

## CHAPTER 1: THE GENERAL FRAMEWORK

### 1.0 Introduction

The most comprehensive prior attempt at a single framework for problems of marriage theory is that of J. H. B. Kemperman (1967). In a series of four simultaneously published papers, Kemperman outlined the general problem in the context of convex spaces and carried out certain applications of interest to population genetics. However, these papers also give clear evidence that the author was aware of other applications.

In the present chapter, we present the principal definitions and results of Kemperman's first paper, which was the most general. Our definitions and theorems will closely follow those of Kemperman, but the explanations given here may differ widely or narrowly from those of the original paper. Therefore, to give specific credit, we shall note with "K. ---" (where --- shows the page number of the original text) material which is direct quotation or close paraphrase (principally definitions or theorems). The reader should assume all other text is the opinion of the present author. However, we strongly recommend a reading of the original materials, if this and all other sections.

### 1.1 Basic Ideas

We assume a population consisting of  $n$  types of males and  $f$  types of females. Using  $i$  to index male types and  $j$  to index female types, we can define a set of ordered pairs  $(i, j)$  in  $M \times F$ , the cross product of the  $M$  of the  $m$  male types and the set  $F$  of the  $f$  female types.

We may define a new set  $T$  as

$$T = \{(i, j) : i \text{ and } j \text{ are a prohibited mating type}\}$$

We call  $T$  the set of prohibited pairs. If  $(i, j) \notin T$ , we call  $(i, j)$  admissible.

Note that in the simplest case  $m = f$ , and for each pair  $(i, j) \notin T$  there is a positive constant  $K$  which is the number of male and female offspring of each pair of individuals of pair-type  $(i, j)$ .

(K. 247) For each admissible type  $(i, j)$  and each  $k = 1, \dots, m$  the expected number  $K_{ijk}$  of male offspring is equal to the expected number of female offspring of type  $k$ . Also,  $f_{ijk} \geq 0$ , and  $\sum_k f_{ijk} = 1$ . (Note that this condition could be done away with, and not affect other development.)

(K. 247) If we have  $p_1, \dots, p_n$  nonnegative numbers, with  $\sum_{i=1}^n p_i = 1$ , and consider a population of  $Np_i$  males of each type  $i$  in a total population of size  $N$ , and  $Nq_j$  females of each type  $j$ , then we call the vector  $p = (p_1, \dots, p_m)$  the type distribution of the population of males, and  $q = (q_1, \dots, q_n)$  the type distribution of females.

Notice that in genetics, the type distribution would refer, for example to relative frequencies of genotypes, while in social anthropology it will refer, for example, to relative frequencies of clan members or lineage members. In fact, any partition of the members of a population will be suitable to define a type distribution of the population into the subsets of the given partition. However, one difficulty we may note (treated more extensively in White, 1963, Chapter 1; Boyd, 1971; and Lorrain, 1974) is that some methods of partitioning the population are constructed from the "viewpoint" of a particular member of the population, while other methods are independent of "viewpoint".

For example, whether or not an individual has a particular given family name is generally not dependent on who provides the list of names (e.g., civil records serve well in many cases), but who is a "first cousin" of whom will provide different partitions of the population relative to each particular individual. We note that this difference in perspective becomes unimportant (e.g., results in isomorphic partitions) in particular cases treated by White, and subsequently others (Courage, 1974, and the above references), but we also note that the framework of the present chapter only applies strictly to populations with partitions independent of "perspective". (Atkins, 1974, has a method for always partitioning a population essentially according to patri- or matri-lineal descent, which is also compatible with the notion of type distribution.)

Kemperman next defines the idea of system of mating, with the disclaimer that he is not interested in "which particular system is most likely to be followed" by a particular group. However, as will be apparent in Chapter 6, that is precisely one question we seek to answer.

DEFINITION 1.1 (K. 248) Let  $S_n$  denote the collection of all type distributions  $p = (p_1, \dots, p_n)$ . By a system of mating we shall mean a rule which associates to each  $p \in S_n$  a set of  $n^2$  nonnegative numbers  $P_{ij}$  ( $i, j = 1, \dots, n$ ) not all zero, satisfying

$$(1) \sum_{j=1}^n P_{ij} \leq p_j \quad \text{and} \quad \sum_{i=1}^n P_{ij} \leq p_i$$

(for  $i = 1, \dots, n$  and  $j = 1, \dots, n$  respectively) and

$$(2) P_{ij} = 0 \text{ if } (i, j) \in T$$

This is equivalent to a system of  $n^2$  nonnegative functions  $G_{ij}(t)$  on  $S_n$  satisfying

$$\sum_{i,j} G_{ij} > 0, G_{ij}(p) \equiv 0 \text{ when } (i, j) \in T.$$

In such a population, the number of pairs of each type  $(i, j)$  is  $C_{ij} = Np_i p_j$  if  $N$  is the number of each sex, and  $P_{ij} = G_{ij}(p)$ .

This notation allows one characterization of random mating, which not incidentally also corresponds to the method used by Darwin to compute expected inbreeding coefficients in a given population. In particular, for  $(i, j) \notin T$ , let  $G_{ij}(p) = P_{ij}$ . One might call this completely random mating if  $|T| = 0$ , but conditionally random mating otherwise.

Kemperman defines, but does not develop, the possibility that if  $w$  is a random variable,  $P_{ij} = G_{ij}(p, w)$ . Note that in population genetics, numerous special cases of  $G_{ij}(p, w)$  have been studied, including subdivision of the population (Wahlund, 1975), and migration (see references in Crow and Kimura, 1974) but most such models essentially regard  $w$  as having an effect on the entire vector  $p$  and study the resulting effect on particular  $P_{ij}$ 's. We can also regard selection as normally studied in population genetics as a case where  $w$  acts as an operator (not always linear) on  $G_{ij}(p)$ . However, we could also interpret selection as introduction of a random variable or particular operator into the computation of  $G_{ij}(p)$ .

Under conditionally random mating, a portion  $Np_{i0}$  of the males and  $Np_{0j}$  of the females remain unmated. For that matter, there may also be unmated individuals under completely random mating, depending on the sex ratio and our assumptions on mating sequence, etc. The unmated proportions of males and of females respectively are given by:

$$(3) P_{i0} = p_i - \sum_{j=1}^n P_{ij}, \quad P_{0j} = p_j - \sum_{i=1}^n P_{ij}$$

for each  $i$  and each  $j$ .

(K. 249) The proportion of unmated individuals will be called the mating ratio denoted by  $c$ , a nonnegative number  $\leq 1$ , equal to

$$(4) 1 - c = \sum_{i=1}^m \sum_{j=1}^f P_{ij} = \sum_{i=1}^m \sum_{j=1}^m P_{ij} = \sum_{i=1}^m \sum_{j=1}^f P_{ij}$$

Kemperman notes that the type of distribution  $p^{(t)}$  of the  $t$ th generation may converge to a limiting distribution. A very simple example of this is the Gada system of Ethiopia as modeled by Hoffman (1971), or any other system with Markovian

behavior. Such a limiting distribution is called the stable distribution, and will "usually satisfy"

$$(5) \sum_{i=1}^m \sum_{j=1}^m G_{ij}(p) f_{ijk} = c p_k$$

(K. 250) where  $k = 1, \dots, n$  and

$$(6) c = \sum_{i=1}^m \sum_{j=1}^m G_{ij}(p), \quad 0 \leq c \leq 1.$$

Kemperman offers two conditions necessary for a vector  $p$  to be a stable type distribution. We list these below.

1.  $\sum_{i=0}^m \sum_{j=1}^m P_{ij} = \sum_{j=1}^m P_j$
2.  $\sum_{i=1}^m \sum_{j=1}^m P_{ij} f_{ijk} = c p_k$

with  $c$  given as in (6). For a given  $c$ , the set of all matrices  $\{P_{ij}\}$  which satisfy these conditions will be denoted by  $\Pi_c$ . (Note that  $P_{ij} = 0$  if  $(i, j) \in \pi$ ). Denote the corresponding set of vectors  $p_c$  by  $P_c$ .

Although we shall not here develop the convex space foundations of marriage theory, we note that Kemperman's developments so far, from the perspective of a fixed value of  $c$ , allow us to consider  $\Pi_c$  and  $P_c$  as compact, convex polyhedral subsets of their respective euclidian spaces. By study of the "extreme points" of such spaces, one gains a complete knowledge of the possible equilibria of particular systems. Kemperman applies these ideas in later sections of the paper, not reviewed here.

Notice that the most significant difference of Kemperman's approach from the usual approaches to population models of any type (social or biological) is precisely the preference to embed ideas in a geometric rather than strictly analytic framework. This difference is not trivial, since it appears that a geometric insight is essential to understanding social anthropology. This point will be more apparent in the chapter on the theory of structural numbers, but we submit that Kemperman's choice of foundations have more than an accidental relationship to his awareness of the possible breadth of application. However, (K.261) Kemperman is also aware of the "usual" Perron-Frobenius results upon which "normal" population models are built.

Condition 2 above has two totally different interpretations depending on whether we are considering social theoretic or genetic applications. If we interpret

genetically, then it essentially says "no selection". However, if we interpret a rule on the social assignment of offspring to parents, then condition 2 allows all parental types not in  $\pi$  an equal chance at the "available" offspring, but says nothing whatever about selection in the genetic sense. In other words, there is nothing in Kemperman's model which prohibits our interpreting the numbers  $f_{ijk}$  as indicating the proportion of the offspring per generation given the labels "produced by  $i$  and  $j$  and put in class  $k$ " independently of which couple actually produced the offspring. Or, in still other words, adoption which gives the family name of foster parents to adopted offspring is permitted. Likewise, "cheating" is also possible, providing no one ever "knows".

We now present a most fundamental theorem, which encourages us to proceed.

THEOREM: Suppose there are no non-trivial prohibitions of mating type, and let  $0 < c \leq 1$  be fixed. Then  $\Pi_c$  is non-empty. (K. 259).

(Essentially, the existence of results at all in chapter 4, "Theory of Structural Numbers", constitutes a proof for a related condition with non-trivial prohibitions, this extends to the special cases discussed by White and others referenced above, where we interpret "individuals" in that theory as "all individuals with an identical genealogy".)

PROOF: The proof relies on the existence of fixed point(s) in the vector space of the vectors  $p$  of type distributions of males, and  $q$  of type distributions of females under the linear transformations imposed by the functions  $G_{ij}(\cdot)$ . See K. 260.

Having now laid out the general framework of the problem, we go on in later chapters and sections to first critique the "usual" demographic model, then to develop our alternative. In particular, we shall present a number of different non-trivial mating systems, and then compute for each the theoretical values of  $c$ . Since we have learned from Kemperman that the stable distributions depend explicitly on  $c$ , the importance of this computation should be obvious.

It is useful to conclude with two notes. First, that if we find a fixed point vector, we have found a transformation

$p \rightarrow p^*$  or  $S_n \times S_n$   
for each vector  $p$  and each fixed point  $p^*$ . In general, this is defined for each  $k$  by the functions

$$(7) p_k^* = \frac{1}{c} \sum_{i=1}^m \sum_{j=1}^m G_{ij}(p) f_{ijk}$$

given a particular  $\pi$ .

We conclude with a short observation on Kemperman's material: to the extent we have presented them, they appear to be little more than vector space concepts applied to a special problem. This is correct, in so far as the observation is based on our summary, but is incorrect from the point of view of the succeeding sections of the original paper. The fact that the theorem quoted above references the set  $\Pi_c$ , and that the particular values in  $F_c$  or  $\Pi$  depend on selection of  $c$  and therefore on the constituents of  $T$ , emphasize this departure from purely vector-space oriented concepts.

## 1.2 Critique of Classical Demography

There are several relationships between the present approach of part 1.1 and classical demography, and, moreover, certain "holes" in classic theory which we hope to uncover here. However, we shall also point to a hole that we believe Kemperman has already filled, which it is not our main purpose at present to discuss at length.

Our critique of classical demography is essentially that it is limited to study of stability of age "structures", in the presence of growth in the total population. Models of zero-growth are an afterthought with respect to the main competency of demographic models to deal with age-structured phenomena. However, study of populations from the viewpoint of absolute totals, or subdivisions other than age structure, have been the main concerns of Social Anthropology.

The relative capacities of demographic computation from different kinds of initial information have been summarized by Keyfitz (1968: 183) in a handy table.

We condense the table, using the following notations:

b:	Overall birth rate	$m(a)$ :	Age specific birth rate
d:	Overall death rate	$\mu(a)$ :	Age specific death rate
r:	Growth rate	$e(a)$ :	age-structure proportions
N:	Population size	$p(a)$ :	stable life schedule

These values are computed or given for a particular population at a particular time. Using stable (age structure) population theory, Keyfitz points out that given particular pairs of these, we may determine all or some of the others. The pairs and what they may compute are summarized in the table below. Boxes with an  $r$ ,  $A$ , or  $T$  show that the pair labeling the intersecting row and column at that part of the table determine one of three conditions:

I: the given pair determines all others, and we may get immediately the stable results on all other variables.

TABLE 1.1 Accessible Results from Different Initial Information, Summarized from Keyfitz, 1968.

	By Age		Intrinsic Rates	
Given	$p(a)$	$m(a)$	$r$	$d, b$
$e(a)$	$\Pi$	$\Pi$	$\Pi$	
$p(a)$	$A$	$A$	$A$	$A$
$r$	$T$	$T$	$T$	$T$
$d$			$T$	$T$

A: the given pair allows computation of asymptotic stability of all other variables.

T: the pair allows computation of population totals only, of variables  $b, r, d$ , and  $N$ .

Notice that from the point of view of the type of information usually available either from "general ethnography" or even from painstaking archaeology, we are happy to obtain estimates of  $N$ , and can occasionally get gross estimates of  $b$  and  $d$ . If records or subsequent restudy are available options, we may also get estimates of  $r$ . In rare circumstances, we can get estimates of one of the age structured variables, usually an approximation of  $p(a)$ .

We can thus notice that there are two reasons why the historic isolation of demography and social anthropology has been justified: most "demographic" data is usually not available from anthropological investigation while the weakest results from demographic theory are in treatment of the sorts of data which are available. In particular, from the demographic viewpoint, all we know uniformly in anthropological studies is the crude total population size. In the absence of age structure and growth rates, demography is of no help.

It is worth noting that there is a small body of literature dealing with what it calls "marriage theory". This literature hopes to deal with the effects of age structure by sex on such things as the availability of mates, and so on, (e.g., Keyfitz, 1968, or McFarland, 1972). This literature has been largely unsuccessful, we believe, because it has adopted the wrong model, or at least because it has modeled the problem incorrectly. We believe the "propensities" for different age groups to mate with each other are simply captured by the values  $P_{ij}$ , or at least the functions  $G_{ij}(p)$ , and that the fertilities of mating types are representable as the  $f_{ijk}$  values, which as fertilities over male and female age classes  $i, j$ , may have some  $f_{ijk} = 0$ , even for allowable  $(i, j)$  pairs. We therefore assert that the problem of

the above mentioned papers has been solved for the social theoretic case of zero growth and stationary class distributions. It has not been solved for the demographic theoretic case of non-zero growth, and stable age distribution. Because of the genetic interpretation always available for the social theoretic case, the problem has also been solved for zero growth, stationary genotype distributions (and other genetic distributions); but has not been solved for the "demographic case", as we have reviewed elsewhere (Weiss and Ballonoff, 1975), when interpreted genetically.