TAXIMETRICS COURSE OUTLINE

This course is designed for students in the biological and social sciences. Its purpose is to present to students in these fields some of the fundamental concepts of modern mathematics and to illustrate the power of these concepts for more productive and objective thinking within their chosen fields. It is not our desire to make mathematicians or even mathematical technicians of our students. It is our desire to expose tomorrow's students to the precision, discipline and generality of mathematical thought. Our higher purpose is less to present the empirical (scientific) discoveries of the past and more to suggest ways of thinking about science in the present and future.

The outline presented here is for a five hour semester course of 14 weeks duration of which one week is reserved for testing and papers. The remaining 13 weeks constitute the 13 units of the outline. The 5 hours of credit are broken into 3 hours of lecture and 4 hours of lab per week. The lectures are of a conceptual and methodological nature while the lab is of a practical and applied nature. The analogue between the conceptual and methodological principles of the lectures and their application to the several disciplines represented by the interests of the students will be made in the lab. Thus the lectures should be given by one more mathematically inclined and the labs should be conducted by one more experienced in empirical science.

The significance of the labs is not to be minimised. They will provide a free form discussion/forum environment for the exchange and development of ideas especially as they bear on the relation between theory and practice. Each student will bring to the lab some problem in data synthesis from his own discipline. Lab periods will be spent discussing these several problems and in applying the principles and concepts of the lectures to their organization and solution. It is expected that student data will be analyzed with computing machines at time appropriate to the development of the course.

The lectures may be characterized by weeks as follows:

1. Introduction. Purposes of the course. History of the development of the ideas and the need for them.
2. Naive set theory and Boolean algebra.
3. Logic and the propositional calculus. Its analogue with set theory.
4. Functions and relations
5. Definition of a character and related considerations.
6. Elementary combinatorics and probability theory.
7. What is empirical science. What is mathematical modeling in empirical science.
8. Information retrieval and computing machines.
9. What is information. How can we measure it. Information content in a character.
10. Interdependence and prediction in characters and classifications
11. The Graph Theory Model for classification.
12. Interpreting the Graph Theory Model.
13. Other models in classification.

Recall the two main purposes of the laboratories:

1. To provide an environment for discussion and exchange of ideas between students and experienced empirical scientists.

2. To provide an opportunity for student to apply the concepts of the course to his own empirical problem in description, characterization, and synthesis of data from his field of interest.

The labs might be characterized by weeks as follows:

1. Introductory discussions developing the history and purpose of the course.
2. Problem and practice sessions in logic and set theory with examples from the represented fields.
3. Delimiting for each student some problem from his field of interest to be maintained as his personal example for the remaining labs.
4. Discussing the related problems of data organization and synthesis.
5. Discussion and construction of descriptive characters for each example.

6-7. Description of chosen study by means of characters constructed.

8. Analyse character analysis output.
9. Analyse character analysis output and redefine characters if necessary.

10-11. Analyse graph output.

12. Each student presents a formal decision and report to the rest of the class.
13. Reach conclusions and write report.

Detailed content of lectures.

1. Need for principles in biology and social sciences. Early work (Sneath, Rogers and Tanimoto, Anderson, others). History of the formation and davelopment of the present group.
2. Define set (two ways), operators, equivalence, containment, bays formulae, product. Discuss axiomatic set theory and contradictions.

3. Define propositions and connectors. Truth tables and truth sets. Quantifiers. Equivalence of connectors and operators. Necessary and sufficient conditions. Logical containment. Schools of proof. Least integer paradox.
4. Numbers, binary connectors. Orderings, number systems, relations. Functions and composition.
5. Equivalence, relations, partitions, definition of character, and classification.
6. Counting techniques and enumeration. Probability space distributions Independence and conditional probability.
7. Induction and deduction. Empirical triangle. Analogues.
8. Introduction to computing machines. The TAXIR system.
9. Quantification of information and its relation to communication and probability. In function and conditional information. Refinement, independence and prediction.
10. Causality and prediction. Similarity coefficients. Purposes of classification.
11. Graph theory model (from our papers)
12. How to interpret Graph Model Theory. Iterative procedure.
13. Other Models: Sokal Michener, Camin Sokal, Wagner et al.

Grading in this course should be based on three factors.

1. The student's enthusiasm and effort as a participant in the course.

2. The quality of each of two papers describing concepts and their application (midterm and final papers).

3. The excellence of the empirical solution to the sample problem as addressed by techniques learned in the course.

TAXIMETRICS COURSE OUTLINE

This course is designed for students in the biological and social sciences. Its purpose is to present to students in these fields some of the fundamental concepts of modern mathematics and to illustrate the power of these concepts for more productive and objective thinking within their chosen fields. It is not our desire to make mathematicians or even mathematical technicians of our students. It is our desire to expose tomorrow's students to the precision, discipline and generality of mathematical thought. Our higher purpose is less to present the empirical (scientific) discoveries of the past and more to suggest ways of thinking about science in the present and future.

The outline presented here is for a five hour semester course of 14 weeks duration of which one week is reserved for testing and papers. The remaining 13 weeks constitute the 13 units of the outline. The 5 hours of credit are broken into 3 hours of lecture and 4 hours of lab per week. The lectures are of a conceptual and methodological nature while the lab is of a practical and applied nature. The analogue between the conceptual and methodological principles of the lectures and their application to the several disciplines represented by the interests of the students will be made in the lab. Thus the lectures should be given by one more mathematically inclined and the labs should be conducted by one more experienced in empirical science.

The significance of the labs is not to be minimised. They will provide a free form discussion/forum environment for the exchange and development of ideas especially as they bear on the relation between theory and practice. Each student will bring to the lab some problem in data synthesis from his own discipline. Lab periods will be spent discussing these several problems and in applying the principles and concepts of the lectures to their organization and solution. It is expected that student data will be analyzed with computing machines at time appropriate to the development of the course.

The lectures may be characterized by weeks as follows:

1. Introduction. Purposes of the course. History of the development of the ideas and the need for them.
2. Naive set theory and Boolean algebra.
3. Logic and the propositional calculus. Its analogue with set theory.
4. Functions and relations
5. Definition of a character and related considerations.
6. Elementary combinatorics and probability theory.
7. What is empirical science. What is mathematical modeling in empirical science.
8. Information retrieval and computing machines.
9. What is information. How can we measure it. Information content in a character.
10. Interdependence and prediction in characters and classifications
11. The Graph Theory Model for classification.
12. Interpreting the Graph Theory Model.
13. Other models in classification.

Recall the two main purposes of the laboratories:

1. To provide an environment for discussion and exchange of ideas between students and experienced empirical scientists.

2. To provide an opportunity for student to apply the concepts of the course to his own empirical problem in description, characterization, and synthesis of data from his field of interest.

The labs might be characterized by weeks as follows:

1. Introductory discussions developing the history and purpose of the course.
2. Problem and practice sessions in logic and set theory with examples from the represented fields.
3. Delimiting for each student some problem from his field of interest to be maintained as his personal example for the remaining labs.
4. Discussing the related problems of data organization and synthesis.
5. Discussion and construction of descriptive characters for each example.

6-7. Description of chosen study by means of characters constructed.

8. Analyse character analysis output.
9. Analyse character analysis output and redefine characters if necessary.

10-11. Analyse graph output.

12. Each student presents a formal decision and report.to the rest of the class.
13. Reach conclusions and write report.

Detailed content of lectures.

1. Need for principles in biology and social sciences. Early work (Sneath, Rogers and Tanimoto, Anderson, others). History of the formation and development of the present group.
2. Define set (two ways), operators, equivalence, containment, bays formulae, product. Discuss axiomatic set theory and contradictions.

3. Define propositions and connectors. Truth tables and truth sets. Quantifiers. Equivalence of connectors and operators. Necessary and sufficient conditions. Logical containment. Schools of proof. Least integer paradox.
4. Numbers, binary connectors. Orderings, number systems, relations. Functions and composition.
5. Equivalence, relations, partitions, definition of character, and classification.
6. Counting techniques and enumeration. Probability space distributions independence and conditional probability.
7. Induction and deduction. Empirical triangle. Analogues.
8. Introduction to computing machines. The TAXIR system.
9. Quantification of information and its relation to communication and probability. In function and conditional information. Refinement, independence and prediction.
10. Causality and prediction. Similarity coefficients. Purposes of classification.
11. Graph theory model (from our papers)
12. How to interpret Graph Model Theory. Iterative procedure.
13. Other Models: Sokal, Michener, Camin-Sokal, Wagner et al.

Grading in this course should be based on three factors

1. The student's enthusiasm and effort as a participant in the course.

2. The quality of each of two papers describing concepts and their application (midterm and final papers).

3. The excellence of the empirical solution to the sample problem as addressed by techniques learned in the course.

BOTANY 431
Spring 1968

A set is any collection of things, concepts, processes, etc.

A set is defined by designating what things (etc.) are in it. There are two ways of defining a set.

1. List: A list of the things in the set to be defined is given. This list is enclosed in wiggly brackets, $\{ \quad \}$, which are used only to define sets.
e.g. $A = \{a, b, c, d, e\}$.
A is the name given to the set of letters as listed above.

2. Rule: A rule is provided which enables us to test things to see if they belong to the set to be defined. The definition by rule is as follows:

$A = \{x/x$ is one of the first five letters in the alphabet$\}$.

This is read "A is the set of typical elements, x, such that x is one of the first five letters of the alphabet."

The first way is convenient when the set to be defined consists of a few elements, or of heterogeneous elements which do not satisfy a simple rule. The second way is convenient when the set to be defined consists of a large number of elements or when a convenient rule exists.

When X is the name of a set, the sentence, $\underline{a} \in X$, read "$\underline{a}$ is an element of the set X," will be true when $\underline{a}$ is one of the things defined to belong to X. When X and Y are names of sets the sentence, $X \subseteq Y$, read "X is a subset of Y," will be true when every element of X is also an element of Y; in other words, "$a \in X$ and $a \notin Y$" is false.

By convention ϕ denotes the set with no elements. For any set X, $\phi \subseteq X$.

$A = B$ means A is another name for B.
$A = B$ means A has exactly the same elements as B.
$A = B$ means $A \subseteq B$ and $B \subseteq A$.

A set may have for its elements anything - in particular a set may have other sets for its elements. A set with only one element is called a singleton; $\{a\}$ is read "singleton $\underline{a}$."

$$\{a\} \neq a .$$

If X is a set, then the symbol, $|X|$, stands for, (is a name for) the number of elements in X. $|X|$ is a number, not a set.

2^X is the name of a set whose elements are subsets of X.

$$2^X = \{y/y \text{ is a subset of } X\}$$

$|2^X| = 2^{|X|}$.

Closed binary operations on sets.

An operation is a proceedure which converts things to other things. Binary means the operation converts two things to one thing.

Closed on sets means the operation converts two sets to some other set.

Union, $\cup$, is a closed binary operation.

If A and B are sets, then $A \cup B$ is a set defined as follows:

$$A \cup B = \{x/x \in A \text{ or } x \in B\}$$

Intersect, $\cap$, is a closed binary operation. If A and B are sets, then $A \cap B$ is a set.

$$A \cap B = \{x/x \in A \text{ and } x \in B\}.$$

Relative complement, -, is a closed binary operation. If A and B are sets, then A - B is a set and

$$A - B = \{x/x \in A \text{ and } x \notin B\}$$

We may construct complicated names for sets by combining these operations. However, whenever ambiguities as to the order in which operations chould take place, parentheses must be used.

$A \cap (B \cup C)$ is unambiguous, $B \cup C$ is the name of the set with which A is intersected.

$(A \cap B) \cup C$ is unambiguous. $A \cap B$ is the name of the set with which C is unioned.

$A \cap B \cup C$ is ambiguous.

Commutativity. $A \cup B = B \cup A$, i.e., the symbol may be turned arounc without changing its value. Thus the operation, $\cup$, is said to be commutative. Similarly $A \cap B$. A - B is not commutative.

Distributivity. In arithmetic we know that whenever x, y, and z are numbers, $x(y + z) = xy + xz$. We say multiplication distributes over addition. We also know that $x + yz \neq (x + y)(x + z)$, and we say that addition does not distribute over multiplication. Addition and multiplication are two closed binary operations on numbers. Intersect and union are two closed binary operations on sets. It is true that

1. Intersect distributes over Union: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
2. Union distributes over Intersect: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Whenever a set S is designated as the totality of things of interest to us within the context of some consideration, and we wish to work only with sets which are subsets of S, i.e. are members of 2^S, we define $\sim A$ to mean $S - A$, and we read $\sim A$ as "the complement of A." $\sim(\sim A) = A$.

Bays formulae
1. $\sim(A \cap B) = \sim A \cup \sim B$
2. $\sim(A \cup B) = \sim A \cap \sim B$.

Boolean Algebra is the study of the subsets of a set, including these rules for discovering different names for the same set. E.g.

$\sim(C \cap \sim(B \cap \sim A))$ is the name of a set whenever C, B and A are subsets of some set S. If we read "=" as "is another name for" the following is an example of manipulations in Boolean algebra.

$$\begin{aligned}
\sim(C \cap \sim(B \cap \sim A)) &= \sim(C \cap (\sim B \cup A)) \\
&= \sim((C \cap \sim B) \cup \sim(C \cap A)) \\
&= \sim(C \cap \sim B) \cap \sim(C \cap A) \\
&= (\sim C \cup B) \cap (\sim C \cup \sim A) \\
&= \sim C \cup (C \cap \sim A) \\
&= \sim(C \cap \sim(B \cap \sim A))
\end{aligned}$$

It will not be necessary to become expert at this kind of manipulation but some knowledge of the rules which permit it is desired.

Numbers: Natural numbers, $\{1, 2, 3, \ldots\} = J$

Integers, $\{\ldots -2, -1, 0, 1, 2, \ldots\} = Z$

Q = Rational numbers = $\{x/x = p/q,\ p \in Z,\ q \in Z,\ q \neq 0\}$

(A =) Algebraic numbers are numbers which when substituted for "x" in sentences such as

$$C_n x^n + C_{n-1} x^{n-1} + \ldots + C_0 x^0 = 0$$

where C_n, C_{n-1} etc. are Rational numbers, make the sentence true.

R = Real numbers = $\{r/r$ is an infinite sequence of integers with one decimal point$\}$

There are many types of numbers; for our purposes, rational numbers will usually suffice. The following inclusions are true:

$$J \subseteq Z \subseteq Q \subseteq A \subseteq R$$

Numbers as names:

Integers and natural numbers will often be used by us as names, or labels or indices. Here we use only the symbolic and order properties of integers. It may be useful, for example, to refer to 5 sets which we might call A_1, A_2, A_3, A_4, A_5.

We may denote these five sets by A_i, $i = 1, 2, \ldots, 5$.

More abstractly we may wish to refer to n sets where n is an integer. In this case we write A_i, $i = 1, 2, \ldots, n$.

This notation is useful in expressing repeated operations, e.g.

$A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5$ may be written $\bigcup_{i=1}^{5} A_i$.

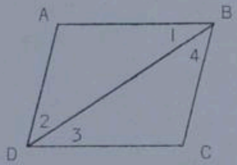
Similarly, $\bigcap_{i=1}^{5} A_i$. We may now express Bays formulae more generally as:

$$\sim \bigcap_{i=1}^{n} A_i = \bigcup_{i=1}^{n} \sim A_i$$

and

$$\sim \bigcup_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \sim A_i.$$

We must be careful not to confuse the numerical prpperties of integers with their naming and sequential properties. You may be amused by a "proof" which was once submitted to me by a student.



ABCD is a parallelogram. The numbers 1, 2, 3, 4 are used here as names for angles ABD, ADB, BDC, and DBC, respectively. We want to prove angle ADC is equal in magnitude to angle ABC. The "proof" goes as follows.

$\angle 2 = \angle 4$	Alternate interior angles
$\angle 3 = \angle 1$	Alternate interior angles
$ADC = \angle 2 + \angle 3 = \angle 5$	
$ABC = \angle 1 + \angle 4 = \angle 5$	

$ADC = \angle 5 = ABC.$

This student's confusion should be clear.

<u>Product:</u> $\times$ Product is a closed binary operation on sets, defined as follows: A and B are sets.

$$A \times B = \{(a,b) / a \in A \text{ and } b \in B\}.$$

This is read "the product of A and B is the set of pairs, (a,b) such that a is an element of A and b is an element of B." The elements of $A \times B$ are ordered pairs of elements. In our discussions (a,b) is not a set. In particular $(a,b) \neq \{a,b\}$.

Propositional Calculus

In our examination of naive set theory we concerned ourselves with a class of things called sets, and with some closed operations (unary and binary) on this class. The result was Boolean Algebra. We now wish to examine the class of propositions (sentences which assert something), together with some closed operations on this class.

Examples of propositions:

The sun is shining. Bot 431 is great.

We will concern ourselves with one unary operation and four binary operations. They are:

1. $\sim$ "not" or "It is not the case that" Unary

   It is not the case that the sun is shining, or The sun is shining.

2. $\wedge$ and. Binary. The sun is shining and Bot 431 is great.
3. $\vee$ or. Binary. The sun is shining or Bot 431 is great.
4. $\leftrightarrow$ means. Binary. The sun is shining means Bot 431 is great.
5. $\rightarrow$ implies. Binary. The sun is shining implies B 431 is great.

Of the four binary operations, 2, 3, 4, 5, only $\rightarrow$, implies, is not commutative.

It is often convenient to have symbolic names for sentences. Letters are usutlly used. Let p, q, and r be names of sentences. $\sim p$, $p \wedge q$, $p \vee q$, $p \leftrightarrow q$, $p \rightarrow q$ would then become names for other sentences.

Statements built from other statements by means of these operations are called compound statements. The truth of a compound statement depends on the truth of the statements that compose it. A convention known as truth tables illustrates this well. The operators themselves are defined using truth tables.

A compound statement made up of only one proposition, such as $\sim p$, has a truth table as shown.

Truth of p	Truth of $\sim p$
T	F
F	T

There is a column for p, where is entered the possibilities for the truth of p, from which the truth of not p is determined.

The truth tables of other compound statements involve two or more distinct propositions. We usually head columns in truth tables with "p" instead of

"Truth of p" and connectors or operators instead of the whole compound statement. This precedent enables us to write

p	$\wedge$	q
T		T
T		F
F		T
F		F

Here each column is four symbols high to allow for the four possible combinations for the truth of p, and the truth of q. In the center column, headed by , will be entered the truth of the compound statement, p q as shown below.

p	$\wedge$	q
T	T	T
T	F	F
F	F	T
F	F	F
1	2	1

The columns labeled 1 were done first, and those labeled 2 were done next.

The truth tables for the other connectors are given.

p	$\vee$	q
T	T	T
T	T	F
F	T	T
F	F	F
1	2	1

p	$\leftrightarrow$	q
T	T	T
T	F	F
F	F	T
F	T	F
1	2	1

p	$\rightarrow$	q
T	T	T
T	F	F
F	T	T
F	T	F
1	2	1

Two compound sentences are the same (i.e. different names for the same sentence) if their truth tables are the same.

Truth tables of more complicated statements may be constructed from these basic five. For example, consider the statement,

$$p \wedge \sim(q \vee \sim r).$$

There are eight possible combinations in which the sentences p, q, r could be true or false, so the columns in the truth table should be 8 symbols high. Again the numbers give the order in which the columns must be built.

p	$\wedge$	$\sim$	$(q$	$\vee$	$\sim$	$r)$
T	F	F	T	T	F	T
T	F	F	T	T	T	F
T	T	T	F	F	F	T
T	F	F	F	T	T	F
F	F	F	T	T	F	T
F	F	F	T	T	T	F
F	F	T	F	F	F	T
F	F	F	F	T	T	F
1 →	5 ←	4	1 →	3 ←	2 ←	1

The last column, 5, gives us the truth of the whole statement. $p \wedge \sim(q \vee \sim r)$ is true only if q is false but p and r are true.

Any sentence which is always true no matter what the truth of its constituents is said to be logically true. A sentence which is always false is called logically false. $p \wedge \sim p$ is logically false. The following are logically true:

$$(p \rightarrow q) \leftrightarrow (q \vee \sim p)$$

$$(p \leftrightarrow q) \leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q)$$

$$\sim (p \wedge q) \leftrightarrow (\sim p \vee \sim q)$$

$$\sim (p \vee q) \leftrightarrow (\sim p \wedge \sim q)$$

If p is a statement, then there are some states of affairs for which p is true, some for which p is false, and some which have nothing to do with p. If S is the set of states of affairs for which p is true or p is false, then p is said to be truth functional on S. The subset P of S on which p is true is the truth set of p. This concept of truth set establishes an analogy between Boolean algebra and propositional calculus. Let q be truth functional on S as well. If P is the truth set of p and Q is the truth set of q, the truth sets of compound statements are as follows.

Statement	Truth Set
$p \wedge q$	$P \cap Q$
$p \vee q$	$P \cup Q$
$p \rightarrow q$	$\sim P \cup Q$
$p \leftrightarrow q$	$(P \cap Q) \cup (\sim P \cap \sim Q)$
$\sim p$	$\sim P$

Necessary and Sufficient Conditions.

If p and q are two statements and p cannot be true unless q is then we say that q is necessary for p. If P and Q are the Truth Sets for p and q, then $P \subseteq Q$.

If p is always true when q is, then we say that q is sufficient for p and we have $Q \subseteq P$. If p and q mean the same thing then q is sufficient for p and q is necessary for p, $Q \subseteq P$, $Q \supseteq P$. Thus $P = Q$. Notice that p is sufficient for q whenever q is necessary for p.

If q is necessary and sufficient for p, q is called a characterization of p or sometimes an equivalent of p.

A relation on a set may be thought of in many ways. Formally, a relation, R, on a set A is a subset R of A $\times$ A. In other words, a relation R on the set A is a collection of ordered pairs (a, b) with $a \in A$, and $B \in A$. Remember $(a, b) \neq (b, a) \neq \{a, b\}$. Less formally a relation may be thought of as a rule which tells when two elements of A are related.

Example: $A = \{x/x$ is an integer$\}$
R = "is divisible by"

6 R 3 may be read: 6 is divisible by 3.

6 R 3 is a true sentence. We say that (6, 3) is in the relation R. If we think of $R \subseteq A \times A$ we may say
$(6, 3) \in R$.

12 R 5 is a false sentence, and we may write $(12, 5) \notin R$.

Example: $A = \{x/x$ is a student of B 431$\}$
R = "is taller than"

Bob R Suzan is a true sentence. (Bob, Suzan) $\in$ R.

Properties of Relations

Reflexive: R is reflexive on A if a R a is a true sentence for all $a \in A$.

Symmetric: R is symmetric on A if $a R b \rightarrow b R a$ is a true sentence for all and b in A.

Antisymmetric: R is antisymmetric on A if $(a R b \wedge b R a) \longrightarrow a = b$ is a true sentence for all a and b in A.

Transitive: R is transitive on A if $(a R b \wedge b R c) \rightarrow a R c$ is a true sentance for all a, b and c in A.

Partial Order: R is a partial order for A if R is Reflexive, Antisymmetric and transitive on A.

Linear Order: R is a linear order for A if R is a partial order for A and (a, b, in A) $\rightarrow (a R b \vee b R a)$ is a true sentence.

R is an equivalence relation on A if R is reflexive, symmetric and transitive on A. An equivalence relation is a generalization of "=".

A partition for A is a Set of Subsets of A, $\{C_1, C_2, \ldots, C_n\}$ with two properties:

1. $\bigcup_{i=1}^{n} C_i = A$ (exhaustive)
2. $C_i \cap C_j = \emptyset$ if $i \neq j$ (exhaustive)

A partition for A defines an equivalence relation, R, on A as follows: a R b is true whenever a and b belong to the same member of the partition and false otherwise. The members of a partition are called the classes of the partition.

Conversely, every equivalence relation R, on A, defines a partition of A. The classes of this partition are called the equivalenceclasses of R.
A classification for A is a series of partitions P_1, P_2, ..., P_m for A such that if C_i is a class in Partition i then for $j > i$, there is exactly one class C_j in Partition j such that $C_i - C_j = \phi$ and all other classes D_j in Partition j are such that $C_i - D_j = C_i$.

In other words a classification for A is a series of Equivalence Relations for A: R_1, R_2 ... R_m such that

$$R_m \supset R_{m-1} \supset \cdots \supset R_2 \supset R_1.$$

A Function is three things,

1. A set called the domain of arguments of the function.
2. A set called the range of values of the function.
3. A rule which associates with each member of the domain a unique member of the range.

<u>Notation</u>. Sometimes the rule is symbolized with a letter, like f, in which case whenever a symool, like "a", is used for an argument, f(a) stands for the value which the rule f assigns to "a".

A character for a Study S of objects to be classified is a function with

1. Domain S
2. Range a small (less than a dozen or so usually) set of descriptions
3. A rule which assigns to each object in S the value in the range which best describes it.

This very general definition of a character gives you much structural freedom in defining characters. However, this means that you are responsible for developing a good technique to take advantage of this freedom.

Functions

I. A relation F is a function if $[(a,b) \in F$ and $(a,c) \in F \Rightarrow b = c]$ is true.

Suppose a relation F is a function.

Then 1. $D = \{x/x$ is the first member of some element of $F\}$ is called the Domain of arguments of F and

2. $R = \{x/x$ is the second member of some element of $F\}$ is called the Range of values of F.

II. Another definition of function is: A function, f, is three things

i. a set D called the Domain of arguments for S

ii. A set R called the Range of values of f

iii. A rule which associates with every element of the Domain some unique element of the Range.

If f is a function and a is an argument then f(a), read "f of a" is a symbolic name for the value which the function f associates with the argument a.

Let $f:D \rightarrow R$ be a function. This notation means that D is the Domain, R is the range and f is the symbol for the functional association.

Suppose $A \subseteq D$, by f(A) we mean $\{x/$for some $a \in A$, $f(a) = x\}$. $f(A) \subseteq R$ and is called the image of A under f. f is said to be onto R is f(d) = R. Otherwise f is said to be into R.

Let f be a function in the relation sense. Then f^{-1} is the relation:

$(a,b) \in f^{-1} \Leftrightarrow (b,a) \in f.$

The inverse of f is f^{-1}. Whenever both f and f^{-1} are functions, we say that f is one-to-one. Otherwise we say that f is many-to-one.

Functional Composition.

If $f:D \rightarrow R$, and $g:D' \rightarrow G'$ and $R \subseteq D'$ we may define a new function $g \circ f:D \rightarrow R'$ as follows: $g \circ F(x) = g(f(x))$.

If f is 1 - 1 then $f^{-1} \circ f(x) = f \circ f^{-1}(x) = x$.

Any function, I for which $I(x) = x$ for all arguments x in its domain is called the identity function, $f^{-1} \circ f = I$.

Whenever $B \subseteq R$ and $f:D \rightarrow R$ we may define $f^{-1}(B) \subseteq D$ even if f is not 1-1.

$f^{-1}(B) = \{x/f(x) \in B\}$ is called the inverse image of B under f.

If $f : D \rightarrow R$ and $R = \{r_1, r_2, \ldots, r_n\}$ then the collection of subsets of D $f^{-1}(\{r_1\})$, r = 1, 2, ..., n is a partition for D.

STRUCTURE OF A CHARACTER

1. A character is a basis for comparison together with a rule for deciding when any two objects in the study are similar or different with respect to the basis for comparison in question. This rule is an equivalence relation the equivalence classes of which constitute the states of the character.

2. A Character is a classification of the study based on a single criterion.

3. A Character is a function, f, with Domain the study to be described, and with Range a small set of descriptions. The rule f associates with each object in the study the description in the range which best describes it. The respective inverse images of each description in the range constitute the states for the character.

INFORMATION. Any system which, at any given time, may exist in one of a number of distinct states contains information proportional to how difficult it is to guess the unknown state of the system without actually observing it. Information quantity is measured in units of yes-no questions. If the guesser asked one who knew about the system in question enough yes-no questions to discover the unknown state, and if the questions were asked in the most clever way possible, the information in the system would be equal in quantity to the average number of such yes-no questions asked.

There are two things which determine this number of yes/no questions (i.e., the quantity of information in a system). These are

1. The number of possible states for the system
2. The probability with which any of these states is known to actually occur.

Any system whose states are known and for which the probability of those states are also known admits a quantification of its information. If K symbolizes such a system H(K) will represent the amount of information in that system.

INFORMATION IN A CHARACTER

In order to measure the information content of a character it is necessary to interpret that character as an information containing system. Think of a character in the function way with the study as the domain and a small set of descriptions as the range. The states of the character become the states of the system when the character is viewed as a system.

Let the character be called K and let K_1 represent the first state, K_2 the second state, etc. The probability of K_1 is $|K_1| / |S|$, the probability of K_2 is $|K_2| / |S|$, etc. We may now express (symbolize) the information in character K as H(K).

Dependence of Information Containing System

Assume we have two information containing systems, the first a fair coin with two states, head and tails, and the second a fair die with six states, one dot, two,dots, ..., and six dots. If we flip the coin and roll the die and observe the coin it doesn't help us guess about the die. Similarly observing the die does not help us guess about the coin. In cases such as this for which observing one system does not help us guess about some other system, we say the two systems are independent. They do not influence each other.

Any two systems which are such that observing the state of one helps us to guess the state of the other are called dependent.

Let A and B be systems. Let A have n possible states and let B have m possible states. By the cartesian product of the two systems, written $A \times B$, we mean the system with $m \cdot n$ states each of which is an ordered pair, (a,b) with "a" a state of A and "b" a state of B.

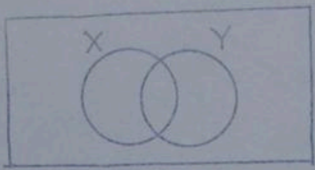
If A and B are independent systems than $H(A) + H(B) = H(A \times B)$. This means that it takes as much information to determine the state of A and the state of B separately as it does to determine both states at once. This is because knowledge of A does not help us guess about B. If A and B were dependent then knowledge of A would help us guess about B. In this case $H(A) + H(B) > H(A \times B)$ because on the left we must discover the states of A and B separately while on the right after discovering the state of A, we now find it easier to discover the state of B. Since, on the right, we still must at least discover the state of A, we know $H(A \times B) \geq H(A)$, and $H(A \times B) \geq H(B)$.

The symbol $H(A/B)$ measures how difficult it is to guess A after observing B. If A and B are independent $H(A/B) = H(A)$. This says that it is just as hard to guess A after observing B as before which is what independent means. In general:

$$H(A/B) = H(A \times B) - H(B).$$

$R(A,B)$ may be thought of as the amount of information in common between A and B. $R(A,B) = H(A) + H(B) - H(A \times B)$. From this it should be apparent that $R(A,B) = R(B,A)$.

We may now envision information in what is called (by me) an heuristic information space. In this space we represent information as subsets. Set X corresponds to the information in A. Set Y corresponds to the information in B.



$R(A,B) = X \cap Y$

$H(A/B) = X - Y$

$H(A \times B) = X \cup Y$

As an indication of dependence we define

$$D(A,B) = \frac{H(A/B) + H(B/A)}{H(A \times B)}$$

$D(A,B) = 1$ if A and B are independent.

If $H(A/B) = 0$, then B is a refinement of A. In the heuristic information space, $Y \geq X$.

If $H(A/B) = 0$ and $H(B/A) = 0$, then A and B are identical. In the heuristic information space $X \subseteq Y$ and $Y \subseteq X$ and thus $Y = X$.

$D(A,B) = 0$ only if $A = B$ (or A is identical with B).

In general $0 \leq D(A,B) \leq 1$ and this indicates the extent of the dependence.

Relations (Continued)

Recall a relation R on S is

1. A subset of $S \times S$, $R \subseteq S \times S$ or
2. A rule "is related to" If a,b are each in S, $(a,b) \in R$ whenever "a is related to b." is true.

The Graph of a Relation

Any relation can be represented by a directed graph. Arrange the elements of S in a convenient way; connect a to b with an arrow from a to b whenever $(a,b) \in R$.

If R is symmetric $(a,b) \in R \rightarrow (b,a) \in R$ in which case the two arrows between a and b may be replaced by a single line.

If R is reflexive then $(a,a) \in R \ \forall a \in S$. In this case the looped arrow which corresponds to (a,a) may be omitted.

If $T \subseteq S$ and R a relation on S then we may define a relation R' on T in a natural way. $(a,b) \in R'$ whenever $a \in T$, $b \in T$, and $(a,b) \in R$. T is said to inherit its relation from the relation R. Notice that $R' \subseteq R \subseteq S \times S$ so that R' may be thought of as a relation on S as well.

Let G be the graph of the relation R on S. Let $T \subseteq S$. Let G' be the graph of R' on T where R' is the relation which T inherits from R. Then G' is called a subgraph of G.

Closures of a Relation

Let R be a relation on S. Q is larger than R if $|Q| > |R|$. Notice that if $Q \supset R$ then $|Q| > |R|$.

The reflexive closure of R is the smallest reflexive relation Q such that $R \subseteq Q$.

R fails to be reflexive on S if there is some $a \in S$ such that $(a,a) \notin R$. Thus the reflexive closure of R is R together with just enough additional pairs of the form (a,a) to make this new relation reflexive.

The symmetric closure of R is the smallest symmetric relation Q such that $Q \supseteq R$.
The transitive closure of R is the smallest transitive relation Q such that $Q \supseteq R$.

Connected Graphs

Let R be a relation on S. Two elements a,b of S are said to be connected if there is a sequence

$$a = p_1, p_2, p_3, \ldots p_m = b$$

of elements in S such that

$$p_i \, R \, p_{i+1} \quad \text{for } i = 1, 2, \ldots m-1.$$

Such a sequence will be called a chain from a to b.

The graph of a symmetric relation on S is said to be connected if any two elements in S are connected.

Let G be the graph of a relation on S. Let G' be a subgraph of G. Let p be the statement "G is connected" and q be the statement "G' is connected." None of the following statements are true in all cases.

$$p \wedge q$$
$$p \wedge \sim q$$
$$\sim p \wedge q$$
$$\sim p \wedge \sim q$$

Let G' be connected. If every other subgraph H, of G, such that $H \supseteq G'$ is not connected, then G' is a maximal connected subgraph of G. (G' is sometimes called a component of G.)

If G is the graph of a relation R on S then the maximal connected subgraphs of G partition S. In this way each relation R on S corresponds to an equivalence relation on S. This equivalence relation will be the symmetric, reflexive, transitive closure of R, and its equivalence classes will be the maximal connected subgraphs of the graph of R.

Bot. 431 Bibliography

GENERAL DISCUSSIONS

R. E. Blackwelder, 1964, Systematics Association Publication #6, pp. 17-28
"Phenetic and Phylogenetic Classification"

E. Mayer, 1965, SSZ 14:2, pp. 73-97
"Numerical Phenetics and Taxonomic Theory"

P. H. A. Sneath, 1956, J. Gen. Microbiol. 17, pp. 184-200
"Some Thoughts on Bacterial Classification"

C. D. Michener and R. R. Sokal, 1957, Evolution 11:2, pp. 130-162
"A Quantitative Approach to a Problem in Classification"

R. R. Sokal, 1962, J. Theor. Biol. 3, pp. 230-267
"Typology and Empiricism in Taxonomy"

G. H. Ball, "Data Analysis in the Social Sciences"

CHARACTERS AND SIMILARITIES

D. C. Eades, 1965, SSZ 14:1, pp 98-99
"The Inappropriateness of the Correlation Coefficient as a Measure of Taxonomic Resemblance."

R. R. Sokal, 1961, SSZ, 10:2, pp. 70-79
"Taxonomic Structure from Randomly and Systematically Scanned Biological Images"

W. R. Lockhart, 1967, Iowa State University
"Hierarchical Characters"

D. W. Goodall, 1966, Biometrics 22:4, pp. 882-907
"A New Similarity Index Based on Probability"

STATISTICAL METHODS (some examples)

J. C. Gower, 1966, Biometrika 53:3 & 4, p 325 seq
"Some distance properties of latent root and vector methods used in multivariate analyses"

F. J. Rolf, 1965, Taxon, 14:8, pp 262-267
"A Randomization Test of the Nonspecificity Hypothesis in Numerical Taxonomy."

R. R. Sokal and C. D. Michener, 1958, U. Kan. Sci. Bull. 38: 22, pp 1409-1438
"A Statistical Method for Evaluating Systematic Relationships"

STRUCTURE OF BIOLOGICAL CLASSIFICATION

A. Rescigno and G. A. Maccacaro, 1960. in Information Theory, Colin Chenny, Ed.
"Information Content of a Biological Classification."

R. C. Buck and D. L. Hull, 1966. SSZ 15:2, pp. 97-111
"Logical Structure of Linnean Hierarchy"

J. H. Camin and R. R. Sokal, 1965, Evolution 19, pp. 311-326
"A Method for Deducing Branching Sequences in Phylogeny."

CLASSIFICATION METHODS

J. Rubin, 1965 IBM Corp. Pub.
"Optimal Classification into Groups"

M. Wirth, G. Estabrook, D. J. Rogers, 1966. SSZ, 15:1 pp 59-69
"A Graph Theory Model for Systematic Biology"

P. H. A. Sneath, 1957, J. Gen. Microbiol. 17, pp 201-226
"Application of Computers to Taxonomy"

W. W. Moss, 1967, SSZ 16:3, pp 117-207
"Some New Analytic and Graphic Approaches to Numerical Taxonomy"

R. E. Bonner, 1964, IBM Jour. of Res. and Dev., 8:1, pp22-
"On Some Clustering Techniques"

D. J. Rogers and T. T. Tanimoto, 1960, Science 132:3434, pp 1115-1118
"A Computer Program for Classifying Plants"

L. Silvestri, et al., 1962, Symposium of the Society for General Microbiology 12: pp 333-360

Brill - 27 February 1968

Simple program in Machine Language

address	main storage	
0 = 0000	0000 0101	data
1 = 0001	0000 0011	data
2 = 0010	1101 : 0000	instructions
3 = 0011	0100 : 0001	instructions
4 = 0100	1011 : 1010	instructions
5 = 0101	~ : ~	instructions
6 = 0110	~ : ~	instructions
7 = 0111	~ : ~	instructions
8 = 1000	~ : ~	instructions
9 = 1001	~ : ~	instructions
10 = 1010	~ : ~	instructions
~	~ : ~	

Instruction word

| op code : address |

Same program in Assembly Language

```
I     DEC  0
J     DEC  0
      CLA  I
      SUB  J
      TZE  JOB2
JOB1  ~~~~~~
      ~~~~~~
      ~~~~~~
      ~~~~~~
      ~~~~~~
JOB2  ~~~~~~
      ~~~~~~
      ~~~~~~
      ~~~~~~
```

Same program in FORTRAN

```
    IF (I.EQ.J) GO TO 2
1   ~~~~~~~~~~~~
    ~~~~~~~~~~~~
2   ~~~~~~~~~~~~
```

Bot. 431 Bibliography

GENERAL DISCUSSIONS

R. E. Blackwelder, 1964, Systematics Association Publication #6, pp. 17-28
"Phenetic and Phylogenetic Classification"

E. Mayer, 1965, SSZ 14:2, pp. 73-97
"Numerical Phenetics and Taxonomic Theory"

P. H. A. Sneath, 1956, J. Gen. Microbiol. 17, pp. 184-200
"Some Thoughts on Bacterial Classification"

C. D. Michener and R. R. Sokal, 1957, Evolution 11:2, pp. 130-162
"A Quantitative Approach to a Problem in Classification"

R. R. Sokal, 1962, J. Theor. Biol. 3, pp. 230-267
"Typology and Empiricism in Taxonomy"

G. H. Ball, "Data Analysis in the Social Sciences"

CHARACTERS AND SIMILARITIES

D. C. Eades, 1965, SSZ 14:1, pp 98-99
"The Inappropriateness of the Correlation Coefficient as a Measure of Taxonomic Resemblance."

R. R. Sokal, 1961, SSZ, 10:2, pp. 70-79
"Taxonomic Structure from Randomly and Systematically Scanned Biological Images"

W. R. Lockhart, 1967, Iowa State University
"Hierarchical Characters"

D. W. Goodall, 1966, Biometrics 22:4, pp. 882-907
"A New Similarity Index Based on Probability"

STATISTICAL METHODS (some examples)

J. C. Gower, 1966, Biometrika 53:3 & 4, p 325 seq
"Some distance properties of latent root and vector methods used in multivariate analyses"

F. J. Rolf, 1965, Taxon, 14:8, pp 262-267
"A Randomization Test of the Nonspecificity Hypothesis in Numerical Taxonomy."

R. R. Sokal and C. D. Michener, 1958, U. Kan. Sci. Bull. 38: 22, pp 1409-1438
"A Statistical Method for Evaluating Systematic Relationships"

STRUCTURE OF BIOLOGICAL CLASSIFICATION

A. Rescigno and G. A. Maccacaro, 1960. in Information Theory, Colin Chenny, Ed.
"Information Content of a Biological Classification."

R. C. Buck and D. L. Hull, 1966. SSZ 15:2, pp. 97-111
"Logical Structure of Linnean Hierarchy"

J. H. Camin and R. R. Sokal, 1965, Evolution 19, pp. 311-326
"A Method for Deducing Branching Sequences in Phylogeny."

CLASSIFICATION METHODS

J. Rubin, 1965 IBM Corp. Pub.
"Optimal Classification into Groups"

M. Wirth, G. Estabrook, D. J. Rogers, 1966. SSZ, 15:1 pp 59-69
"A Graph Theory Model for Systematic Biology"

P. H. A. Sneath, 1957, J. Gen. Microbiol. 17, pp 201-226
"Application of Computers to Taxonomy"

W. W. Moss, 1967, SSZ 16:3, pp 117-207
"Some New Analytic and Graphic Approaches to Numerical Taxonomy"

R. E. Bonner, 1964, IBM Jour. of Res. and Dev., 8:1, pp22-
"On Some Clustering Techniques"

D. J. Rogers and T. T. Tanimoto, 1960, Science 132:3434, pp 1115-1118
"A Computer Program for Classifying Plants"

L. Silvestri, et al., 1962, Symposium of the Society for General Microbiology 12: pp 333-360

Estabrook Syllabus p 10-17
Botany 431 (Class Flowering Plants)

Spring 1968

A relation on a set may be thought of in many ways. Formally, a relation, R, on a set A is a subset R of A $\times$ A. In other words, a relation R on the set A is a collection of ordered pairs (a, b) with $a \in A$, and $B \in A$. Remember $(a, b) \neq (b, a) \neq \{a, b\}$. Less formally a relation may be thought of as a rule which tells when two elements of A are related.

Example: $A = \{x/x \text{ is an integer}\}$

R = "is divisible by"

6 R 3 may be read: 6 is divisible by 3.

6 R 3 is a true sentence. We say that (6, 3) is in the relation R. If we think of $R \subseteq A \times A$ we may say

$(6, 3) \in R$.

12 R 5 is a false sentence, and we may write $(12, 5) \notin R$.

Example: $A = \{x/x \text{ is a student of B 431}\}$

R = "is taller than"

Bob R Suzan is a true sentence. (Bob, Suzan) $\in$ R.

Properties of Relations

Reflexive: R is reflexive on A if a R a is a true seetence for all $a \in A$.

Symmetric: R is symmetric on A if $a R b \rightarrow b R a$ is a true sentence for all and b in A.

Antisymmetric: R is antisymmetric on A if $(a R b \wedge b R a) \longrightarrow a = b$ is a true sentence for all a and b in A.

Transitive: R is transitive on A if $(a R b \wedge b R c) \rightarrow a R c$ is a true sentence for all a, b and c in A.

Partial Order: R is a partial order for A if R is Reflexive, Antisymmetric and transitive on A.

Linear Order: R is a linear order for A if R is a partial order for A and (a, b, in A) $\rightarrow$ (a R b $\vee$ b R a) is a true sentence.

R is an equivalence relation on A if R is reflexive, symmetric and transitive on A. An equivalence relation is a generalization of "=".

A partition for A is a Set of Subsets of A, $\{C_1, C_2, \ldots, C_n\}$ with two properties:

1. $\bigcup_{i=1}^{n} C_i = A$ (exhaustive)

2. $C_i \cap C_j = \phi$ if $i \neq j$ (exhaustive)

A partition for A defines an equivalence relation, R, on A as follows: a R b is true whenever a and b belong to the same member of the partition and false otherwise. The members of a partition are called the classes of the partition.

Conversely, every equivalence relation R, on A, defines a partition of A. The classes of this partition are called the equivalenceclasses of R. A classification for A is a series of partitions $P_1, P_2, \ldots, P_m$ for A such that if C_i is a class in Partition i then for $j > i$, there is exactly one class C_j in Partition j such that $C_i - C_j = \phi$ and all other classes D_j in Partition j are such that $C_i - D_j = C_i$.

In other words a classification for A is a series of Equivalence Relations for A: $R_1, R_2 \ldots R_m$ such that

$$R_m \supset R_{m-1} \supset \cdots \supset R_2 \supset R_1.$$

A Function is three things,

1. A set called the domain of arguments of the function.
2. A set called the range of values of the function.
3. A rule which associates with each member of the domain a unique member of the range.

Notation. Sometimes the rule is symbolized with a letter, like f, in which case whenever a symool, like "a", is used for an argument, f(a) stands for the value which the rule f assigns to "a".

A character for a Study S of objects to be classified is a function with

1. Domain S
2. Range a small (less than a dozen or so usually) set of descriptions
3. A rule which assigns to each object in S the value in the range which best describes it.

This very general definition of a character gives you much structural freedom in defining characters. However, this means that you are responsible for developing a good technique to take advantage of this freedom.

Functions

I. A relation F is a function if $[(a,b) \in F$ and $(a,c) \in F \Rightarrow b = c]$ is true.
Suppose a relation F is a function.

Then 1. $D = \{x/x$ is the first member of some element of $F\}$ is called the Domain of arguments of F and

2. $R = \{x/x$ is the second member of some element of $F\}$ is called the Range of values of F.

II. Another definition of function is: A function, f, is three things

i. a set D called the Domain of arguments for S

ii. A set R called the Range of values of f

iii. A rule which associates with every element of the Domain some unique element of the Range.

If f is a function and a is an argument then f(a), read "f of a" is a symbolic name for the value which the function f associates with the argument a.

Let $f:D \to R$ be a function. This notation means that D is the Domain, R is the range and f is the symbol for the functional association.
Suppose $A \subseteq D$, by f(A) we mean $\{x/\text{for some } a \in A, f(a) = x\}$. $f(A) \subseteq R$ and is called the image of A under f. f is said to be onto R is $f(d) = R$. Otherwise f is said to be into R.

Let f be a function in the relation sense. Then f^{-1} is the relation:

$(a,b) \in f^{-1} \Leftrightarrow (b,a) \in f$.

The inverse of f is f^{-1}. Whenever both f and f^{-1} are functions, we say that f is one-to-one. Otherwise we say that f is many-to-one.

Functional Composition.

If $f:D \to R$, and $g:D' \to G'$ and $R \subseteq D'$ we may define a new function $g \circ F:D \to R'$ as follows: $g \circ F(x) = g(f(x))$.

If f is 1 - 1 then $f^{-1} \circ f\ (x) = f \circ f^{-1}(x) = x$.

Any function, I for which $I(x) = x$ for all arguments x in its domain is called the identity function, $f^{-1} \circ f = I$.

Whenever $B \subseteq R$ and $f:D \to R$ we may define $f^{-1}(B) \subseteq D$ even if f is not 1-1.
$f^{-1}(B) = \{x/f(x) \in B\}$ is called the inverse image of B under f.
If $f : D \to R$ and $R = \{r_1, r_2, \ldots, r_n\}$ then the collection of subsets of D $f^{-1}(\{r_1\})$, $r = 1, 2, \ldots, n$ is a partition for D.

STRUCTURE OF A CHARACTER

1. A character is a basis for comparison together with a rule for deciding when any two objects in the study are similar or different with respect to the basis for comparison in question. This rule is an equivalence relation the equivalence classes of which constitute the states of the character.

2. A Character is a classification of the study based on a single criterion.

3. A Character is a function, f, with Domain the study to be described, and with Range a small set of descriptions. The rule f associates with each object in the study the description in the range which best describes it. The respective inverse images of each description in the range constitute the states for the character.

INFORMATION. Any system which, at any given time, may exist in one of a number of distinct states contains information proportional to how difficult it is to guess the unknown state of the system without actually observing it. Information quantity is measured in units of yes-no questions. If the guesser asked one who knew about the system in question enough yes-no questions to discover the unknown state, and if the questions were asked in the most clever way possible, the information in the system would be equal in quantity to the average number of such yes-no questions asked.

There are two things which determine this number of yes/no questions (i.e., the quantity of information in a system). These are

1. The number of possible states for the system
2. The probability with which any of these states is known to actually occur.

Any system whose states are known and for which the probability of those states are also known admits a quantification of its information. If K symbolizes such a system H(K) will represent the amount of information in that system.

INFORMATION IN A CHARACTER

In order to measure the information content of a character it is necessary to interpret that character as an information containing system. Think of a character in the function way with the study as the domain and a small set of descriptions as the range. The states of the character become the states of the system when the character is viewed as a system.

Let the character be called K and let K_1 represent the first state, K_2 the second state, etc. The probability of K_1 is $|K_1| / |S|$, the probability of K_2 is $|K_2| / |S|$, etc. We may now express (symbolize) the information in character K as H(K).

Dependence of Information Containing System

Assume we have two information containing systems, the first a fair coin with two states, head and tails, and the second a fair die with six states, one dot, two,dots, ..., and six dots. If we flip the coin and roll the die and observe the coin it doesn't help us guess about the die. Similarly observing the die does not help us guess about the coin. In cases such as this for which observing one system does not help us guess about some other system, we say the two systems are independent. They do not influence each other.

Any two systems which are such that observing the state of one helps us to guess the state of the other are called dependent.

Let A and B be systems. Let A have n possible states and let B have m possible states. By the cartesian product of the two systems, written $A \times B$, we mean the system with $m \cdot n$ states each of which is an ordered pair, (a,b) with "a" a state of A and "b" a state of B.

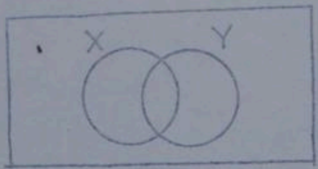
If A and B are independent systems than $H(A) + H(B) = H(A \times B)$. This means that it takes as much information to determine the state of A and the state of B separately as it does to determine both states at once. This is because knowledge of A does not help us guess about B. If A and B were dependent then knowledge of A would help us guess about B. In this case $H(A) + H(B) > H(A \times B)$ because on the left we must discover the states of A and B separately while on the right after discovering the state of A, we now find it easier to discover the state of B. Since, on the right, we still must at least discover the state of A, we know $H(A \times B) \geq H(A)$, and $H(A \times B) \geq H(B)$.

The symbol $H(A/B)$ measures how difficult it is to guess A after observing B. If A and B are independent $H(A/B) = H(A)$. This says that it is just as hard to guess A after observing B as before which is what independent means. In general:

$$H(A/B) = H(A \times B) - H(B).$$

$R(A,B)$ may be thought of as the amount of information in common between A and B. $R(A,B) = H(A) + H(B) - H(A \times B)$. From this it should be apparent that $R(A,B) = R(B,A)$.

We may now envision information in what is called (by me) an heuristic information space. In this space we represent information as subsets. Set X corresponds to the information in A. Set Y corresponds to the information in B.



$R(A,B) = X \cap Y$

$H(A/B) = X - Y$

$H(A \times B) = X \cup Y$

As an indication of dependence we define

$$D(A,B) = \frac{H(A/B) + H(B/A)}{H(A \times B)}$$

$D(A,B) = 1$ if A and B are independent.

If $H(A/B) = 0$, then B is a refinement of A. In the heuristic information space, $Y \geq X$.

If $H(A/B) = 0$ and $H(B/A) = 0$, then A and B are identical. In the heuristic information space $X \subseteq Y$ and $Y \subseteq X$ and thus $Y = X$.

$D(A,B) = 0$ only if $A = B$ (or A is identical with B).

In general $0 \leq D(A,B) \leq 1$ and this indicates the extent of the dependence.

Relations (Continued)

Recall a relation R on S is

1. A subset of $S \times S$, $R \subseteq S \times S$ or
2. A rule "is related to" if a,b are each in S, $(a,b) \in R$ whenever "a is related to b." is true.

The Graph of a Relation

Any relation can be represented by a directed graph. Arrange the elements of S in a convenient way; connect a to b with an arrow from a to b whenever $(a,b) \in R$.

If R is symmetric $(a,b) \in R \rightarrow (b,a) \in R$ in which case the two arrows between a and b may be replaced by a single line.

If R is reflexive then $(a,a) \in R \ \forall a \in S$. In this case the looped arrow which corresponds to (a,a) may be omitted.

If $T \subseteq S$ and R a relation on S then we may define a relation R' on T in a natural way. $(a,b) \in R'$ whenever $a \in T$, $b \in T$, and $(a,b) \in R$. T is said to inherit its relation from the relation R. Notice that $R' \subseteq R \subseteq S \times S$ so that R' may be thought of as a relation on S as well.

Let G be the graph of the relation R on S. Let $T \subseteq S$. Let G' be the graph of R' on T where R' is the relation which T inherits from R. Then G' is called a subgraph of G.

Closures of a Relation

Let R be a relation on S. Q is larger than R if $|Q| > |R|$. Notice that if $Q \supset R$ then $|Q| > |R|$.

The reflexive closure of R is the smallest reflexive relation Q such that $R \subseteq Q$.

R fails to be reflexive on S if there is some $a \in S$ such that $(a,a) \notin R$. Thus the reflexive closure of R is R together with just enough additional pairs of the form (a,a) to make this new relation reflexive.

The symmetric closure of R is the smallest symmetric relation Q such that $Q \supseteq R$.
The transitive closure of R is the smallest transitive relation Q such that $Q \supseteq R$.

Connected Graphs

Let R be a relation on S. Two elements a,b of S are said to be connected if there is a sequence

$$a = p_1, p_2, p_3, \ldots p_m = b$$

of elements in S such that

$$p_i \; R \; p_{i+1} \quad \text{for} \quad i = 1, 2, \ldots m-1.$$

Such a sequence will be called a chain from a to b.

The graph of a symmetric relation on S is said to be connected if any two elements in S are connected.

Let G be the graph of a relation on S. Let G' be a subgraph of G. Let p be the statement "G is connected" and q be the statement "G' is connected." None of the following statements are true in all cases.

$p \wedge q$

$p \wedge \sim q$

$\sim p \wedge q$

$\sim p \wedge \sim q$

Let G' be connected. If every other subgraph H, of G, such that $H \supseteq G'$ is not connected, then G' is a maximal connected subgraph of G. (G' is sometimes called a component of G.)

If G is the graph of a relation R on S then the maximal connected subgraphs of G partition S. In this way each relation R on S corresponds to an equivalence relation on S. This equivalence relation will be the symmetric, reflexive, transitive closure of R, and its equivalence classes will be the maximal connected subgraphs of the graph of R.