INFORMATION THEORETIC APPROACH TO MIGRATION ANALYSIS

P. Krishnan, University of Alberta

I. INTRODUCTION

Geographic and social mobilities of a population are of interest not only to policy decision makers but also to population scientists. Spatial or social, mobility is a complex phenomenon. Although, in this paper, our attention is mainly focussed on the spatial aspects of population movements, some social analogues are given room in the discussions.

The term "migration" connotes the geographical aspects of population mobility and involves "a change in place of abode, or place of usual residence". Migration is operationally defined as a "change of residence from one civil division to another".

There are a wide range of issues surrounding various measures of migration. For a good discussion of some of these issues, the U.N. monograph (United Nations, 1970) may be referred to. Since the estimation of a proper risk population for the computation of migration rates is not easy, various authors (Hamilton, 1965; Shryock, 1964; Thomlinson, 1962) have made a number of suggestions. In view of these difficulties, comparisons between regions, or nations in regard to the incidence of migration become less meaningful. The approach presented here would be able to partially solve some of these difficulties and open up new avenues of research in this area.

2. ENTROPY

Following Shannon, the expected information of a message on the occurrence of any one of the mutually exclusive events $E_i (i=1, 2, \ldots n)$ with prior probability $p_i (i=1, 2, \ldots n)$ is

$$H_n = \sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

H is the entropy of the probability distribution $P=(p_1, \ldots p_n)$. The meaning of entropy has been extended to social sciences in a series of studies by Theil and others. We follow here Theil's (1972) interpretation of entropy in terms of "integration-segregation", "inverse of concentration" or "dividedness" in its broadest sense. The only demographic application of this concept that this writer is aware of is due to Berry and Schwind (1969). Berry and Schwind show interest in testing different types of gravity models using entropy measures for a deeper insight into migration. This paper tries to introduce other notions into the field with the help of entropy.

3. MIGRATION DIVIDEDNESS

The population of a locality (region, province, nation) can be divided into two mutually exclusive groups: (a) migrants and (b) non-migrants. Depending on the definition employed and purpose at hand, this categorization may be labelled "stayers-movers". The movers may be thought of as separable into several non-overlapping sub-categories. Let p_s be the proportion of stayers according to place of birth (residence) data and P_m the proportion of movers. Then

$$P_s + p_m = 1 \qquad (2)$$

If p_m is decomposable into its components (subgroup proportions),

$$p_s + \sum_{i=1}^{n-1} p_{m_i} = 1 \qquad (3)$$

The entropy of this stayer-mover distribution is

$$H_2 = -[p_s \log p_s + p_m \log p_m] \qquad (4)$$

or

$$H_n = -\sum_{i=1}^{n-1} p_{m_i} \log p_{m_i} - p_s \log p_s \qquad (5)$$

Mathematically, the entropy vanishes when the population consists of only one group and is a maximum when the groups equally share the total population. But in practice, it is difficult to find a population (ignoring isolated island communities) which has only migrants. Hence in practice, zero entropy implies that the population is composed of stayers alone. Large values of entropy indicate that the migrant groups into which the population can be divided are numerous.

The entropy decomposition theorem can be utilized to set up the relationship between H_n and H_2 in (5).

Migration entropy permits us to assess the extent to which a population is divided into "stayers" and "movers". One may argue that the proportion of stayers in a population would serve the purpose. True! But the cut-off point for comparison purposes is usually arbitrary. It also does not consider separately the movements out to various other places. Entropy, on the other hand, is a smooth and continuous function, does not have the problem of "arbitrariness", and makes use of the data on all types of movements.

4. MIGRATION INEQUALITY

The notion of income inequality is extendable to migration studies. Let a nation be divisible into n provinces. If p_i and $q_i (i=1, 2, \ldots n)$ are respectively the population and migration shares of the ith province, and (p_i, q_i) are pair wise equal, there is no migration inequality with respect to the provinces. Inequality arises if the population and the migration shares are not compatible. Following Theil (1967), we can define the migration inequality measure as the expected information of the message which transforms the population shares to migration shares. Hence

$$H_n = \sum_{i=1}^{n} q_i \log \left(q_i / p_i\right) \qquad (6)$$

If $q_i=p_i$, H_n vanishes yielding the "egalitarian" situation. A large value of H_n indicates high degree of inequality.

Dividing the migrants by race, education, religion, native tongue, etc., one could develop measures of inequality specific for the stratification characteristic.

5. TESTING FOR GOODNESS OF FIT

Various types of models have been suggested to account for migration. Whether the models are of the gravity or economic type, the regression technique has been conventionally used for testing purposes. Markov models or their modifications are usually tested for on the assumption of steady state. We suggest that minimum entropy can be employed for testing the goodness of fit of migration models.

The motivation for this suggestion comes from Kerridge (1961). Suppose we have two finite discrete distributions $P=(p_i, \ldots p_n)$, $p_i \geq 0$ $\Sigma p_i=1$, and $Q=(q_i, \ldots q_n)$, $q_i \geq 0$, $\Sigma q_i=1$. If we assign the

distribution Q to an experiment with the space partitioned into n mutually exclusive events where the true underlying distribution is P, then according to Kerridge,

$$I_n = \sum_{i=1}^{n} p_i \log (p_{i/q_i}) \qquad (7)$$

is a measure of the error made by the observer. For another interpretation see Kullback (1959). Hence that model which has a minimum of I_n is preferrable to others. It is easily shown that

$$I_n = \tfrac{1}{2}\Sigma \frac{p_i(p_i-q_i)(p_i-3q_i)}{q_i^2} \qquad (8)$$

6. ILLUSTRATION

An illustration based on data from the 1971 Canadian census is presented in this section. Table 1 provides the discrete distribution of populations in 1971 by their place of residence in 1966 for each province/territory. The maximum value entropy is $\log_e 13 = 2.56495$. The entropy values for the provinces/territories are shown in Table 2.

The Yukon and the North West Territories have the second largest and the largest of entropies. Obviously the Territories do not have a high proportion of stayers and have to depend mostly on migrants for economic activity

Quebec has the lowest and Newfoundland the second lowest entropies. One can state these two provinces do not have (receive) a high proportion of migrants. The maritime provinces as a whole seem to attract migrants as do the western provinces. Among the western provinces, Saskatchewan is the least and Alberta the most attractive. Ontario is more attractive to migrants as compared to Quebec or Newfoundland but not as attractive as the remaining provinces.

From Table 2, one can note that the rankings of the provinces by entropy and stayer-proportions do differ. If the coming in of numerous types of migrants is taken as an indicator of the attractiveness of a province, then the entropy measure can be employed as a quantitative measure of migrant attraction. Thus Quebec and Newfoundland are the least attractive while Yukon and N.W.T. are the most attractive provinces/territories.

Table 3 shows the population shares of the different provinces/territories in Canada with their immigration shares. The data are based on the 1971 place of residence data. It is evident that a province (Ontario) which accounts for 35.7 percent of population has 53.3 percent of the immigrants. The other province which exceeds in its share of migration is B.C. Alberta is close to the egalitarian line. All other provinces fall for below the egalitarian line. The entropy measure of inequality is calculated as 10.5 percent.

REFERENCES

Berry, B. J. L. and P. J. Schwind (1969), "Information and entropy flows", Geographical Analysis, 1, 5-14.

Hamilton, C. H. (1965), "Practical and mathematical considerations in the formulation and selection of migration rates", Demography, 2, 429-443.

Kerridge, D. F. (1961), "Inaccuracy and inference" Journal of the Royal Statistical Society, Series B, 23, 184-194.

Kullback, S. (1959), Information Theory and Statistics, New York: John Wiley and Sons.

Shryock, A. S. (1964), Population Mobility within the United States, Chicago: Community and Family Study Center, University of Chicago.

Theil, H. (1972), Statistical Decomposition Analysis, Amsterdam: North-Holland Publishing Company.

Theil, H. (1967), Economics and Information Theory New York: American Elsevier.

Thomlinson, R. (1962), "The determination of a base population for computing migration rates", Milbank Memorial Fund Quarterly, 40, 356-366.

United Nations (1970), Methods of Measuring Internal Migration, Population Studies, No. 47, New York: The United Nations.

Table 1

POPULATION DISTRIBUTION BY RESIDENCE IN 1966 AND 1971, CANADA

RESIDENCE IN 1971	RESIDENCE IN 1966												
	Nfld.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Yukon	N.W.T.	Other Countries
Nfld.	.7593	.0003	.0033	.0016	.0293	.0860	.0029	.0025	.0058	.0058	.0002	.0019	.0551
P.E.I.	.0186	.4547	.1000	.0716	.0413	.1674	.0219	.0052	.0101	.0180	.0005	0	.0820
N.S.	.0268	.0170	.5378	.0648	.0475	.1465	.0142	.0065	.0156	.0194	.0006	.0008	.1025
N.B.	.0176	.0134	.0771	.5364	.0941	.1258	.0142	.0048	.0114	.0121	.0008	.0013	.0910
Que.	.0021	.0006	.0041	.0075	.7949	.0460	.0044	.0014	.0031	.0045	.0001	.0002	.1369
Ont.	.0107	.0021	.0154	.0111	.0594	.5941	.0142	.0048	.0114	.0121	.0008	.0013	.0910
Man.	.0036	.0014	.0089	.0073	.0232	.0989	.5255	.0878	.0386	.0339	.0006	.0020	.1692
Sask.	.0017	.0007	.0037	.0029	.0094	.0410	.0565	.7119	.0634	.0365	.0004	.0015	.0703
Alta.	.0021	.0013	.0087	.0057	.0205	.0623	.0461	.1109	.5040	.0735	.0017	.0047	.1585
B.C.	.0022	.0009	.0094	.0048	.0258	.0730	.0415	.0461	.0908	.5240	.0030	.0017	.1767
Yukon	.0076	.0007	.0090	.0028	.0236	.0686	.0340	.0644	.1684	.3548	.1123	.0291	.1240
N.W.T.	.0116	.0033	.0176	.0160	.0440	.1227	.0600	.0886	.2448	.1155	.0138	.1601	.1018

Source: Computed from 1971 Census of Canada

Table 2

MIGRATION ENTROPY FOR CANADIAN PROVINCES, 1971

PROVINCE	ENTROPY	RANK	PROPORTION OF STAYERS	RANK
Nfld.	0.8201	2	0.7593	2
P.E.I.	1.7321	10	0.4547	10
N.S.	1.5815	8	0.5378	5
N.B.	1.5667	7	0.5364	6
Que.	0.7448	1	0.7949	1
Ont.	1.0757	3	0.5941	4
Man.	1.5331	5	0.5255	7
Sask.	1.1279	4	0.7119	3
Alta.	1.5959	9	0.5040	9
B.C.	1.5396	6	0.5240	8
Yukon	1.9398	11	0.1123	12
N.W.T.	2.1650	12	0.1601	11

Source: Author's computations and 1971 Census of Canada.

Table 3

MIGRATION INEQUALITY: CANADA, 1971

PROVINCE	PER CENT POPULATION SHARE	PER CENT MIGRATION SHARE
Nfld.	2.4	0.49
P.E.I.	0.5	0.18
N.S.	3.7	1.60
N.B.	2.9	1.09
Que.	27.9	16.71
Ont.	35.7	53.26
Man.	4.6	3.82
Sask.	4.3	1.42
Alta	7.6	7.30
B.C.	10.1	13.93
Yukon	0.08	0.11
N.W.T.	0.16	0.11

Source: 1971 Census of Canada.