of disjoint sets A_i belonging to $\mathfrak{a}$ such that $\mu_1(A_i) > 0$ and such that the only sets of $\mathfrak{a}$ which have positive measure are sets which contain at least one A_i.

No non-null subset of an A_i can be measurable, for suppose $A \subset A_i$, $A \neq \phi$ and A is measurable. The $(A_i - A)$ is measurable and $\mu(A_i) = \mu(A) + \mu(A_i - A)$. But $\mu(A) = 0$ and $\mu(A_i - A) = 0$ because neither A or $A_i - A$ contain a set A_j. However $\mu(A_i) > 0$ which leads to a contradiction.

Because of the results of the above arguments, we may regard the subsets A_i as points of W_1 when we construct the measurable space $\mathfrak{A} \times \mathfrak{B}$.

Let $A_0 = -(\bigcup_{i=1}^{\infty} A_i)$; then A_0 is measurable and $\mu(A_0) = 0$ and $W_1 = \bigcup_{i=0}^{\infty} A_i$.

Let E by any measurable subset of $W_1 \times W_2$ such that $\mu(E) > 0$. We form measurable subsets of $W_1 \times W_2$ from E by taking sets of the form

$$E_0 = [(A_0 \times W_2) \cap E]$$

$$E_1 = [(A_1 \times W_2) \cap E]$$

$$\vdots$$

$$E_n = [(A_n \times W_2) \cap E]$$

$$\vdots$$

The sets E_i are disjoint and $\bigcup_{i=0}^{\infty} E_i = E$.

Because the measure μ is additive we get

$$\mu(\bigcup_{i=0}^{\infty} E_i) = \sum_{i=0}^{\infty} \mu(E_i) = \mu(E).$$

Since $\mu(E) > 0$ by hypothesis, at least one of the sets $\mu(E_i)$ must have positive measure.

Because $\mu(E_0) = \mu((A_0 \times W_2) \quad E) < \mu(A_0 \times W_2) = \mu_1(A_0) = 0$, there is an $i \neq 0$ such that $\mu(E_i) > 0$. Now we may regard the sets A_i as points and by using the theorem that almost every section of a measurable set is a measurable set we find that

$$E_1 = ((A_1 \times W_2) \cap E) = (A_1 \times B)$$

where B is a measurable set in $\mathfrak{B}$.

$$0 < \mu(E_1) = \mu(A_1 \times B) \leq (W_1 \times B) = \mu_2(B),$$

hence $\mu_2(B) > 0$.

Thus, $\mu_1 \times \mu_2(E_1) = \mu_1 \times \mu_2(A_1 \times B) = \mu_1(A_1) \cdot \mu_2(B) > 0$ and $\mu_1 \times \mu_2(E) \geq \mu_1 \times \mu_2(E_1) > 0$. We have established that if $\mu(E) > 0$ then $\mu_1 \times \mu_2(E) > 0$. Since the measures involved here are strictly finite this is sufficient to insure that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$.

Q.E.D.

COROLLARY 3.12. If one of the spaces W_1, say W_1, has at most a countable number of points then the original measure μ is absolutely continuous with respect to the product measure $\mu_1 \times \mu_2$.

PROOF. This is obvious since the measure μ_1 must be discrete.

Q.E.D.

COROLLARY 3.13. If one of the two random variables X and Y takes on at most a countable number of values with probability one, then μ is absolutely continuous with respect to $\mu_1 \times \mu_2$.

PROOF. Let X be the random variable which takes on the values $x_1, x_2, \ldots$ with probability one. Then the sets $[x_1]$, $[x_2]$ ... are identified with the sets $A_1, A_2, \ldots$ and the result is immediate.

Q.E.D.

THEOREM 3.14. If the random variables X and Y are such that μ is absolutely continuous with respect to Lebesgue measure then μ is absolutely continuous with respect to $\mu_1 \times \mu_2$.

PROOF. The proof shall consist of showing that if $\mu(E) > 0$ then $\mu_1 \times \mu_2(E) > 0$ where E is a Lebesgue measurable set. In this theorem $W_1 = W_2 = (-\infty,+\infty)$.

Since μ is absolutely continuous with respect to Lebesgue measure, there exists a measurable function f(x,y) such that

$$\mu(E) = \int_E f(x,y)dxdy$$

for all measurable sets E.

By definition and with the use of Fubini's theorem we get

$$\mu_1(A) = \mu(A \times W_2) = \int_A \int_{W_2} f(x,y)dydx$$

$$= \int_A f_1(x)dx$$

and

$$\mu_2(B) = \int_B \int_{W_1} f(x,y)dxdy = \int_B f_2(y)dy.$$

The measure $\mu_1 \times \mu_2$ is determined on $\mathcal{A} \times \mathcal{B}$ by

$$\mu_1 \times \mu_2 = \int f_1(x) \cdot f_2(y)dxdy$$

and hence $\mu_1 \times \mu_2$ is absolutely continuous with respect to Lebesgue measure.

Suppose that E is a Lebesgue measurable set in $W_1 \times W_2$ such that

$$\mu(E) = \int_E f(x,y)dxdy > 0.$$

Let F be a set such that $f(x,y) > 0$ for all (x,y) in F. Then since F is measurable $E \cap F = G$ is measurable and

$$\mu(G) = \int_{E\cap F} f(x,y)dxdy = \int_E f(x,y)dxdy > 0.$$

Let $G_x = ([y] : (x,y)$ belong to $G)$, that is the x sections of the set G. Let A be a subset of W_1 such that G_x for every x in A is non-measurable or that G_x has Lebesgue measure 0. Then it is well known that the set $\bigcup_{x \text{ in } A} G_x = H$ has Lebesgue measure zero. See [20] volume II page 76.

Let $G_y = ([x] : (x,y)$ in $G)$ and let B be a subset of W_2 such that G_y for every y in B is non-measurable or G_y has Lebesgue measure zero. Then $\bigcup_{Y \text{ in } B} G_y = I$ has Lebesgue measure zero.

Let $(G\cap{-H})_x$ designate the x-sections of $(G\cap{-H})$. Now $f(x,y)$ is strictly positive on G and hence is strictly positive on every x-section of $(G\cap{-H})$. Since the Lebesgue measure of each non-null x-section of $(G\cap{-H})$ is positive we have

$$f_1(x) = \int_{W_2} f(x,y)dy \geq \int_{(G\cap-\,H)_x} f(x,y)dy > 0$$

for each non-null x section. Hence for all x for which there exists a y such that (x,y) belongs to $(G\cap{-H})$, $f_1(x) > 0$.

In a similar fashion we may prove that $f_2(y) > 0$ for all y for which there exists a point x such that (x,y) belongs to $(G\cap{-I})$.

From the above it is obvious that $f_1(x) \cdot f_2(y) > 0$ for all (x,y) belonging to $(G\cap - I\cap - H) \subset E$.

Since the Lebesgue measure of I and H are zero, it is clear that $(G \cap - I \cap - H)$ will have positive Lebesgue measure. Thus, combining the facts that $(G \cap - I \cap - H)$ has positive Lebesgue measure and that $f_1(x) \cdot f_2(y) > 0$ on the set $(G \cap - H \cap - I)$ we find that

$$\mu_1 \times \mu_2(E) \geq \mu_1 \times \mu_2(G \cap - H \cap - I)$$

$$= \int_{G \cap - H \cap - I} f_1(x)\, f_2(y)\, dxdy > 0.$$

Hence $\mu(E) > 0$ implies that $\mu_1 \times \mu_2(E) > 0$.

Q.E.D.

THEOREM 3.15. If X and Y are two random variables such that $\mu = \lambda\alpha + (1-\lambda)\gamma$ where α is a probability measure absolutely continuous with respect to Lebesgue measure, γ is a discrete probability measure, and λ is a real number such that $0 \leq \lambda \leq 1$, then μ is absolutely continuous with respect to $\mu_1 \times \mu_2$.

PROOF. If $\lambda = 0$ or if $\lambda = 1$ then the result follows by Theorem 3.14 or by Theorem 3.11. Hence, we may assume that $0 < \lambda < 1$.

Let α_1 be the marginal of α on W_1 and α_2 be the marginal of α on W_2. Let γ_1 be the marginal of γ on W_1 and γ_2 the marginal of γ on W_2.

Using the definition of marginal measures we find that

$$\mu_1 = \lambda\alpha_1 + (1 - \lambda)\gamma_1$$

and

$$\mu_2 = \lambda\alpha_2 + (1 - \lambda)\gamma_2 .$$

Let A be a measurable subset of W_1 and B be a measurable subset of W_2. Then

$$\mu_1 \times \mu_2(A \times B) = \mu_1(A) \cdot \mu_2(B)$$

$$= (\lambda\alpha_1(A) + (1-\lambda)\gamma_1(A)) \cdot (\lambda\alpha_2(B) + (1-\lambda)\gamma_2(B)).$$

Hence

$$\mu_1 \times \mu_2(A \times B) = \lambda^2\alpha_1(A)\alpha_2(B)$$

$$+ \lambda(1-\lambda)(\alpha_1(A)\gamma_2(B) + \gamma_1(A)\alpha_2(B))$$

$$+ (1-\lambda)^2\gamma_1(A) \cdot \gamma_2(B)_1 = \lambda^2\alpha_1 \times \alpha_2(A \times B)$$

$$+ \lambda(1-\lambda)(\alpha_1 \times \gamma_2(A \times B) + \gamma_1 \times \alpha_2(A \times B))$$

$$+ (1-\lambda)^2\gamma_1 \times \gamma_2(A \times B).$$

The above equality will hold for all measurable rectangles in $W_1 \times W_2$. Hence by a process similar to that used in Theorem 3.9 we find that

$$\mu_1 \times \mu_2(E) = \lambda^2\alpha_1 \times \alpha_2(E)$$

$$+ \lambda(1-\lambda)(\alpha_1 \times \gamma_2(E) + \gamma_1 \times \alpha_2(E)) + (1-\lambda)^2\gamma_1 \times \gamma_2(E)$$

for all measurable E in $W_1 \times W_2$.

Suppose that $\mu_1 \times \mu_2(E) = 0$. Then since $0 < \lambda < 1$, we find that

$$\alpha_1 \times \alpha_2(E) = 0$$

and

$$\gamma_1 \times \gamma_2(E) = 0.$$

By Theorem 3.14 or Theorem 3.11 α is absolutely continuous with respect to $\alpha_1 \times \alpha_2$ and γ is absolutely continuous with respect to $\gamma_1 \times \gamma_2$. Hence

$$\alpha(E) = 0$$

and

$$\gamma(E) = 0.$$

Thus $\mu(E) = 0$. Since all the measures involved are finite valued this is sufficient to show that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$.

Q.E.D.

It is worth noting at this point that by Theorems 3.11, 3.14, 3.15 the most common types of two-dimensional random vectors are included in the class we are considering. That is, the discrete, continuous, and mixed types are in this class.

For the remainder of this chapter we are going to prove a number of isolated and special results which will show the usefulness of considering the special class of bivariate probability measures which are absolutely continuous with respect to the product of their marginals.

THEOREM 3.16. If μ_1 is discrete and if x_i is such that $\mu_1(x_i) > 0$ then $R(x_i,y)$ exists and $R(x_i,y) \le (\frac{1}{\mu_1(x_i)} - 1)$ almost everywhere μ_2.

PROOF. $R(x_i,y)$ exists by Theorem 3.11. Let $\{x_i\}$ $(i = 1, 2, 3, \ldots)$ be the set of all x_i such that $\mu_1(x_i) > 0$. Then by Theorem 3.1

$$\sum_{i=1}^{\infty} R(x_i,y)\mu_1(x_i) = 0 \text{ for almost all } y$$

and

$$R(x,y) \ge -1 \text{ almost everywhere } \mu_1 \times \mu_2 .$$

Suppose that $R(x_i,y) > (\frac{1}{\mu_1(x_i)} - 1)$ for all y in a set B belonging to $\mathfrak{B}$ such that $\mu_2(B) > 0$. Let E be the set of $W_1 \times W_2$ such that $R(x_i,y) < -1$. Let $E_i = (x_i \times W_2) \cap E$, then each E_i is a set of the form $(x_i \times B_i)$ where B_i is in $\mathfrak{B}$. $\mu_1 \times \mu_2(E_i) = 0$ and hence $\mu_2(B_i) = 0$ for each i. Thus $\mu_2(\bigcup_{i=1}^{\infty} B_i) = 0$. Let $B_0 = \{B \cap -(\bigcup_{i=1}^{\infty} B_i)\}$; then $\mu_2(B_0) > 0$ and $R(x_i,y) > (\frac{1}{\mu_1(x_i)} - 1)$ for all y in B_0. Further, if y belongs to B_0 then $R(x_i,y) \ge -1$ for all x_i.

Then

$$R(x_i,y)(\mu_1(x_i)) > (1 - \mu_1(x_i)) \qquad (1)$$

and since μ_1 is a probability measure we have the following set of inequalities

$$\sum_{i=1}^{\infty} (-1)\mu_1(x_i) \geq -1 \tag{2}$$

and

$$\sum_{\substack{j=1 \\ i \neq j}}^{\infty} R(x_j,y)\mu_1(x_j) \geq \sum_{\substack{j=1 \\ j \neq i}}^{\infty} (-1)\mu_1(x_j)$$

for all y in B_0. Combining (1) and (2) we get the following:

$$R(x_i,y)\mu_1(x_i) + \sum_{j=1}^{\infty} (-1)\mu_1(x_j) > 0$$

or

$$R(x_i,y)\mu_1(x_i) + \sum_{\substack{j=1 \\ j \neq i}}^{\infty} (-1)\mu_1(x_j) > 0 .$$

Using inequality (3) we may transform this last inequality into

$$R(x_i,y)\mu_1(x_i) + \sum_{\substack{j=1 \\ j \neq i}}^{\infty} R(x_i,y)\mu_1(x_j) > 0$$

or

$$\sum_{i=1}^{\infty} R(x_i,y)\mu_1(x_i > 0$$

for all y in B_0. But B_0 is a set of positive μ_2 measure and hence this is a contradiction.

Q.E.D.

COROLLARY 3.17. If μ_1 is finite discrete then $R(x,y)$ is bounded almost everywhere $\mu_1 \times \mu_2$.

PROOF. If $x_1, \ldots, x_n$ are the points for which $\mu_1(x_i) > 0$, then by Theorem 3.8

$$R(x_1,y) \le \left(\frac{1}{\mu_1(x_1)} - 1\right) \text{ almost everywhere } \mu_2$$

$$\vdots$$

$$R(x_n,y) \le \left(\frac{1}{\mu_1(x_n)} - 1\right) \text{ almost everywhere } \mu_2 \; .$$

Hence $R(x,y) \le \max_i \left\{\left(\frac{1}{\mu_1(x_i)} - 1\right)\right\}$ almost everywhere $\mu_1 \times \mu_2$.

Q.E.D.

It is possible to construct an example where the function $R(x,y)$ actually takes on the maximal value given by the theorem, which shows that the bound given is the best possible.

The following theorem gives bounds on the conditional expectation in terms of the ordinary expectation when one of the marginals is discrete.

THEOREM 3.18. If the marginal μ_1 of μ on W_1 is discrete and if

$$E[h(y)] = \int_{W_2} h(y) \, d\mu_2 \text{ exists}$$

then

$$E[h(y)|x_i] \geq E[h(y)] - E[h^+(y)] - (\frac{1}{\mu(x_i)} - 1)E[h^-(y)]$$

$$E[h(y)|x_i] \leq E[h(y)] + E[h^-(y)] + (\frac{1}{\mu(x_i)} - 1)E[h^+(y)]$$

where the x_i are the countable number of points such that $\mu_1(x_i) > 0$, $h^+(y) = \max(h(y),0)$ and $h^-(y) = \max(-h(y),0)$.

PROOF.

$$E[h(y)|x_i] = \int_{W_2} h(y)(1 + R(x_i,y))\, d\mu_2$$

$$= E[h(y)] + E[h(y) \cdot R(x_i,y)]$$

where each of the expected values exist either by hypothesis or by Corollary 3.7.

Let

$$R^+(x_i,y) = \max(0, R(x_i,y))$$

$$R^-(x_i,y) = \max(0,-R(x_i,y)),$$

then

$$E[h(y)R(x_i,y)] \leq E[h^-(y)R^-(x_i,y)] + E[h^+(y)R^+(x_i,y)].$$

By elementary properties of integrals we get

$$E[h^-(y)R^-(x_i,y)] \leq \text{ess sup}\{R^-(x_i,y)\} \cdot E[h^-(y)]$$

and

$$E[h^+(y)R^+(x_i,y)] \leq \text{ess sup}\{R^+(x_i,y)\} \cdot E[h^+(y)],$$

where $\text{ess sup } f(y) = \inf\{M: \mu_2(y: f(y) > M) = 0\}$.

Now by Theorem 3.1 and Theorem 3.16 ess sup $R^-(x_1,y) \leq 1$ and ess sup $R^+(x_1,y) \leq (\frac{1}{\mu_1(x_1)} - 1)$. Thus

$$E[h(y)R(x_1,y)] \leq E[h^-(y)] + (\frac{1}{\mu_1(x_1)} - 1)E[h^+(y)]$$

and

$$E[h(y)|x_1] \leq E[h(y)] + E[h^-(y)] + (\frac{1}{\mu_1(x_1)} - 1)E[h^+(y)].$$

A similar proof will show that

$$E[h(y)|x_1] \geq E[h(y)] - E[h^+(y)] - (\frac{1}{\mu_1(x_1)} - 1)E[h^-(y)].$$

Q.E.D.

THEOREM 3.19. If μ_1 is discrete and if $h(y)$ is a function on W_2 such that the expectation $E[h(y)]$ exists then the conditional expectation of $h(y)$ given x_1 takes the form

$$E[h(y)|x_1] = E[h(y)] + \mathfrak{L}_{x_1}(h(y))$$

where $\mathfrak{L}_{x_1}$ is a continuous linear functional over the space of μ_2 integrable functions on W_2.

PROOF. As in Theorem 3.18 we find that

$$E[h(y)|x_1] = E[h(y)] + E[h(y) \cdot R(x_1,y)]$$

for all functions $h(y)$ whose expectation exists.

By definition

$$E[h(y) \cdot R(x_1,y)] = \int_{W_2} h(y) \cdot R(x_1,y) \, d\mu_2 .$$

By Theorem 3.1 and Theorem 3.8 for fixed x_1, the

$$\text{ess sup } \{|R(x_1,y)|\} \leq \max \{1, \frac{1}{\mu(x_1)} - 1\}.$$

Hence $R(x_1,y)$ is bounded in the ess sup norm and thus by a standard result of classical function space theory

$$\int_{W_1} (\ ) \cdot R(x_1,y)\, d\mu_2 = \mathfrak{L}_{x_1}(\ )$$

defines a continuous linear functional over the μ_2 integrable functions on W_2. See [19], page 249. Thus

$$E[h(y)|x_1] = E[h(y)] + \mathfrak{L}_{x_1}(h(y)).$$

Q.E.D.

Using the notation and techniques of this chapter we will derive results about some standard parameters of mathematical statistics.

THEOREM 3.20. If X and Y are two random variables whose covariance exists and such that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ then

$$\text{covariance of X and Y} = \text{cov.}(X,Y) = \int_{W_1}\int_{W_2} x \cdot y\, R(x,y)\, d\mu_1 \times \mu_2 .$$

PROOF. It is a standard result of mathematical statistics that

$$\text{cov}\,(X,Y) = E[X\cdot Y] - E[X]E[Y].$$

As in the proof of Corollary 3.6 we have

$$E[X\cdot Y] = \int_{W_1}\int_{W_2} (x\cdot y)(1 + R(x,y))\, d\mu_1 \times \mu_2$$

$$= \int_{W_1}\int_{W_2} (x \cdot y)\, d\mu_1 \times \mu_2 + \int_{W_1}\int_{W_2} x \cdot y\, R(x,y)\, d\mu_1 \times d\mu_2 .$$

By Fubini's theorem

$$\int_{W_1}\int_{W_2} x \cdot y\, d\mu_1 \times \mu_2 = \int_{W_1} x\, d\mu_1 \cdot \int_{W_2} y\, d\mu_2$$

$$= E[X]E[Y].$$

Hence

$$E[X \cdot Y] = E[X]E[Y] + \int_{W_1}\int_{W_2} x \cdot y\, R(x,y)\, d\mu_1 \times \mu_2$$

and it is immediate that

$$\operatorname{cov}(X,Y) = \int_{W_1}\int_{W_2} x \cdot y\, R(x,y)\, d\mu_1 \times \mu_2 .$$

Q.E.D.

<u>COROLLARY 3.21</u>. If X and Y are two random variables which meet the conditions of Theorem 2.11 and are such that $E[X^2]$ and $E[Y^2]$ exist then the correlation coefficient $\rho(X,Y)$ takes the form

$$\rho(x,y) = \frac{\int_{W_1 \times W_2} x \cdot y\, R(x,y)\, d\mu_1 \times \mu_2}{\sqrt{E[(x - E[x])^2]\ \ E[(y - E(y))^2]}} .$$

PROOF. This result follows immediately from the definition of the correlation coefficient

$$\rho(x,y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{E[(x - E(x))^2] \quad E[(y - E(y))^2]}} .$$

Q.E.D.

For the class of bivariate distributions which we are considering the regression curves take an especially simple form.

THEOREM 3.22. If X and Y are such that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ and if $E[y]$ exists then the regression curve of y on x, $m_1(x)$ takes the form

$$m_1(x) = E[y|x] = E[y] + E[y \cdot R(x,y)]$$

PROOF. By Corollary 3.5

$$m_1(x) = E[y|x] = \int_{W_2} y(1 + R(x,y)) \, d\mu_2$$

$$= E[y] + E[y \cdot R(x,y)]$$

where all the expectations exist either by hypothesis or by Corollary 3.7.

Q.E.D.

Similarly $m_2(y)$, the regression curve of x on y, takes the form $E[x|y] = E[x] + E[x \cdot R(x,y)]$ when $E[x]$ exists.

This leads to the following interesting theorem.

THEOREM 3.23. If X and Y are two random variables which meet the condition of Theorem 3.22 and are such that their covariance exists then a sufficient condition for $\text{cov}\ (X,Y) = 0$ is that the regression curve $m_1(x) = E[y|x]$ be equal to a constant almost everywhere μ_1.

PROOF. Suppose that

$$E[y|x] = c \text{ almost everywhere } \mu_1;$$

then

$$E[y|x] = E[y] + E[y \cdot R(x,y)] = c. \qquad (4)$$

It is well known that

$$E[E[y|x]] = E[Y];$$

thus

$$E[y] = E[c].$$

Hence we have that $c = E[y]$, and from this and (4) it follows that

$$E[y \cdot R(x,y)] = 0 \text{ almost everywhere } \mu_1. \qquad (5)$$

By Theorem 3.11

$$\text{cov}\ (X,Y) = \int_{W_1 \times W_2} x \cdot y\, R(x,y)\, d\mu_1 \times \mu_2,$$

which by Fubini's theorem is equal to

$$\int_{W_1} x \int_{W_2} y \cdot R(x,y)d\mu_2 d\mu_1 = \int_{W_1} x \cdot E[y \cdot R(x,y)]d\mu_1$$

$$= 0 \text{ by } (5).$$

Q.E.D.

Continuing our discussion of the regression curve we obtain the result contained in the next theorem.

THEOREM 3.24. If X and Y are two random variables such that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ and if $E[x \cdot E[y|x]]$, $E[y \cdot E[x|y]]$, $E[x]$, and $E[y]$ exist then

$$E[x \cdot E[y|x]] = E[y \cdot E[x|y]].$$

PROOF.

$$E[x \cdot E[y|x]] = \int_{W_1} x \cdot E[y|x]d\mu_1$$

$$= \int_{W_1} x(E[y] + E[y \cdot R(x,y)])d\mu_1$$

by Theorem 3.22.

This last integral is equal to

$$\int_{W_1} x\, E[y]d\mu_1 + \int_{W_1} x \int_{W_2} y \cdot R(x,y)\, d\mu_2\, d\mu_1$$

which by an application of Fubini's theorem takes the form

$$E[x] \cdot E[y] + \int_{W_2} y \int_{W_1} x\, R(x,y) d\mu_1 \, d\mu_2 = \int_{W_2} \Big[y$$

$$x\, E[x] + y\, E[x \cdot R(x,y)] \Big] d\mu_2$$

$$= E[y(E[x] + E[x \cdot R(x,y)])]$$

$$= E[y \cdot E[x|y]].$$

Q.E.D.

In the next chapter we will discuss the simplest type of functions which are members of the class R(x,y).

CHAPTER 4
SIMPLEST TYPE OF DEPENDENCE

In this chapter we are going to study bivariate distributions where the function R(x,y) has the simplest possible form. That is

$$R(x,y) = f(x) \cdot g(y)$$

where

$$\int_{W_1} f(x)\, d\mu_1 = 0$$

and

$$\int_{W_2} g(y)\, d\mu_2 = 0.$$

If either f(x) = 0 almost everywhere μ_1 or g(y) = 0 almost everywhere μ_2 then R(x,y) = 0 almost everywhere $\mu_1 \times \mu_2$. Hence we will always assume that f(x) and g(y) are non-zero on some set of positive measure.

Certain special types of such functions have been considered before. (See Morgenstern [18], Gumbel [9], [10], [11] and Farlie [3], [4], [5].) However this is the first examination of such functions in their complete generality.

Before deriving a characterization of such special bivariate distributions we will need the following lemma.

LEMMA 4.1. If $R(x,y) = f(x) \cdot g(y)$ then $|f(x)|$ is bounded almost everywhere μ_1 and $|g(y)|$ is bounded almost everywhere μ_2.

PROOF. By construction

$$\int_{W_1} f(x)d\mu_1 = 0,$$

and since $f(x)$ is different from zero on a set of positive measure $f(x)$ must be strictly positive on some set A_1 and strictly negative on some set A_2 such that $\mu_1(A_1) > 0$ and $\mu_1(A_2) > 0$. Thus there will exist sets A'_1 and A'_2 with positive measure such that $f(x) \geq E_1 > 0$ on A'_1 and $f(x) \leq E_2 < 0$ on A'_2.

Suppose that $g(y) > (-\frac{1}{E_2})$ on a set B_1 such that $\mu_2(B_1) > 0$. Then $R(x,y) = f(x)\ g(y) < -1$ on the set $(A'_2 \times B_1)$. But $\mu_1 \times \mu_2(A'_2 \times B_1) = \mu_1(A'_2) \cdot \mu_2(B) > 0$, and since by Theorem 3.1 $R(x,y) \geq -1$ almost everywhere $\mu_1 \times \mu_2$ this leads to a contradiction. Hence $g(y) \leq (-\frac{1}{E_2})$ almost everywhere μ_2. A similar argument will show that $g(y) \geq (-\frac{1}{E_1})$ almost everywhere μ_2. Thus $|g(y)|$ is bounded almost everywhere μ_2.

By the same procedure it is possible to show that $|f(x)|$ is bounded almost everywhere μ_1.

Q.E.D.

The conclusion of the above lemma can be written in classical function space terminology as $f(x)$ belongs to L_∞ with respect to μ_1 and $g(y)$ belongs to L_∞ with respect to μ_2.

The next theorem is the central result of this chapter.

THEOREM 4.2. If a bivariate measure μ with marginals μ_1 and μ_2 meets the conditions of Theorem 2.3, then the conditional expectation $E[h(y)|x]$ of functions $h(y)$ such that $E[h(y)]$ exists has the form

$$E[h(y)|x] = E[h(y)] + f(x)\mathfrak{L}(h(y))$$

almost everywhere μ_1 where $\mathfrak{L}(h(y))$ is a continuous linear functional over the μ_2 integrable functions $h(y)$ on W_2 if and only if μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ and the Radon-Nikodym derivative of μ with respect to $\mu_1 \times \mu_2$ is given by $1 + R(x,y) = 1 + f(x)\ g(y)$.

PROOF. Assume that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ and $R(x,y) = f(x)\ g(y)$. Then the defining equation for the conditional expectation of $h(y)$ given x is by Corollary 3.5.

$$E[h(y)|x] = \int_{W_2} h(y)(1 + f(x)\ g(y))d\mu_2.$$

Hence we have that

$$E[h(y)|x] = E[h(y)] + f(x) \int_{W_2} h(y) \cdot g(y)\, d\mu_2 .$$

By Lemma 4.1 g(y) belongs to L_∞ with respect to μ_2. Thus, using a result from classical function space theory we find that

$$\int_{W_2} (\quad)\, g(y)\, d\mu_2 = \mathfrak{L}(\quad)$$

defines a continuous linear functional on the μ_2 integrable functions on W_2. See [19], page 249. Hence we have that

$$E[h(y)|x] = E[h(y)] + f(x)\, \mathfrak{L}(h(y)).$$

This finishes the proof of sufficiency.

Suppose that for any function h(y) integrable over W_2 with respect to μ_2, the conditional expectation of h(y) takes the form

$$E[h(y)|x] = E[h(y)] + f(x)\, \mathfrak{L}(h(y))$$

where $\mathfrak{L}(\quad)$ is a continuous linear functional. Since the measure μ_2 is finite, for every continuous linear functional $\mathfrak{L}(\quad)$ on the μ_2 integrable functions there exists a function g(y) in L_∞ with respect to μ_2 such that

$$\mathcal{L}(\quad) = \int_{W_2} (\qquad) \; g(y) \; d\mu_2 .$$

We define a signed measure μ' on the space $(W_1 \times W_2, \mathfrak{A} \times \mathfrak{B})$ by

$$\mu' = \int (1 + f(x) \; g(y)) d\mu_1 \times \mu_2 .$$

The above integral represents the indefinite integral and is thus a set function on the $\mu_1 \times \mu_2$ measurable sets. Since $\mu_2[\;|x]$ defines a probability measure on W_2, $f(x) \cdot g(y) \geq -1$ almost everywhere $\mu_1 \times \mu_2$. For suppose there exists a set E such that $\mu_1 \times \mu_2(E) > 0$ and $f(x) \cdot g(y) < -1$ on E. Let $E_x = \{y|(x,y)$ is in $E\}$; then since $\mu_1 \times \mu_2(E) > 0$ there will exist a set A of positive μ_1 measure such that if x is in A, $\mu_2(E_x) > 0$. Since $f(x) \cdot g(y) < -1$ on E we find that

$$\mu_2(E_x|x) = \mu_2(E_x) + \int_{E_x} f(x) \; g(y) \; d\mu_2 < 0.$$

Hence for a set of positive μ_1 measure $\mu_2(\;|x)$ does not define a probability measure on W_2. This is a contradiction.

Since

$$\int_{W_1} E[h(y)|x] \; d\mu_1 = E[h(y)]$$

and

$$E[c|x] = c,$$

we find that

$$\int_{W_1} f(x)\ d\mu_1 = 0$$

and

$$\int_{W_2} g(y)\ d\mu_2 = 0.$$

Hence $R(x,y) = f(x) \cdot g(y)$ meets the hypothesis of Theorem 3.1 and μ' is a probability measure with marginals μ_1 and μ_2.

By Corollary 3.4

$$\mu_2'(\ \ |x) = \int (1+f(x)\ g(y))d\mu_2 = \mu_2(\ \ \ )+f(x)\int g(y)d\mu_2.$$

We have that

$$\mu_2(\ \ |x) = \mu_2(\ \ \ ) + f(x) \int g(y)\ d\mu_2$$

where the integral represents the indefinite integral and is thus a set function. Thus by Theorem 3.9

$$\mu = \mu'.$$

Q.E.D.

<u>COROLLARY 4.3</u>. If $W_2 = (-\infty,+\infty)$ then for all h(y) integrable with respect to μ_2,

$$E[h(y)|x] = E[h(y)] + f(x)\mathfrak{L}(h(y)),$$

where $\mathfrak{L}(h(y))$ is a continuous linear functional on the μ_2 integrable function, if and only if μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ and $R(x,y) = f(x)\ g(y)$.

<u>PROOF</u>. This follows from Theorem 4.2 and Theorem 2.4.

Q.E.D.

<u>COROLLARY 4.4</u>. If μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ then for all h(y) integrable with respect to μ_2

$$E[h(y)|x] = E[h(y)] + f(x)\mathfrak{L}(h(y))$$

if and only if $R(x,y) = f(x) \cdot g(y)$.

<u>PROOF</u>. This follows from Theorem 4.2 and Theorem 3.2.

Q.E.D.

It seems reasonable in view of the preceeding theorem to regard bivariate distributions where $R(x,y) = f(x) \cdot g(y)$ as the simplest generalization of independence. The following theorem and its corollary give further support to this statement.

<u>THEOREM 4.5</u>. Let X and Y be two random variables such that μ is absolutely continuous with respect to Lebesgue measure and such that $R(x,y) = f(x) \cdot g(y)$. Then if $F_1(x) = 0$ for $x < 0$ and $F_2(y) = 0$ for $y < 0$ the random

variables X and Y are independent if and only if the density function of $Z = X + Y$ is equal to the convolution of the density functions $f_1(x)$ and $f_2(y)$.

PROOF. Let $Z = X + Y$; then, as is well known, the density function of Z is given by

$$f_3(z) = \int_{-\infty}^{\infty} f_1(x)\ f_2(z-x)(1 + R(x,z-x))\ dx \qquad (1)$$

If X and Y are independent then $R(x,y) = 0$ and from (1) it follows that

$$f_3(z) = \int_{-\infty}^{\infty} f_1(x)\ f_2(z-x)\ dx.$$

Thus the density of Z is equal to the convolution of the two density functions.

Assume that X and Y are not independent. That is that $f(x) \cdot g(y) \neq 0$ on some set of positive $\mu_1 \times \mu_2$ measure. The measures μ_1 and μ_2 are determined on the real line by

$$\mu_1(A) = \int_A f_1(x)\ dx$$

and by

$$\mu_2(B) = \int_B f_2(y)\ dy.$$

Since $f(x) \cdot g(y) \neq 0$ on a set of positive $\mu_1 \times \mu_2$ measure, it follows that $f(x) \neq 0$ on a set of μ_1 positive measure and $g(y) \neq 0$ on a set of μ_2 positive measure. Hence we get that $f(x) \cdot f_1(x) \neq 0$ on a set of positive Lebesgue measure and $g(y) \cdot f_2(y) \neq 0$ on a set of positive Lebesgue measure.

By hypothesis $f_1(x) = 0$ for $x < 0$ and $f_2(y) = 0$ for $y < 0$. Hence (1) becomes in this case

$$f_3(z) = \int_0^z (f_1(x)\ f_2(z-x) + f_1(x)\ f(x)\ f_2(z-x) g(z-x)) dx$$

or

$$f_3(z) = \int_0^z f_1(x) f_2(z-x) dx + \int_0^z f_1(x) f(x) f_2(z-x) g(z-x) dx.$$

The existence of all the above integrals is established by standard theorems about convolutions. See for example [1].

Let

$$h(z) = \int_0^z f_1(x)\ f(x)\ f_2(z-x)\ g(z-x)\ dx.$$

Hence $h(z)$ is equal to the convolution of the two functions $f_1(x) \cdot f(x)$ and $f_2(y) \cdot g(y)$. Each of these functions is non-zero on a set with positive Lebesgue measure. Hence by a famous theorem due to Titchmarsh [26] $h(z) \neq 0$ on a set with positive Lebesgue measure. Thus if X and Y are not independent the density function of their sum is not equal to the convolution of their density functions.

Q.E.D.

It is obvious that a similar theorem will hold if X and Y take on only negative values with probability one.

COROLLARY 4.6. Let X and Y be two random variables which meet the hypothesis of Theorem 4.5. Then X and Y are independent if and only if the characteristic function of $Z = X + Y$ is equal to the product of their characteristic function.

PROOF. It is a standard theorem of mathematical statistics that if two X and Y are independent then the characteristic function of their sum $Z = X + Y$ is equal to the product of their characteristic functions.

Let $\varphi_1(t)$ be the characteristic function of X, $\varphi_2(s)$ be the characteristic function of Y, and $\varphi_3(w)$ be the characteristic function of Z. From the proof of Theorem 4.2 and the definition of the characteristic function we get

$$\varphi_3(w) = \int_{-\infty}^{\infty} f_3(z) \exp(iwz)\, dz$$

$$= \int_{-\infty}^{\infty} \exp(iwz) \int_{0}^{z} f_1(x) f_2(z-x)(1+f(x)g(z-x))\,dx\, dz$$

or

$$\varphi_3(w) = \int_{-\infty}^{\infty} \exp(iwz) \int_0^z f_1(x) f_2(z-x)dx \; dy + \int_{-\infty}^{\infty} \exp(iwz)h(z)dz$$

where h(z) is the function defined in Theorem 4.2.

By a standard theorem on Fourier integrals

$$\int_{-\infty}^{\infty} \exp(iwz) \int_0^z f_1(x) \; f_2(z-x)dx \; dz = \varphi_1(w) \cdot \varphi_2(w).$$

By Theorem 4.5, $h(z) = 0$ almost everywhere Lebesgue measure if and only if X and Y are independent. Hence, if X and Y are not independent $h(z) \neq 0$ on a set of positive Lebesgue measure, and by the uniqueness assertion of the Fourier integral theorem

$$\int_{-\infty}^{\infty} \exp(iwz) \; h(z) \; dz \neq 0.$$

Thus if X and Y are not independent

$$\varphi_3(w) \neq \varphi_1(w) \; \varphi_2(w).$$

Q.E.D.

It is well known that the above results do not hold for general bivariate distributions.

If we consider integral valued random variables of this class we obtain a similar theorem.

<u>THEOREM 4.7</u>. If X and Y are two random variables which take on the integral values {0, 1, 2, . . .} with probability one, have probability functions $P_1(x)$, $P_2(y)$, and if $R(x,y) = f(x)\, g(y)$, then X and Y are independent if and only if the probability function $P_3(z)$ of $Z = X + Y$ is equal to the convolution of the probability functions $P_1(x)$ and $P_2(y)$.

<u>PROOF</u>. The probability function of $Z = X + Y$ is given by

$$P_3(r) = \sum_{i=0}^{r} (1+f(i)g(r-i))P_1(i)P_2(r-i) \text{ for } r = 0,1,\ldots$$

Hence

$$P_3(r) = \sum_{i=0}^{r} P_1(i)P_2(r-i) + \sum_{i=0}^{r} f(i)g(r-i)P_1(i)P_2(r-i).$$

If X and Y are independent then $f(x) \cdot g(y) = 0$ and thus the second sum vanishes. Hence $P_3(z)$ is equal to the convolution of $P_1(x)$ and $P_2(y)$.

If X and Y are not independent then there exists an (i,j) such that $f(i)\, P_1(i) \neq 0$ and $g(j)\, P_2(j) \neq 0$. Let $m =$ minimum $\{i\}$ where $f(i)\, P_1(i) \neq 0$ and $k =$ minimum $\{j\}$ where $g(j)\, P_2(j) \neq 0$. Then we find that

$$\sum_{i=0}^{k+m} f(i)\, P_1(i) \cdot g(k+m-i) \cdot P_2(k+m-i) = f(m)P_1(m)g(k)P_2(k) \neq 0.$$

It follows thus that

$$P_3(k + m) \neq \sum_{i=0}^{k+m} P_1(i) \cdot P_2(k + m - i)$$

and hence $P_3(z)$ is not equal to the convolution of $P_1(x)$ and $P_2(y)$.

Q.E.D.

COROLLARY 4.8. If X and Y meet the conditions of Theorem 4.7 then they are independent if and only if the generating function of their sum $Z = X + Y$ is equal to the product of their generating functions.

PROOF. Let $G_1(s)$, $G_2(t)$ and $G_3(w)$ be the generating functions of X, Y and Z respectively. It is well known that if X and Y are independent, then the generating function of their sum is equal to the product of their generating functions.

The generating function of Z is defined to be

$$G_3(s) = \sum_{i=0}^{\infty} P_3(i)s^i.$$

By Theorem 4.7, we find that

$$G_3(s) = \sum_{i=0}^{\infty}(\sum_{j=0}^{i}(1 + f(j)\ g(i-j)\ P_1(j)\ P_2(i-j))s^i$$

$$= \sum_{i=0}^{\infty}(\sum_{j=0}^{i} P_1(j)\ P_2(i-j))s^i$$

$$+ \sum_{i=0}^{\infty}(\sum_{j=0}^{i} f(j)\ g(i-j)\ P_1(j)\ P_2(i-j))s^i.$$

It is a well-known result that

$$\sum_{i=0}^{\infty}\left(\sum_{j=0}^{i} P_1(j)\ P_2(i-j)\right)s^i = G_1(s)\cdot G_2(s).$$

$G_3(s)$ is a power series in s which converges in the closed interval [-1,+1]. By Theorem 4.7 there will exist at least one r such that

$$\sum_{j=0}^{r} f(j)\ g(r-j)\ P_1(j)\ P_2(r-j) \neq 0.$$

Hence in the power series expansion of $G_3(s)$ the coefficient of s^r will be different from the coefficient of s^r in the power series represented by $G_1(s)\cdot G_2(s)$. Hence by the uniqueness of the expansion of a function in a power series

$$G_3(s) \neq G_1(s)\cdot G_2(s).$$

Q.E.D.

We are now going to consider the covariance and correlation coefficient of two random variables whose joint distribution is of this simple form.

<u>THEOREM 4.9</u>. If X and Y are two random variables such that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ and $R(x,y) = f(x)\cdot g(y)$ then if their covariance exists it is given by

$$\text{cov}\ (X,Y) = E[x\cdot f(x)]\cdot E[y\cdot g(y)].$$

<u>PROOF</u>. By Theorem 3.20

$$\operatorname{cov}(X,Y) = \int_{W_1}\int_{W_2} x \cdot y\, R(x,y)\, d\mu_1 \times \mu_2$$

$$= \int_{W_1} x \cdot f(x)\, d\mu_1 \int_{W_2} y \cdot g(y)\, d\mu_2$$

$$= E[x \cdot f(x)]E[y \cdot g(y)].$$

Q.E.D.

COROLLARY 4.10. If X and Y are two random variables which meet the conditions of Theorem 4.9, then the correlation coefficient ρ of X and Y if it exists is given by

$$\rho = \frac{E[x \cdot f(x)]\, E[y \cdot g(y)]}{\sqrt{E[(X - E[X])^2]E[(Y - E[Y])^2]}}\,.$$

PROOF. This result follows immediately from Theorem 4.9 and the definition of the correlation coefficient.

Q.E.D.

COROLLARY 4.11. If X and Y are two random variables which meet the conditions of Theorem 4.9, then $\operatorname{cov}(X,Y) = 0$ if and only if $E[x \cdot f(x)] = 0$ or $E[y \cdot g(y)] = 0$.

PROOF. This result follows immediately from Theorem 4.9.

Q.E.D.

We are now going to investigate the regression curves for this special case of two-dimensional random variables.

THEOREM 4.12. Let X and Y be two random variables such that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ and $R(x,y) = f(x)\, g(y)$. Then if $E[h(y)]$ and $E[m(x)]$ exist the regression curves $E[h(y)|x]$ and $E[m(x)|y]$ are given by

$$E[h(y)|x] = E[h(y)] + f(x)\, E[h(y) \cdot g(y)]$$

$$E[m(x)|y] = E[m(x)] + g(y)\, E[m(x) \cdot f(x)].$$

PROOF. As in the proof of Theorem 3.22, we find that

$$E[h(y)|x] = E[h(y)] + \int_{W_2} h(y)\, f(x)\, g(y)\, d\mu_2$$

$$= E[h(y)] + f(x)\, E[h(y) \cdot g(y)].$$

A similar proof holds for $E[m(x)|y]$.

Q.E.D.

We can write the conclusion to Theorem 4.12 in the following way. The regression curves of μ_2 integrable functions take the form

$$E[h(y)|x] = \mathcal{L}_1(h(y)) + f(x)\mathcal{L}_2(h(y))$$

where $\mathcal{L}_1(\ )$ and $\mathcal{L}_2(\ )$ are two linear operators on the μ_2 integrable functions. We thus see that all the regression curves for all μ_2 integrable functions have the same general shape

$$E[h(y)|x] = K_1 + f(x)\ K_2$$

where K_1 and K_2 are two constants whose values are determined by $\mathcal{L}_1(\ )$ and $\mathcal{L}_2(\ )$ respectively.

The following theorem gives even more remarkable evidence of the importance of the regression curves for the class of two-dimensional random variables we are considering in this chapter.

THEOREM 4.13. Let X and Y be two random variables such that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ and $R(x,y) = f(x)\ g(y)$. Then if X and Y have non-zero covariance the bivariate measure μ is uniquely determined by μ_1, μ_2, $E[x|y]$ and $E[y|x]$.

PROOF. By Theorem 4.12

$$E[x|y] = E[x] + g(y)\ E[x \cdot f(x)]$$

$$E[y|x] = E[y] + f(x)\ E[y \cdot g(y)].$$

By hypothesis $\text{cov}(X,Y) \neq 0$ and hence by Corollary 4.11, $E[y \cdot g(y)] \neq 0$ and $E[x \cdot f(x)] \neq 0$.

We now set

$$E[x \cdot f(x)] = 1.$$

The choice of 1 in the above equality is arbitrary; we could choose any non-zero real number. Then $E[x]$ is a known number and we thus are able to find $g(y)$. Then $E[y]$ and $E[y \cdot g(y)]$ are known numbers and we are able to determine $f(x)$.

The following argument shows that the f(x) determined from the second equation must satisfy

$$E[x \cdot f(x)] = 1.$$

By Theorem 3.24

$$E[y \cdot E[x|y]] = E[y \cdot E[y|x]].$$

We also have that

$$E[y \cdot E[x|y]] = E[y]\ E[x] + E[y \cdot g(y)]\ E[x \cdot f(x)]$$

$$E[x \cdot E[y|x]] = E[y]\ E[x] + E[y \cdot g(y)]\ E[x \cdot f(x)].$$

From these equations it follows that once $E[x \cdot f(x)]$ is fixed for the equation

$$E[x|y] = E[x] + g(y)\ E[x \cdot f(x)],$$

the f(x) determined by the other regression curve must have the same property.

Q.E.D.

The primary use of two-dimensional random variables of the class considered in this chapter has been for the construction of examples and the evaluation of methods of measuring the association between two random variables. For specific examples the reader is referred to Crosby [2], Farlie [3], [4], [5] and Gumbel [9], [10], [11].

CHAPTER 5
FINITE DISCRETE MEASURES

In this chapter we are going to examine the set of finite discrete two-dimensional measures with given marginals as a convex set. Then we will use the association of these measures with the functions $1 + R(x,y)$ to establish some elementary properties of this set of measures.

DEFINITION 5.1. A set U is said to be convex if μ' and μ'' are in the set U implies that $\mu = \lambda\mu' + (1-\lambda)\mu''$ belongs to the set U where λ is a real number such that $0 \le \lambda \le 1$.

DEFINITION 5.2. An element of a convex set U is said to be an extreme point of the set U if $\mu = \lambda\mu' + (1-\lambda)\mu''$ where λ is a real number such that $0 \le \lambda \le 1$ and μ' and μ'' belong to U implies that $\mu = \mu'$ or $\mu = \mu''$.

LEMMA 5.1. The set of bivariate measures U with the same marginals μ_1 and μ_2 forms a convex set.

PROOF. Let μ' and μ'' belong to U. Then by definition of the marginal measures

$$\mu'(A \times W_2) = \mu_1(A)$$

and

$$\mu''(A \times W_2) = \mu_1(A)$$

for all sets A in $\mathfrak{a}$.

Let λ be a real number such that $0 \leq \lambda \leq 1$. Then $\mu = \lambda\mu' + (1 - \lambda)\mu''$ is a measure on $\mathcal{A} \times \mathcal{B}$ and

$$\begin{aligned}\mu(A \times W_2) &= \lambda\mu'(A \times W_2) + (1 - \lambda)\mu''(A \times W_2) \\ &= \lambda\mu_1(A) + (1 - \lambda)\mu_1(A) \\ &= \mu_1(A).\end{aligned}$$

Hence μ is a measure with marginal μ_1 on W_1. A similar computation shows that μ has the marginal μ_2 on W_2.

Q.E.D.

<u>LEMMA 5.2</u>. The set of functions $F = \{1 + R(x,y)\}$ measurable over $(W_1 \times W_2, \mathcal{A} \times \mathcal{B})$ which define probability measures with respect to $\mu_1 \times \mu_2$ with marginals μ_1 and μ_2 is a convex set.

<u>PROOF</u>. It will be sufficient to show that if $R'(x,y)$ and $R''(x,y)$ are two functions which have the properties of Theorem 3.1, then the function

$$R(x,y) = \lambda R'(x,y) + (1 - \lambda)R''(x,y),$$

where λ is a real number such that $0 \leq \lambda \leq 1$, will have the same properties.

$R(x,y)$ is measurable by elementary properties of measurable functions. Since $R'(x,y) \geq -1$ and $R''(x,y) \geq -1$ almost everywhere $\mu_1 \times \mu_2$, it follows that $R(x,y) \geq -1$ almost everywhere $\mu_1 \times \mu_2$.

We have that

$$\int_{W_1} R'(x,y)d\mu_1 = 0 \text{ almost everywhere } \mu_2$$

and

$$\int_{W_1} R''(x,y)d\mu_1 = 0 \text{ almost everywhere } \mu_2 .$$

Hence

$$\int_{W_1} R(x,y)d\mu_1 = \int_{W_1} (\lambda R'(x,y) + (1 - \lambda)\ R''(x,y))d\mu_1$$

$$= 0 \text{ almost everywhere } \mu_2 .$$

Similarly, we find that

$$\int_{W_2} R(x,y)d\mu_2 = 0 \text{ almost everywhere } \mu_1 .$$

By Theorem 3.1 these conditions are sufficient to insure that the measure

$$\mu = \int (1 + R(x,y))d\mu_1 \times \mu_2$$

will be a probability measure with the marginals μ_1 and μ_2, and hence $1 + R(x,y)$ will be a member of the set of functions F.

Q.E.D.

The next theorem demonstrates the usefulness of considering the set of functions F.

THEOREM 5.3. If the set of probability measures U on $(W_1 \times W_2, \mathfrak{A} \times \mathfrak{B})$ with marginals μ_1 on W_1 and μ_2 on W_2 is such that if μ belongs to U then μ is absolutely continuous with respect to $\mu_1 \times \mu_2$, then there is a one-to-one correspondence between the extreme points of the convex set U and extreme points of the convex set F of all functions which determine such measures on $\mu_1 \times \mu_2$.

PROOF. Let μ be an extreme point of the set U. Then by hypothesis

$$\mu = \int (1 + R(x,y))d\mu_1 \times \mu_2$$

where $(1 + R(x,y))$ is some function in F.

Suppose that

$$(1 + R(x,y)) = \lambda(1 + R'(x,y)) + (1-\lambda)(1 + R''(x,y))$$

where λ is a real number such that $0 \leq \lambda \leq 1$ and $1 + R'(x,y)$ and $(1 + R''(x,y))$ belong to F.

Then

$$\mu = \int (\lambda(1 + R'(x,y)) + (1-\lambda)(1 + R''(x,y)))d\mu_1 \times \mu_2$$

$$= \lambda\mu' + (1 - \lambda)\mu''.$$

Where

$$\mu' = \int (1 + R'(x,y))d\mu_1 \times \mu_2$$

and

$$\mu'' = \int (1 + R''(x,y))d\mu_1 \times \mu_2 .$$

μ' and μ'' belong to U. Hence by hypothesis $\mu = \mu'$ or $\mu = \mu''$. Assume that $\mu = \mu'$, then

$$\mu = \int (1 + R'(x,y))d\mu_1 \times \mu_2$$

and by the uniqueness assertion of the Radon-Nikodym theorem

$(1 + R(x,y)) = (1 + R'(x,y))$ almost everywhere $\mu_1 \times \mu_2$. Hence $(1 + R(x,y))$ is an extreme point of the set F.

Conversely, let $(1 + R(x,y))$ be an extreme point of the set F and let

$$\mu = \int (1 + R(x,y))d\mu_1 \times \mu_2 .$$

Suppose that $\mu = \lambda\mu' + (1-\lambda)\mu''$ where λ is a real number such that $0 \le \lambda \le 1$, μ' and μ'' belong to U. Then by hypothesis

$$\mu' = \int (1 + R'(x,y))d\mu_1 \times \mu_2$$

and

$$\mu'' = \int (1 + R''(x,y))d\mu_1 \times \mu_2$$

where $1 + R'(x,y)$ and $1 + R''(x,y)$ belong to F. We thus have that

$$\mu = \int (\lambda(1 + R'(x,y)) + (1-\lambda)(1 + R''(x,y)))d\mu_1 \times \mu_2,$$

and from the uniqueness assertion of the Radon-Nikodym theorem it follows that

$$(1 + R(x,y)) = \lambda(1 + R'(x,y)) + (1-\lambda)(1 + R''(x,y))$$

almost everywhere $\mu_1 \times \mu_2$.

Since $(1 + R(x,y))$ is an extreme point of F, we have that

$$(1 + R(x,y)) = (1 + R'(x,y)) \quad \text{or}$$

$$(1 + R(x,y)) = (1 + R''(x,y))$$

almost everywhere $\mu_1 \times \mu_2$. Hence $\mu = \mu'$ or $\mu = \mu''$ and μ is an extreme point of the set U.

Q.E.D.

The preceding theorem sets up a one-to-one correspondence between the extreme points of the convex set of measures with given marginals and the extreme points of the convex set of Radon-Nikodym derivatives of such measures.

COROLLARY 5.4. If one of the marginals is discrete then there exists a one-to-one correspondence between the extreme points of the set U and the extreme points of the set F.

PROOF. This follows immediately from Theorem 5.3 and Theorem 3.11.

Q.E.D.

We are now going to use the above results to prove some interesting theorems where both the marginals μ_1 and μ_2 are finite discrete.

THEOREM 5.5. The set of two-dimensional probability measures with finite discrete marginals μ_1 and μ_2 is a convex set with a finite number of extreme points.

PROOF. By Theorem 5.3 and Corollary 5.4 it will be sufficient to prove that there exist a finite number of extreme points of the convex set of functions F.

Let $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$ be the set of points where $\mu_1(x_i) > 0$ and $\mu_2(y_i) > 0$. We associate the functions $1 + R(x_i, y_i)$ with matrices of the form

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \cdot & & & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & & \cdot \\ a_{n1} & a_{n2} & & a_{nm} \end{bmatrix},$$

where each $a_{ij} = 1 + R(x_i, y_j)$.

It is an obvious consequence of Theorem 3.16 that

$$(1 + R(x_i,y_j)) \leq \min\left(\frac{1}{\mu_1(x_i)}, \frac{1}{\mu_2(y_j)}\right) .$$

From this fact and from the properties of the function $1 + R(x_i,y_j)$ it is clear that there will be a one-to-one correspondence between the functions $1 + R(x_i,y_j)$ and the $n \times m$ matrices such that

$$0 \leq a_{ij} \leq \min\left(\frac{1}{\mu_1(x_i)}, \frac{1}{\mu_2(y_j)}\right)$$

$$\sum_{i=1}^{n} a_{ij}\mu_1(x_i) = 1 \text{ for all } j$$

and

$$\sum_{j=1}^{m} a_{ij}\mu_2(y_j) = 1 \text{ for all } i.$$

Let M represent the set of all $n \times m$ matrices which meet the above requirements. A straightforward computation shows that M will be a convex set and there will be a one-to-one correspondence between the extreme points of M and the extreme points of the set F.

Let M_1 be the set of all $n \times m$ matrices such that

$$0 \leq a_{ij} \leq \frac{1}{\mu_1(x_i)} \text{ for all } i$$

and that

$$\sum_{i=1}^{n} a_{ij}\mu_1(x_i) = 1 \text{ for all } j.$$

Let M_2 be the set of all n x m matrices such that

$$0 \leq a_{ij} \leq \frac{1}{\mu_2(y_j)} \quad \text{for all } j$$

and that

$$\sum_{j=1}^{m} a_{ij}\mu_2(y_j) = 1 \text{ for all } i.$$

If a matrix belongs to the set M it is obvious that it will belong to the set M_1 and the set M_2. Hence it will belong to the set $M_1 \cap M_2$. If a matrix belongs to $M_1 \cap M_2$ an elementary argument shows that it will belong to M. Hence $M = M_1 \cap M_2$.

It is obvious that M_1 and M_2 are convex sets. We will show that M_1 has a finite number of extreme points.

Let m_1 be an extreme point of the set M_1. Then each of the columns of m_1 must be of the form

$$\begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{\mu(x_i)} \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \end{bmatrix}$$

The ith entry is equal to $\frac{1}{\mu(x_i)}$. All other entries are zero. The entry which is different from zero will depend on the column. For suppose not, that is there is some column, say column j, which is not of this form. Then it is clear that there will be at least two entries in this column which are non-zero. Column j will have the form

$$c_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \cdot \\ \cdot \\ \cdot \\ a_{\ j} \\ \cdot \\ \cdot \\ \cdot \\ a_{nj} \end{bmatrix}$$

where $a_{ij} > 0$ and $a_{kj} > 0$.

Let

$$c'_j = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{\mu_1(x_1)} \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} .$$

The only non-zero element is $a'_{ij} = \frac{1}{\mu_1(x_1)}$.

Let

$$c''_j = \frac{1}{(1 - a_{ij}\mu_1(x_i))} (c_j - a_{ij}\mu_1(x_i)c'_j).$$

Then let

$$m_1 = (c_1, c_2, \ldots, c_j, \ldots, c_m)$$

$$m_1' = (c_1, c_2, \ldots, c'_j, \ldots, c_m)$$

$$m_1'' = (c_1, c_2, \ldots, c''_j, \ldots, c_m).$$

A simple computation shows that m_1' and m_1'' belong to M_1. We find that

$$m_1 = (1 - a_{ij}\mu_1(x_i))m_1'' + a_{ij}\mu_1(x_i)m_1'$$

where $m_1 \neq m_1''$ and $m \neq m_1'$, and this is a contradiction to the hypothesis that m is an extreme point of the set M_1.

Thus we have shown that all columns of the extreme points of the set M_1 are of the form given above. There are n^m such matrices and thus the convex set M_1 will have a finite number of extreme points.

If we replace columns by rows and $\mu_1(x_i)$ by $\mu_2(y_j)$, a similar argument will show that M_2 will have a finite number of extreme points.

We now associate with each of the sets M, M_1, and M_2 subsets E, E_1, and E_2 in $n \cdot m$ dimensional Euclidean space. That is, E will be a subset of $n \cdot m$ dimensional Euclidean space made up of elements of the form

$$(a_{11}, a_{12}, \ldots, a_{1m}, a_{21}, a_{22}, \ldots, a_{2m}, \ldots, a_{nm})$$

such that

$$0 \le a_{ij} \le \min\left(\frac{1}{\mu_1(x_i)}, \frac{1}{\mu_2(y_j)}\right)$$

$$\sum_{i=1}^{n} a_{ij}\mu_1(x_i) = 1 \text{ for all } j$$

and

$$\sum_{j=1}^{m} a_{ij}\mu_2(y_j) = 1 \text{ for all } i.$$

E_1 and E_2 will be defined in an analogous manner. A simple argument shows that E, E_1, and E_2 will be convex sets and that there will be a one-to-one correspondence between the extreme points of E, E_1, and E_2 and the extreme points of M, M_1, M_2 respectively.

It is an elementary result in the theory of convex sets in Euclidean space that the intersection of two convex sets each with a finite number of extreme points will have a finite number of extreme points. It is possible using elementary set theoretic arguments to show that

$$E = E_1 \cap E_2.$$

Hence, since E_1 and E_2 have a finite number of extreme points, E will have a finite number of extreme points. Since there is a one-to-one correspondence between the extreme points of the set E and the set M, M will have a finite number of extreme points. Hence F and U will have a finite number of extreme points.

Q.E.D.

The proof of the above theorem actually contains an important result. This is the association of the set of two-dimensional probability measures with finite discrete marginals with convex polyhedrons in Euclidean space. It is hoped that in the future this association will lead to some new results in probability theory.

The next theorem gives a method of obtaining an extreme point in a special case.

THEOREM 5.6. If μ_1 and μ_2 are discrete and if the function $1 + R(x_i,y_j)$ is such that for each x_i with $\mu_1(x_i) > 0$ there is only one point y_{j_i} with $\mu_2(y_{j_i}) > 0$ such that $1 + R(x_i,y_j) > 0$ then $1 + R(x_i,y_j)$ is an extreme point of the set F.

PROOF. Suppose that

$$(1 + R(x_i,y_j)) = \lambda(1+R'(x_i,y_j))+(1-\lambda)(1+R''(x_i,y_j))$$

where λ is a real number such that $0 \le \lambda \le 1$ and $1 + R'(x_i,y_j)$ and $1 + R''(x_i,y_j)$ belong to F. If $\lambda = 0$ then

$$(1 + R(x_i,y_j)) = (1 + R''(x_i,y_j))$$

and if $\lambda = 1$ then

$$(1 + R(x_i,y_j)) = (1 + R'(x_i,y_j)).$$

Hence we may assume that $0 < \lambda < 1$.

Since $1 + R(x_i,y_j) \geq 0$ for all (x_i,y_j) with $\mu_1(x_i)\mu_2(y_j) > 0$ and $1 + R(x_i,y_j)$ in F, we find that

$$1 + R(x_i,y_k) = 0 \text{ if } Y_k \neq Y_j.$$

Hence $\mu_2(y_k) > 0$ implies that

$$1 + R'(x_i,y_k) = 0$$

and

$$1 + R''(x_i,y_k) = 0 \text{ if } y_k \neq y_j$$

for all x_i such that $\mu_1(x_i) > 0$. From the above and from Theorem 3.1, it follows that

$$(1 + R(x_i,y_j))\mu_1(x_i) = 1,$$

$$(1 + R'(x_i,y_j))\mu_1(x_i) = 1,$$

and

$$(1 + R''(x_i,y_j))\mu_1(x_i) = 1$$

for all (x_i,y_j) such that $\mu_1(x_i)\mu_2(y_j) > 0$. Hence

$$1 + R(x_i,y_j) = 1 + R'(x_i,y_j) = 1 + R''(x_i,y_j)$$

almost everywhere $\mu_1 \times \mu_2$.

Q.E.D.

The following theorem is equivalent to a classical result in combinatorial mathematics. (See [23], page 58.)

<u>THEOREM 5.7</u>. Let $W_1 = W_2 = \{1, 2, \ldots, n\}$ and let $\mu_1(i) = \mu_2(j) = \frac{1}{n}$ for all i and j in W_1 and W_2. Then the set of bivariate measures U on $W_1 \times W_2$ with marginals μ_1 and

μ_2 has exactly n! extreme points. These extreme points correspond to the n! one-to-one functions between W_1 and W_2.

PROOF. There are n! one-to-one functions between W_1 and W_2. Let h be one of these functions. Then if $h(i) = j$ we set $1 + R(i,j) = n$ and $1 + R(i,k) = 0$ if $k \neq j$. This function $1 + R(i,j)$ associated with h obviously belongs to F. By Theorem 5.6 this function associated with h will be an extreme point of the set F. Thus we have n! such functions, all different and all of which are extreme points. We call this set of extreme points of F, E.

Let $1 + R_1(i,j)$ be a function in F which is not a member of E. Then there is at least one j such that $0 < 1 + R_1(i,j) < n$ for some i. Hence there must be another i_1 such that $0 < 1 + R_1(i_1,j) < n$.

Since $0 < 1 + R_1(i_1,j) < n$ there must exist a j_2 such that $0 < 1 + R_1(i_1,j_2) < n$. We can continue this process alternating i's and j's. At some step m such that $m \leq 2n - 1$, we will have that $i_m = i_k$ or $j_m = j_k$ where $k < m$. Assume that $i_m = i_k$. The proof in the other case will be similar.

Let

$$\delta = \min\{1 + R_1(i_k,j_{k+1}), \ldots, 1 + R_1(i_m,j_{m-1}),$$
$$n - (1 + R_1(i_k,j_{k+1})), \ldots, n - (1 + R_1(i_m,j_{m-1}))\}.$$

By our choice of the points $(i_k,j_{k+1}), \ldots (i_m,j_{m-1}), \delta > 0$.

Let

$$1 + R_2(i,j) = 1 + R_1(i,j) - \delta$$

$$\text{for } (i_k, j_{k+1}),\ (i_{k+2}, j_{k+3}) \ \cdots\ (i_{m-2}, j_{m-1})$$

$$1 + R_2(i,j) = 1 + R_1(i,j) + \delta$$

$$\text{for } (i_{k+2}, j_{k+1}),\ (i_{k+4}, j_{k+3}) \ \cdots\ (i_m, j_{m-1})$$

$$1 + R_2(i,j) = 1 + R_1(i,j) \text{ otherwise.}$$

Let

$$1 + R_3(i,j) = 1 + R_1(i,j) + \delta$$

$$\text{for } (i_k, j_{k+1}),\ \cdots\ (i_{m-2}, j_{m-1})$$

$$1 + R_3(i,j) = 1 + R_1(i,j) - \delta$$

$$\text{for } (i_{k+2}, j_{k+1}),\ \cdots\ (i_m, j_{m-1})$$

$$1 + R_3(i,j) = 1 + R_1(i,j) \text{ otherwise.}$$

By the choice of δ and the points $(i_k, j_{k+1}), \ldots, (i_m, j_{m-1})$, $1 + R_2(i,j)$ and $1 + R_3(i,j)$ will be members of the set of functions F. Also we have that

$$1 + R_1(i,j) = \tfrac{1}{2}(1 + R_2(i,j)) + \tfrac{1}{2}(1 + R_3(i,j)).$$

Hence $1 + R_1(i,j)$ cannot be an extreme point of the set F. It follows that there are exactly n! extreme points of the set F.

Q.E.D.

We are now going to use some known results about bivariate distribution functions to determine some extreme points of the convex set U. For the results of the remainder of this chapter the assumption that μ is absolutely continuous with respect to $\mu_1 \times \mu_2$ is not used.

LEMMA 5.8. The set D of distribution functions $F(x,y)$ with marginals $F_1(x)$ and $F_2(y)$ forms a convex set.

PROOF. This follows immediately from the definition of $F_1(x)$, $F_2(y)$ and convexity.

Q.E.D.

By a theorem due to M. Fréchet [7] it is known that

$$H_0(x,y) = \max[F_1(x) + F_2(y) - 1, 0] \quad (1)$$

$$H_1(x,y) = \min[F_1(x), F_2(y)] \quad (2)$$

are members of the set D such that if $F(x,y)$ belongs to D then

$$H_0(x,y) \leq F(x,y) \leq H_1(x,y) \text{ for all } x \text{ and } y.$$

THEOREM 5.9. The distribution functions $H_0(x,y)$ and $H_1(x,y)$ defined by 1 and 2 are extreme points of the convex set of bivariate distribution functions D with marginals $F_1(x)$ and $F_2(y)$.

PROOF. Suppose that

$$H_0(x,y) = \lambda H(x,y) + (1 - \lambda)\, G(x,y)$$

where λ is a real number such that $0 \leq \lambda \leq 1$ and $H(x,y)$ and $G(x,y)$ belong to D.

It is known that

$$H_0(x,y) \leq H(x,y) \text{ for all } x \text{ and } y$$

and

$$H_o(x,y) \le G(x,y) \text{ for all } x \text{ and } y.$$

By a straightforward argument it follows from this that $H_o(x,y) = H(x,y)$ or $H_o(x,y) = G(x,y)$. Hence $H_o(x,y)$ is an extreme point of the set D.

Since $H_1(x,y) \ge F(x,y)$ for all $F(x,y)$ in D and all x and y, a similar argument shows that $H_1(x,y)$ is an extreme point of the set D.

Q.E.D.

<u>THEOREM 5.10</u>. If $F(x,y)$ is an extreme point of the convex set D, then the Lebesgue-Stieltjes measure defined by $F(x,y)$ is an extreme point of the bivariate measures U having marginals μ_1 and μ_2.

<u>PROOF</u>. By hypothesis

$$\mu = \int dF(x,y)$$

where $F(x,y)$ is an extreme point of the set D of distribution functions. Suppose that

$$\mu = \lambda\mu' + (1 - \lambda)\mu''$$

where $0 \le \lambda \le 1$, and where μ' and μ'' belong to U. We then have that

$$\mu'(\ ) = \int dF'(x,y)$$

and

$$\mu''(\ ) = \int dF''(x,y)$$

where $F'(x,y)$ and $F''(x,y)$ belong to the set D. The integrals in the above equations are indefinite integrals and represent set functions. It follows that

$$\mu = \lambda \int dF'(x,y) + (1-\lambda) \int dF''(x,y).$$

By a well-known property of the Lebesgue-Stieltjes integral the above equation is found to be equivalent to

$$\mu = \int d\{\lambda F'(x,y) + (1-\lambda)F''(x,y)\}.$$

Since with each measure μ in U there is associated a unique $F(x,y)$ we have that

$$F(x,y) = \lambda F'(x,y) + (1-\lambda)F''(x,y).$$

Hence by hypothesis

$$F(x,y) = F'(x,y) \quad \text{or} \quad F(x,y) = F''(x,y).$$

Thus

$$\mu = \mu' \quad \text{or} \quad \mu = \mu''$$

and μ is an extreme point of the set U.

Q.E.D.

COROLLARY 5.11. The Lebesgue-Stieltjes measure associated with $H_0(x,y)$ and $H_1(x,y)$ as defined in (1) and (2) are extreme points of the set U.

PROOF. This follows immediately from Theorems 5.9 and 5.10.

Q.E.D.

The above theorems and corollary allow us to always find at least two extreme points where $W_1 = W_2 = (-\infty,+\infty)$.

SELECTED BIBLIOGRAPHY

1. S. Bochner, Lectures on Fourier Integrals, Princeton University Press, Princeton, New Jersey, 1959.

2. D. Crosby, A method of constructing some examples in probability theory, Univ. of Arizona, 1966. (unpublished set of notes)

3. D. J. G. Farlie, The performance of some correlation coefficients for a general bivariate distribution, Biometrika, 47 (1960), 307-323.

4. D. J. G. Farlie, The asymptotic efficiency of Daniels' generalized correlation coefficient, J. R. Statist. Soc. B. 23 (1961), 128-142.

5. D. J. G. Farlie, The asymptotic efficiency of Daniels' generalized correlation coefficient, Biometrika. 50 (1963), 499-504.

6. M. Fréchet, Sur les tableaux de corrélation dont les marges sont données, Ann. Univ. Lyon, Sect. A(3), 14 (1951), 53-77.

7. M. Fréchet, Sur les tableaux de corrélation dont les marges sont données, Comptes Rendus Acad. Sci. 242 (1956), 2426-2428.

8. M. Fréchet, Les tableaux dont les marges sont données, Trabajos de Estadistica. 11 (1960), 3-18.

9. E. J. Gumbel, Distribution a plusieurs variables dont les marges sont données, Comptes Rendus Acad. Sci. 246 (1958), 2717-2720.

10. E. J. Gumbel, Bivariate logistic distributions, J. Amer. Statis. Assoc. 55 (1960), 698-707.

11. E. J. Gumbel, Bivariate logistic distributions, J. Amer. Statis. Assoc. 56 (1961), 335.

12. P. R. Halmos, Measure Theory, D. Van Nostrand, Princeton, New Jersey, 1950.

13. G. K. Kagleson, Polynomial expansions of bivariate distributions, Ann. Math. Statist. 35 (1964), 1208-1215.

14. H. O. Lancaster, The structure of bivariate distributions, Ann. Math. Statist. 29 (1958), 719-736.

15. H. O. Lancaster, Correlation and canonical forms of bivariate distributions, Ann. Math. Statist. 34 (1963), 532-538.

16. H. O. Lancaster, Correlation and complete dependence of random variables, Ann. Math. Statist. 34 (1963), 1315-1321.

17. M. Loéve, Probability Theory, 3rd ed., D. Van Nostrand, Princeton, New Jersey, 1955.

18. D. Morgenstern, Einfache Beispiele zweidimensionaler Verteilungen, Mitt. Math. Stat. 8 (1956), 234-235.

19. M. E. Munroe, Introduction to Measure and Integration, Addison-Wesley, Reading, Mass., 1953.

20. I. P. Natanson, Theory of Functions of a Real Variable, Rev. ed., 2 vols., Ungar, New York, 1955.

21. H. R. Pitt, Integration, Measure and Probability, Oliver and Boyd, Edinburgh and London, 1963.

22. F. Riesz and B. Sz.-Nagy, Functional Analysis, Ungar, New York, 1955.

23. H. Ryser, Combinatorial Mathematics, Mathematical Association of America, New York, 1963.

24. S. D. Silvey, On a measure of association, Ann. Math. Statist. 35 (1964), 1157.

25. M. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de Statistique de L'Université de Paris, VIII, 3, (1959), 229-231.

26. E. C. Titchmarsh, The zeros of certain integral functions, Proc. London Math. Soc. 25 (1926), 283-302.