1982

Entropy Maximization, Information Minimization And Population Potential: Joint Considerations

James Albert Pooler

ENTROPY MAXIMIZATION, INFORMATION MINIMIZATION AND POPULATION POTENTIAL: JOINT CONSIDERATIONS

by

James A. Pooler

Department of Geography

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Faculty of Graduate Studies

The University of Western Ontario

London, Ontario

April, 1982

ABSTRACT

The dissertation presents joint considerations of the concepts of (1) entropy maximization and population potential and (2) information minimization and population potential.

In order to calculate population potential it is necessary to determine the probability that a randomly selected individual at one location will have a spatial interaction at another location. The methods of maximum entropy and minimum information provide a means for estimating such probabilities. Population potential is closely related to, and derivable from, spatial interaction and this allows an indirect connection to be made between population potential and the methods of maximum entropy and minimum information. This is, therefore, a reconsideration of the potential concept within the context of these contemporary methods of probability estimation and it addresses the question of whether it is possible to derive, and practicable to employ, entropy maximizing and information minimizing population potential functions.

The amount of potential interaction between places is a function of both the attenuating effect of distance and the emissivity of the origin, and an attempt is made to demonstrate that maximum entropy is of relevance to the former and minimum information to the latter. In the

former case a quadratic gamma family of distance response functions for potentials is derived, and in the latter case a method is presented for allowing for changes in the emissivity of the origin by sequentially adding alternative prior probability distributions.

ACKNOWLEDGEMENTS

In the first years of this author's graduate career Robert McDaniel provided intellectual stimulation, sound guidance, and an attentive ear. He remains a close ally, and a good friend, and I would like to offer him my sincere thanks.

The Michigan experience wouldn't have been an experience if it were not for the presence there of Waldo Tobler, Gunnar Olsson and Joao Francisco de Abreu. I consider myself lucky to have been able to experience the unique and insightful 'Tobler Approach' to things geographical and I would like to thank him for providing me with a glimpse of the spatial world he sees. Gunnar Olsson, also sometimes known as H.C.E., took all of the 810'ers down a rabbit hole and through the wonder land of language. Many ladders were built, some were thrown away, but in the end the dream, just like Alice's, ended. Even Finnegans wake. The adventures are not forgotten, however, and an enormous debt of gratitude is owed. Francisco was the best of friends and I would like to thank him for the many, many hours of sincere conversation and debate. Peter Hoag, Joe Lau, Adrian Pollock, and Mike Watts should not go unmentioned.

I would like to thank my principal (and principle) advisor, William Warntz, for his encouragement, interest, advice, good humor, consolation, friendship and, to use his own words (written some twenty years ago about a graduate student he once knew), "patience and forbearance with what at times must have been a 'trying young man'". It was he who first piqued my interest in spatial analysis and he has since lent many years of unyielding support. My debt to him shall never be repaid.

Mike Goodchild has contributed enormously to this study. He has taken a genuine interest in the work and has guided me at every step of the way, in spite of my inventive mathematics and split infinitives. He has also demonstrated patience above and beyond the call of duty and I offer him my sincere thanks.

Throughout the years Bill Coffey has been one of my closest co-conspirators in the game of graduate life. From the highs of the controversy surrounding the intellectual conspiracy to form the macro-lab, to the lows of the Elbow Room requiems, we struggled together against the Philistines. We were much smarter then than we are now, and in a few more years we may know nothing at all. We're both stubborn, and I think that's why we get along so well.

Thanks are also due to Don Janelle for his support, concern, and friendship over the years.

A special note of appreciation goes to parents and in-laws who have waited patiently, uncritically, and silently over the years for the thesis, a job, and a grandchild. Sometimes the unspoken word says the most. We thank them for their support, concern, and love.

In addition to all of the above members of the cast, a note of gratitude also goes out to Mr. and Mrs. George Hatherell, who provided excellent living accommodations while this was being written, and especially to Jackie, Rob, Carrie, and Sean Mackintosh, who provided so much worry free time by giving the best of loving care to Sarah.

Finally, the deepest and most sincere expression of gratitude must go to Priscilla who has sacrificed so much for the selfish interests of her husband. Without her support and love this thesis would never have been written and it is for that reason that I dedicate it to her.

London, Ontario
August, 1981

James Pooler

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Population Potential

The concept of population potential was initially formulated and popularized by Princeton University astrophysicist John Q. Stewart (1941, 1942, 1947). Assume first that we have a map which has been partitioned into zones of known population size and on which we have identified the zone centroids and, possibly, other control points. Stewart asserted that the "influence" v at a point j of a population P in zone i, is a decreasing function of the distance r between the zone centroid and the other point

$$v_{ij} = P_i/r_{ij} \quad . \tag{1.1.1}$$

The population potential V at j is obtained by summing over all contributing zones such that

$$V_j = \sum_i P_i/r_{ij} \quad , \tag{1.1.2}$$

where V_j is in units of persons per distance unit. The summation is repeated for each control point and a contour map of isopotentials can then be produced via standard methods of interpolation. Potential is calculable at any spatial scale. Coffey (1977) for example has employed the concept at the urban scale, while Warntz (1965) has constructed potentials at the world level.

Equation (1.1.2) is normally considered to represent a numerical approximation to the definite integral

$$V_j = \int_{(\theta)} \frac{1}{r} \, d \, \underline{da} \quad , \qquad (1.1.3)$$

where d is the population density of an infinitesimal element of area da and r is the distance from this element to point j. The integration is extended to all elements of the surface (θ) where the density is not zero. McCalden (1972, 1975) has reported on the difficulties involved in attempting to evaluate analytically, integral forms of potential.

It is common to employ a self-potential term v_{jj} in the calculation of equation (1.1.2) as a measure of a population's influence on itself. The reason for such a procedure is that otherwise v_{jj} equals infinity since P_i is being divided by zero. One common method (there are others) of calculating v_{jj} is to assume that the region associated with j is circular with uniform density and to take the distance to itself as one-half of the radius. Thus if $a = \pi r^2$ where a is area, then $r = (a/\pi)^{1/2}/2$. Self-potential, given a value for area, is therefore

$$v_{jj} = 2 \, P_j / (a/\pi)^{1/2} \quad . \qquad (1.1.4)$$

The quantity V_j was described by Stewart (1942), and by Warntz and Wolff (1971), as a measure or index of the influence of population at a distance. As Stewart (1947, 221) stated: "where the influence of people sums up to large values we have 'highlands' and 'peaks' of influence. Such points are nearer to more people and all kinds of sociological activities are expected to be at a high level there."

Reviews of the concept of population potential can be found in Carrothers (1956), Isard et al (1960), Olsson (1965) and Warntz and Wolff (1971). The role of the potential concept in the growth of systematic Anglo-American human geography is discussed in Johnston (1979).

1.2 Entropy Maximization

Since its introduction into geography the entropy concept has been employed in two basic ways: as a device for estimating the form of a probability distribution, given a limited amount of information, and as a measure on an existing probability distribution.

Assume, in the former case, that we have a population (in the statistical sense) of size X, where x_i represents the number of members to be assigned to cell i. Define

$$p_i = \frac{x_i}{X} \qquad (1.2.1)$$

as the probability that a randomly selected member will be assigned to cell i. It is given that

$$\sum_i p_i = 1 \quad , \qquad (1.2.2)$$

and it is known also that

$$\sum_i p_i y_i = \bar{y} \quad , \qquad (1.2.3)$$

where $\bar{y}$ is the mean value of some measure y_i on n cells ($i=1,\ldots n$). The most likely form for a probability distribution which assigns members of the population to cells is one which maximizes the entropy (see Shannon and Weaver, 1949)

$$H = -\sum_i p_i \ln p_i \quad . \qquad (1.2.4)$$

It is well known (see Jaynes, 1957; Tribus, 1961; Wilson, 1970; Webber, 1976, for example) that the maximum of equation (1.2.4), with constraints (1.2.2) and (1.2.3) is

$$p_i = \exp(-\lambda - by_i) \qquad (1.2.5)$$

where λ and b are parameters which ensure that the constraints are satisfied.

Equation (1.2.5) is an expression which indicates the statistically most likely form for a probability distribution given only the information in constraints (1.2.2) and (1.2.3). In human geography the methodology outlined above has been widely used as the basis for deriving density models, location models, and in particular, spatial interaction models (see Wilson 1970, 1974).

In the case of the second type of usage of the entropy concept—as a measure on an existing probability distribution—the value of H in equation (1.2.4) is calculated directly from the data. The higher the value of H, the more dispersed the distribution. The lower the value of H the greater the tendency to uniformity where, for a perfectly uniform probability distribution, H achieves its maximum value

$$H = \ln n \quad . \tag{1.2.6}$$

When calculated in this direct way H has been variously interpreted as a measure of concentration, uncertainty, information, dispersion, and dividedness. This mode of application of the entropy concept has also been quite popular over the years in human geography (see Getis and Boots, 1978).

A guide to the entropy maximizing methodology has been written by Senior (1979). An overall review of the concept from a geographic perspective is in Haynes, Phillips, and Mohrfeld (1980).

1.3 Information Minimization

Assume, that in addition to constraints (1.2.2) and (1.2.3), some additional a priori information is available which is suspected to be of relevance in the assignment of individuals to cells. Define s_i as some such measure on a cell, having a total value of S over all cells, and

$$q_i = \frac{s_i}{S} \tag{1.3.1}$$

as the probability that s_i will be found in a randomly selected cell. The statistically most likely form for probability distribution p_i, subject to constraints (1.2.2), (1.2.3) and (1.3.1), is the one which minimizes the Kullback (1959) information

$$I(q:p) = \sum_i p_i \ln \frac{p_i}{q_i} \qquad (1.3.2)$$

where q and p are prior and posterior probability distributions respectively. The minimum of this function subject to the three constraints is

$$p_i = q_i \exp(-\lambda - by_i) \qquad (1.3.3)$$

where the prior q_i will ensure that the assignment of individuals to cells takes place in direct proportion to the measure s_i on those cells.

This is the method of minimum information and it has also been employed in human geography as the basis for deriving spatial models (see Webber, 1979, for example). It can be shown that the entropy is a special case of the information when the prior is uniform and as such it has been argued that the Kullback information statistic constitutes a generalization of, and improvement on, the Shannon entropy (see Hobson and Cheng, 1973).

The Kullback statistic, like the entropy, can also be employed as a measure on existing probability distributions. In particular it is a measure of the information to be gained in the transition from prior to posterior probability distributions; the greater the difference between the distributions, the higher the value of I(q:p).

The use of I(q:p) as a measure on probability distributions is extensively discussed in Theil (1967; 1972). The method of minimum information is treated in Snickars and Weibull (1977) and especially in Webber (1979).

1.4 Purpose of the Study

The purpose of this study is to undertake joint investigations of (1) entropy maximization and potential, and (2) information minimization and potential. The problem of calculating the potential at a point j is, in the first instance, one of determining the probability that a randomly selected individual at i will have a spatial interaction at j. The entropy maximizing and information minimizing formalisms provide methods for estimating such probabilities. Population potential is derivable from interaction and this therefore allows an indirect connection to be established between potential and the methods of maximum entropy and minimum information. Insofar as the potential concept has continued to stand on a rather shaky foundation for some thirty years now—based, as will be shown, on a questionable physical analogy—a reconsideration of the concept within the context of these contemporary methods of probability estimation seems warranted.

The amount of interaction between places, within the context of population potential, is a function of both the attenuating effect of the distance between them and the size or emissivity of the origin. An attempt will be made here to demonstrate that the method of maximum entropy is of relevance to the former and the method of minimum information to the latter.

The organization of the discussion is as follows. The concept of potential is considered in greater detail in the second chapter, and the concepts of maximum entropy and minimum information are discussed more fully in the third chapter. In the fourth chapter, the joint consideration of the concepts is undertaken with the derivation of entropy maximizing and information minimizing interaction models and, indirectly, potential functions. The fifth chapter presents a consideration of some of the problems having to do with the operationalization of the derived functions and with the interpretation of results. The sixth and final chapter presents a discussion of some of the more general issues arising from the joint investigations as well as a critical review of results.

There is no empirical investigation included as such, since the principal concern is not with the ability of any particular form of the potential function to predict the size of other variables (as has often traditionally been the case). Rather the central issue is considered to be the question of whether it is possible, and practicable, to derive and develop potential functions within the entropy maximizing and information minimizing frameworks.

CHAPTER 2

POPULATION POTENTIAL

2.1 The Origins of the Potential Concept

Stewart's original derivation of population potential seems to have been based primarily on empirical data. In an analysis of data on the geographical distribution of college students he (1941, 89) found that

> The number of undergraduates or alumni of a given college who reside in a given area is directly proportional to the total population of that area and inversely proportional to the distance from the college.

The discovery of this empirical regularity led Stewart (1941, 89) to state a general definition of population potential:

> I define the "potential" of the population of a given area, at a given point, as the population of the area, in millions, divided by the average distance, in miles, from the point to the area.

Such a definition was apparently not only a result of Stewart's empirical findings but also due in part to his familiarity with analogous concepts in the physical sciences. As he (1947, 461) wrote:

> The evident tendency of people to congregate in larger and larger cities represents an attraction of people for people that turns out to have a mathematical as well as merely verbal resemblance to Newton's law of gravitation. Lagrange in 1773 found that where the attraction of several planets at once was under consideration, a new

> mathematical coefficient, not used by Newton, simplified the calculations. This coefficient amounted to a measure of the gravitational influence of a planet of mass m at a distance d, and it was as simple as possible, merely m/d. Later mathematical physicists, Laplace and Poisson, further elaborated the m/d concept in celestial mechanics. Not until 1828 did Green find that similar measures existed of the influence of an electric charge e, and of a magnet pole, p, at a distance; namely e/d and p/d respectively. To these quantities the name "potentials" was given—the gravitational potential, the electrostatic potential, the magnetic potential.

Stewart (1947, 461) continued

> In 1939 evidence was uncovered which suggested that the influence of people at a distance could be expressed by a similar coefficient, namely P/r—P being the number of people, and r their distance away. For this coefficient the name 'potential of population' was at once suggested because of the physical analogies.

The Newtonian law of gravitation to which Stewart referred is the inverse-square law describing the gravitational force F_{ij} between two masses m_i and m_j separated by a distance r

$$F_{ij} = G \frac{m_i m_j}{r^2} \qquad (2.1.1)$$

where G is the gravitational constant and where the force, defined as the "weight" due to gravitation, is measured in newtons (see Klein, 1971). The gravitational potential at a point is a measure of the work or energy required to move a unit of mass from infinity to the point (see Kellogg, 1929; MacMillan, 1930, for example). The potential can be derived from the law of gravitational attraction as follows.

Since the concern is with a single unit mass (i.e., $m_j = 1$) the equation for gravitational force (2.1.1) can be rewritten

$$F_{ij} = G \frac{m_i . 1}{r^2} \quad . \qquad (2.1.2)$$

The unit mass is to be moved from infinity toward m_i and it is desired to measure the amount of work done at location α (which is between infinity and the mass). The amount of work done is measured as a force times a distance when the force is constant (over a very small distance $\underline{dr}$, force is constant). Therefore at some arbitrary point r, the amount of work done W in moving a very small distance, from r to r-$\underline{dr}$ is

$$\underline{d}W = -G \frac{m_i}{r^2} \underline{d}r \quad . \qquad (2.1.3)$$

To get the total work done in moving from infinity to α it is necessary to integrate equation (2.1.3) over that range such that

$$\int_{\infty}^{\alpha} \underline{d}W = -\int_{\infty}^{\alpha} G \frac{m_i}{r^2} \underline{d}r \quad . \qquad (2.1.4)$$

Evaluating this expression we have

$$W = -Gm_i \int_{\infty}^{\alpha} \frac{dr}{r^2} ,$$

$$= +Gm_i \left| \frac{1}{r} \right|_{\infty}^{\alpha} ,$$

$$= + \frac{Gm_i}{\alpha} - \frac{Gm_i}{\infty}$$

$$= + \frac{Gm_i}{\alpha} \qquad (2.1.5)$$

as the gravitational potential at location α.

This brief description of the relationship between Newton's law of gravitational attraction and the concept of gravitational potential due to Lagrange has illustrated how the inverse-square formulation is transformed into the inverse distance formulation. We can now go on to investigate such a transformation in the case of the gravity model of spatial interaction and the index of population potential in geography.

The traditional geographical "gravity model" is sometimes written as an analogue of the inverse-square expression for gravitational force given above (equation 2.1.1), that is

$$T_{ij} = k \frac{P_i P_j}{r_{ij}^2}, \qquad (2.1.6)$$

where T_{ij} represents the amount of interaction, or number of trips, between places i and j (see, for example, Kolars and Nystuen, 1974, 69; Taylor, 1977, 287; Wilson, 1967, 16). It might be suggested that the Stewart form of the population potential function can be derived directly from this gravity model simply by dividing through by P_j. It can be seen however that it is also necessary to take the intermediate steps described above if one wishes to derive the Stewart potential from equation (2.1.6). This gravity model (2.1.6) is an inverse-distance-squared function whereas the Stewart form of potential is an inverse-distance function.

The usual approach, instead, has been to rewrite the gravity model with a generalized exponent b on the distance term such that

$$T_{ij} = k \frac{P_i P_j}{r_{ij}^b} \qquad (2.1.7)$$

and then to derive the potential function from this (see Yeates, 1974). Thus, a potential function can be derived from equation (2.1.7) if the equation is employed to estimate the propensity for interaction from all i's to a single j. In this case we can write

$$\frac{P_1 P_j}{r_{1j}^b} + \frac{P_2 P_j}{r_{2j}^b} + \frac{P_3 P_j}{r_{3j}^b} + \ldots + \frac{P_{nj}}{r_{nj}^b} = \sum_i \frac{P_i P_j}{r_{ij}^b} . \qquad (2.1.8)$$

Insofar as P_j is common throughout the summation we can divide through to get

$$\frac{P_1}{r_{1j}^b} + \frac{P_2}{r_{2j}^b} + \frac{P_3}{r_{3j}^b} + \ldots + \frac{P_n}{r_{nj}^b} = \sum_i \frac{P_i}{r_{ij}^b} , \qquad (2.1.9)$$

where

$$\sum_i \frac{T_{ij}}{P_j} = V_j = \sum_i \frac{P_i}{r_{ij}^b} \quad . \qquad (2.1.10)$$

This is a popular approach to the derivation of a potential function, however it is important to specify that this derivation is not consistent with the potential function employed by Stewart; to derive the function in this manner and then set b equal to one is purely arbitrary.

In summary it can be noted that to state that "Newtonian theory suggests an exponent of 1.0" (Rich, 1980, 20) or that "a classical Newtonian interaction model must be assumed in order to derive equation [(1.1.2)]" (Sheppard, 1979a, 442) is misleading. The Stewart demographic potential function is not derived or derivable from the Newtonian law of gravitational attraction unless the intermediate steps outlined above (2.1.5) are carried out.

We have seen two approaches to the derivation of potential measures. In one case the potential was gotten by dividing out the destination term in an interaction model based on the Newtonian inverse-square law of gravitational attraction. In the other case the derivation was based on the original definition of potential as a measure of work done in a physical system. In commenting on these two approaches Goodchild (1979, 87) noted that

> Unfortunately [the former] is inconsistent with [the latter], since it would require

> the power of distance to be the same in both potential and interaction equations, rather than differing by one. It is easy to see why it has become popular, however, since the word 'potential' is much closer to its common meaning. In the physical analogy 'potential' is the potential to do work [and] not in any sense potential interaction. Since there is no obvious social analogy to the concept of physical work, it is in fact difficult to see how the idea of a physical analogy ever arose.

For the remainder of this study we will accept the definition of potential as potential per capita spatial interaction. This is discussed in greater detail in the section which follows.

2.2 Interpretations of Potential

Although the derivation given in equations (2.1.7) to (2.1.10) does not lead to the Stewart potential function it is nevertheless very useful in helping to illustrate the meaning of, and interpretation to be given to, the potential concept. As a result of the division by P_j in equation (2.1.8) it can be said that equation (2.1.10) becomes a statement concerning the propensity for interaction on a per capita basis at j. This is an important property of a potential function and one which distinguishes a potential function from an interaction model. It is also an area where some ambiguity seems to exist.

Sheppard (1979a, 441), for example, has described the omission of the P_j term in potential functions as if it were a theoretical weakness. In discussing the Stewart form of potential he noted that the assumption is made that "characteristics of the destination do not influence the level of spatial interaction" and that such an assumption is "clearly untenable." He (1979a, 422) stated further that "the problem is solved by inserting P_j in the numerator of [the equation]." The statement that "no geographer has made the adjustment" (Sheppard, 1979a, 442) is however testimony to the fact that potential is commonly regarded as something other than a model of spatial interaction.

The amount of interaction between places is not only a function of the "emissiveness" of the origin but also of the "attractiveness" of the destination. Since a potential function contains no destination term it obviously has little relevance to the modelling of observed spatial interaction. It is only reasonable to employ a potential function in modelling spatial interaction when one is considering spatial interaction to a single place and, as we will see, even in such instances it would normally be preferable to employ an origin constrained interaction model or a Lowry-type location model (see Wilson, 1970). An unconstrained potential function can be employed but will only provide results which are "proportional to" the amount of interaction.

In the context of spatial interaction, potential is probably best defined as a measure or index of the "potential for interaction" or "possible interaction." As Rich (1980, 3) has stated

> Whereas the gravity model is concerned with analysing or predicting an observed pattern of spatial flows, the potential model is more concerned with the opportunity for interaction . . . than with the interaction itself.

It is in this sense that potential has commonly been defined as an index of accessibility or influence, and as a measure of the intensity of possible contact or social intensity. Whichever word or phrase one chooses to use it is perhaps most important to note that in every case the potential is being described as an index or measure rather than a model.

The definition of the basic Stewart potential function as an index is perhaps better understood if it is compared to the concept of population density. The practice of dividing population by area is well known and widely accepted, although this was by no means always the case (see Stewart and Warntz, 1958). Potential can be defined as a generalized or weighted population density. Craig (1972, 10) makes this explicit by noting that if population density d_i is expressed in terms of population and area a_i

$$d_i = P_i / a_i \quad , \qquad (2.2.1)$$

then population can be written with respect to density and area

$$P_i = d_i a_i \quad , \qquad (2.2.2)$$

and potential can be given as

$$V_j = \sum_i d_i a_i / r_{ij} \quad . \qquad (2.2.3)$$

In Craig's (1972, 10) words, this expresses population potential as a "weighted average of the population densities".

Population potential has been repeatedly tested against a wide variety of variables considered to be representative of the influence of population or social intensity. An extensive, though not complete list of such variables was given by Warntz and Wolff (1971, 236):

1. Telephone calls, telegrams, and mail
2. Bank checks
3. Bus, railway, and airline passengers
4. Visitors to fairs
5. Hotel registrations
6. Marriage licenses
7. Obituary notices
8. College attendance
9. Areas of cities
10. Rural population densities
11. Land values
12. Highway and railway network densities and alignments
13. Bank deposits and the "velocity" of money
14. Information flows and decision making
15. Administrative areas
16. Taxes
17. Patents
18. Business failures
19. Alcoholism and mental health

20. Farm sizes
21. Commodity prices

(see also Stewart, 1950; Stewart and Warntz, 1958; and Neft, 1966). These comparisons, normally by regression analysis, often showed positive and strong correlations between the log of population potential and the log of such variates. In addition, power laws associated with the slopes of the regression lines were often discovered to hold true over time. Rural population density, for example, tends to be proportional to the square of the population potential in the United States for a number of census years (Stewart, 1947; Stewart and Warntz, 1958).

Potential maps can and have also been used as purely descriptive devices as well. They have, for example, been employed to illustrate the impact of the entry of the U.K., Ireland, Norway, and Denmark into the European Economic Community (Clark <u>et al</u>, 1969), and to illustrate the impact of a new town on its region (Batty, 1976a, 101).

Potential, then, is not primarily a model of spatial interaction but rather an index or measure of aggregate accessibility. It can also be defined, following Warntz and Wolff (1971, 216) as a general measure of relative position or location. It is related to spatial interaction in the sense that it describes the propensity for interaction on a <u>per capita</u> basis at the destination.

What are the possible uses for such an index? It has been seen that although potential can be employed as a descriptive device, one interest has historically been with finding the relationship between potential and other social and economic variables. Warntz and Wolff (1971, 236), for example, suggested that the "strong correlations" between potential and variables such as those listed above indicate that "the potential-of-population surface represents the spatial 'structuring' for these kinds of phenomena." The implication of such a statement is that population potential is assumed to be more highly correlated with such variables than is population density. Stated another way it could be said that certain quantities and levels of activities are expected to vary as a function of the degree of accessibility of the area in which they exist; proximity to the aggregate population is hypothesized to be a better explanatory variable than local population density.

A word of warning is in order here. One would naturally expect many of the types of variables listed above, such as marriage licences, land values, or highway network densities, to be positively correlated with population densities. If the goal is to study the correlation between potential and such dependent variables, then the correlation between the density variable from which the potential was constructed, and the dependent variable, should also be investigated. Goodchild et al (1981) found, for example, that the observed correlation between log United States population density and log population potential for 1975 (using the Stewart form for potential) is 0.899. This suggests that many of the variables which were found to be correlated with potential, as listed above, may also be highly correlated, or even more so, with population density. Goodchild et al (1981) suggest that

> the correlation between potential and density is combined with the correlation between the two densities. The significance of the spatial arrangement of [variable] pairs can be tested by repeating the analysis with randomized pairs. Randomization of [one variable] alone will test whether the relationship between [the variables] is significant.

In general it might be said that the goal should be to determine how much of the correlation is due to local density and how much is attributable to spatial proximity to the remainder of the population.

The usually strong correlations between population densities and population potentials do, however, suggest that potential can be regarded as a generalized population density and employed as such. As Craig (1972, 11) has noted, if one wishes to compare the population densities of different places, population potential

> is directly relevant to the kind of comparison for which the 'unweighted'

> or 'crude' population densities are used; not just as an undefined measure of pressure or accessibility, but as a weighted sum of population densities which will often be more meaningful than the comparison of unweighted densities which are so dependent on the area used.

Another possible usage for potential functions, which relates to their being defined as indices of accessibility, is as terms in other models. Thus even though potential is not directly concerned with modelling interaction, it can be employed as an origin or destination term <u>within</u> an interaction model. Imagine, for example, that it was desired to model recreational trips to parks where the interest was in trip makers who were staying in a different park each night. In this case a potential might be employed as a destination term which is representative of aggregate accessibility to camping facilities. Perhaps the most common application of potential functions in this manner is in spatial interaction shopping models where potential is employed to represent accessibility to stores and facilities. Both of these types of usage require a redefinition of the P_i term. This subject will be considered in greater detail below and this will also illustrate additional possible uses of potential functions.

This review of the meaning of, and possible uses for potential is far from complete. It would also be possible for example, to include discussions of the topological properties of potential surfaces (Warntz and Woldenberg,

1967) and the gradients and flows associated with such properties (Warntz, 1966). The role of ratios of income and population potentials as indicators of urban, social, and economic integration could have been considered (Dutton, 1970; Coffey, 1977) as could the use of *per capita* income and income potential in the identification of economic sub-systems or regions and associated population "weights" (Warntz, 1965; Pooler and de Abreu, 1979). Various thermodynamic concepts and analogies could also have been discussed (Fein, 1970; Warntz, 1973a). A detailed consideration of these issues goes beyond the scope of the present study but it can nevertheless be suggested that in some cases it may be appropriate to reconsider such issues in the context of the arguments to be presented in the remainder of this study.

In the broadest terms it can be said that potential is a general index or measure of accessibility and relative location which is useful whenever such an index is required. Potential is, however, just one of a number of measures on areal distributions. Some of these will be reviewed in the section which follows. The purpose of the review is to illuminate, by contrast, the nature of potential.

2.3 Other Measures on Areal Distributions

Potential was described in the preceding discussion as a general measure of position. In this regard the highest peak of a potential surface takes on a special significance as the most accessible point of an areally distributed population. Related summary measures on areal distributions exist and will be briefly reviewed here. The discussion will be restricted to measures which are of a mathematical form which is applicable to "grouped" or spatially partitioned density data. It will also be restricted to the consideration of four basic measures of average position (or spatial central tendency) and their associated moments. Additional measures of average position, and measures of dispersion around the averages, are available but will not be considered here (we draw freely on Warntz and Neft, 1960, and Neft, 1966; these sources should be consulted for more detailed treatments).

One measure on an areal distribution is the <u>mean center</u> which can be defined as the position where

$$\sum_i P_i \, r_{ij}^2 \, , \tag{2.3.1}$$

is a minimum. In other words this is "the point where the sum of the squares of the distances to the individuals comprising the population will be a minimum" (Warntz and Neft, 1960, 48). The mean center can be described as the "balancing point" or "center of gravity" of a distribution. If the population density is assumed to be uniform, this measure will indicate the <u>center of area</u> (Neft, 1966, 29).

Another measure of average position is the median center. In linear statistics the median is that value about which "the algebraic sum of the deviations is a minimum" (Warntz and Neft, 1960, 49). If the deviations are regarded as distances, the median center of a two-dimensional distribution is defined as the minimum of

$$\sum_i P_i \, r_{ij} \quad . \qquad (2.3.2)$$

In general terms the median center is the position of minimum aggregate travel and can be considered to represent an optimum location for retail and service facilities, meeting places, and so on (although there are also other definitions of this).

A third measure on an areal distribution is the modal center which is defined, in general terms, as the highest value on a "smoothed" density surface. The precise value of the modal center "can be determined only by means of the mathematical formula which describes the continuous curve of 'closest possible fit' for the given frequency distribution" (Warntz and Neft, 1960, 50).

A fourth measure of average position on an areal distribution is known as the harmonic mean center and is the location of the minimum of

$$\frac{1}{\sum_i P_i / r_{ij}} \quad . \qquad (2.3.3)$$

This will be recognized as the reciprocal of population potential and obviously where equation (2.3.3) is at a minimum the potential will be at a maximum. Neft (1966, 35) noted that "this relationship also adds to the usefulness of the peak of potential, since the fact that this peak has certain statistical properties akin to the harmonic mean can...be utilized."

With the exception of the modal center, the measures of average position which were introduced above are definable not just at their minima, but also at any point on a surface. It follows from this that these measures can be mapped. Warntz and Neft (1960), and Neft (1966), have termed such mappings as being illustrative of areal moments.

The use of the term "moment" arises from the fact that the equations for the mean, median, and harmonic mean centers are analogous to a class of statistical measures on linear frequency distributions known as moments. A general formula for the moment about any point j can be written as

$$M_{bj} = \frac{\sum_i P_i \, r_{ij}^{\,b}}{P} , \qquad (2.3.4)$$

where P is the total population and b is both the exponent on distance and the number of the moment. Since P is a constant for a given population it can be ignored. It can be seen that by setting b equal to two, one, and minus one, equations analogous to those for the mean, median, and harmonic mean respectively, result.

In the case of the mean center or "center of gravity" the associated second moment can be defined

$$M_{2j} = \sum_i P_i \, r_{ij}^{2} \quad . \qquad (2.3.5)$$

The contours of the second moment—in person-miles-squared—will map as concentric circles around the center of gravity (regardless of the distribution of population).

In order to map the first areal moment, which corresponds with the median center, define

$$M_{1j} = \sum_i P_i \, r_{ij}^{1} \quad , \qquad (2.3.6)$$

as the value, in persons times miles, at any point. The resulting contour map will represent "aggregate travel distance". If any value on the map is divided by the total population, the resulting value at that point "represents the arithmetic mean distance, in miles, required to move every individual in the population to that point by shortest distance" (Warntz and Neft, 1960, 64).

The contour map which corresponds with the harmonic mean center is based on the inverse first moment

$$M_{-1j} = \sum_i P_i \, r_{ij}^{-1} \quad , \qquad (2.3.7)$$

and this, of course, is identical to a map of population potential (provided that the self-contribution is treated the same in each case). Again, this will map as concentric rings only for circularly symmetrical population distributions (although local minima and maxima are possible).

Neft (1966) has argued that there are a number of desirable properties which make the peak of potential (or harmonic mean center) the most useful of these measures of average position. Whether this is true is difficult to answer in a definitive way, however, since the usefulness of any of these measures of average position will depend upon the particular situation in which they are to be employed. Nevertheless it can be pointed out, for the sake of contrast, that the peak of a potential surface is not a minimum aggregate travel point, nor a balancing or center of gravity point (although they may coincide in some cases).

The discussion in this and the preceding sections has had to do, directly or indirectly, with some of the properties of potentials and potential surfaces. Such a discussion would be incomplete, however, if it did not include the subject of the next section, a review of the relationship between spatial aggregation and potential.

2.4 Spatial Aggregation and Potential

In discussing the relationship between spatial aggregation and potential, at least four effects due to zone size and/or configuration can be identified. The first of these relates to the "smoothness" of the potential surface (where the word "smooth" is being employed in a general pedagogic and conceptual sense, that is, it will not be given a mathematical definition except to say, as above, that potential can be regarded as a distance weighted density). Stewart (1947, 222) originally pointed out, and Goodchild et al (1981) have recently reiterated; that a potential surface based on the Stewart form of the potential function can be regarded as a smoothed or filtered version of the density surface from which it was constructed:

> The potential index is a distance-weighted sum and can be visualized as an operator or filter applied to the density surface; because weight decreases with distance, V_j represents local density values more than distant ones. The smoother the density surface, then, the higher we might expect the log-log correlation to be with potential (Warntz, 1965, p. 16). Thus one interpretation of the observed correlation is that it indicates a degree of smoothness in the density surface.
>
> (Goodchild et al, 1981, p. 342)

A second and closely related issue concerns the effect of the regional configuration. In general, the size of administrative areas or data collection units tends to vary inversely with the density of the data. For example where population densities are high, counties, states, provinces, census tracts, and so on, tend to be relatively small, and vice versa. In calculating United States population potentials, for example, whether at the state or county level, the density of control points generally decreases as one moves westward. This will have some effect on the results, though this has never been investigated. A similar problem arises when constructing potentials for urban areas since, for the most part, administrative areas tend to increase in size with distance from the center of the city. Dalvi and Martin (1976) investigated empirically the effects of both aggregation level and zone configuration on accessibility measures for the city of London, and found that both had a significant effect upon the value of the accessibility index.

A third issue relating to spatial aggregation levels concerns the role of the self-potential. Although the concept was briefly mentioned in Chapter 1 it has been ignored subsequently. Such an approach to the concept is the norm. It is usually treated as a computational problem, where the goal is simply to avoid dividing by zero. Yet, as Goodchild (1979, 88) has pointed out, the self-potential can be, depending on the zone size, the dominant term in the sum. In computing 1970 United States population potentials for example, using equation (1.1.2) for the potentials, equation (1.1.4) for the self-potentials, and data at the state level, the self-potential comprises, on the average, 17% of the totals by state, with California being the largest at 61% of the total for the state. This assumes a uniform population density within each state. For assumptions other than uniform these proportions can be much higher. As Court (1966, 41) wrote:

> The self-potential of New York City's population, assumed to be uniformly distributed over a circle equal to the entire legal area, is greater than the potential imposed on New York City by the remaining 171.5 million residents of the coterminous United States. A more concentrated distribution, such as the conical, would imply that New York City's residents are more than twice as important to it as all the rest of the country!

Insofar as the self-potential comprises a significant portion of the sum, it can be argued that it changes the very nature and intent of the potential function. In particular, it would appear that the self-potential can surreptitiously play the role of a destination term, making the potential function analogous to a spatial interaction model. It can be suggested that this may in fact have accounted for much of the early success in the comparison of population potentials with spatial interaction data at the state level of aggregation (see the list in section 2.2).

The problem of the self-potential is highly dependent on zone size, with the severity of the bias being directly related to the relative size of the zones. Warntz (1979, 11) has argued that provided sufficiently small areal units are employed in the calculation of potential, such as the

3,068 United States counties, the self-potential is "generally of very limited importance." The reasoning here is that by using smaller areal units the contribution from the self-potential is diminished in comparison with the contribution from the remaining 3,067 counties. Furthermore it can be noted that as the size of the areal units decreases, the calculation of potential more closely approximates the integral form of the function (as given in equation (1.1.3)). The relation between the spatial aggregation level and role of the self-potential is, however, an issue that has yet to be systematically investigated in detail.

The fourth issue that can be discussed under the heading of spatial aggregation involves the classification of variables as being intensive or extensive. An extensive variable can be defined as being space occupying, while an intensive variable is said to have no external spatial extent. Variables such as population size are extensive while variables such as potential or accessibility are intensive (see Broadbent, 1970).

In its standard form, a potential function is essentially an "incoming" quantity; the summation is over i at j. Repeating here for convenience, the usual form is

$$V_j = \sum_i P_i/r_{ij} \quad . \qquad [(1.1.2)]$$

Suppose however that we are also interested in the total amount of potential or per capita interaction "outgoing" from each i (for example, to be employed as an unconstrained origin term representing "emissiveness" in a destination constrained trip distribution model). The summation is therefore over j at i

$$V_i^o = \sum_j P_i/r_{ij} \quad , \qquad (2.4.1)$$

where V_i^o denotes outgoing potential. There is a problem which restricts the usefulness of equation (2.4.1) in its present form.

Consider first the usual incoming case. Imagine that there are two identical regions i and j, each containing 100 people, 10 miles apart. The contribution from i to j, using the Stewart form for potential, is 100÷10 or 10 persons per mile. Imagine now that region i is divided into equal halves i and i' each containing half of the population. The contribution from each i to j is now 50÷10 or 5 persons per mile, and the sum at j remains at 10 persons per mile. Thus in the case of incoming potential at j our results are not adversely affected by halving region i. The same does not hold true for the outgoing case however.

Imagine now that we have restored region i to its singular form and we are concerned with the value of the outgoing potential from i. The contribution from i to j remains as before, at 10 persons per mile. Consider the effect however if <u>region j</u> is now split into halves j and j'. The contribution from i to j is 10 persons per mile and the contribution from i to j' is 10 persons per mile. Thus the sum of the outgoing potential from i is now 20 persons per mile, a result which is inconsistent with what is expected. If we were to continue to partition region j, the amount of outgoing potential from i would continue to increase in spite of the fact that the size of the contribution in each of the j regions would remain the same, save for changes in distance.

In general, the sum of the outgoing potential using equation (2.4.1) as it stands, is a function of the number of control points being employed. In the case of the usual "incoming potential," equation (1.1.2) remains valid because population is an extensive variable and increasing the number of regions decreases the number of people in each. However, in the case of equation (2.4.1) results are inconsistent because potential is an intensive variable and changing the number of regions by partitioning or grouping has no effect on the value for potential in each.

There is a solution to this problem. It is necessary, in the case of equation (2.4.1), to multiply by the area of each j region such that

$$V_i^o = \sum_j a_j \; P_i / r_{ij} \quad . \tag{2.4.2}$$

This makes the output of this equation invariant with respect to the zoning system. If this is not obvious consider again the hypothetical example above, assuming now that the area of region j before partitioning is 10 square miles. Before region j is halved, the total potential outgoing from i, using equation (2.4.2), is 10x100÷10=100 persons per mile. Again region j is halved, and the total outgoing potential using equation (2.4.2) remains (5x100÷10) + (5x100÷10)=100 persons per mile. The result is therefore invariant with respect to the zoning system and although the absolute values are now larger (100 rather than 10) this is unimportant because potential is a relative quantity. If it is desired that the absolute values remain the same, this can be accomplished by dividing the entire right hand side of equation (2.4.2) by the total area

$$V_i^o = \frac{\sum_j a_j \; P_i / r_{ij}}{a} \quad . \tag{2.4.3}$$

In the example above, this gives the total potential outgoing from i as 10 persons per mile regardless of the partitioning of region j.

Equation (2.4.3) gives results which are comparible with those of equation (1.1.2). It should be noted however that the issue of the outgoing potential is important only when it is desired to compare or jointly employ ingoing and outgoing potentials. If the concern is with outgoing potential alone, the fact that it is a function of the number of control points will affect only the absolute values of V_i^o.

Having briefly reviewed the effects of spatial aggregation and partitioning on potentials we can now go on to consider some alternative definitions of the population and/or distance terms in potential functions. It should be noted that most of the arguments in the preceding sections apply to potential functions regardless of how the terms are defined.

2.5 Alternative Definitions of Terms

No doubt owing to its broad conceptual nature, the concept of population potential began to be generalized to include variables other than population shortly after its introduction. For the most part, the functions employed to generate these measures of potential were identical to population potential; any changes consisted primarily of redefining the population or distance terms.

Harris (1954) for example defined a market potential

$$M_j = \sum_i E_i / c_{ij} \quad , \qquad (2.5.1)$$

where E_i represents retail sales and c_{ij} represents transport cost over land (including terminal costs). This index was considered to represent the accessibility of any location j to the total market. It has also been described as economic potential (Vickerman, 1974, 679).

A generalization proposed by Warntz (1959) was a commodity supply space potential

$$Y_j = \sum_i X_i/r_{ij} \quad , \qquad (2.5.2)$$

where X_i represented commodity output over some time period (see also Tegsjö and Oberg, 1966). Spatial demand, in the same study, was represented by income potential

$$U_j = \sum_i Z_i/r_{ij} \quad , \qquad (2.5.3)$$

where Z_i represents total income at i (see also Stewart and Warntz, 1958; Warntz, 1965). Equation (2.5.3) can be considered to represent a weighted population potential since total income can be given as the product of population and per capita income.

A more recent example comes from Inhaber and Przednowek (1974) and Inhaber (1975) who calculated a scientific potential

$$J_j = \sum_i K_i/r_{ij} \quad , \qquad (2.5.4)$$

where K_i represented the number of primary authors at i who had published in major scientific journals. This scientific potential was defined as "a measure of the proximity of scientists (or scientific activity) to a given point" (Inhaber and Przednowek, 1974, 46).

These few examples serve to point out that it is possible to define many types of potentials simply by changing the definition of the population term. Of course, the distance term can also be redefined in a number of ways, for example, as cost distance, time distance, spherical distance, city block distance, and so on.

In addition to redefining the distance term with respect to the way distance is measured, it is also possible to employ alternative functional forms representing the effect of distance on per capita interaction. This is the subject of the following discussion.

2.6 Alternative Distance Response Functions

In the Stewart form of potential, the effect of distance on per capita interaction is an inverse function of simple linear distance. This is, however, just one among many possible ways in which the effect of distance can be mathematically defined; many other distance response functions (Amson, 1972b) have been suggested. Raising the distance variable to some power, such that

$$V_j = \sum_i P_i / r_{ij}^b \quad , \qquad [(2.1.10)]$$

has been commonplace since the advent of the potential formulation. One of Stewart's (1942) early papers, for example, reported the results of a least squares estimation of a power function

$$\log u_j = k + \log P_i + b \log r_{ij} \quad . \qquad (2.6.1)$$

The analysis was of the geographical origins of undergraduates u_j at Princeton, Yale, Harvard and M.I.T. and it was found that the mean estimate for b was "close to negative unity" (Stewart, 1942, 69).

An alternative approach to estimating the b value for the power function potential has been to determine a value for the power term using spatial interaction data and then to use that value to construct potentials from density data. Carroll (1955) for example, employed telephone calls and inter-city travel data to estimate a b which was then employed in the calculation of population potential. Although potentials based on simple distance are probably the most widely known, the practice of using interaction data to determine a power which is then used to calculate potential has been commonplace over the years (see, for example, Hansen, 1959; Ray, 1965; Lakshmanan and Hansen, 1965).

Interesting variations of the simple power function were suggested early on by Anderson (1955) and Carrothers (1956). Anderson proposed that the value for the power term be inversely related to population size, that is $b_i = f(1/P_i)$. The reasoning here is that the larger the size of the populations, the greater the transport facility between them. Carrothers, on the other hand, suggested that the power term ought to be inversely related to distance itself. In this case $b_{ij} = f(1/r_{ij})$, implying that "an extra unit of distance added to a long movement is of less importance than an extra unit added to a short movement" (Carrothers, 1956, 97). (This is also implied by b=1).

Another formulation which has been employed is the negative exponential where

$$V_j = \sum_i P_i \exp(-br_{ij}) \quad . \qquad (2.6.2)$$

Although it is not clear who first suggested its inclusion as a distance response function in the calculation of potential, the negative exponential is now widely recognized in the literature (Ingram, 1970; Vickerman, 1974; Weibull, 1976).

Ingram (1970) reasoned that the functions given by equations (2.1.10) and (2.6.2) both declined too rapidly near the origin. He suggested, as an alternative, a modified normal or Gaussian function of the form

$$V_j = \sum_i P_i \exp(-r_{ij}^2/\ell) \quad , \qquad (2.6.3)$$

where ℓ is a constant which is related to the spatial dispersion of a given set of points (see also Echenique et al, 1969). The use of this function is a response to the noticeable absence of a distance decay effect at short distances (Olsson, 1965, 52).

To the author's knowledge equations (2.1.10), (2.6.2), and (2.6.3), cover most of what has been done in redefining distance response functions in potential measures of the form being considered. The list of employable functions need not stop here however. Virtually any function which is monotonically decreasing with increasing values of r_{ij} is a candidate for inclusion in a potential equation.

It is possible, for example, to write a composite function which in effect combines the negative exponential and power functions. Such a formulation could employ the gamma distribution to define a potential

$$V_j = \sum_i P_i \exp(-b_1 r_{ij})\, r_{ij}^{-b_2} , \qquad (2.6.4)$$

with an exponentially damped power function (see Wilson and Kirkby, 1975; Openshaw and Connolly, 1977).

Another way to combine power and exponential functions is to use a modified Weibull distribution function where

$$V_j = \sum_i P_i \exp(-b_1 r_{ij}^{b_2}) \qquad (2.6.5)$$

(see Tribus, 1969, 155).

Other possibilities can be based on the Goux typology of distance transformations (Taylor, 1971). These might include potentials based on the normal function

$$V_j = \sum_i P_i \exp(-br_{ij}^2) \quad , \qquad (2.6.6)$$

square root exponential function

$$V_j = \sum_i P_i \exp(-br_{ij}^{0.5}) \quad , \qquad (2.6.7)$$

or the log-normal function

$$V_j = \sum_i P_i \exp(-b(\log r_{ij})^2) \quad . \qquad (2.6.8)$$

The latter five equations represent just a few examples of the many possible distance response functions which could be employed to define potentials. Additional functions could be borrowed from those which have been employed in modelling urban population densities (Zielinski, 1979) and spatial interaction (Openshaw and Connolly, 1977). One might also include stepped, non-global, stepwise discontinuous, and families of deterrence functions (Openshaw and Connolly, 1977).[1]

The use of many of the functions above is more complex than was the case for the power functions. A discussion of parameter space search techniques for such functions can be found in Batty (1976a, Chs. 5-9).

[1] It should be emphasized that this brief review does not pretend to be exhaustive of the literature on potential or accessibility. Related work has been published, for example, by Wachs and Kumagai (1973), Weibull (1976; 1980), Dalvi and Martin (1976), Sheppard (1979a), and Tobler (1979). However the emphasis in those works is not primarily on

All of the potential equations which have been reviewed in this section are distinguished from one another by the form of their distance response functions. It follows that they can all be represented by a single equation

$$V_j = \sum_i P_i \, f(r_{ij}) \quad , \qquad (2.6.9)$$

where $f(r_{ij})$ is any non-increasing function. Equation (2.6.9) can be said to represent a general form of potential. The Stewart form of potential can be considered to be a special case of equation (2.6.9).[2]

In conclusion it can be said that none of the potentials which have been described in this section have been independently derived or given any a priori basis. The distance response functions employed in equations (2.6.2) to (2.6.8) were simply borrowed from the field of spatial interaction modelling. This borrowing can be said to have been done, for the most part, without regard to the manner in which the functions can be derived, and without regard to the possible empirical consequences of employing such

---

[2] In light of the social physics derivation of the Stewart potential function given earlier in this chapter, and in light of the analogy with the definition of potential in the physical sciences, it would probably be appropriate to term the Stewart formulation "the potential" and to designate the other functional forms by some alternative label such as "indices of accessibility." In the interests of clarity and economy of discussion, however, we shall continue to employ the word "potential" for any function of the form of (2.6.9).

functions. Only the Stewart formulation has any a priori basis at all, and even in this it is necessary to rely on a questionable physical analogy.

It was pointed out in the introduction that the entropy maximizing and information minimizing methodologies provide a basis for making a priori estimates of the functional form of a probability distribution. It was also suggested that in order to calculate the potential it is necessary to determine the probability that a randomly selected individual at one location will have a per capita spatial interaction at another location. We will consider in greater detail, therefore, how the entropy and information formalisms can be employed to derive probability distribution estimates. This is the subject of the next chapter.

CHAPTER 3

ENTROPY MAXIMIZATION AND INFORMATION MINIMIZATION

3.1 The Entropies

It is commonly argued that there are in geography two general schools of thought on the nature of the entropy concept (see, for example, Wilson, 1970; Marchand, 1972; Sheppard, 1975; Coffey, 1979). The first, described by Sheppard (1975, 1) as the descriptive school, combines concepts from thermodynamics and systems theory and regards entropy as an intrinsic property of a spatial system. The approach is exemplified in Fein (1970), Warntz (1973b), Chapman (1977) and Coffey (1979) among many others. By and large the arguments within this school of thought are based on analogies with physical science. The second school of thought, pioneered by Wilson (1967), argues that "entropy-maximizing methods provide a useful and practical model-building tool" (Wilson, 1970, 125). In this view, entropy maximization is employed as a method of estimating the form of a probability distribution on the basis of limited prior knowledge. The dichotomy concerning the entropy concept in geography follows quite naturally from the historical evolution of the concept in the physical sciences. It can be argued however that the two views of entropy are in fact instances of the same basic concept.

Historically the concept of entropy dates back to the coining of the word by Clausius in 1850 when it was associated with transformations from work effects to heat effects in thermodynamics. In 1872, Boltzmann used the entropy function in the context of statistical mechanics and it became known as the Boltzmann H-function. In 1948 Shannon (Shannon and Weaver, 1948) developed information theory with entropy as a central component. The analogy between entropy in information theory and entropy in statistical mechanics was recognized at the time and, as Tribus (1969, 110) pointed out "there was a considerable debate as to whether this function was the same or merely an analog to the entropy of Clausius." It remained for Jaynes (1957) to show that "the function had deeper meaning than had been supposed . . . . and that the two entropies were, in fact, examples of the same idea and not merely analogies" (Tribus, 1969, 110).

The similarity between the apparently different views of entropy can be attributed to the fact that they are devices for measuring and describing distributions. Clausius' view of entropy was as a device for studying the distribution of energy. Boltzmann employed it to study the distribution of molecules in a gas. Shannon was concerned with the distribution of information. Jaynes contribution was to generalize the concept in order to show that it was applicable to the study of any distribution

The dichotomy concerning entropy in geography arises not from the concept itself but rather from the fact that the two schools of thought take different approaches to the concept. The statistical school of thought takes a quantitative approach to the concept in a manner which is analogous to the approach taken in statistical mechanics and information theory. The descriptive school, however, takes a qualitative approach to the entropy concept, and it is this which opens the door to confusion.

Those practitioners who take a purely descriptive approach to the entropy concept generally discuss the concept in the context of "open systems," "closed systems," "energy," "environment," and in particular "order." Thus it might be argued that an "open system" draws "energy" from its "environment" and thereby becomes more "ordered." All of this is then likened to a process whereby the "entropy of the system" (the degree of order) has increased. It is the proponents of this school of thought who have perpetuated the myth that there are different views or definitions of entropy, since they apparently do not see any connection between such broad ideas and the entropy of information theory. Coffey (1979, 187) for example argues that "as used in information theory, entropy has no relation to its use in the second law of thermodynamics" (see also Marchand, 1972).

Metaphysical discussions of the entropy of social systems, as exemplified in the preceding paragraph, retain their apparent distinction from the entropy of information theory only insofar as they remain at the level of analogy. As soon as any attempt is made to quantify these broad concepts and relationships, it becomes immediately apparent that the entropy concepts are identical.

Consider especially the word "order." It is indeed valid to equate the entropy concept with the degree of order in a system. Suppose that the "system" of interest is the United States and that we are concerned with studying the degree of order in this system. How could we measure this? One obvious approach would be to consider the distribution of population among the states, where a uniform distribution is disordered (high entropy) and a sharply peaked distribution (everyone resides in California) is highly ordered (low entropy). This usage of the words "entropy" and "order" is entirely consistent with the usage in classical thermodynamics, statistical mechanics, and information theory. In the latter case, entropy is considered to be a measure of "information" and the interpretation of this is exactly the same as the interpretation of order; a sharply peaked distribution (highly ordered) contains more information than a broad, flat (disordered) distribution in the sense that we are more certain about the nature of the peaked distribution. For example, the statement that "everyone in the United States lives in California" contains more information than the statement that "the distribution of population in the United States is uniform." In the former case the reader knows exactly where each individual lives, whereas in the latter case he knows only that equal numbers of people are expected to be found in regions of equal area, and there are literally billions of possible arrangements of individuals that will satisfy this condition. Stated another way it can be said, in the former case, that "John Doe lives in California" whereas in the latter case all that is known is that "John Doe lives in one of n equal area regions." It is in this sense that a sharply peaked distribution contains more information than a flat distribution. This example also illustrates why entropy can also be regarded as a measure of uncertainty. In the case of the flat, uniform distribution we are more uncertain about the location of an individual.

Regardless of whether we regard entropy as a measure of order, uncertainty, or information; regardless of whether we are looking at the distribution of temperature, molecules, information, or population; and regardless of whether we take Clausius, Boltzmann, Shannon, or Jaynes as our source, the reasoning behind the concept remains the same throughout; entropy is simply a measure of the orderliness of a distribution at a point in time. A process can be "entropic," but entropy per se is not a process. We strongly disagree with assertions such as Coffey's (1979, 337) that "the three formulations of entropy, the classical, the statistical, and the informational are quite distinct" as well as his (1979, 339) claim that the original entropy concept associated with classical thermodynamics has been "mutilated." Just the opposite is true. The three formulations are all instances of the same basic concept, and it is precisely for this reason that the entropy formulation is so interesting. Coffey's (1979) dissection of the entropy concept is at odds with his plea for unifying concepts.

It was pointed out earlier that it was Jaynes (1957) who was instrumental in generalizing the entropy concept and demonstrating that it could be employed in the study of any probability distribution. Insofar as many of the phenomena of interest to geographers can be treated as probability distributions, there would seem to be a natural connection between geography and entropy. Doubts can be expressed about the approach of the descriptive school of thought insofar as the metaphysical analogies can be so readily replaced by actual probability distributions. We agree with Wilson (1970, 122) who argued that

> Entropy . . . is not some Platonic property of an urban system which can be discussed without further definition. There is a measure of entropy associated with any probability distribution, and, conversely, whenever the concept of entropy is used it should be used only as the entropy of a probability distribution, which it should be possible to set out explicitly.

In the sections which follow we shall attempt to outline the concept of the entropy of a probability distribution, and the associated methods of entropy maximization, and information minimization, in greater detail.

3.2 A Simple Example

There have been a number of review articles published on the role of the entropy concept in geography (Gould, 1972; Cesario, 1975; Webber, 1977; Senior, 1979). In each case it can be suggested that the examples employed by the authors are either non-geographic or, if geographic, are too complex to be followed easily. In the present discussion we shall attempt to use the simplest possible geographical example. That so abstract a concept can be illuminated with so simple an example is perhaps testimony to its elegance.

We will see below that there are two mathematical definitions of entropy commonly employed in geography; one is from statistical mechanics and the other from information theory. If any rationale is required for yet another 'pedagogic discussion' of entropy maximization it can be noted that while Gould (1972), Cesario (1975), and Senior (1979) all based their discussions on the definition of entropy from statistical mechanics, it is the definition from information theory that is now almost exclusively employed in geography. Webber (1977) treated this latter definition pedagogically, but at a more advanced level.

Imagine that we have a simple linear town composed of three equal area zones. Four people work in zone 1 and it is our task to assign them to homes in zones 2 and/or 3 in an unbiased way. This example can be regarded as being equivalent to a location model, a spatial interaction model, or population density model.

Given only the information that the four workers are to be assigned to residences, how should we proceed? In the absence of any other information it might seem that an intuitively unbiased assignment would be to locate 2 workers in each of the two zones, for this would give probabilities, as in the case of an unbiased coin toss, of .5 for each region. We quote at length from Jaynes (1957, 622) on the inappropriateness of such an assignment:

> The problem of specification of probabilities in cases where little or no information is available, is as old as the theory of probability. Laplace's "Principle of Insufficient Reason" was an attempt to supply a criterion of choice, in which one said that two events are to be assigned equal probabilities

> if there is no reason to think otherwise. However, except in cases where there is an evident element of symmetry that clearly renders the events "equally possible," this assumption may appear just as arbitrary as any other that might be made . . . . Since the time of Laplace, this way of formulating problems has been largely abandoned owing to the lack of any constructive principle which would give us a reason for preferring one probability distribution over another in cases where both agree equally well with the available information.

It should be emphasized that this is not to say that the assignment of equal probabilities is no longer intuitively desirable, but rather that we require some "constructive principle" which will allow us to make such an assignment in a systematic way. Such a constructive principle can be found in the entropy concept. Quoting again from Jaynes (1957, 622):

> Just as in applied statistics the crux of a problem is often the devising of some method of sampling which avoids bias, our problem is that of finding a probability assignment which avoids bias, while agreeing with whatever information is given. The great advance provided by information theory lies in the discovery that there is a unique, unambiguous criterion for the "amount of uncertainty" represented by a discrete probability distribution, which agrees with our intuitive notion that a broad distribution represents more uncertainty than does a sharply peaked one, and satisfies all other conditions which make it reasonable.

The unique and unambiguous measure of uncertainty to which Jaynes refers is entropy. The reasoning behind the use of this measure can be illustrated within the context of our assignment problem.

The first column of Table 3.2.1 is a complete enumeration of all possible assignments of <u>numbers</u> of workers to zones 2 and 3. These are <u>macrostates</u>. It is obvious that such an enumeration leads us no closer to finding an unbiased method of assigning workers to homes.

Consider now however, as an alternative approach, the possibility of identifying the workers individually and considering all possible combinations of individual assignments. If the workers are identified as a, b, c, and d we have, as possible distributions, those listed in the second column of Table 3.2.1. These are <u>microstates</u> and it is the relation between the microstates and the macrostates that suggests a solution to the assignment problem. From the third column of Table 3.2.1 it is apparent that the assignment of two workers to each of the two regions can occur in the greatest number of ways. Wilson (1970, 3) points out that the macrostate which has the greatest number of microstates associated with it is, in a statistical sense, the most probable. Such probabilities are given in the fourth column of Table 3.2.1 and it can be seen that the uniform distribution of workers has the highest probability of occurence. Thus we have a "constructive principle" which allows us to assign <u>a priori</u>

TABLE 3.2.1

MACROSTATES AND MICROSTATES IN A SIMPLE LOCATION MODEL

Distribution of Workers (Macrostates)		All Possible Assignments of Individual Workers (Microstates)		Number of Ways of Macrostates Occuring (n)	Probability of Occurence of Macrostate n/Σn
4	0	abcd	0	1	.063
0	4	0	abcd	1	.063
		abc	d		
3	1	abd	c	4	.250
		adc	b		
		bcd	a		
		d	abc		
1	3	c	abd	4	.250
		b	acd		
		a	bcd		
		ab	cd		
		ac	bd		
2	2	ad	cb	6	.375
		bc	ad		
		bd	ac		
		cd	ab		

probabilities in an unbiased way.[1]

In order to further illuminate the reasoning behind this point of view it may be useful to consider a more intuitively familiar example. You are asked to place a bet on the total number of dots turning up in a single role of two fair dice (see Cesario, 1975). How would you place your bet?

All possible outcomes of the dice tossing problem are enumerated in Table 3.2.2 (which has a format identical to Table 3.2.1). It is apparent that the outcome which can occur in the greatest number of ways is a "7". Provided that the payoff is the same for every outcome we would therefore expect the rational gambler to bet on "7." As Cesario (1975, 41) argued "a rational gambler would bet on that outcome which has the greatest chance of occuring, that is, the most 'probable' outcome. Because each die is assumed to be fair, the most probable outcome is that sum which can occur in the greatest number of ways."

---

[1] Of course, the validity of this statement, and the entropy methodology as a whole, can be said to be contingent upon making the a priori assumption that all microstates are equally likely. This can be interpreted, in turn, as a condition which is identical to the 'Principle of Insufficient Reason' and thus it may be argued that the entropy maximizing argument is a circular one. However, if the microstates are not a priori equally likely, the method of minimum information allows such prior information to be explicitly incorporated into the analysis and this will be discussed in greater detail in section 3.5 below.

TABLE 3.2.2

MACROSTATES AND MICROSTATES IN THE ROLL OF TWO DICE

Sum of Dots on Two Dice (Macrostates)	No. of Dots on each of the Dice (Microstates)		No. of Ways of Macrostates Occuring (n)	Probability of Macrostate n/Σn
2	1	1	1	.028
3	1	2	2	.056
	2	1		
4	1	3	3	.084
	3	1		
	2	2		
5	1	4	4	.111
	4	1		
	2	3		
	3	2		
6	1	5	5	.139
	5	1		
	4	2		
	2	4		
	3	3		
7	1	6	6	.167
	6	1		
	5	2		
	2	5		
	4	3		
	3	4		
8	2	6	5	.139
	6	2		
	3	5		
	5	3		
	4	4		
9	3	6	4	.111
	6	3		
	4	5		
	5	4		
10	4	6	3	.084
	6	4		
	5	5		
11	5	6	2	.056
	6	5		
12	6	6	1	.028

It was stated in the quotation from Jaynes above that entropy can be considered to be a measure of the "amount of uncertainty" represented by a distribution. In the context of our assignment problem the meaning of this phrase can be better understood if we compare the uniform (2-2) distribution of workers to the maximally peaked (4-0) distribution. In the latter case we are absolutely certain about the microstate (abcd-0) whereas in the former case we are least certain about the microstates. It is in this sense that entropy can be described as a measure of uncertainty, where the desirable solution is the one which maximizes our uncertainty about the microstates and which therefore is the least biased from a subjective point of view.

The results which we have discovered here by exhaustive means can also be found using the entropy. An equation for the Shannon (Shannon and Weaver, 1948) or information theory definition of entropy was given in the first chapter as

$$H = -\sum_i p_i \ln p_i \quad . \qquad [(1.2.4)]$$

Let T_i represent the number of trips to each zone and T the total number of trips. Then

$$p_i = \frac{T_i}{T} \quad , \qquad (3.2.1)$$

is the probability that a randomly selected worker will be assigned to a particular zone. Note that although we are considering trips, this problem does not involve a trip <u>matrix</u> since we are restricting the assignment of workers to those from zone 1 to zones 2 and 3. This is analogous to considering a single row of a trip matrix. If we were concerned with trips between all zones, equation (1.2.4) would have to be summed additionally over j such that

$$H = -\sum_i \sum_j p_{ij} \ln p_{ij} \quad , \qquad (3.2.2)$$

where

$$p_{ij} = \frac{T_{ij}}{T} \quad . \qquad (3.2.3)$$

Equation (1.2.4) (or (3.2.2)) represents the definition of entropy from information theory. The definition from statistical mechanics is

$$W = \frac{T!}{\prod_i T_i} \qquad (3.2.4)$$

which will give the same results as equation (1.2.4) when T_i is large (see Webber, 1976, 278). Equations (1.2.4) and (3.2.4) can both be employed to find the most likely macrostate.

Equation (3.2.4) can be employed in order to calculate the number of microstates associated with each macrostate. For the macrostates in our assignment problem, for example, we have (where 0! = 1)

Macrostates		Number of Ways of Occuring
4 0	0 4	$\frac{4!}{0!\ 4!} = 1$
3 1	1 3	$\frac{4!}{1!\ 3!} = 4$
2	2	$\frac{4!}{2!\ 2!} = 6$

and it can be seen that these results agree exactly with those in Table 3.2.1. In the case of the information theory form of the entropy equation (1.2.4), we do not get the number of microstates directly, but instead we find the maximum value of the equation. In the assignment problem example we have

Macrostates		Value for Entropy H
4 0	0 4	$-\left[(\frac{4}{4}\ \ln\ \frac{4}{4}) + (\frac{0}{4}\ \ln\ \frac{0}{4})\right] = 0.000$
3 1	1 3	$-\left[(\frac{3}{4}\ \ln\ \frac{3}{4}) + (\frac{1}{4}\ \ln\ \frac{1}{4})\right] = 0.562$
2	2	$-\left[(\frac{2}{4}\ \ln\ \frac{2}{4}) + (\frac{2}{4}\ \ln\ \frac{2}{4})\right] = 0.693$

Again it is apparent that the solution, the maximum value for H, occurs in the case of the uniform distribution. Equations (3.2.4) and (1.2.4) are devices for finding the statistically most likely form of a distribution, that is, they identify the macrostate which has the greatest number of microstates associated with it and which can therefore occur in the greatest number of ways.

It was pointed out that equations (1.2.4) and (3.2.4) give the same results when T_i is large and this leads one to question what the difference is between the two. It has already been noted that one definition has its source in information theory (1.2.4) and the other in statistical mechanics (3.2.4) so on this basis the distinction is an historical one. In mathematical terms the two are different in that (3.2.4) relies on Stirling's approximation in the derivation of results while (1.2.4) does not. In other words, if it is desired to find the maximum of equation (3.2.4), Stirling's approximation

$$\ln T_i! = T_i \ln T_i - T_i \quad , \qquad (3.2.5)$$

is normally employed to estimate the factorial terms (this is why the stipulation concerning large T_i is made, since the larger the T_i the more accurate the approximation). Wilson (1970, 8) has noted that this is not a very important distinction however since "we only use the [derivative] of Stirling's approximation and the derivatives coincide in the two cases." In more general terms the distinction between the two forms of entropy has been described by Wilson (1970, 8) as being "analogous to the objective and subjective views of probability." According to Jaynes (1957, 622)

> The 'objective' school of thought regards the probability of an event as an objec-

> tive property of that event, always capable in principle of empirical measurement by observation of frequency ratios in a random experiment. In calculating a probability distribution the objectivist believes that he is making predictions which are in principle verifiable in every detail, just as are those of classical mechanics. The test of a good objective probability distribution p(x) is: does it correctly represent the observable fluctuations of x?

The subjective school of thought, on the other hand

> regards probabilities as expressions of human ignorance; the probability of an event is merely a formal expression of our expectation that the event will or did occur, based on whatever information is available. To the subjectivist, the purpose of probability theory is to help us in forming plausible conclusions in those cases where there is not enough information available to lead to certain consequences; thus detailed verification is not expected. The test of a good subjective probability distribution is does it correctly represent our state of knowledge as to the value of x?

Wilson (1970, 9) has argued that the entropy function from statistical mechanics (3.2.4) is "essentially objective" while the entropy from information theory (1.2.4) is "essentially subjective."

Regardless of the historical, mathematical, and philosophical differences between the two definitions of entropy they do provide the same results (when T_1 is large) and can be mathematically derived from one another (see Wilson, 1970, 8). More importantly, however, there seems to be some concensus that the entropy function from information theory is broader (Sheppard, 1975, 4), more flexible (Wilson, 1970, 8), more fruitful (Webber, 1977, 254), and more convenient and rigorous (Senior, 1979, 206).

From this point forward we shall confine our attention to the information theory definition of entropy, equation (1.2.4). This will be done not only in the interests of economy of discussion, and in light of the preceding arguments, but also because the majority of reviews of the entropy concept have focused their attention on the definition of entropy from statistical mechanics.

In the present section the macrostate with the largest number of possible microstates was identified, first, by exhaustively looking at every possible arrangement and, secondly, by calculating the value of both entropy statistics on every possible macrostate in order to find the maximum. The next two sections describe the procedures which are employed to find the maximum, and hence the most probable distribution, without having to calculate the value of the entropy statistic on every macrostate.

All of the results to be presented can be extended to the entropy as given by equation (3.2.4). A mathematical proof that (1.2.4) is a unique and unambiguous measure of uncertainty is given in Shannon and Weaver (1948, 116), Jaynes (1957, 630), and Wilson (1970, 131).

3.3 Constrained Entropy Maximization

Repeating here for convenience, we have a mathematical function

$$H = -\sum_i p_i \ln p_i \quad , \qquad [(1.2.4)]$$

and it was shown in the example in the preceding section that the statistically most probable distribution is associated with the maximum of this function. In the example of the assignment problem we found the maximum by calculating every possible value for H. Since this approach will be impractical for large numbers of macrostates we require a method to find the maximum of the function by analytic means.

The usual approach to such a maximization problem is well known and involves taking the first derivative of the function, setting it equal to zero, and solving for p_i. This usual method of maximization cannot be directly employed for the entropy function however since it is applicable only to unconstrained functions.

The constraint to which we refer is

$$\sum_i p_i = 1 \quad . \qquad [(1.2.2)]$$

This is a normalizing constraint which ensures that the probabilities sum to one. While working through the example we could ensure that this constraint was satisfied simply by making sure that all workers were allocated to regions 2 and/or 3. For more complex problems, this constraint must be explicitly incorporated into the maximization procedure.

In order to maximize the constrained entropy function it is necessary to employ the method of Lagrange multipliers. The method, stated simply, allows a constrained maximization problem to be transformed into an equivalent unconstrained problem. This means that the maximum of an unconstrained Lagrangian function can be found by the conventional methods of differential calculus (see McAdams, 1970, 145; Wilson and Kirkby, 1975, 280; and Senior, 1979, 202-203: Senior warns that the discussion of the Lagrangian method in Gould (1972, 696ff.) is misleading).

In order to transform the constrained entropy function into an unconstrained Lagrangian function it is necessary to multiply constraint (1.2.2) by a Lagrangian multiplier λ (we shall actually use $\lambda+1$ for convenience of notation). Combining the entropy function and the constraint, multiplied by $(\lambda+1)$, we form a Lagrangian

$$L = -\sum_i p_i \ln p_i + [(\lambda-1)(1-\sum_i p_i)] \quad . \qquad (3.3.1)$$

We can now find the maximum of this function by conventional means.

Taking the first partial derivative of (3.3.1) with respect to p_i and setting the result equal to zero we have

$$\frac{\partial L}{\partial p_i} = -\ln p_i - 1 - \lambda + 1 = 0 \quad . \tag{3.3.2}$$

Solving for p_i we get

$$\ln p_i + \lambda = 0 \quad . \tag{3.3.3}$$

and therefore

$$p_i = \exp(-\lambda) \quad . \tag{3.3.4}$$

This is the maximum entropy distribution or, in other words, the macrostate probability distribution which is the most probable given the information we have. Stated another way, we have identified a method for finding the maximum value of H without calculating every possible value of H. This becomes clearer if it is noted that, by substituting (3.3.4) into (3.3.1) we get

$$\lambda = \ln n \quad , \tag{3.3.5}$$

and hence

$$p_i = \frac{1}{n} \quad . \tag{3.3.6}$$

(see Tribus, 1969, 128).

In the context of the example we have been using the above result indicates that the probability of location in each zone is one over the number of zones, or

$$p_i = \frac{1}{n} = \frac{1}{2} = .5$$

which agrees with the earlier results. Note also that, in the case of a uniform distribution, the entropy can be directly calculated as

$$H = \ln n \qquad [(1.2.6)]$$

(see Tribus, 1969, 128). In our example this gives

$$H = \ln 2 = .693$$

which agrees exactly with the result obtained earlier using equation (1.2.4) directly. This result, as Tribus (1969, 128) noted, "demonstrates that the principle of insufficient reason is a special case of the application of the principle of maximum entropy." In other words, it has been shown again that in the absence of any other information the maximum entropy approach provides a method for assigning the intuitively preferable, uniform probability distribution.

It is perhaps surprising that we have come so far in this geographic discussion of the entropy concept without having mentioned the role of distance. This is the subject of the next section.

3.4 Distance Constrained Entropy Maximization

In order to introduce the idea of entropy maximization under a distance constraint we return to the problem of assigning four workers to two regions. In previous sections it was emphasized that the assignment was taking place in the absence of any other information. Now we wish to introduce some additional information in the form of an average distance, and evaluate the consequences for the entropy solution.

We shall specify that the average distance the four workers travel in commuting one-way from jobs to homes is 1.25 miles. We know also that the distance from region 1 to region 2 (and region 2 to region 3) is one mile and the distance from region 1 to region 3 is two miles. If we consider the average distance associated with each macrostate we get results as follows

Macrostate		Average Distance
4	0	$\frac{(4 \cdot 1) + (0 \cdot 2)}{4} = 1$ mile
0	4	$\frac{(0 \cdot 1) + (4 \cdot 2)}{4} = 2$ miles
3	1	$\frac{(3 \cdot 1) + (1 \cdot 2)}{4} = 1.25$ miles
1	3	$\frac{(1 \cdot 1) + (3 \cdot 2)}{4} = 1.75$ miles
2	2	$\frac{(2 \cdot 1) + (2 \cdot 2)}{4} = 1.5$ miles

The solution to the problem is obvious. The only distribution or macrostate which satisfies the distance constraint is three workers to region 2 and one worker to region 3. In most practical modelling situations, of course, the solution is not this simple; there may be many macrostates which will satisfy a single distance constraint. In other words it can be said, following Wilson (1970, 3), that a mean distance constraint is a "higher-level macrostate description" and because of this it contains "less information than the trip distribution macrostate description." As a result, many different macrostate distributions can exhibit the same mean distance.[1]

Our problem is to find the most probable macrostate which also satisfies some distance constraint. This can be accomplished using the method outlined in the preceding section. The distance constraint can be written in general as

$$\sum_i p_i r_i = \bar{r} \tag{3.4.1}$$

which says that the sum of the probabilities times the distances should equal some average distance $\bar{r}$. We wish to maximize the Shannon entropy function (1.2.4), this time subject to the normalization constraint (1.2.2) and the mean distance constraint (3.4.1).

Form a Lagrangian by adding together the entropy function and the constraints

[1]It can also be noted, on a more technical level, that since trips are discrete it may never be possible to exactly satisfy a mean distance constraint unless fractions of trips are allowed.

$$L = -H + [(\lambda-1)(1-\sum_i p_i)] + b[\bar{r}-\sum_i p_i r_i] \quad , \quad (3.4.2)$$

where b is a Lagrangian multiplier which will ensure that constraint (3.4.1) is satisfied. Taking the first partial derivative of (3.4.2) with respect to p_i and setting the result equal to zero we have

$$\frac{\partial L}{\partial p_i} = -\ln p_i - 1 - \lambda + 1 - br_i = 0 \quad . \quad (3.4.3)$$

Solving for p_i we have

$$-\ln p_i - \lambda - br_i = 0 \quad , \quad (3.4.4)$$

and then

$$p_i = \exp(-\lambda)\exp(-br_i) \quad . \quad (3.4.5)$$

This is the maximum entropy distribution consistent with the information in the normalization constraint (1.2.2) and the mean distance constraint (3.4.1); it identifies the macrostate probability distribution which has the largest number of microstates associated with it and which also satisfies the constraints. In the context of the example we have been using, the p_i term on the left hand side of the equation can be interpreted as the probability that a worker will live in region i.

Equation (3.4.5) can be put in a more familiar form as follows. Repeating here for convenience it is desired that

$$\sum_i p_i = 1 \qquad [(1.2.2)]$$

In order to solve for $\exp(-\lambda)$ we first substitute (3.4.5) into (1.2.2)

$$\sum_i \exp(-\lambda)\exp(-br_i) = 1 \qquad (3.4.6)$$

Rearranging we have

$$\exp(-\lambda) = \frac{1}{\sum_i \exp(-br_i)} \qquad (3.4.7)$$

where, to simplify the notation we define

$$A = \exp(-\lambda) \qquad (3.4.8)$$

Repeating here for convenience it can be noted again that p_i was defined as

$$p_i = \frac{T_i}{T} \qquad [(3.2.1)]$$

If p_i in equation (3.4.5) is replaced by equation (3.2.1) and $\exp(-\lambda)$ in equation (3.4.5) is replaced by A (equation (3.4.7) we have, as a final result

$$T_i = A\, T \exp(-br_i) \quad , \qquad (3.4.9)$$

where

$$A = \frac{1}{\sum_i \exp(-br_i)} \qquad (3.4.10)$$

The practical use of equations (3.4.9) and (3.4.10) can be illustrated with the data from the assignment problem.

Before this can be done, however, some estimate for the value of the parameter b is required.

In the usual approach to the calibration of models such as this, iterations are carried out, and the parameter b is repeatedly readjusted, until the model value for b matches the observed value (see Baxter, 1973, for example). During the first iteration of the calibration procedure some preliminary or starting value for b is therefore required as a point of departure. Such a value for b can be arrived at if it is noted that the mean value of a continuous function $y = \exp(-br_{ij})$ can be derived by integration as follows

$$\bar{r} = \frac{\int_0^\infty r_{ij} \exp(-br_{ij})\, \underline{d}r_{ij}}{\int_0^\infty \exp(-br_{ij})\, \underline{d}r_{ij}} = \frac{\frac{1}{b^2}}{\frac{1}{b}} = \frac{1}{b} \qquad (3.4.11)$$

(a complete evaluation of the integrals is presented in Masser and Brown, 1978, 67-68). It follows from this that a reasonable first approximation to b in the discrete (model) case is

$$b = \frac{1}{\bar{r}_{obs}} \qquad (3.4.12)$$

where $\bar{r}_{obs}$ is the observed, as opposed to model, mean distance.

In the "calibration" of our simple assignment problem it will suffice to employ the inverse of the mean trip length as the value for b in order to demonstrate how the constraints operate. In the example this gives b = 1/1.25 miles = 0.8. Using this value for b we have for trips from region 1 to region 2

$$T_{12} = \frac{4 \cdot \exp(-0.8 \cdot 1)}{\exp(-0.8 \cdot 1) + \exp(-0.8 \cdot 2)} = 2.76 ,$$

and from region 1 to region 3

$$T_{13} = \frac{4 \cdot \exp(-0.8 \cdot 2)}{\exp(-0.8 \cdot 1) + \exp(-0.8 \cdot 2)} = 1.24 .$$

These results, when rounded-off to integers, agree exactly with the desired results for the distance constrained assignment problem (a refinement of the calibration procedure which guarantees integer solutions has been proposed by Charnes et al, 1976). Note also that 2.76 + 1.24 = 4.00, which satisfies the normalization constraint. The average distance associated with the integer solution is 1.25 miles and this is also in agreement with the desired results.

These calculations are, of course, intended only to be illustrative of the use of the entropy maximizing interaction model. In actual practice (for more complex problems) the A's and b would be repeatedly readjusted in an iterative algorithm until the constraints were satisfied (see Baxter, 1973; Wilson, 1974; Batty, 1976a; Senior, 1979).

We have seen once again in this section that the entropy maximizing formulation consists of identifying the maximum of the entropy function subject to whatever additional information is known. The macrostate or probability distribution associated with this maximum can be defined as the most probable since it has the greatest number of microstates associated with it and can occur in the greatest number of ways. Tribus (1969, 123) presents a proof which indicates that the Shannon entropy is neither a local saddle point nor a minimum but rather an unconditional global maximum. Furthermore, Wilson (1970, 20-22) presents evidence to the effect that the maximum of the Shannon entropy function is a very sharp maximum (see also Jaynes, 1968, 231).

It was indicated earlier that the validity of the entropy maximizing method can be called into question on the basis that *a priori* equally likely microstates must be assumed. We shall see in the next section that there is an extension of the method which overcomes the problem.

3.5 The Method of Minimum Information

In the example which has been employed throughout this chapter the assumption has been made that characteristics of the destination zones have no effect on the assignment of workers to those zones (i.e., all zones are of equal size, each contains the same number of homes, etc.). In the language of information theory this is the assumption of *a priori* equally likely microstates, that is, in the absence of such prior information, the probability that a worker will be assigned to any particular zone is equal over all zones. This was demonstrated in section 3.3 where the maximization was carried out subject only to the normalizing constraint and the result was a uniform distribution of workers. Snickars and Weibull (1977, 144) have commented on the assumption of equally likely prior microstates:

> Now, if there really is no a priori information available, then, in the spirit of Laplace, this assumption as to prior probabilities is warranted. However, if there is some a priori information about the specific conditions at hand in an application this assumption may be unreasonable. This means that the assumption in the entropy derivation is biased in relation to the a priori information in question.

Such a bias can be demonstrated within the context of the example. Suppose that some prior information is introduced in the form of the areas of the destination zones. One would naturally expect such information to have an effect on the assignment of workers to zones in the sense that it would be expected that the assignment of workers to zones would be in direct proportion to the size of the zones. Information concerning such unequal prior probabilities is not, however, taken into account within the entropy maximizing framework and the result will be biased with respect to this a priori information.

It was indicated in the introduction that the statistically most likely form for a probability distribution p_i which takes information concerning some prior probability distribution q_i into account, is the one which minimizes the Kullback (1959) information

$$I(q:p) = \sum_i p_i \ln \frac{p_i}{q_i} \qquad [(1.3.2)]$$

The prior q_i is defined with respect to some measure on the cells (see equation 1.3.1) and, in the context of the example we have been discussing, it can be defined with respect to the areas a_i of the destination zones such that

$$q_i = \frac{a_i}{\sum_i a_i} \, , \qquad (3.5.1)$$

where, by definition

$$\sum_i q_i = 1 \, . \qquad (3.5.2)$$

It is desired to find the minimum of the Kullback information (1.3.2) subject to the information that

$$\sum_i p_i = 1 \, , \qquad [(1.2.2)]$$

and

$$q_i (i = 1,\ldots,n) \text{ known} . \quad (3.5.3)$$

Form a Lagrangian in order to maximize $-I(q:p)$

$$L = -\sum_i p_i \ln \frac{p_i}{q_i} + [(\lambda-1)(1-\sum_i p_i)] , \quad (3.5.4)$$

and then

$$\frac{\partial L}{\partial p_i} = -\ln p_i - 1 + \ln q_i - \lambda + 1 = 0 . \quad (3.5.5)$$

Solving for p_i, we obtain

$$p_i = q_i \exp(-\lambda) , \quad (3.5.6)$$

and, since $\exp(-\lambda)$ is a constant, we can substitute in equation (3.5.1) for q_i to get

$$p_i = \frac{a_i}{\sum_i a_i} = q_i . \quad (3.5.7)$$

Thus, when the only information available is that concerning the prior, the method of minimum information results in an assignment which is in direct proportion to the prior, in this particular case, the areas of the zones.

If the information concerning the mean distance to be travelled is now reintroduced, we can also find the minimum of the Kullback information subject to the usual normalization constraint (1.2.2), the known prior (3.5.3) and

$$\sum_i p_i r_i = \bar{r} . \quad [(3.4.1)]$$

In this case we have

$$L = -I(q:p) + [(\lambda-1)(1-\sum_i p_i)] + b[\bar{r}-\sum_i p_i r_i] , \qquad (3.5.8)$$

and, minimizing with respect to p_i,

$$\frac{\partial L}{\partial p_i} = -\ln p_i + \ln q_i - \lambda - br_i = 0 , \qquad (3.5.9)$$

which leads to

$$p_i = q_i \exp(-\lambda)\exp(-br_i) \quad . \qquad (3.5.10)$$

This is identical to the maximum entropy result (3.4.5) reported in the last section except that the prior probability distribution has now appeared in the model. Substituting in the normalization constraint (1.2.2) to solve for $\exp(-\lambda)$ gives a more familiar result

$$p_i = \frac{q_i \exp(-br_i)}{\sum_i q_i \exp(-br_i)} , \qquad (3.5.11)$$

which will assign workers to zones in direct proportion to the size of zones and in inverse proportion to an exponential function of distance.

Suppose that in the example assignment problem, zone 2 contains 0.3, and zone 3 contains 0.7, of the total area of the two destination zones (remembering that zone one is only an origin). The resulting assignment of workers to zones, using equation (3.5.11) is

$$T_{12} = \frac{0.3 \cdot 4 \cdot \exp(-0.8 \cdot 1)}{[0.3 \cdot \exp(-0.8 \cdot 1)] + [0.7 \cdot \exp(-0.8 \cdot 2)]} = 1.95,$$

and

$$T_{13} = \frac{0.7 \cdot 4 \cdot \exp(-0.8 \cdot 2)}{[0.3 \cdot \exp(-0.8 \cdot 1)] + [0.7 \cdot \exp(-0.8 \cdot 2)]} = 2.05.$$

Once again, two workers (in integer form) have been assigned to each of the two destination zones. In comparison with the maximum entropy results in the previous section, it can be seen that the attenuating effect of distance has now been partially offset by the variation in the size of the zones. Note also that the mean distance constraint has not been satisfied; if this were desired, then it would be necessary to adjust b iteratively until a correspondence between the observed mean and the model mean was attained (although in this particular example there are so few trip-makers that this would be redundant unless non-integer trips were allowed).

The method of minimum information thus provides a means of making an estimate of the statistically most likely form for a probability distribution p_i when prior probabilities are known and are not equally likely. Although we have used the example of a prior defined with respect to the areas of the zones here, it will be seen in the next chapter that there are a large number of different types of data which are permissible choices for the prior and these can include properties of the origins and/or destinations, as well as the flows between them. In addition, these can be defined from either empirical observation or descriptive models. It quickly becomes apparent that ~~the method of~~ minimum information is an important generalization of minimum entropy, not only because of the resolution of the problem of unequally likely priors, but also because of the flexibility and increased range for potential application it introduces.

Earlier in this chapter, entropy was described as a measure of the information in a probability distribution (i.e., it was said that a peaked distribution contains information). It should be noted that this definition and usage of the word "information" is different from, and inconsistent with, the definition of information in the Kullback formulation. The difference between the two definitions has formed the subject of considerable debate and controversy in the field of information theory and is much too complex to be thoroughly discussed and reviewed here. The interested reader is referred to Hobson and Cheng (1973), and especially Webber (1979), for more detailed treatments. The difference between the definitions does not have any direct effect on the results to be obtained from the methods of maximum entropy and minimum information.

Before we go on to consider the derivation of potential functions within the entropy maximizing and information minimizing frameworks, we will briefly review some of the alternative ways in which the Shannon and Kullback functions have been employed in human geography.

3.6 Extensions of the Methods

In this discussion of entropy maximization and information minimization the derivation of uniform and negative exponential probability distributions has been illustrated. The use is not limited to these two cases however, and Tribus (1969) has shown that a number of alternative functional forms can be derived by changing the constraints. In this regard, Tribus (1969) has derived truncated and normal Gaussian, incomplete gamma, beta, Cauchy, Weibull, Bose-Einstein, and Poisson distributions using the maximum entropy formulation. Consideration of the use of additional constraints in a geographical context will be considered in greater detail in the next chapter.

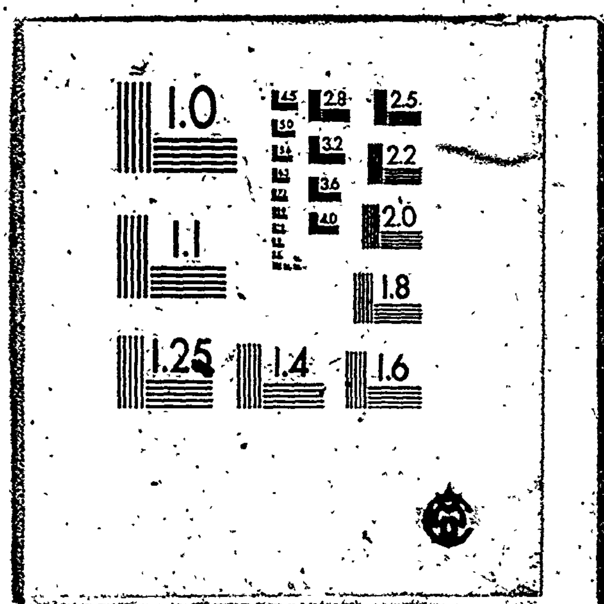
We have seen in this chapter how the methods of constrained entropy maximization and information minimization can be employed to derive probability distributions in an unbiased manner. Such probability distributions can form the basis not only for spatial interaction models, but also for location and density models (see Webber, 1979). In fact the entropy/information formalisms can be extended to any geographical problem which can be formulated in terms of a probability distribution (see, for example, Leopold and Langbein, 1962). The functions can be employed either to derive probability distributions, as has been emphasized here, or to measure existing distributions (see, for example, Curry, 1964, 1972; Medvedkov, 1967, 1970; Berry and Schwind, 1969; Gurevich, 1969; Chapman, 1970, 1973; Semple and Golledge, 1970; Semple and Griffin, 1971; Semple and Gauthier, 1972; Garrison and Paulson, 1973; Semple, 1973; Batty, 1974b; Haynes and Enders, 1975; Perin, 1975; Semple and Demko, 1977; and Getis and Boots, 1978).

There are other aspects of the entropy/information concepts which also could have been considered here. For example, arguments have been made that the Shannon entropy function can be considered to be a special case of the more general Kullback function (this will be demonstrated below), and that the latter is to be preferred because a number of mathematical inconsistencies which arise when the Shannon measure is extended to continuous space do not arise for the equivalent Kullback measure (see Hobson and Cheng, 1973; Batty, 1974b). Batty (1972; 1974a; 1976b) has extended this development into the geographical context by defining a spatial entropy

$$H = -\sum_i p_i \ln \frac{p_i}{a_i} , \qquad (3.6.1)$$

where a_i represents the area of zones. He (1974, 5) argued that

2



> equation [(3.6.1)] is more useful in spatial analysis than equation [(1.2.4)] because the effects of partitioning a spatial system in different ways can be compared in absolute terms using equation [(3.6.1)]. Furthermore, the spatial entropy statistic can be used in comparisons between different regions.

(see also Curry, 1972). Angel and Hyman (1976) have employed the continuous form of the Shannon measure to develop spatial interaction models written in continuous terms, and Bussière and Snickars (1970) have employed it to produce urban population density models. Charnes et al (1972) have employed the discrete Kullback measure to derive a spatial interaction model in information theoretic terms (see also Phillips et al, 1976, and Charnes et al, 1976). More advanced discussions of the Shannon and Kullback functions can be found in March and Batty (1975). A discussion of alternative measures of information is in Walsh and Webber (1977).

Another topic which was not discussed in this review is in the relation between entropy maximizing, information minimizing, and maximum likelihood procedures (see Batty and Mackie, 1972; Wilson, 1974; and Webber, 1979). Wilson and Kirkby (1975, 291) for example, noted that "if maximum likelihood methods are used for parameter estimation . . . the equations to be solved for these parameters turn out to be just the entropy maximizing constraint equations." Similarly, another overlooked issue was in the relations between entropy maximization, information minimization, and linear and geometric programming (see Wilson and Senior, 1974; Dinkel et al, 1977; Webber, 1979, 141, for example).

Finally, it can be noted that this discussion has also ignored the question of whether the entropy/information formulations provide a "theoretical foundation" for the models which result from them. Such a claim was made in Wilson's (1967) initial introduction of the entropy maximizing concept but more recently Sheppard (1976; 1979b) has provided arguments to the contrary, suggesting that the technique is primarily a method of hypothesis testing. In order to resolve this question it would appear to be necessary to decide what it is that constitutes "theory." Wilson, for example, described his approach as a "statistical theory" while Sheppard (1979b) regards the work of Smith (1975), involving individual choice behavior, as exemplary of theory. We shall leave the question open to debate for now, referring the reader to Beckmann and Golob (1972), Hansen (1972), Fisch (1977), and Cesario (1979) for alternative and/or critical points of view.

The purpose of this chapter has not been to present an exhaustive review but rather, as was suggested at the outset, to provide an introduction to the concepts of entropy maximization and information minimization as a foundation for the next chapter.

CHAPTER 4

MAXIMUM ENTROPY AND MINIMUM INFORMATION POTENTIALS

To briefly reiterate, it can be pointed out again that while the second chapter was concerned with potential, and the third with entropy maximization and information minimization, the present discussion is addressed to the relations between them and, in particular, to the derivation of potential functions. Problems of calculation and interpretation which result from employing entropy and information based potential functions will be discussed in Chapter 5.

4.1 Potentials and Trip Distributions

In the second chapter it was shown how potential functions can be derived from gravity, or trip distribution models. Given a model of the form

$$T_{ij} = k \frac{P_i P_j}{r_{ij}^b} \quad , \qquad [(2.1.7)]$$

it was indicated that potential can be obtained by dividing each side by P_j, such that

$$\frac{T_{ij}}{P_j} = k \frac{P_i}{r_{ij}^b} \quad . \qquad (4.1.1)$$

Since potential is a sum at a point, a summation sign can be added to each side (while dropping the k for convenience) to get

$$\sum_i \frac{T_{ij}}{P_j} = \sum_i \frac{P}{r_{ij}^{b}} \qquad [(2.1.10)]$$

This demonstrates, once again, that population potential is summed per capita spatial interaction at the destination, that is,

$$V_j = \frac{\sum_i T_{ij}}{P_j} , \qquad (4.1.2)$$

which can also be written

$$V_j = \sum_i \left[\frac{T_{ij}}{P_j}\right] , \qquad (4.1.3)$$

since P_j is constant on each zone.

The equations above describe the relationship between population potential and spatial interaction and it can be seen by equation (4.1.3) that the most significant determinant of the form of the potential function will be the trip distribution model. Wilson (1971) has identified a family of trip distribution models, any of which could be inserted in equation (4.1.3) to form a potential function. However, as was indicated in the second chapter, the principal concern in this study is with potential functions of the traditional Stewart-Warntz type and, in particular, with the form of their distance response functions. In other words the focus is on potential functions of the form

$$V_j = \sum_i P_i \, f(r_{ij}) \qquad [(2.6.9)]$$

and the goal is to derive a trip distribution model which, when inserted in equation (4.1.3), will produce the desired result. The task of examining the types of potentials which would obtain from inserting a variety of trip distribution models, such as the members of the Wilson family, into equation (4.1.3) is left for future research.

By first deriving a suitable trip distribution model, and then inserting the result in equation (4.1.3), we are taking an <u>indirect</u> approach to the maximum entropy and minimum information derivation of potential functions. It should be noted that the author's own attempts to derive potential functions <u>directly</u> from maximum entropy have proven unsuccessful. This can be attributed, at least in part, to the fact that when one attempts to define an estimate of probability for potentials, the quantities employed are not analogous to those normally used in entropy maximization. For example, if one defines a probability of <u>per capita</u> interaction of the form

$$p_i = \frac{x_i}{X} , \qquad [(1.2.1)]$$

both the numerator and denominator of the right-hand-side are unknowns. In the usual situation, as is the case with trip distribution modelling, the denominator of an equation such as (1.2.1) is known a priori and a normalizing constraint is required to ensure that the sum of the numerator exactly equals the given denominator. In the case of potentials, however, the probabilities will sum to one without a normalizing constraint since the denominator is an a posteriori result and is simply the sum of the numerator. Whether such problems can be handled successfully within the entropy maximizing framework remains to be seen, and for the moment this issue will also be left as a matter for future consideration.

4.2 Potentials Based on Maximum Entropy Trip Distributions

It was stated above that it is desired to derive a trip distribution model which can be inserted in equation (4.1.3) to produce a result which is of the form of the potential given by equation (2.6.9). It can be seen easily that the only trip distribution model which will produce the desired result is

$$T_{ij} = k\ P_i\ P_j\ f(r_{ij}) \quad . \qquad (4.2.1)$$

This model, although included by Wilson (1971,2) in the maximum entropy family, is not derivable from maximum entropy since it is totally unconstrained. As was indicated in the previous chapter, in order to be able to employ the method of maximum entropy to derive a trip distribution model it is necessary to introduce constraints. A constrained form of the potential given by equation (2.6.9) is

$$V_j = \sum_i \left[\frac{P_i \ f(r_{ij})}{\sum_j f(r_{ij})} \right] , \qquad (4.2.2)$$

or alternatively

$$V_j = \sum_i A_i \ P_i \ f(r_{ij}) , \qquad (4.2.3)$$

where

$$A_i = \frac{1}{\sum_j f(r_{ij})} . \qquad (4.2.4)$$

These functions will be discussed in greater detail below, and it will be shown that the only trip distribution model which can be inserted in equation (4.1.3) to get a potential of the form of equation (4.2.2) is

$$T_{ij} = A_i \ P_i \ P_j \ f(r_{ij}) , \qquad (4.2.5)$$

where

$$A_i = \frac{1}{\sum_j f(r_{ij})} . \qquad (4.2.6)$$

Trip distribution model (4.2.5) is not wholly derivable from maximum entropy since it contains one unconstrained term P_j, but it is, nevertheless, at least partially based on the method. The point is that if it is desired to employ the method of maximum entropy as a basis for deriving potential functions then it is necessary that those functions, and the trip distribution models on which they are founded, be constrained. As will be discussed in greater detail in section 4.3 below, normalized potential functions, and the balancing factors associated with them (such as equation (4.2.4)) are also important with respect to the empirical results to be obtained.

Consider a country (or some other area of interest) partitioned into n zones. Each zone has a total of P_i interactors where T_{ij} is the number who will have an interaction with the jth zone. Define

$$p_{ij} = \frac{T_{ij}}{P} \tag{4.2.7}$$

as the probability that a randomly selected interactor will be located in i and have an interaction at j, where

$$P = \sum_{ij}\sum T_{ij} = \sum_i P_i \quad . \tag{4.2.8}$$

Unlike the usual trip distribution model, which is constrained with respect to known numbers of trip-makers entering and leaving zones, it is necessary to constrain this particular model with respect to the total population of each origin zone. Thus

$$\sum_j T_{ij} = P_i \quad , \tag{4.2.9}$$

which indicates that it is assumed that every resident of each zone is making a trip. This is, of course, an unrealistic assumption for a trip distribution model, but is not unreasonable for an index like population potential. Dividing each side of constraint (4.2.9) by P will put it in probability form

$$\sum_j p_{ij} = \frac{P_i}{P} \quad . \tag{4.2.10}$$

Given that we are now considering a matrix of spatial interaction (unlike the previous chapter) the Shannon entropy needs to be summed over i and j, such that

$$H = -\sum_{ij} p_{ij} \ln p_{ij} \quad . \qquad [(3.2.2)]$$

Suppose we wish to maximize the entropy subject only to the constraint that $\sum_{ij} p_{ij} = 1$. Form a Lagrangian

$$L = -\sum_{ij} p_{ij} \ln p_{ij} + [(\lambda-1)(1-\sum_{ij} p_{ij})] \tag{4.2.11}$$

where, once again, λ is chosen to satisfy the normalization constraint. Differentiating (4.2.11) and setting the result equal to zero we have

$$\frac{\partial L}{\partial p_{ij}} = -\ln p_{ij} - 1 - \lambda + 1 = 0 \quad , \tag{4.2.12}$$

and solving for p_{ij} results in

$$p_{ij} = \frac{1}{n^2} . \quad (4.2.13)$$

Equation (4.2.13) indicates that in the absence of any prior information beyond the normalization constraint, the most likely form for the probability distribution is uniform. This agrees with results obtained in the third chapter.

In order to consider the above result within the context of potential we can put $T_{ij}/P = p_{ij}$ (equation (4.2.7)) into result (4.2.13) for p_{ij} to get

$$T_{ij} = \frac{P}{n^2} . \quad (4.2.14)$$

Inserting (4.2.14) for T_{ij} in the definition of potential (4.1.3) we have

$$V_j = \sum_i \frac{\frac{P}{n^2}}{P_j} , \quad (4.2.15)$$

or, since P and n^2 are both constant,

$$V_j \propto \frac{1}{P_j} . \quad (4.2.16)$$

The potential in any zone will therefore be proportional to the inverse of the population of that zone when the trip distribution is uniform. This demonstrates once again that it is necessary to include some additional information if the result of the entropy maximizing procedure is to be of further interest.

The requirement for additional information raises one problematic issue associated with potential functions (and also illustrates a further difference between such functions and trip distribution models). In the example in the previous chapter it was shown how some additional information can be introduced in the form of an average distance travelled in the system. Obviously, such data are not readily available in the context of potential because the concern is with potential (per capita) spatial interaction rather than observed interaction. The absence of such data is, however, primarily an empirical problem (and will be discussed as such below). As Tribus (1969, 122) pointed out, it does not necessarily impede the entropy maximizing procedure:

> ...we may not always know the mean value [i.e., $\bar{r}$]. But we may know that such a mean value exists. This information therefore may be put into the formalism to determine the FORM of the probability distribution, though it will not determine the numerical value of the constants appearing in the probability distribution.

Consider a function of distance $f(r_{ij})$. Suppose that the total of this function weighted by the number of trips is known. Writing this total as R, we have

$$\sum_i \sum_j T_{ij} \, f(r_{ij}) = R \quad . \qquad (4.2.17)$$

Dividing each side of equation (4.2.17) by the total number of interactions such that

$$\sum_{ij}\sum \frac{T_{ij}}{P} f(r_{ij}) = \frac{R}{P} \quad , \qquad (4.2.18)$$

will allow us to express equation (4.2.17) in probability form

$$\sum_{ij}\sum p_{ij} f(r_{ij}) = \bar{R} \quad , \qquad (4.2.19)$$

where $\bar{R}$ denotes the mean of the generalized distance. Having introduced this additional information we can now maximize the entropy (3.2.2) subject to constraints (4.2.10) and (4.2.19).

Form a Lagrangian by adding together the entropy and the constraints

$$L = -H + [(\lambda_i - 1)((P_i/P) - \sum_j p_{ij})]$$

$$+ b\,[\bar{R} - \sum_{ij}\sum p_{ij} f(r_{ij})] \quad , \qquad (4.2.20)$$

where b is the Lagrangian multiplier associated with constraint (4.2.19). Differentiating and setting equal to zero in order to find the maximum we have

$$\frac{\partial L}{\partial p_{ij}} = -\ln p_{ij} - 1 - \lambda_i + 1 - bf(r_{ij}) = 0 \quad , \qquad (4.2.21)$$

and solving for p_{ij} we get

$$p_{ij} = \exp[-\lambda_i - bf(r_{ij})] \quad . \qquad (4.2.22)$$

To solve for $\exp(-\lambda_i)$ substitute equation (4.2.22) into the normalization constraint (4.2.10)

$$\sum_j \exp[-\lambda_i - bf(r_{ij})] = \frac{P_i}{P} \quad , \qquad (4.2.23)$$

which yields, on rearrangement

$$\exp(-\lambda_i) = \frac{P_i}{P \sum_j \exp[-bf(r_{ij})]} \quad . \qquad (4.2.24)$$

Substituting equation (4.2.24) into result (4.2.22) for $\exp(-\lambda_i)$ we have

$$p_{ij} = \frac{P_i \exp[-bf(r_{ij})]}{P \sum_j \exp[-bf(r_{ij})]} \quad , \qquad (4.2.25)$$

and finally inserting T_{ij}/P for p_{ij} (equation (4.2.7)) in equation (4.2.25) we get

$$T_{ij} = \frac{P\ P_i \exp[-bf(r_{ij})]}{P \sum_j \exp[-bf(r_{ij})]} \quad . \qquad (4.2.26)$$

The P's cancel, leaving us with

$$T_{ij} = \frac{P_i \exp[-bf(r_{ij})]}{\sum_j \exp[-bf(r_{ij})]} \qquad (4.2.27)$$

which can also be written as

$$T_{ij} = A_i\ P_i \exp[-bf(r_{ij})] \quad , \qquad (4.2.28)$$

where

$$A_i = \frac{1}{\sum_j \exp[-bf(r_{ij})]} \qquad (4.2.29)$$

Equation (4.2.27) is very similar to what was specified earlier to be the desired type of trip distribution model (equations (4.2.5) and (4.2.6)). One difference is that the model above contains the parameter b. Another, more significant difference, is that trip distribution model (4.2.27) does not contain a term for the population at the destination P_j. Before we go on to insert such a term, we can examine very briefly the potential function which will result from using model (4.2.27).

Repeating here for convenience, population potential was defined as

$$V_j = \sum_i \frac{T_{ij}}{P_j} \qquad [(4.1.3)]$$

Inserting result (4.2.27) into this definition we get

$$V_j = \sum_i \frac{\dfrac{P_i \exp[-bf(r_{ij})]}{\sum_j \exp[-bf(r_{ij})]}}{P_j}, \qquad (4.2.30)$$

or

$$V_j = \sum_i \frac{P_i \exp[-bf(r_{ij})]}{P_j \sum_j \exp[-bf(r_{ij})]}, \qquad (4.2.31)$$

which demonstrates clearly the role the destination term will play if a trip distribution model such as (4.2.27) is employed.

In order to get the desired potential function (equation (4.2.2)) it is necessary to insert P_j in the numerator of trip model (4.2.27) such that

$$T_{ij} = \frac{P_i \; P_j \; \exp[-bf(r_{ij})]}{\sum_j \exp[-bf(r_{ij})]} \quad . \qquad (4.2.32)$$

This adjustment can be made only on the basis of the gravity model analogy, but is consistent with the usual approach to the problem (for example, Wilson, 1967, 1971). If trip model (4.2.32) is inserted in the definition for potential (4.1.3), the P_j's cancel and we get

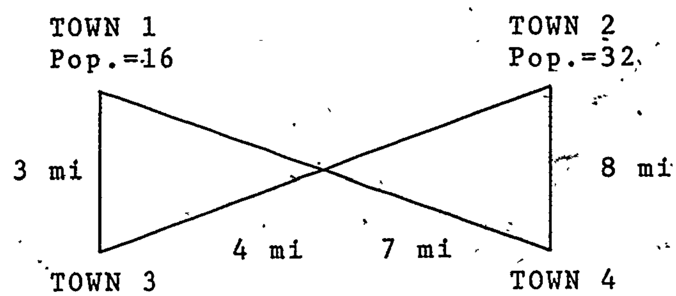
$$V_j = \sum_i \left[\frac{P_i \; \exp[-bf(r_{ij})]}{\sum_j \exp[-bf(r_{ij})]} \right] \; , \qquad (4.2.33)$$

which is very similar to what was stated to be the desired result, with the exception that the parameter b appears in what is now a negative exponential distance response function. If this constrained potential is employed in order to calculate potentials the grand sum of the potential will equal exactly the total population size. A simple example of the calculation of potentials with an equation such as (4.2.33) is presented in Table 4.2.1. It can be noted that the constrained potential can also be written

TABLE 4.2.1

AN EXAMPLE OF THE CALCULATION OF THE CONSTRAINED POTENTIAL
(using linear distance)

$$V_j = \sum_i \left[\frac{P_i \exp(-br_{ij})}{\sum_j \exp(-br_{ij})} \right]$$



Given the populations at Towns 1 and 2, and the distances, it is desired to calculate the constrained potential at Towns 3 and 4. The sum of potential at Towns 3 and 4 should be equal to the total population size (i.e., 48). The parameter b is set equal to 1.0. The calculations are as follows:

$$V_3 = \frac{16e^{-3}}{e^{-3}+e^{-7}} + \frac{32e^{-4}}{e^{-4}+e^{-8}} = 15.71 + 31.42 = 47.13$$

$$V_4 = \frac{16e^{-7}}{e^{-3}+e^{-7}} + \frac{32e^{-8}}{e^{-4}+e^{-8}} = .29 + .58 = .87$$

$$\text{Total } V = 48.00$$

$$V_j = \sum_i A_i P_i \exp[-bf(r_{ij})] \quad , \qquad (4.2.34)$$

where

$$A_i = \frac{1}{\sum_j \exp[-bf(r_{ij})]} \quad . \qquad (4.2.35)$$

Equation (4.2.33) can be said to represent the statistically most likely form for a potential function given only the information in the normalization and distance constraints on the trip distribution model. In Jaynes (1968, 231) words

> The distribution [(4.2.33)] is the one which is, in a certain sense, spread out as uniformly as possible without contracting the given information... it agrees with what is known but expresses a "maximum uncertainty" with respect to all other matters

It was pointed out above that since a potential function is an index of potential or possible <u>per capita</u> interaction, average distance data will not normally be available for the calculation of the balancing factor b. Although this issue will be discussed in greater detail below, it can be pointed out for now that in light of the results obtained so far, b can be determined, as an <u>a posteriori</u> result, from

$$\sum_{ij} p_{ij} \; f(r_{ij}) = \bar{R} = \frac{\sum_{ij} r_{ij} \exp[-bf(r_{ij})]}{\sum_{ij} \exp[-bf(r_{ij})]} \qquad (4.2.36)$$

which simply expresses the mean distance with respect to the form of the distance response function (see Webber, 1976, 279).

Before we go on to consider the derivation of some alternative distance response functions for potentials it is necessary first to digress briefly in order to discuss the role of normalizing constraints in the calculation of potentials.

4.3 Normalizing Constraints and Potentials

Normalizing constraints are employed in spatial interaction models in ~~order to~~ ensure that the model number of trips leaving or arriving at origins and destinations matches the observed number. Such a requirement is unnecessary for potential functions quite simply because they are not intended to be employed in modelling spatial interaction; a potential function is an index of potential or possible per capita interaction at the destination. However, the use of normalizing constraints and the balancing factors associated with them in potential functions will have an effect on the resulting form of the potential surface. Suppose, for example, that we were to construct one potential map using an unconstrained negative exponential function

$$V_j = \sum_i P_i \exp(-br_{ij}) \quad , \qquad [(2.6.2)]$$

and another using the constrained negative exponential function

$$V_j = \sum_i A_i P_i \exp(-br_{ij}) \quad , \qquad (4.3.1)$$

where

$$A_i = \frac{1}{\sum_j \exp(-br_{ij})} \quad . \qquad (4.3.2)$$

The map is to be constructed on a grid of equally spaced control points with equal P_i at each. One potential function is constrained with respect to population size, and the other is not, and because of this, the two resulting potential surfaces—if constructed using the same data, and the same value for b—would differ. The reason for this difference is to be found in the effect of the balancing factors (equation (4.3.2)).

If we were to calculate the total distance from each point to all other points on the bounded uniform grid of control points, the points central to the grid would have the smallest totals by virtue of their centrality. On a graph showing the relationship between distance and the negative exponential distance response function (Figure 4.3.1) the sum of $\exp(-br_{ij})$ (the denominator of equation (4.3.2)) will be larger when distances are small ((a) on the graph) and smaller when distances are large ((b) on the graph).

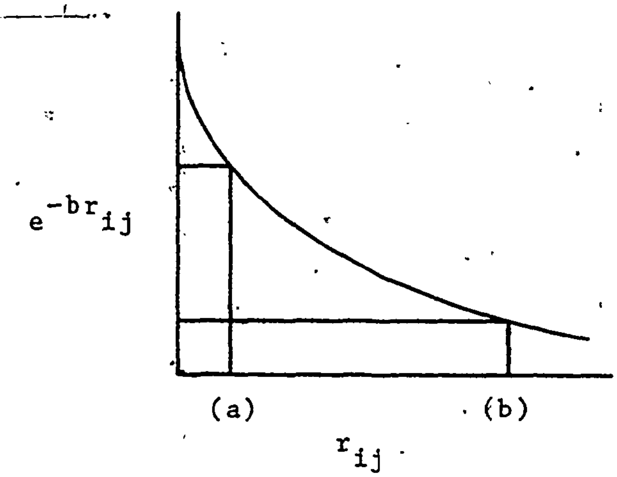


Figure 4.3.1 Distance and the Negative Exponential Distance Response Function

The balancing factor (4.3.2) is written with respect to the inverse of this sum, however, which means that when distances are smaller (at the center of the map) the value of the A_i's will be lower. Conversely, when distances are larger (at the periphery of the map) the value of the A_i's will be larger. The effect of these differences in the relative values of the A_i's, when calculating the V_j's, will be that at the center of the region of interest the values of the V_j's will be lower than they would have been otherwise, since the larger A_i's at the periphery of the region will tend to increase the size of the denominator in the constrained potential. Similarly, at the periphery of the region, the values of the V_j's will be larger than they would have been otherwise, since the smaller A_i's at the center of the region will tend to decrease the size of the denominator in the constrained potential. Stated another way it can be said that if one map were constructed using the unconstrained function (2.6.2), and the other with the constrained function (4.3.1), the latter map, all other things being equal, will tend to be "flatter" than the former.

In the case of the usual unconstrained potential, the greatest number of *per capita* interactions occur over the shortest distances and, as a result, the potential surface associated with a bounded uniform distribution of population will be peaked at the center. This is a desirable property of the potential surface if the boundary is a real one (such as a coastline) and there are no P_i outside the boundary. If, however, the boundary is an arbitrary one (such as a county or state), and there are P_i outside of the boundary, then it is *not* desirable that the potential of the uniform distribution inside the boundary be peaked at the center. In fact, it is this property of unconstrained potentials which can be said to be one of their major weaknesses; the potential on an arbitrarily bounded uniform distribution of population with P_i outside the boundary should, in theory, be uniform.

Insofar as the balancing factors in a constrained potential function will have the effect of "flattening" what would otherwise be a centrally peaked potential surface, it would appear that their inclusion in potential functions is a first step in overcoming this deficiency. The task of determining the exact degree of flattening is largely an empirical one and is left for future research. For the moment, however, it can be noted that if b is set equal to zero in the constrained potential function, (4.3.1) and (4.3.2), then

$$V_j = \sum_i \left(\frac{P_i}{n} \right) , \qquad (4.3.3)$$

and V_j will be constant on each zone.

In the second chapter a number of possible forms for the distance response function in a potential equation were discussed. To this point, however, we have only considered negative exponential and generalized forms of potential. It will be seen in the next section that there are a number of other distance response functions which can be obtained using the entropy maximizing approach.

4.4 A Family of Potential Functions

Within the field of urban population density modelling, Amson (1972a) has proposed, and Zielinski (1979; 1980) has reviewed, a "family" of density models based on the quadratic gamma function. This function expresses the population density d(r) at a distance from the center of a city d_o as

$$d(r) = d_o \exp(-b_1 r - b_2 r^2)\ r^{-b_3} \qquad (4.4.1)$$

The terms b_1, b_2, and b_3 are the parameters of the function and setting them equal to zero individually and in pairs results in the other models of the family. Setting b_3 equal to zero (now writing only the distance response function) we get the quadratic negative exponential function

$$\exp(-b_1 r - b_2 r^2) \quad , \qquad (4.4.2)$$

setting b_1 equal to zero gives the normal gamma function

$$\exp(-b_2 r^2)\ r^{-b_3} \quad , \qquad (4.4.3)$$

and setting b_2 equal to zero gives the gamma function

$$\exp(-b_1 r)\ r^{-b_3} \quad . \qquad (4.4.4)$$

Setting b_2 and b_3 equal to zero we get the negative exponential function

$$\exp(-b_1 r) \quad , \qquad (4.4.5)$$

setting b_1 and b_3 equal to zero gives the normal function

$$\exp(-b_2 r^2) \quad , \qquad (4.4.6)$$

and, finally, setting b_1 and b_2 equal to zero results in the power function

$$r^{-b_3} \quad . \qquad (4.4.7)$$

Given that a number of members of this quadratic gamma family were employed within potential functions in the second chapter, it would seem useful to consider employing all of them as a basis for defining a family of potential functions. Moreover, it would seem to be especially desirable to attempt to derive all of the members of such a family of distance response functions on the basis of maximum entropy.

It has been shown that the form of the distance response function resulting from the entropy maximizing procedure depends upon the nature of the constraint(s) specified. It follows from this that it should be possible to derive the quadratic gamma function, and hence the members of the family, if the appropriate constraints can be determined.

Note first that the quadratic gamma distance response function in equation (4.4.1) can be rewritten in a manner which makes it more obvious that the function consists of three separate parts, that is

$$\exp(-b_1 r)\exp(-b_2 r^2)\ r^{-b_3} \quad . \qquad (4.4.8)$$

We have seen that the first term in this function results from entropy maximization if a constraint is placed on the mean distance (rewriting and renumbering for convenience)

$$\sum_{i}\sum_{j} p_{ij}\ r_{ij} = \bar{r} \quad . \qquad (4.4.9)$$

An appropriate constraint is now also required for the second or middle term in equation (4.4.8). Such a constraint has been shown by Tribus (1969, 131) to be that associated with the variance of the distribution

$$\sum_{ij}\sum p_{ij}\ (r_{ij} - \bar{r})^2 = \sigma^2$$

which can also be written

$$\sum_{ij}\sum p_{ij}\ r_{ij}^2 = \overline{r^2} + \sigma^2 \qquad (4.4.10)$$

where σ is the standard deviation (and σ^2 the variance). It has also been suggested (see Wilson, 1970, 35) that the inverse power function—the third term in (4.4.8)—can be derived via entropy maximization if a constraint is placed on the log of the mean distance such that[1]

$$\sum_{ij}\sum p_{ij}\ \ln r_{ij} = \overline{\ln r} \qquad (4.4.11)$$

---

[1] Note that this constraint can also be written in the form of the geometric mean, that is, if

$$\sum_{ij} p_{ij}\ \ln p_{ij} = \overline{\ln r}$$

then

$$\ln(\prod_{ij} r_{ij}^{p_{ij}}) = \overline{\ln r}$$

$$\ln(\prod_{ij} r_{ij}^{p_{ij}}) = \sum_{ij} \ln r_{ij}/n^2$$

$$\ln(\prod_{ij} r_{ij}^{p_{ij}}) = \frac{1}{n^2} \ln \prod_{ij} r_{ij}$$

$$\ln(\prod_{ij} r_{ij}^{p_{ij}}) = \ln(\prod_{ij} r_{ij})^{\frac{1}{n^2}}$$

$$\prod_{ij} r_{ij}^{p_{ij}} = (\prod_{ij} r_{ij})^{\frac{1}{n^2}} .$$

In order to derive a trip distribution model with a quadratic gamma distance response function from the maximum entropy formalism we can maximize the function in the usual way, now forming a Lagrangian using constraints (4.2.10), (4.4.9), (4.4.10), and (4.4.11) as follows

$$L = -H + [\lambda_i - 1)((P_i/P) - \sum_j p_{ij})] + b_1[\bar{r} - \sum_{ij}\sum p_{ij}\, r_{ij}]$$
$$+ b_2[\overline{r^2} + \sigma^2 - \sum_{ij}\sum p_{ij}\, r_{ij}^2]$$
$$+ b_3[\overline{\ln r} - \sum_{ij}\sum p_{ij} \ln r_{ij}] \quad . \qquad (4.4.12)$$

Maximizing we have

$$\frac{\partial L}{\partial p_{ij}} = -\ln p_{ij} - 1 - \lambda_i + 1 - b_1 r_{ij}$$
$$-b_2 r_{ij}^2 - b_3 \ln r_{ij} = 0 \quad , \qquad (4.4.13)$$

and solving for p_{ij} gives

$$p_{ij} = \exp(-\lambda_i)\exp(-b_1 r_{ij})\exp(-b_2 r_{ij}^2)\; r_{ij}^{-b_3} \quad . \qquad (4.4.14)$$

Substitution in normalizing constraint (4.2.10) to solve for $\exp(-\lambda_i)$ will lead to a result of

$$p_{ij} = \frac{P_i \exp(-b_1 r_{ij})\exp(-b_2 r_{ij}^2)\; r_{ij}^{-b_3}}{P \sum_j \exp(-b_1 r_{ij})\exp(-b_2 r_{ij}^2)\; r_{ij}^{-b_3}} \quad . \qquad (4.4.15)$$

Equation (4.4.15) is a probability distribution with a quadratic gamma distance response function. This probability distribution can be turned into a trip distribution model, and then into a potential function, by repeating the steps outlined in section 4.2. This leads to a quadratic gamma potential function of the form

$$V_j = \sum_i A_i P_i \exp(-b_1 r_{ij}) \exp(-b_2 r_{ij}^2)\, r_{ij}^{-b_3}, \quad (4.4.16)$$

where

$$A_i = \frac{1}{\sum_j \exp(-b_1 r_{ij}) \exp(-b_2 r_{ij}^2)\, r_{ij}^{-b_3}}. \quad (4.4.17)$$

It is now possible to define a maximum entropy family of potential functions by dropping the distance constraints on the entropy function one at a time and in pairs. In the interests of economy of discussion the balancing factors A_i will not be written out. In each case below, the A_i's will be equal to one over the sum over j of the distance response function (as in equation (4.4.17)).

If the only constraints on the entropy function are (4.4.9) and (4.4.10) the maximization will result in a quadratic negative exponential potential function

$$V_j = \sum_i A_i P_i \exp(-b_1 r_{ij}) \exp(-b_2 r_{ij}^2) \quad . \qquad (4.4.19)$$

Maximizing the entropy subject only to constraints (4.4.10) and (4.4.11) will result in a normal gamma potential function

$$V_j = \sum_i A_i P_i \exp(-b_2 r_{ij}^2) r_{ij}^{-b_3} \quad , \qquad (4.4.20)$$

and similarly, maximizing subject only to constraints (4.4.9) and (4.4.11), will produce a gamma potential function

$$V_j = \sum_i A_i P_i \exp(-b_1 r_{ij}) r_{ij}^{-b_3} \quad . \qquad (4.4.21)$$

If constraints (4.4.10 and (4.4.11) are dropped, and the maximization of the entropy function is carried out subject only to constraint (4.4.9), the result will be a negative exponential potential function

$$V_j = \sum_i A_i P_i \exp(-b_1 r_{ij}) \quad , \qquad (4.4.22)$$

while maximizing subject to constraint (4.4.10) alone will give a normal potential function

$$V_j = \sum_i A_i P_i \exp(-b_2 r_{ij}^2) \quad . \qquad (4.4.23)$$

Finally, maximizing the entropy subject only to constraint (4.4.11) will result in a power form of potential function

$$V_j = \sum_i A_i P_i r_{ij}^{-b_3} \qquad (4.4.24)$$

which is similar to the original Stewart formulation when b_3 is set equal to one.

The seven potential functions which have been set out above (equations (4.4.16) to (4.4.24) can be said to form a coherent quadratic gamma family which is based on the maximum entropy formalism. In the approach of Amson (1972) and Zielinski (1979, 1980), the members of the quadratic gamma family are obtained by dropping the parameters of the distance response function one at a time and in pairs, whereas in the maximum entropy approach the alternative family members are obtained by dropping the constraints on the entropy function one at a time and in pairs. The latter approach can be said to be preferable since it provides some rationale for choosing and interpreting any particular family member. In the former approach, the dropping of parameters would appear to be somewhat arbitrary, whereas in the maximum entropy approach an explicit connection is made between the parameters and their constraints. A summary of the relationship between the constraints and distance response functions is presented in Table 4.4.1.

It remains to provide a common sense or intuitive interpretation of the meaning of the various distance

TABLE 4.4.1

A SUMMARY OF THE QUADRATIC GAMMA FAMILY OF POTENTIALS SHOWING THE RELATIONSHIP BETWEEN CONSTRAINTS AND DISTANCE RESPONSE FUNCTIONS RESULTING FROM ENTROPY MAXIMIZATION

Distance Constraint	Distance Response Function	Name
$\sum_j P_{ij} g(r_{ij}) = \overline{g(r)}$	$V_j = \sum_i A_i P_i f(r_{ij})$	(following Zielinski, 1979)
(1) $\bar{r}$, $\overline{r^2}$, $\overline{\ln r}$	$e^{-b_1 r_{ij}} e^{-b_2 r_{ij}^2} r_{ij}^{-b_3}$	Quadratic Gamma
(2) $\bar{r}$, $\overline{r^2}$	$e^{-b_1 r_{ij}} e^{-b_2 r_{ij}^2}$	Quadratic Negative Exponential
(3) $\overline{r^2}$, $\overline{\ln r}$	$e^{-b_2 r_{ij}^2} r_{ij}^{-b_3}$	Normal Gamma
(4) $\bar{r}$, $\overline{\ln r}$	$e^{-b_1 r_{ij}} r_{ij}^{-b_3}$	Gamma
(5) $\bar{r}$	$e^{-b_1 r_{ij}}$	Negative Exponential
(6) $\overline{r^2}$	$e^{-b_2 r_{ij}^2}$	Normal
(7) $\overline{\ln r}$	$r_{ij}^{-b_3}$	Inverse Power

response functions. For the functions based on a single constraint ((4.4.22), (4.4.23), and (4.4.24)) this is not too difficult. The negative exponential potential function (4.4.22) can be interpreted to be the statistically most likely form for a potential function when spatial interaction varies inversely with distance. Similarly, for the power form of the potential function (4.4.24)—which is based upon the constraint (4.4.11) being placed upon the log of distance—it can be said that interaction varies logarithmically or less than linearly with distance. As Wilson (1970, 35) noted:

> Suppose people perceive travel costs not as we measure them, but as the logarithm of what we measure. Such an assumption would apply if the cost of travelling 50 miles was perceived to be less by the traveller who was committed to 200 miles anyway than to the traveller who was going 50 miles in total. Then r_{ij} is replaced by $\ln r_{ij}$, and $\exp(-br_{ij})$ becomes $\exp(-b\ln r_{ij})$, which is r_{ij}^{-b}. Thus, if models fit better with inverse power functions than with negative exponential functions, this tells us something about the way travellers perceive costs.

For the normal potential function (4.4.23) the maximization is carried out subject only to the constraint on the variance. The important thing to note for purposes of interpretation is that the constraint (4.4.10) is written with respect to the square of the distance. Thus it can be said that when interaction varies as the square of the distance, the statistically most likely form for the distribution of *per capita* interaction is identical in form to the normal curve. It can also be noted that equation (4.4.23) is illustrative of the "function-of-a-function" concept (discussed by Wilson and Kirkby, 1975, 80) in the sense that the normal distance response function can be regarded as the exponential of the quadratic term $-br_{ij}^2$.

The remaining members of the family of potential functions set out above are more difficult to interpret in an intuitive manner since they are composite functions; equations (4.4.19), (4.4.20) and (4.4.21) each contain two terms, while equation (4.4.16) contains three terms. Perhaps the best that can be said is that these composite functions should be regarded as combining the properties of the individual functions from which they are composed. For example, the gamma function (4.4.21), as Zielinski (1980, 144) has stated "can be seen as an amalgamation of the negative-exponential and inverse-power [functions], and has the merits of both in its flexibility." In the context of the present approach, it can also be said that the composite functions indicate that there are combinations of constraints operative in the system. Moreover, it can be noted that the degree to which any of the composite functions tend to approximate the individual functions they contain, or the individual constraints on which they are based, depends largely on the numerical values obtained for the parameters in the fitting of the functions. For example, in the case of the gamma function, when b_1 is small and b_3 is large, the distribution will approximate the power function, while when b_3 is small and b_1 is large, it will tend to act like the negative exponential function. In general it can be said that in the case of the composite functions, the issue of meaning or interpretation is partly replaced by the empirical question of which function provides the best fit to a given distribution. The matter of interpretation is then a function of the parameter values obtained.

This section has presented a discussion of the indirect, maximum entropy derivation of potential functions. In the next section we shall consider the manner in which the principle of minimum information can be employed.

4.5 Potentials Based on Minimum Information Trip Distributions

In Chapter 3 the method of minimum information was introduced as an extension and refinement of the maximum entropy formalism. This section explores some of the possible applications of the method of minimum information in the context of population potential.

Repeating here for convenience, the Kullback (1959) information gain is

$$I(q:p) = \sum_i p_i \ln \frac{p_i}{q_i} , \qquad [(1.3.2)]$$

where p_i and q_i are posterior and prior probabilities respectively. Minimizing this function is analogous to maximizing the entropy function except that the Kullback function allows any additional information concerning non-constant prior distributions to be taken into account explicitly. In order to consider minimum information and potential jointly, it will be useful to first discuss some of the sorts of prior information which may or may not be of relevance to per capita interaction. It should be noted that this discussion is with respect to potential functions of the form which were derived earlier in the chapter. Many, if not all of the types of information discussed below are of relevance to trip distribution models, and in the case of other forms of potential (based on alternative trip models) these arguments may not apply.

In the third chapter the example employed in order to illustrate the method of minimum information concerned the assignment of workers to homes in zones, where the zones were of unequal area. It was noted previously that potential is an intensive variable (see section 2.4) and, as such, the size of the destination zone has no effect on the quantity; splitting the destination zone, for example, has no effect on the potential contributions in each of the two new zones. It follows from this that there is no rationale for attempting to employ minimum information to take prior information concerning destination zone size into account. Similarly there is no reason to consider the areas of the originating zones *per se* as being of relevance in the estimation of potentials.

In a discussion of the method of minimum information, Snickars and Weibull (1977, 145) described a number of possible choices for the prior in spatial models, among them, population distributions at past times, and historical travel patterns. It is apparent that data on observed or historical travel patterns are not permissible choices for the prior in minimum information potentials of the form being considered here since potential consists of *possible per capita* interaction rather than *observed* spatial interaction. Similarly, population distributions at past times, when considered as being in the destination zones, also are not of relevance since, by definition, characteristics of the destination do not influence the amount of *per capita* spatial interaction. Population distributions at past times in the origin zones would also seem to have little or no pertinence to the computation of present day potentials.

It is apparent that many of the types of priors which are normally employed in spatial model building—such as those concerning the areas of the zones or historically observed travel patterns—have no relevance to the minimum information estimation of the form of potential functions. This is not surprising, for as we have seen, the type of potential functions under discussion here have a number of properties and characteristics which set them apart from trip distribution and location models. However there is a manner in which appropriate priors can be chosen for these potentials and this is based on the characteristics of the origins.

The potential at a point is a function of both the characteristics of the originating zone and the attenuating effect of distance between zones. The latter aspect was treated in the previous section, while the former has barely been considered. It is argued here that there are characteristics of the originating zones, besides their populations, which may be of relevance to the amount of <u>per capita</u> interaction emanating from them, and that prior probability distributions can be defined on the basis of such characteristics. In the calculation of population potential, for example, some properties of the origin zone which may be of relevance to the amount of <u>per capita</u> interaction, in addition to the population size, might be the number of automobiles there, the income of the population, energy costs in the zone, and so on. Such data would seem to be an appropriate choice for priors within the context of potential; they represent a priori information which may have an effect on the amount of per capita spatial interaction emanating from a zone. Stated another way, if such data are known, and are nonuniform, the microstates are not all a priori equally likely, and the method of minimum information, rather than maximum entropy, should be employed.

Thus it is argued that while entropy maximization is of relevance to the study of the effects of distance in potential functions, the method of minimum information is applicable to the study of the emissivity of the originating zones. In particular it is suggested that a method of sequential information minimizing (see Webber, 1979) can be fruitfully employed as an organizing framework within which the characteristics of the originating zones can be systematically investigated.

In the second chapter it was noted that a traditional concern in the study of potentials has been with the correspondence or correlation between potential and other variables deemed to be representative of accessibility or social intensity. Such a concern can be construed as an attempt to make a spatial prediction on the basis of limited prior information, that is, given the populations, say, and the distances, it is desired to see how well the value of some other phenomenon, such as the value of land, can be predicted. The discussion in the previous section indicated that one method of attempting to improve such predictions is to take into account possible variations in the attenuating effect of distance by considering alternative constraints on distance. Another manner in which an attempt can be made to improve on the predictions, however, is to take into account possible variations in the effect of the origin zone. Such an approach to this problem has been commonplace and is implicit, for example, in the procedure of replacing the population term by some other term deemed to be more relevant to a particular problem (such as retail sales (see Harris, 1954)), or in weighting the population term by some quantity which is seen to be of significance (such as <u>per capita</u> income (see Warntz, 1959, 1965, 1979)). The method of minimum information provides an organizing framework within which such adjustments and weightings can be handled in a systematic and sequential way. The significance of the method is that it allows an explicit connection to be made between the constraints deemed to be operative in some spatial system and the size of the origin term in the potential function. This is precisely analogous to the manner in which the connection was made earlier between constraints on distance and the form of the distance response function. In fact, and as we will see below, the method of maximum entropy is a special case of the method of minimum information when the prior is uniform. The method of minimum information allows normalization and distance constraints, and information concerning non-constant prior probability distributions, to be incorporated within a single formalism.

A version of the Kullback information gain which is appropriate for a matrix of interaction is

$$I(q:p) = \sum_{ij}\sum p_{ij} \ln \frac{p_{ij}}{q_{ij}} \quad . \qquad (4.5.1)$$

It is desired to minimize this function subject to whatever additional information is known. The pieces of prior information we have can be expressed as follows (now renumbering the constraints for convenience):

$$\sum_j p_{ij} = \frac{P_i}{P} \quad , \qquad (4.5.2)$$

$$\sum_{ij}\sum p_{ij} r_{ij} = \bar{r} \quad ; \qquad (4.5.3)$$

and

$$q_{ij} (i, j=1,\ldots,n) \text{ known.} \qquad (4.5.4)$$

Here p_{ij} is defined as it was earlier in the chapter, that is,

$$p_{ij} = \frac{T_{ij}}{P} \quad , \qquad [(4.2.7)]$$

and q_{ij} is defined with respect to any nonconstant prior distribution as measured on the originating zones, that is

$$q_{ij} = \frac{s_{ij}}{\sum_i s_{ij}} \quad , \qquad (4.5.5)$$

where s_{ij} is the amount of some phenomenon in zone i considered to have an effect on the number of interactions between it and j. The prior probability distribution q_{ij} might, for example, be defined with respect to the number of automobiles in the origin zones, that is, it is suspected that this will have an effect on the number of interactions with j. The prior can also be defined with respect to some measure <u>between</u> zones, which is considered to facilitate spatial interaction, for example, s_{ij} could be the number of direct road, rail, and/or air route connections between i and j (expressed as a proportion of the total number of connections between all zones).

It is now desired to find the minimum of the Kullback information (4.5.1) subject to the information in constraints (4.5.2), (4.5.3), and (4.5.4). Form a Lagrangian to maximize $-I(q:p)$

$$L = -I(q:p) + [(\lambda_i - 1)((P_i/P) - \sum_j p_{ij})]$$

$$+ b(\bar{r} - \sum_i \sum_j p_{ij} \, r_{ij}) \quad , \qquad (4.5.6)$$

where λ_i and b are multipliers associated with constraints (4.5.2) and (4.5.3) respectively. Differentiating equation (4.5.6) with respect to p_{ij} and setting it equal to zero gives

$$\frac{\partial L}{\partial p_{ij}} = -\ln p_{ij} - 1 + \ln q_{ij} - \lambda_i + 1 - br_{ij} = 0 \ . \quad (4.5.7)$$

Rearranging equation (4.5.7) and taking antilogs results in

$$p_{ij} = q_{ij} \exp(-\lambda_i) \exp(-br_{ij}) \quad . \quad (4.5.8)$$

To solve for $\exp(-\lambda_i)$ substitute in the normalizing constraint (4.5.2) in the usual way to obtain

$$\exp(-\lambda_i) = \frac{P_i}{P \sum_j q_{ij} \exp(-br_{ij})} \quad , \quad (4.5.9)$$

and substitute equation (4.5.9) into equation (4.5.8) for $\exp(-\lambda_i)$ to get

$$p_{ij} = \frac{P_i \ q_{ij} \exp(-br_{ij})}{P \sum_j q_{ij} \exp(-br_{ij})} \quad (4.5.10)$$

Equation (4.5.10) represents the probability of interaction between zones and it can be seen that the prior probability distribution is now included in the function.

Substituting T_{ij}/P for p_{ij} (equation (4.2.7)) in equation (4.5.10) we have

$$T_{ij} = \frac{P \ P_i \ q_{ij} \exp(-br_{ij})}{P \sum_j q_{ij} \exp(-br_{ij})} \quad (4.5.11)$$

where, upon cancelling the P's, and again inserting P_j in the numerator of the right-hand-side on the basis of the gravity model analogy, we get a trip distribution model

$$T_{ij} = \frac{P_i \; P_j \; q_{ij} \; \exp(-br_{ij})}{\sum_j q_{ij} \; \exp(-br_{ij})} \quad . \qquad (4.5.12)$$

In order to turn this trip model into an expression concerning per capita spatial interaction at the destination, it is inserted into the definition of potential (repeating here for convenience)

$$V_j = \sum_i \frac{T_{ij}}{P_j} \quad , \qquad [(4.1.3)]$$

to give

$$V_j = \sum_i \left[\frac{P_i \; q_{ij} \; \exp(-br_{ij})}{\sum_j q_{ij} \; \exp(-br_{ij})} \right] \qquad (4.5.13)$$

It should be noted again that it is only by the device of inserting P_j in the trip model that it cancels out of the final result. Had P_j not been inserted in the trip distribution model then it would appear in the potential as follows

$$V_j = \frac{\sum_i \frac{P_i \; q_{ij} \; \exp(-br_{ij})}{\sum_j q_{ij} \; \exp(-br_{ij})}}{P_j} \quad . \qquad (4.5.14)$$

To put potential function (4.5.13) in more familiar form we can separate out the balancing factor such that

$$V_j = \sum_i A_i \; P_i \; q_{ij} \; \exp(-br_{ij}) \quad , \qquad (4.5.15)$$

where

$$A_i = \frac{1}{\sum_j q_{ij} \exp(-br_{ij})} \tag{4.5.16}$$

Equations (4.5.15) and (4.5.16) are identical to equations (4.2.34) and (4.2.35) which were derived by maximizing the entropy, except that they contain the prior q_{ij} (and the distance response functions are different). The assignment of <u>per capita</u> interaction with this function will be directly proportional to the prior probability distribution (times the population) and inversely proportional to an exponential function of distance.

There is a slightly different and interesting approach which can be taken with equation (4.5.13). Snickars and Weibull (1977), and Batty (1978, 142) have noted that since $q_{ij} = s_{ij}/\sum_i s_{ij}$, a result such as equation (4.5.13) can also be written

$$V_j = \sum_i \left[\frac{P_i \; s_{ij} \exp(-br_{ij})}{\sum_j s_{ij} \exp(-br_{ij})} \right] , \tag{4.5.17}$$

or

$$V_j = \sum_i A_i \, P_i \, s_{ij} \exp(-br_{ij}) \quad , \tag{4.5.18}$$

where

$$A_i = \frac{1}{\sum_j s_{ij} \exp(-br_{ij})} \quad . \tag{4.5.19}$$

These equations will give the same results as those to be obtained from equations (4.5.15) and (4.5.16) even though the prior is now expressed with respect to the actual data s_{ij} (rather than as a probability).

A further aspect of writing the results in the form of equation (4.5.17) is that it allows us to redefine the population term in a potential function using the method of minimum information. If s_{ij} is defined with respect to any variable measured on a per capita basis in the origin zone, this will transform the P_i term into the units of the other variable. For example, if s_{ij} is defined as per capita income in the origin zone, then because the product of population and per capita income is total income ($P_i \, s_{ij} = Z_i$) equation (4.5.17) can be written directly as income potential

$$U_j = \sum_i A_i \, Z_i \, \exp(-b r_{ij}) \quad , \qquad (4.5.20)$$

where

$$A_i = \frac{1}{\sum_j s_{ij} \exp(-b r_{ij})} \qquad [(4.5.19)]$$

The same results hold for any s_{ij} measured in per capita terms; if s_{ij} is measured as automobiles per capita, for example, the origin term in the potential function will be numbers of automobiles.

The principle of minimum information thus provides a method whereby additional information concerning the origin zones, or connections between zones, can be included in potential functions. The choice of the prior distribution, of course, may be argued to be no less arbitrary than simply deciding to weight population by some other variable as has traditionally been the case. However, minimum information does provide an organizing framework within which such additional information can be systematically incorporated. Moreover, Webber (1979, 140) has argued that minimum information provides a vehicle for "sequential model building" and this has application in the context of potential. Webber (1979, 140) has described the sequential approach as follows:

> The information minimizing research process...represents a simple strategy: Once a social system has been identified, a small number of constraints is then used in the first model of that system; if the results are insufficiently accurate, additional constraints are involved. The process of adding constraints continues until the predictions are acceptable, but only those constraints are added which affect the predictions.

Sequences of information minimizing potential functions can be developed following this approach and this would allow for the inclusion of sufficient prior information such that some desired level of predictability could be reached. Some of the types of variables which might be considered as priors relevant to potentials are: agricultural production; automobiles, air flights, buses, trains; telephones and mail shipments; manufacturing production and shipments; income; energy costs; numbers of direct road, rail, and air connections; numbers of contiguous zones (join counts); numbers of intervening opportunities; and so on. A form of the Kullback information statistic which could incorporate m such variables is

$$I(q:p) = \sum_{i}\sum_{j} p_{ij} \ln \frac{p_{ij}}{q^1_{ij}\, q^2_{ij}\, q^3_{ij} \cdots q^m_{ij}} \quad , \qquad (4.5.21)$$

or, more generally

$$I(q:p) = \sum_{i}\sum_{j} p_{ij} \ln \frac{p_{ij}}{\prod_m q^m_{ij}} \quad . \qquad (4.5.22)$$

Minimizing equation (4.5.22) subject to normalization constraint (4.2.10) and the distance constraint

$$\sum_{i}\sum_{j} p_{ij}\, f(r_{ij}) = \bar{R} \quad , \qquad [(4.2.19)]$$

and following the other intermediate steps outlined above, will result in a generalized potential function

$$V_j = \sum_i \left[\frac{P_i \prod_m q^m_{ij} \exp[-bf(r_{ij})]}{\sum_j \prod_m q^m_{ij} \exp[-bf(r_{ij})]} \right] \quad , \qquad (4.5.23)$$

which says that the probability of <u>per capita</u> interaction between zones is directly proportional to m prior distributions and inversely proportional to an exponential function of distance. When it is remembered that three types of constraints on distance lead to seven different distance response functions, and when these are then considered in combination with m prior probability distributions, it is apparent that the method of minimum information provides us with a framework for deriving a very large number of potentials.

The minimum information potential function (4.5.23), like the potential functions derived from maximum entropy, is statistically the most likely to occur, given the information in the constraints. In fact, it can be demonstrated that information minimizing is equivalent to entropy maximizing when the prior is uniform. Define a prior

$$q_j = \frac{a_j}{\sum_j a_j} , \qquad (4.5.24)$$

where a_j is the size of the destination zone (in, say, a location model). Then minimizing the Kullback information statistic subject to normalization and distance constraints will result in a model of the form

$$p_{ij} = \frac{a_j \; f(r_{ij})}{\sum_j a_j \; f(r_{ij})} , \qquad (4.5.25)$$

where, to quote Batty (1978, 142) "it is clear that if $[a_j]$ is uniform, that is, if $[a_1 = a_2 = \ldots = a_n]$, then the zone size effect cancels from the model."

The equivalency of minimum information and maximum entropy in the case of a uniform prior can also be demonstrated in another manner. Define the uniform prior

$$q_{ij} = \frac{1}{n} \quad . \tag{4.5.26}$$

Inserting this into the Kullback function (4.5.1) we have

$$I(q:p) = \sum_{ij}\sum p_{ij} \ln p_{ij} n \quad ,$$

$$= \ln n + \sum_{ij}\sum p_{ij} \ln p_{ij} \quad . \tag{4.5.27}$$

Because ln n is a constant, minimizing $I(q:p)$ can be seen to be equivalent to maximizing H, that is

$$\min[I(q:p) - \ln n] = \max - \sum_{ij}\sum p_{ij} \ln p_{ij} \quad , \tag{4.5.28}$$

where the right-hand-side is the Shannon entropy (see Batty, 1978, 142; Webber, 1979, 120). This illustrates that entropy maximization is a special case of information minimization in which the prior is uniform (see also Hobson and Cheng, 1973).

This chapter has presented the indirect derivation of potential functions through the methods of maximum entropy and minimum information. A general discussion of the usefulness and practicality of these derivations and results will be reserved for the moment. In the next chapter the discussion focuses instead on some of the more practical issues associated with the use and operationalization of the sorts of potential functions which have been presented here.

CHAPTER 5

OPERATIONALIZATION AND INTERPRETATION

Before we consider the methods of calculation and empirical meaning of the sorts of potential functions which have been defined above, we will first reconsider the issue of the self-potential, now within the context of the maximum entropy methodology. This is the subject of the discussion which follows.

5.1 Stewart's Self-Potential

The calculation of the self-potential was described in the first chapter as a problem which arises when one wishes to compute the contribution of potential per capita interaction of the population within a region, to the total per capita interaction (or potential) at its centroid. It was pointed out that when the Stewart form of potential (1.1.2) is being employed, it is necessary to use an equation such as (1.1.4) in order to avoid dividing by zero when i equals j. In the second chapter, it was also suggested that perhaps the best approach to resolving the self-potential problem, when possible, is to reduce the aggregation level of the data. Decreasing the sizes of the areal units not only reduces the size of the self-potential contribution in comparison to the contribution from all other regions, but also results in computations which more closely approximate the integral form of the

Stewart function (1.1.3). Nevertheless, it is not always possible or convenient to reduce the level of aggregation, and even if such data are available, it still remains necessary to calculate the self-potential, however small its relative magnitude.

In discussing the problem of the self-potential it should be pointed out first that Stewart's method of assuming a region to be circular, and then dividing the population by half the radius is not an arbitrary one. In this case the self-potential is given by

$$v_{jj} = 2P/r \quad , \tag{5.1.1}$$

where r, the radius, is found from

$$r = (a/\pi)^{1/2} \quad . \tag{5.1.2}$$

This method is based upon the mathematical fact that the potential at the center of a disk having uniform density can be shown to be equal to the population divided by half the radius, as will be shown below (following Warntz et al, 1971, App. I).

Consider a circular area of radius r with a uniform distribution of population and imagine that this area has been partitioned into n concentric rings such that each ring has the same width, that is $(r_j - r_{j-1})$ is constant. Here r_j denotes the radius of a ring, where the total radius is

$$r = \sum_j (r_j - r_{j-1}) \quad . \qquad (5.1.3)$$

If the number of rings is large, an approximation to the integral of the potential of population at the center of the circular region can be found by dividing the population within each ring by the distance to the center and summing over all rings. Thus

$$v_{jj} = \sum_j \frac{d\,\pi r_j^2 - d\,\pi r_{(j-1)}^2}{(r_j + r_{j-1})/2} \quad , \qquad (5.1.4)$$

where, once again, d denotes population density. If density is held constant, (5.1.4) can be simplified, such that

$$v_{jj} = 2\pi d \sum_j \frac{r_j^2 - r_{j-1}^2}{r_j + r_{j-1}} \quad . \qquad (5.1.5)$$

Expanding the numerator of the right hand term in the right hand side gives

$$v_{jj} = 2\pi d \sum_j \frac{(r_j + r_{j-1})(r_j - r_{j-1})}{r_j + r_{j-1}} \qquad (5.1.6)$$

and upon cancelling, we have

$$v_{jj} = 2\pi d \sum_j (r_j - r_{j-1}) \quad . \qquad (5.1.7)$$

Recalling that d = P/a, where P is total population and 'a' is total area, and substituting into (5.1.3) results in

$$v_{jj} = \frac{2\pi r\,P}{a} \quad . \qquad (5.1.8)$$

Noting that $a = \pi r^2$ we have, as a final result

$$v_{jj} = \frac{2\pi r P}{\pi r^2} = \frac{2P}{r} , \qquad [(5.1.1)]$$

for the potential at the center of a circular region with uniform density.

There are two assumptions to be made in employing Stewart's method of calculating self-potential as it has been outlined above. The first is the assumption of circularity. This is not too severe an assumption since, as Stewart and Warntz (1958, 144) noted "For an irregular area, a, not too different in shape from a circle, the same formula approximately applies, if by r we understand the value determined by setting $\pi r^2 = a$, again even though 'a' be irregular." Moreover it can be noted that for elongated shapes, Stewart and Warntz (1958) suggested that the potential at the center may be approximated by treating the shape as being elliptical, and calculating the appropriate self-potential on this basis. This can be accomplished by using the ratio of the semimajor to semiminor axis of the elliptical region, in conjunction with the table provided by Stewart and Warntz (1958).

The second assumption implicit in the Stewart method of calculating self-potential is that which takes the distribution of population to be uniform over the region. There will obviously be many cases where such an assumption cannot be considered to be valid, and it was in response to this deficiency that Court (1966) suggested an alternative approach. He (1966, 4) wrote

> The potential produced by a county's population at some central point can be computed only if the exact distance of each person from the point is known or assumed. Such information cannot be obtained in any but the simplest cases. Instead, the manner in which the population is distributed over the area must be assumed, or estimated. Such assumptions require mathematical models of the population.

We have already seen, in the previous chapter, that there are a number of possible models for the distribution of population (i.e., the quadratic gamma family, for example). Court (1966) himself investigated eight models (uniform, radially constant, conical, Rayleigh, negative exponential, radial negative exponential, radial linear decrease, and half-normal). For the purposes of brevity of discussion, we shall only consider the negative exponential in further detail here. There are a number of reasons for this. Not only is this probably the best known and most often used of the population density models but it is also derivable from the maximum entropy formalism. Furthermore, as far as this author is aware, the negative exponential has been the only non-uniform population model ever actually used in order to compute self-potentials (see, for example, Warntz et al, 1971; Warntz, 1979).

5.2 Entropy Maximization and Court's Self-Potential

In order to employ Court's (1966) procedure for measuring the self-potential it is necessary to obtain a model for the distribution of population over the region of interest. Bussière and Snickars (1970), and Bussière (1972), have demonstrated that the negative exponential population density model can be derived via the entropy maximizing method.[1] Define a probability

$$p(r,\theta) = \frac{d(r)}{P} \qquad (5.2.1)$$

where d(r) "is a stochastic variable representing the mean surface density at a distance r from the centre" (Bussière, 1972, 89). The term $p(r,\theta)$ can be said to represent the probability of a randomly selected inhabitant of a region with total population P, being located at distance r, angle θ, from the center d_o of the region. It is assumed in this derivation that the region of interest is circular and the distribution of population is symmetrical about the center. Following Bussière (1972) define a normalization constraint

$$\int_0^\infty 2\pi r\, p(r,\theta)\, dr = 1 \qquad (5.2.2)$$

[1] The derivation by Bussière and Snickars is in continuous terms and hence there is no need to employ the method of minimum information in order to take zone sizes into account. If, however, one wishes to derive a population density model for discrete zones, where the zones are of unequal size, then the method of minimum information should be employed (see Batty, 1974a, 1974b, for example).

and an average distance constraint

$$\int_0^\infty 2\pi r \; p(r,\theta) \; r \; \underline{dr} = \bar{r} \qquad (5.2.3)$$

where $\bar{r}$ represents the mean distance of the population from the center. The most likely form for the probability density function $p(r,\theta)$ is one which maximizes the entropy

$$H = -\int_0^\infty 2\pi r \; p(r,\theta)[\ln p(r,\theta)] \; \underline{dr} \qquad (5.2.4)$$

subject to constraints (5.2.2) and (5.2.3). Bussière (1972) shows the result of the maximization to be

$$p(r,\theta) = \frac{b^2}{2\pi} \exp(-br) \qquad (5.2.5)$$

Substitution in (5.2.1) gives

$$d(r) = \frac{b^2 P}{2\pi} \exp(-br) \quad , \qquad (5.2.6)$$

or, in more familiar notation,

$$d(r) = d_o \exp(-br) \quad , \qquad (5.2.7)$$

where

$$d_o = \frac{b^2 P}{2\pi} \qquad (5.2.8)$$

Equations (5.2.6) and (5.2.7) represent the statistically most likely form for the distribution of population around the center of a circular region when the constraint is placed on the mean distance and when the effect of distance is linear with distance. This maximum entropy result can now be employed in order to find the self-potential at the center of the region following the method of Court (1966).[2]

In his discussion of the self-potential associated with the negative exponential population model, Court (1966) presented the population function in rectangular coordinates such that

$$p(x,y) = \frac{b^2}{2\pi} \exp(-b \sqrt{x^2 + y^2}) \qquad (5.2.9)$$

where "the coefficient $b^2/2\pi$ is required for the zeroth moment to be unity; it has been widely interpreted, erroneously, as a parameter independent of b" (Court, 1966, 16). Except for the coordinate system, equation (5.2.9) is identical to the maximum entropy result (5.2.5).

In order to calculate the self-potential at the center of the circular region with a negative exponential distribution of population it is necessary to first define the radial probability density by integrating around the circle and, in the case of equation (5.2.9), transforming to polar coordinates. In this case, equation (5.2.5), or

---

[2] The results and arguments to be presented are not original but are primarily a restatement of Court's (1966) original work. They are included here partly for the sake of completeness, and partly because they were never published and are difficult to find. Professor Court noted (personal communication) that the ditto copy of the paper which he sent to this author was the only one he had remaining. The author gratefully acknowledges Professor Court's assistance.

equation (5.2.9), can be written

$$p(r) = rb^2 \exp(-br) \quad . \qquad (5.2.10)$$

where p(r) is the probability that a randomly chosen individual will be at distance r. This is the radial probability density, obtained by integrating around the circle, and may be visualized "as the result of sweeping a board around the bivariate distribution, piling up along one radius all the population that originally was distributed over the plane" (Court, 1966, 8).

Court argued that if the probability density function p(r) is given, the self-potential at the center v_o (now using polar coordinates) can be found by integrating such that

$$v_o = P \int r^{-1} p(r) \, \underline{d}r \qquad (5.2.11)$$

where "the integration is over the domain of the variable: if p(r) represents a population distribution within a circle of radius a, the integration is from zero to a. If p(r) is a discrete distribution, the integration is equivalent to a summation" (Court, 1966, 9). Conceptually, this procedure is equivalent to that outlined above in the case of the calculation of the potential at the center of a uniform circular distribution.

The self-potential is now obtained by integrating over the radial probability density. Thus, if equation (5.2.10) is substituted for p(r) in equation (5.2.11) we have

$$v_o = P \int_0^\infty r^{-1}\ rb^2 \exp(-br)\ \underline{d}r \quad , \qquad (5.2.12)$$

or

$$v_o = P \int_0^\infty b^2 \exp(-br)\ \underline{d}r \quad . \qquad (5.2.13)$$

Because b^2 is a constant it can be moved outside of the integration

$$v_o = P\ b^2 \int_0^\infty \exp(-br)\ \underline{d}r \quad . \qquad (5.2.14)$$

The integral of exp(-br) is 1/b and hence the result is

$$v_o = Pb \qquad (5.2.15)$$

(see Court, 1966, 17). This is an estimate of the self-potential at the center of a circular region with a monocentric, negative exponential distribution of population, which indicates that the self-potential is equal to the total population times a parameter b. A numerical value for the self-potential cannot be computed, however, until some estimate is obtained for this parameter.

Court (1966) noted that, in the case of the negative exponential model, the percentage of the total population contained in a circle of radius c (where c=r/2) can be calculated from

$$P(c) = 100[1 - (bc + 1) \exp(-bc)] \quad . \qquad (5.2.16)$$

Court (1966, 17) also noted that, given equation (5.2.16), "tables of the incomplete gamma function, or suitable chisquare tables, yield values of bc corresponding to selected values of P(c). For P(c) = 95 percent, bc = 4.745 ..." In other words, if this value for bc is inserted in equation (5.2.16), the result will be

$$P(c) = 100[1 - (4.745 + 1) \exp(-4.745)] = 95\%$$

Having determined this value for bc, Court (1966, 17) concluded that "if the people inside a circle of radius r represent 95 percent of a population having a negative exponential distribution, the self-potential of the entire population, including the five percent outside the circle, is

$$v_o = \frac{4.745\ P}{r} \quad ." \qquad (5.2.17)$$

If it is wished to compare this value with the potential of a uniform distribution (equation 5.1.1), then 4.745/2 = 2.373 and

$$v_o = 2.373\ \frac{2P}{r} \qquad (5.2.18)$$

Thus it can be said that for equally sized, circular regions having the same total populations, the self-potential associated with the assumption of a negative exponential population distribution is 2.373 times greater than the self-potential associated with the assumption of a uniform population distribution. Commenting on these results, Warntz <u>et al</u> (1971, 10) noted that

> This is a substantial difference and affects urban peak potentials considerably. As the negative exponential character of the spatial distribution of urban population densities is by no means irrefutable, such approximations should not be taken too seriously; nonetheless they do provide, we believe, a better approximation than do self-potentials calculated from the uniform density model, even though it has its own difficulties.

The maximum entropy approach does provide some rationale for using the negative exponential population model although, as we have seen, this depends largely on the nature of the constraints which are specified. Nevertheless, it is possible to employ the entropy maximizing formalism in order to derive the form of the population distribution, and then to use the procedures devised by Court (1966) in order to calculate the potential at the center of the distribution. These latter procedures, it can be noted, can be employed in order to calculate the self-potentials associated with forms other than the negative exponential (such as the quadratic gamma family members) though we shall not consider them here.

It can also be pointed out that an alternative solution to the self-potential problem in the context of entropy maximization is simply to consider the results of section 4.2 at a lower level of aggregation. Repeating here for convenience, it was shown that the maximum entropy and statistically most likely form for a function describing potential *per capita* interaction is

$$V_j = \sum_i \left[\frac{P_i \exp[-bf(r_{ij})]}{\sum_j \exp[-bf(r_{ij})]}\right] \qquad [(4.2.33)]$$

If sufficiently disaggregate data are available, there is no reason why this function cannot also be employed to calculate self-potential. Suppose, for example, that it were desired to calculate United States population potentials using equation (4.2.33) at the state level of aggregation. If county level data were also available, it would be possible to employ equation (4.2.33) and such data to calculate the potential at the control point representing the "center" of each state, and to employ this quantity as the self-potential. This approach would, of course, again leave us with the problem of calculating the self-potential of the county in which the control point is located although, as was indicated earlier, the relative size of the contribution is diminished. And obviously, if such disaggregate data were available, the logical approach would be to use them to calculate all of the potentials, rather than just the self-potential. This discussion has been included partly for the sake of completeness *vis a vis* the maximum entropy approach, and partly for the purpose of illustrating that there are alternative methods of calculating the self-potential. The latter point could be important if it were desired to empirically test and compare the various methods of computing self-potential.

Consideration of the types of approaches to dealing with the self-potential problem outlined above is, of course, predicated on the assumption that there is a problem of dividing by zero when $i = j$. Within the quadratic gamma family of functions this situation occurs only for those members containing a power function: (1) the quadratic gamma, (3) the normal gamma, (4) the gamma, and (7) the inverse power. For the remaining three members of the family—(2) the quadratic negative exponential, (5) the negative exponential, and (6) the normal—the problem of dividing by zero is non-existent since they do not contain a power term. The latter three functions contain only exponential terms and at a distance of zero, the exponential is finite, that is, by definition $e^0 = 1$. But although the purely exponential functions avoid the technical problem of dividing by zero, they do introduce what may be a conceptual problem since they will result in a self-potential which is simply equal to the total population size of the region. If this is unacceptable, then it will be necessary to employ one of the methods discussed above.

Finally, it can be noted that there is an additional and alternative approach to the calculation of potential which avoids the problem of the self-potential entirely.

This method, suggested by Goodchild (in conversation), is to place a sufficiently fine grid over the map of interest, and then to calculate the potential at the intersections of the grid. In other words, the summation would be over the i's representing areal centroids in the usual way, but the summation would be at j's defined as the grid intersections (assuming, of course, that no data point coincides with a grid point). Although this approach will be slightly more complicated (in terms of defining the grid), it has a number of advantageous properties. In addition to avoiding the self-potential problem, the grid method allows a choice to be made with respect to the resolution level of the results, provides a uniform distribution of destination control points, and, in light of these properties, can simplify the contouring process (whether automated or manual).

Having considered self-potentials within the context of the entropy maximizing paradigm, we can now go on to discuss some of the issues and problems associated with the calculation of maximum entropy and minimum information potential functions. In particular the point will be made that employing a potential function with an "adjustable" distance response function (that is, with parameters) can result in a potential surface which is virtually identical to the density surface from which it was constructed. The discussion will focus, once again, on the negative exponential potential function (2.6.2), though the arguments to be presented apply to the other members of the quadratic gamma family of potential functions.

5.3 Calculating the Potentials

One fundamental difference between the potential functions derived earlier in the preceding chapter and the Stewart form of potential, is that all of the former contain one or more adjustable parameters while the latter does not. In the derived potential functions, b is a parameter—the slope of the function—which is associated with some mean distance. The introduction of such a parameter in a potential function leads immediately to computational problems.

In the calibration of spatial interaction models there are two basic approaches to estimating the value of the parameter on distance. In the first case, data on the total or mean trip length (or cost) are available and are employed as a calibration target. The calibration process involves the adjustment of b in an iterative procedure until the model distance matches the observed distance (see Hyman, 1969; Baxter, 1973). In the second case, one or more measures of goodness of fit, such as the correlation coefficient or standard deviation, are employed as the calibration target. In this situation, the b parameter (and perhaps others) are adjusted iteratively until some satisfactory degree of fit between observed numbers of trips and model numbers of trips is achieved (see Batty, 1976a). Both of these methods of calibration assume the existence of data concerning the numbers and lengths of observed trips and, because of this, neither can be directly applied to the computation of a potential function. An indirect approach is required in each case and we shall now consider these in greater detail.

Suppose we are interested in constructing a potential surface based on the negative exponential potential function (4.4.22). Some value for b, the slope of the function, is required. The first method of calculation assumes the existence of data on mean trip length. Insofar as a potential surface is constructed from a density surface, such data are never directly available and an indirect approach is required. It is possible, for example, to employ the Stewart form of the function to find an average distance

$$\frac{\sum_i \sum_j v_{ij} \, r_{ij}}{\sum_i \sum_j v_{ij}} = \bar{r} \qquad (5.3.1)$$

With the Stewart form of the potential function it can be noted again that

$$v_{ij} = P_i / r_{ij} \qquad [(1.1.1)]$$

and

$$\sum_{ij}\sum v_{ij} = V \quad . \tag{5.3.2}$$

Hence equation (5.3.1) can be simplified to

$$\frac{\sum_{ij}\sum P_i}{V} = \bar{r} \quad , \tag{5.3.3}$$

or, since summing P_i over both i and j is equivalent to adding the total population to itself n-1 times

$$\frac{P(n-1)}{V} = \bar{r} \quad . \tag{5.3.4}$$

Equation (5.3.4) indicates that if the sum of population, the number of regions, and the sum of potential are known, an average distance of potential *per capita* interaction can be specified (such simplifications are of course not possible for other forms of the distance response function; it should also be noted that this measure is exclusive of the distance associated with self-potentials).

The average distance associated with the Stewart form of potential can then be employed to derive an estimate for b in the negative exponential potential function. In the case of the negative exponential it is normally assumed that the b value equals the inverse of the mean distance (see Brown and Masser, 1978, 67). Thus

$$b = \frac{1}{\bar{r}} \quad , \tag{5.3.5}$$

and for potential this can be given directly as the inverse of equation (5.3.4).

$$b = \frac{V}{P(n-1)} \quad . \qquad (5.3.6)$$

The resulting value for b can then be directly employed to construct a potential surface based on a negative exponential distance response function. This is obviously a very roundabout method of parameter estimation and would probably be useful only when one wishes to directly compare potential surfaces resulting from the Stewart and negative exponential forms having the same mean distance.

The second method of parameter estimation, in the case of spatial interaction modelling, assumes the existence of data on numbers of trips or interactions between regions. Again these data are not available from the density data employed in the construction of a potential surface. Furthermore it can be pointed out that were such data available they would, by definition, represent observed spatial interaction. Hence, it would be more appropriate to employ a spatial interaction model which contains a term representing destination attractiveness. These same arguments would apply to the practice of using interaction data to estimate b in the power function potential as described in section 2.6.

It is nevertheless possible to employ statistical procedures such as least squares in the calculation of potential functions although once again, an indirect approach is required and, as we shall now see, the use of such procedures raises serious questions about the nature and purpose of potential. Suppose that we are interested in the relation between population potential and the value of land. As was suggested in section 2.2, the usual though often unstated hypothesis is that it is expected that such a dependent variable is more highly correlated with population potential than with population density. With the Stewart form of potential, testing this hypothesis is a relatively straightforward matter; indices are constructed, correlations calculated, and conclusions can be drawn. If the hypothesis is supported, arguments can be made to the effect that accessibility, as well as population density, plays a rôle in determining land value. With a potential function with adjustable parameters, however, complications are introduced.

In order to employ a potential function such as the negative exponential in the absence of data on mean distances, it is necessary to fit the function directly to the dependent variable. In the case of equation (4.4.22), it would therefore be possible, for example, to adjust b iteratively until the highest possible correlation was achieved between land values and potentials. This procedure would have to be repeated each time an additional dependent variable is to be examined and it can be suggested that in virtually every case it would be expected that the correlation between potential and the dependent variable will exceed the correlation between the density and the dependent variable as a result of the introduction of an adjustable parameter in the potential function.

As just one example of the effect of such a parameter it can be noted that log population density against log Stewart form of potential gives a Pearson correlation of 0.89 for 1970 United States population data by states (this is identical to the result reported by Goodchild <u>et al</u> (1981) for 1975 data). For an unconstrained negative exponential potential function (equation (2.6.2)), using the average distance defined by equation (5.3.4) as a calculation target, the correlation between population potential and population density drops to 0.73. However if the b parameter is adjusted to maximize the correlation, the coefficient rises, even with a crude method of adjustment (ten percent reductions), to over 0.96. With the use of more refined techniques the correlation between log population density and log population potential would exceed 0.96 when a negative exponential function is employed to calculate the potentials. The introduction of an adjustable slope parameter allows the researcher to increase or decrease the slope of the potential function in order to make the surface which results from it more closely resemble some density surface. This, in turn, leads to problems concerning the empirical meaning of the surface which is the subject of the discussion which follows.

5.4 Interpreting the Results

It was pointed out in Section 2.4 that the conventional Stewart form of a potential surface can be regarded as an averaged or smoothed version of the density surface from which it was constructed. It can further be suggested that it is this very property of a potential surface which distinguishes it from the density surface and which makes it of interest. Such reasoning is, for example, implicit in the hypothesis stated above concerning land values. However, in a potential function with an adjustable slope parameter, the greater or steeper the slope, the more closely the potential surface resembles the density surface. Consider the correlation between population potential and population density described above, as an illustration of this phenomenon. It was reported that the correlation, with more refined methods of calculation, would exceed 0.96. This of course suggests that the negative exponential "potential surface" merely replicates the population density surface; the calculation of potential is, in effect, redundant in such a situation.

The role played by the slope parameter is undoubtedly the major complication introduced by the inclusion of an "adjustable" distance response function in a potential equation. In the situation where the Stewart form of the function is being employed, no analogous problem exists because the potential surface is, by definition, smoother than the density surface. In the situation where the best fit between a potential surface and some dependent variable is achieved with a steeply sloping distance response function however, the implication is that the original density distribution may be a better predictor of the dependent variable than the potential. Stated another way it can be said that the greater the effect of the distance response function, the greater the correspondence between potential and the density variable employed to calculate the potential.

A much more severe version of this problem can occur when a steeply sloping distance response function is employed in combination with a standard measure of self-potential. In was indicated above that the self-potential is usually the dominant term in the potential summation. If such self-potentials are combined with a distance response function which minimizes the contribution from other regions, the resulting potential surface will be very similar to the original density surface, differing, for the most part, only to the extent that the self-potential on a region differs from the population density on that region. Thus if we write a power function potential with the self-potential explicitly included

$$V_j = \sum_i P_i \, r_{ij}^{-b} + v_{jj} \quad , (i \neq j) \qquad (5.4.1)$$

it can be noted that as b becomes large, $\sum_i P_i \, r_{ij}^{-b}$ approaches zero, and the potential approximates the self-potential

$$V_j \simeq v_{jj} \qquad (5.4.2)$$

The only significant difference between the potential surface and the original density surface will be in the degree to which the self-potentials differ from the densities. If we write density as

$$d_j = \frac{P_j}{a_j} \quad , \qquad [(2.2.1)]$$

and self-potential, as defined in the first chapter, as

$$v_{jj} = \frac{2P_j}{\sqrt{\frac{a_j}{\pi}}} = \frac{3.54 \; P_j}{\sqrt{a_j}} \quad , \qquad [(1.1.4)]$$

the similarity becomes obvious; density varies with area while self-potential varies with the square root of area. If we were to employ a constant distance to evaluate self-potential, as has sometimes been the case (see below), the self-potential will simply be a constant proportion of the population size of each region.

The problems associated with slope parameters and self-potentials in the calculation of potential functions have not gone entirely unnoticed by other researchers. Houston (1969) for example, studied the effect of both the b value and the self-potential in a power function market potential for the USSR. Using constant distances r_{jj} to calculate the self-potentials he found, in the case of the following pairs of parameters—b = 1.25, r_{jj} = 1 km; b = 1.5, r_{jj} = 1, 8 km; b = 1.75, r_{jj} = 1, 8, 25 km; and b = 2.0, r_{jj} = 1, 8, 25, and 50 km—that the self-potential was never less than 97 percent of the total market potential. He (1969, 235) concluded that "the findings here prompt the question of how many potential studies, in the quest for higher correlations and oblivious to the combined effect of b and r_{jj}, have managed only to replicate the distribution of the mass variable." Houston (1969) also cites a study by Ray (1965) as an example of this type of occurence, noting that Ray's values for b of 1.42 and r_{jj} of 5 miles resulted, in Houston's own study, in a market potential in which self-potential accounted, on the average, for 97 percent of the total market potential with only one point, out of 128, at less than 90 percent. This would suggest that the correlation between market potential and manufacturing which Ray (1965) found for southern Ontario was, in essence, a correlation between manufacturing and the retail sales data which were employed to construct the market potential.

In addition to these complications which are introduced as a result of allowing the distance response function to vary, there are those which will result from also allowing the population or origin term to vary by adding priors. The sequential approach to adding priors has already been discussed and need not be repeated here. It can be said, however, that allowing the relative magnitude of the origin term to be changed and adjusted, quite simply adds another variable to the analysis. If the goal is to study, say, the correlation between potential and some other quantity, the consideration of a sequence of priors will further complicate the testing procedures. In other words, if we were to first set the origin term equal to P_i and then find the distance response function and b value(s) which gives the highest coefficient of correlation, it would not be appropriate to accept this result as given and then go on to consider a sequence of priors independently. Rather, each time a new prior was added in the attempt to further improve the goodness-of-fit, it would also be necessary to once again evaluate all of the distance response functions since some alternative distance response function, or some alternative b value(s), might now provide better results.

In conclusion it can be said that there are a number of problems associated with the calculation of potential functions where the distance response function contains an adjustable slope parameter and where alternative priors can be added. Such problems would be further compounded for functions with more than one parameter (a discussion of parameter space search in the case of more than one parameter is contained in Batty, 1976a). It has been shown that not only are there practical problems, but also that there is a very real danger of constructing mathematically fallacious and redundant potentials. This is not to argue that the use of alternative distance functions and sequences of priors is considered to be inappropriate. Rather the purpose is to suggest that as much attention should be given to the correlation between the two densities as to the correlation between the dependent density variable and the potential.

Before proceeding, it can also be noted that although the introduction of additional constraints and hence additional parameters will normally improve the fit of any particular function, this not only implies the use of more complex nonlinear fitting procedures, increased computational costs, and reduced efficiency, but also results in the loss of one degree of freedom in the estimation for each additional parameter introduced (see Andrulis, 1981, 240). Moreover, as Zielinski (1980, 144) has noted, "with each parameter added, a model inevitably gains in accuracy and loses in meaning."

5.5 Summary

A summary of some of the important properties of potential, and potential functions, as discussed throughout this study, and particularly in the present chapter, is as follows:

1. Potential describes spatial interaction only to the extent that the interaction can be considered to be on a per capita basis at the destination.
2. Potential is an index of aggregate accessibility; the usual hypothesis is that it will correlate strongly with variables which can be expected to vary as a function of accessibility.
3. Constraining the origin term in a potential function has an effect on the resulting form of the potential surface; all other things being equal, an origin constrained function will produce a "flatter" surface than an unconstrained function.
4. The number and/or configuration of control points can affect the nature of a potential surface; the fewer or less dense the points, the smoother the surface.

5. The self-potential can, depending on the level of spatial aggregation, be the dominant term in the sum; in such cases this suggests that it can surreptitiously play the role of a destination term.
6. Potential is an intensive variable; it is therefore necessary to multiply by the area of the destination region when calculating the amount of potential outgoing from a point.
7. In the calculation of a potential equation which has a slope parameter in the distance response function, complications result from the use of density rather than interaction data; in particular, an additional density variable or data on mean distances are required as calculation targets.
8. Steeply sloping distance response functions can result in redundant potential surfaces which replicate the density surfaces from which they were constructed; this problem is exacerbated when the self-potential is included in the sum.
9. The definition of potential as an index of aggregate accessibility implies that a potential surface should be "smoother" than the density surface from which it was constructed; this will not necessarily be true of functions with adjustable slope parameters but is, by definition, true of the Stewart form of potential.

Given this discussion of some of the practical problems associated with the empirical use of the derived potential functions, we can now go on to consider some of the more general issues which result from the joint consideration of entropy maximization, information minimization, and potential.

CHAPTER 6

REVIEW AND DISCUSSION

6.1 Summary of Results

Throughout this study entropy maximization and information minimization have been treated and discussed as if they were separate entities. This distinction is somewhat artificial however and has been maintained primarily because it corresponds with the historical evolution of the methods in geography. It was demonstrated that maximum entropy is in fact a special case of minimum information and it follows from this that this study could have been organized principally around the method of minimum information, with maximum entropy treated as a special case.

It can be said, therefore, that the study undertook to derive probability distribution estimates, and trip models, for potentials by minimizing the Kullback (1959) information (now renumbering for convenience)

$$I(q:p) = \sum_{i}\sum_{j} p_{ij} \ln \frac{p_{ij}}{\prod_{m} q_{ij}^{m}} \qquad (6.1.1)$$

where p_{ij} is the posterior probability of spatial interaction between zones i and j and q_{ij} is the prior probability of spatial interaction between i and j. The minimization was carried out subject to constraints of the form

$$\sum_j p_{ij} = \frac{P_i}{P} \quad , \qquad (6.1.2)$$

$$\sum_i \sum_j p_{ij} \; f(r_{ij}) = \bar{R} \quad , \qquad (6.1.3)$$

and

$$q_{ij} = (i, j = 1, \ldots, n) \text{ known} \quad . \qquad (6.1.4)$$

The result of such a minimization, after following the intermediate steps discussed in Chapter 4, was shown to be a potential function

$$V_j = \sum_i A_i \; P_i \; \prod_m q_{ij}^m \; \exp[-bf(r_{ij})] \quad , \qquad (6.1.5)$$

where

$$A_i = \frac{1}{\sum_j \prod_m q_{ij}^m \; \exp[-bf(r_{ij})]} \quad ; \qquad (6.1.6)$$

This is a generalized potential equation which allows for changes in the effect of distance and in the emissivity of the originating zones. The form of the distance response function $\exp[-bf(r_{ij})]$ depends on the form of the constraint(s) placed on distance in equation (6.1.3). The population term P_i can be weighted, or redefined, by making the appropriate choice of the prior probability distribution q_{ij} in constraint (6.1.4).

If the prior is uniform, or equivalently, if there is no relevant prior information available, then probability distribution estimates can be gotten by maximizing the Shannon entropy

$$H = -\sum_{ij} p_{ij} \ln p_{ij} \quad , \qquad (6.1.7)$$

subject only to constraints (6.1.2) and (6.1.3). This results first, in a trip model, and then in a potential function

$$V_j = \sum_i A_i P_i \exp[-bf(r_{ij})] \quad , \qquad (6.1.8)$$

where

$$A_i = \frac{1}{\sum_j \exp[-bf(r_{ij})]} \qquad (6.1.9)$$

Given this brief summary of results we will now proceed, in the remaining sections, to consider some of the attendant problems and issues which arise in the context of minimum information, maximum entropy and potential.

6.2 Critical Retrospective

In looking back, perhaps the most appropriate thing that can be said (to paraphrase Wilson (1970, 69)) is that the concept of potential does not easily lend itself to entropy maximizing or information minimizing assumptions. We have seen that the constraints normally employed in entropy and information modelling present problems when considered within the context of potential and we have seen that this, in turn, leads to pragmatic difficulties in the use and operationalization of the resulting potential functions. There were two general sorts of problems, both closely related to two of the three types of constraints considered.

Distance Constraints — The various distance response functions which can be derived for potential equations by the methods of maximum entropy and minimum information were shown to result from constraints being placed on the mean distance. It was argued that actual data concerning average distances need not be known a priori for this information to be included in the maximization or minimization. This was fortunate since such data are generally not available for potentials; the average distance of possible per capita spatial interaction is not known a priori but is an a posteriori result. It follows from this that the potentials cannot be calculated so as to satisfy some known average distance, as is the case with trip distribution models. Rather, they must be calculated so as to maximize or minimize some measure of fit between the potential and some other variable. This means that the derived potential equations cannot be employed to construct potential surfaces unless some additional variable is employed as a calculation target or unless the b values in the distance response functions are arbitrarily set.

This is an important property of the types of potential equations which have been presented here (other than the Stewart form) yet as far as this author is aware it has never been discussed critically in the literature. Many authors, for example, have proposed negative exponential potentials but none have discussed extensively the problems of calculation and interpretation which this entails. Dalvi and Martin (1976), for example, constructed accessibility indicies for London using a negative exponential function, where the value for b was gotten from a transportation study done at a different point in time and at a different level of aggregation. They (1976, 40) noted the inconsistencies of such an approach, but used the values anyway. Similarly, Weibull (1976, 372) employed travel time to work in Stockholm, apparently also gotten from an earlier study, in order to estimate accessibility. Vickerman (1974, 683) introduced a negative exponential potential equation, but then used r_{ij} and r_{ij}^2 in his actual calculations, noting that "limitations in available computing time prevented the consideration of "alternative functional forms such as the exponential, $\exp(-br_{ij})$, in which the constant b has to be calibrated." Finally, Ingram (1970) solved the problem of having to estimate b by calculating potential as

$$V_j = \sum_i P_i e^{-r_{ij}} , \qquad (6.2.1)$$

where the b was simply omitted (see also equation 2.6.3).

These few examples serve to point out that the problem of dealing with parameters on distance in potential and accessibility formulations has generally been solved either by borrowing values from transportation studies and interaction models or by ignoring it. As was suggested in the second chapter, these problems seem to arise from the borrowing of distance response functions from trip distribution models. The consideration of potential within the entropy maximizing and information minimizing formalisms has demonstrated explicitly why this borrowing leads to problems; in the case of a trip distribution model the mean distance is known _a priori_, whereas in the case of potentials it is necessary to _assume_, for the purposes of deriving the functions, that a mean distance is known _a priori_, when in fact it is an _a posteriori_ result.

There is no simple solution to the problem of dealing with parameters on distance in potential and accessibility indices, except to say that a variety of approaches are available. If, for example, it were desired to construct a potential surface using a negative exponential distance response function, yet the only available estimate for b was based on a transportation study from some other time period (or at a different level of aggregation) then it probably would be preferable to use the Stewart form instead. If, on the other hand, it were desired to employ retail potential, say, as the destination term in an origin-constrained shopping model, then it would be possible to test a variety of distance response functions for such a potential by including the tests as a part of the overall shopping model calibration. Similarly, if the interest is in the correlation between potential and other variables, the testing of alternative functions and the calculation of b parameters presents no extraordinary problems. The calculation of multiple parameter distance response functions is difficult, but is no more difficult for potentials than it is for trip distribution models (see Openshaw and Connolly, 1977).

Prior Probability Distributions — Another type of constraint with which it was necessary to deal in the entropy maximizing/information minimizing analysis of potentials were those having to do with prior probability distributions. This presented no serious problems, but it was shown that many of the types of priors which are normally of relevance in spatial models are not permissible choices within the context of the type of potentials being discussed here. This demonstrated once again that the concept of potential does not easily lend itself to the assumptions of maximum entropy and minimum information.

If nothing else, it can be said that this study has illustrated, by negation, some of the basic properties of, assumptions in, and requirements of, the entropy maximizing and information minimizing methods. The consideration of potentials within these frameworks has forced us to examine, in detail, assumptions and constraints which are usually accepted and employed without question.

In considering the discussion in this and the previous chapter, the argument could be made that, at least for practical purposes, the Stewart form of potential may be preferable as a general index of aggregate accessibility. It is, for the most part, a simple and unambiguous index, and provides results with the desired property of smoothness. There are no parameters to be adjusted and therefore there is no danger of constructing a redundant potential surface. In addition, a surface can be constructed with it in the absence of additional data on mean distances or densities.

On the other hand, there are also advantages which can be attributed to the other forms of potential, provided sufficient precautions are taken so as not to unknowingly construct a redundant potential surface. Much better fits could normally be expected when the primary interest of the researcher is in the type of correlation studies that have been discussed here. In addition, it can be noted that although a potential surface constructed with an adjustable slope parameter can be "less smooth" than a Stewart surface, it can also be "more smooth". Thus if some variable is very highly correlated with accessibility, potentials constructed on the basis of the Stewart index may fail to adequately capture the relationship; an even smoother surface may be required.

All things considered it is difficult, if not impossible, to determine which form of potential is preferable in any particular situation without knowing the nature and purpose of the research. If it is desired, for example, simply to illustrate graphically the impact of a new town on a region, the simple Stewart form of potential is no doubt adequate. If however the intention is to study the correlation between potential and other dependent variables, a variety of distance response functions, including simple Euclidean distance, should probably be tested, along with a sequence of alternative priors.

The potential functions presented in this study were derived via the methods of minimum information and maximum entropy. There are still a few general issues which remain to be considered, concerning functions or models derived from these formalisms. Some of these issues were anticipated in earlier chapters. We will see that all of them, in one way or another, again have to do with constraints.

6.3 Some Related Issues

In describing the method of maximum entropy Jaynes (1968, 232) stated that

> If the information incorporated into the maximum-entropy analysis, includes all the constraints actually operative in the random experiment, then the distribution predicted by maximum entropy is overwhelmingly the most likely to be observed experimentally, because it can be realized in overwhelmingly the greatest number of ways.

The question of whether all of the relevant information has been included in the constraints is an important and difficult one, and criticisms have been levelled at the entropy formalism precisely for this reason. In particular it has been argued that if information concerning some mean is available, then information about related measures, such as the variance, also will be available and should be included. As Sheppard (1976, 747) argued, in a discussion of trip distribution modelling,

> if the expected cost of a trip is known then it is likely that this has been calculated from the distribution of trip costs for all journeys, so that other moments should be available. These should also be included as prior information, since travel behaviour depends just as much on them as on the mean.

Sheppard (1976, 747) went on to suggest that if such information is included in the maximization, the solution

> no longer even superficially resembles the gravity model in the traditional sense, since it involves many other parameters. This observation reinforces the argument that maximum entropy should not be used and justified because it produces identifiable results, such as the gravity model; it has the quite independent justification as an inferential technique.

The issue of whether all of the relevant information has been included in the constraints is really of empirical significance only when the maximum entropy/minimum information prediction fails to agree with the empirically observed distribution (in the case of potential this will be some surrogate variable representing accessibility, such as land value). If such a disagreement occurs it can be said, following Jaynes (1968, 232) that "the observed deviations then provide a clue as to the nature of the new constraints" and it is for this reason that Jaynes went on to argue that "the principle of maximum entropy is most useful to us in just those cases where it fails to predict the correct experimental facts." This line of reasoning is, of course, very similar to the method of sequential information minimizing discussed earlier (see Webber, 1979, 131).

Another important and closely related issue concerns the problem of how to explain, interpret, and provide a rationale for the constraints which are employed. The sequential approach to information minimizing provides a method for determining which constraints improve predictions, but it still remains necessary to determine empirically which are important and to interpret the results. In discussing the tendency to concentrate on deductive model generation within the minimum information framework, Batty (1981, 148), for example, made the statement that:

> I believe that all the really exciting challenges in this field now lie in showing how constraints might be identified from data, in the meaning of different levels of spatial information, in parameter invariance and interpretations, and so on. Indeed it appears that the inverse problem, that of determining what constraints are important, rather than what model will result, represents the real frontier.

Similarly, Goodchild (1979, 88) noted that

> the solution to an information-minimizing or entropy-maximizing strategy can be regarded as a null hypothesis, the most likely state of the system given the constraints. Thus it is the deviations from the maximum entropy solution which are interesting, since they represent the effects of additional constraints and unequal probabilities which the system imposes, in other words information not previously known to the researcher. A good fit on the other hand indicates that there is nothing in the system unknown to the researcher. How good then is the fit of [the] model? This is clearly the most important topic for any continued research.

Webber (1979, 140) also anticipated this in a discussion of the sequential approach to minimum information spatial modelling, when he stated that

> the paradigm represents a method whereby the constraints which operate upon a social system are inferred from given data, and more elementary theory must then be developed to explain the origins of the constraints—to explain why it is that the expectations of particular variables take on given values.

These arguments apply to constraints in the form of priors added, as well as the various distance constraints which can be employed. In the discussion of the calculation of minimum information potentials, it was noted that each time a new prior is added it is also necessary to re-evaluate all of the distance response functions and parameters on distance. Thus it could be said that the "sequential" approach to evaluating constraints, when considered within the context of both priors and distance constraints, is more of a two-stage, iterative procedure, alternating back and forth between priors and constraints on distance.

A further issue concerning functions and models derived via the method of minimum information involves redundancy in the constraints. In particular it would appear that there are situations wherein the information contained in one or more priors is already contained in the constraint on the mean. The mean distance employed in calibrating a location model, for example, is normally an observed mean. If the characteristics of the destinations, such as the number of houses there, have an effect on the number of individuals travelling to them, then this will be reflected in the actual number of trips taking place and, therefore, in the observed mean distance. The observed mean, in other words, already contains information concerning the effect of the characteristics of the destinations on the number of trips; to minimize the Kullback information subject to a constraint on the observed mean distance and a prior based on the properties of the destinations is, in such a situation, redundant. That this is so can be demonstrated within the context of the simple assignment problem example which was employed in the third chapter.

In that example, the maximum entropy assignment of workers was three to the second zone and one to the third zone, a distribution in which the average distance travelled was 1.25 miles. The minimum information solution assigned two workers to each of the two zones and this resulted in an average distance travelled of 1.50 miles. The reason for the difference in the assignments was that in the minimum information case, probabilities relating to the areas of destination zones two and three (0.3 and 0.7 respectively) were taken into account. Suppose, however, in the case of the maximum entropy model, that it had been known a priori, as a matter of empirical observation, that the mean distance travelled was not 1.25 miles but was 1.50 miles. Then maximizing the entropy subject only to normalization and distance constraints, where

$$\sum_i p_i r_i = 1.50 \quad ,$$

would have led to the desired result of two workers assigned to each zone (that is, if this distance constraint is satisfied exactly, the distribution of workers must be two in each zone—see section 3.4). If the mean distance travelled is based on observed trips, then those trips will reflect empirically the fact that the zones differ in size. To minimize the information statistic subject to both a mean distance of 1.50 and a prior based on zone size, is redundant since the same information is being included twice in the minimization. It should be pointed out, however, that this procedure will not normally lead to incorrect numerical results.

In the maximum entropy model, b will be adjusted iteratively until the model mean distance equals the observed mean distance. Two workers will be assigned to each zone and the model mean distance will be 1.50. In the minimum information model, the same procedure will be followed and the same results obtained, except that the resulting value for b will now be lower since it has to "absorb" the extra trips generated by including the areas of destinations. For example, in the entropy model (equations 3.4.9 and 3.4.10) the results will be

$$T_{12} = \frac{4 \cdot \exp(-0.0 \cdot 1)}{\exp(-0.0 \cdot 1) + \exp(-0.0 \cdot 2)} = 2.00 ,$$

and

$$T_{13} = \frac{4 \cdot \exp(-0.0 \cdot 2)}{\exp(-0.0 \cdot 1) + \exp(-0.0 \cdot 2)} = 2.00 ,$$

where the value for b of 0.00 is no longer the inverse of the mean distance, but is chosen (for illustrative purposes) to provide a solution which is equal to the desired integer result. In order to achieve the same integer result using the minimum information model (equation 3.5.11) it is necessary to adjust the value for b and to recalculate (rounding to two decimals) as follows

$$T_{12} = \frac{0.3 \cdot 4 \cdot \exp(-0.85 \cdot 1)}{[0.3 \cdot \exp(-0.85 \cdot 1)] + [0.7 \cdot \exp(-0.85 \cdot 2)]}$$
$$= 2.00$$

and

$$T_{13} = \frac{0.7 \cdot 4 \cdot \exp(-0.85 \cdot 2)}{[0.3 \cdot \exp(-0.85 \cdot 1)] + [0.7 \cdot \exp(-0.85 \cdot 2)]}$$
$$= 2.00 .$$

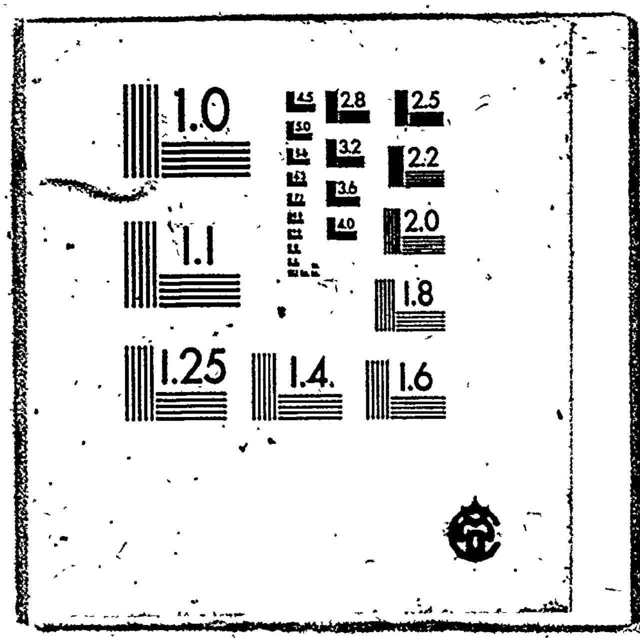
In the minimum information model the b value of -0.85 is lower than the previous entropy model value of 0.00; this is appropriate since the slope of the distance response function must be steeper in the minimum information model in order to offset the effect of the zone sizes.

The fact that b = 0.00 in the maximum entropy model also demonstrates, in another way, how constraints can be redundant. If the $\bar{r}$ value employed as a calibration target is that associated with a uniform distribution, this provides no additional information and the effect of distance drops out of the problem; the assignment is made simply in direct proportion to the sizes of the zones which, in the case of the entropy formalism, are assumed to be equal.

The problem of redundancy in constraints and priors in minimum information functions and models is easily handled within the sequential method of evaluation. If redundant priors are added, they will not improve predictions, and hence are to be dropped anyway. If a prior is added which is suspected to be redundant, but turns out to improve the predictions, then some additional information has been added (for example, that observed trips do not take place in direct proportion to the areas of zones). What is perhaps the most important aspect of the problem is that it does demonstrate that it is necessary to re-evaluate distance response functions and parameters on distance each time a prior is added. If this is not undertaken, then redundant priors can enter the analysis by erroneously improving predictions. In the example above, for instance, if the model results were first calculated with respect to the mean distance and *then* the prior was added, without a re-evaluation of b, the same information would have been included twice in the analysis and the results would have been different. In general it can be said that within the sequential, minimum information approach, it is necessary to ensure at every step of the way, that *all* constraints are being satisfied. This complicates the empirical evaluation of such models and functions, but is absolutely necessary if the results are to be considered valid.

Within the context of potentials the same general arguments concerning redundancy in priors and constraints hold true. However, since the properties of the destination do not effect the amount of *per capita* interaction, and since the mean distance associated with that interaction is an *a posteriori* result, the problem of including such information does not arise. Rather, priors in potentials will be redundant when different properties of the originating zones contain overlapping information, for example, when the proportion of telephones or automobiles corresponds with the relative population size. Such redundancies can be handled by sequentially adding

3 OF / DE 3



and evaluating priors or, alternatively, by studying the correlations between the priors and the population beforehand as a basis for making the decision whether to include a particular prior. There will be similar redundancies with priors based on "between-zone" measures, such as the number of roads, and number of rail lines; these can be treated in the same manner.

Has all of the relevant information been included in the constraints? What is the meaning of the constraints that have been included? How are the constraints related to one another? These are the principal questions which have been addressed briefly in the present discussion; there are, no doubt, many other similar and equally significant ones which can be raised about the methods of minimum information and maximum entropy. Even though this discussion has touched on only a few such questions, it may be said to have substantiated once again Batty's (1981, 148) claim that it is the issue of the constraints, rather than what model will result, which represents "the real frontier."

When the method of maximum entropy was introduced to geography by Wilson in 1967 (although see Curry (1964)) claims were made to the effect that a theoretical basis for the gravity model had been found. It is this issue, now considered with respect to potentials, which forms the subject of the next and concluding section.

6.4 Conclusion

It was stated at the outset that the goal of this study was to consider the derivation of potential functions within the context of the entropy maximizing and information minimizing formalisms. Along the way it was necessary first to review the origins, meaning, and properties of potential functions, and then to introduce and describe the maximum entropy and minimum information techniques. In the fourth chapter entropy maximization, information minimization, and potential were considered jointly and, in the fifth chapter, a number of peripherally related but important issues were discussed.

Except for a brief mention at the end of the third chapter, the issue of whether the maximum entropy and minimum information approaches provide a theoretical foundation for the functions derived from them has, for the most part, been ignored. This issue may be largely irrelevant in the context of the present study since it has been repeatedly argued that a potential function is an index, not a model, and one wonders whether an index requires a theoretical foundation. This may be seen to be analogous, for example, to asking whether the closely related measure of population density required a theoretical basis.

It can be suggested that exaggerated and perhaps overly enthusiastic claims accompanied the introduction of the entropy maximizing formalism in spatial analysis. Wilson (1967, 256), for example, claimed that the technique "does offer a sound theoretical basis for the gravity model" while Gould (1972, 691), (in a discussion of the absence of normalization constraints in the traditional gravity model) argued that Wilson's work "raises the gravity model phoenixlike from the ashes of such absurdity, and places it on a secure theoretical foundation for the first time." More recent evaluations of the method have been more restrained and/or critical. Sheppard (1976, 747), for example, has stated that "a theoretical basis for the gravity model cannot be suggested simply because a gravity-like solution can be reached by the specification of certain prior information" while Beckmann (quoted in Webber, 1976, 290) claimed that the proponents of maximum entropy "have become skilled virtuosos in subjecting every problem to this approach and inventing *ad hoc* constraints to make entropy fit the case." It would seem obvious that the usefulness and validity of the approach lies somewhere in the middle ground between the two extremes.

The question of whether the maximum entropy and minimum information approaches provide a theoretical basis for the potential concept would probably not arise at all, were it not for the connection between potential functions and gravity models, as outlined in the second chapter. In other words, it has traditionally been argued that the gravity model needs a sound theoretical base and, since a potential function can be derived from the gravity model, the question of theory seems to carry over automatically. Potential, as it has been defined in this study, is an index, and as such it can be argued that it does not require a theoretical rationale. Potential is, however, a weighted index, and it can also be argued that it is in the weighting of the distance and origin terms that the entropy maximizing and information minimizing approaches play a valid and important role within the context of potential; not as a theoretical justification, but as tools or concepts which allow the connection between the form of the potential function and the constraints on potential *per capita* spatial interaction, to be made explicit.

In reviewing the many alternative forms which were suggested for potential functions in the second chapter (section 2.6), it was argued that many of the distance response functions were simply borrowed from the field of spatial interaction modelling with little or no regard to their meaning and interpretation. The derivation of potential functions within the entropy maximizing framework can be said to be a step toward overcoming this deficiency in the sense that it establishes a mathematical tie between the actual or perceived effect of distance on potential *per capita* interaction and the appropriate form of the distance response function. Stated another way it can be said that maximum entropy may not provide a theoretical foundation for potential but it does provide a mathematical and statistical rationale for the form of the distance response function by establishing a connection between it and the constraints. This can be said to be of significance since, in effect, it changes the nature of the problem of measuring and defining accessibility or locational attractiveness. To paraphrase Amson (1972a, 165), who was discussing urban population distributions: The problem of determining the form of a potential function has been replaced by a quite different one: that of determining the effect of distance. To know the latter is then to know the former.

Similarly, the method of minimum information provides a basis not only for deriving and interpreting distance response functions but also for weighting and redefining the origin term. This also establishes a connection between the constraints which are deemed to be operative in the system of interest and the resulting form of the potential function. Minimum information, in other words, provides a method for tying together certain prior information about the origin zones, expressed as prior probabilities, and the effect of the zones on *per capita* spatial interaction. In addition, it was suggested that "between-zone" characteristics, other than distance, can also be incorporated through the method of minimum information.

In short it can be said that the method of maximum entropy, and the more general method of minimum information, have their strengths as well as their weaknesses. We have seen that they can be employed to derive generalized models and functions and, in particular, to derive generic potential functions with many possible specific forms. We have also seen that the methods provide an organizing framework within which such functions and models, and the constraints associated with them can be handled. On the other hand it has been argued that there are a number of problems which can be raised, and criticisms which can be levelled, concerning the choice, use, and interpretation of constraints. This has been particularly evident in the present study wherein the attempt has been to treat a spatial index with tools normally intended to be employed for spatial model building; the result was a demonstration of the demands imposed by the methods of maximum entropy and minimum information in terms of the types of data and information they require, and in terms of the assumptions associated with them.

REFERENCES CITED

Amson, J.C. (1972a). "The Dependence of Population Distribution on Location Costs." Environment and Planning, Vol. 4, pp. 163-181.

Amson, J.C. (1972b). "Equilibrium Models of Cities: 1. An Axiomatic Theory." Environment and Planning, Vol. 4, pp. 429-444.

Anderson, J.R. (1955). "Intermetropolitan Migration: A Comparison of the Hypotheses of Zipf and Stouffer." American Sociological Review, Vol. 20, pp. 287-291.

Andrulis, J. (1981). "Urban Population Density: A Comment." Environment and Planning A, Vol. 13, pp. 239-241.

Angel, S., and G.M. Hyman (1976). Urban Fields. London: Pion.

Batty, M. (1972). "Entropy and Spatial Geometry." Area, Vol. 4, pp. 230-236.

Batty, M. (1974a). "Spatial Entropy." Geographical Analysis, Vol. 6, pp. 1-31.

Batty, M. (1974b). "Urban Density and Entropy Functions." Geographical Paper No. 36, Department of Geography, University of Reading.

Batty, M. (1976a). Urban Modelling: Algorithms, Calibrations, Predictions. Cambridge: Cambridge University Press.

Batty, M. (1976b). "Entropy in Spatial Aggregation." Geographical Analysis, Vol. 8, pp. 1-21.

Batty, M. (1978). "Speculations on an Information Theoretic Approach to Spatial Representation." In Spatial Representation and Spatial Interaction, edited by I. Masser and P. Brown, pp. 115-147. Leiden: Martinus Nijhoff.

Batty, M. (1981). Review of M.J. Webber, Information Theory and Urban Spatial Structure. Regional Studies, Vol. 15, pp. 147-148.

Batty, M., and S. Mackie (1972). "The Calibration of Gravity, Entropy, and Related Models of Spatial Interaction." Environment and Planning, Vol. 4, pp. 205-233.

Baxter, R.S. (1973). "Entropy Maximizing Models of Spatial Interaction." Computer Applications, Vol. 1, pp. 57-83.

Beckmann, M., and T.F. Golob (1972). "A Critique of Entropy and Gravity in Travel Forecasting." In Traffic Flow and Transportation, edited by G. Newell. New York: American Elsevier.

Berry, B.J., and P.J. Schwind (1969). "Information and Entropy in Migrant Flows." Geographical Analysis, Vol. 1, pp. 5-14.

Broadbent, T.A. (1970). "Notes on the Design of Operational Models." Environment and Planning, Vol. 2, pp. 469-476.

Brown, P.J.B., and I. Masser (1978). "An Empirical Investigation of the Use of Broadbent's Rule in Spatial System Design." In Spatial Representation and Spatial Interaction, edited by I. Masser and P. Brown, pp. 51-69. Leiden: Martinus Nijhoff.

Bussière, R. (1972). "Static and Dynamic Characteristics of the Negative Exponential Model of Urban Population Distributions." In Patterns and Processes in Urban and Regional Systems, edited by A.G. Wilson, pp. 83-113. London: Pion.

Bussière, R., and F. Snickars (1970). "Derivation of the Negative Exponential Model by an Entropy Maximising Method." Environment and Planning, Vol. 2, pp. 295-301.

Carroll, J.D. (1955). "Spatial Interaction and the Urban-Metropolitan Description." Papers and Proceedings of the Regional Science Association, Vol. 1., pp. 1-14.

Carrothers, G.A.P. (1956). "An Historical Review of the Gravity and Potential Concepts of Human Interaction." Journal of the American Institute of Planners, Vol. 22, pp. 94-102.

Cesario, F.J. (1975). "A Primer on Entropy Modelling." Journal of the American Institute of Planners, Vol. 41, pp. 40-48.

Cesario, F.J. (1979). "Much Ado about Entropy." Geographical Analysis, Vol. 11, pp. 189-196.

Chapman, G.P. (1970). "The Application of Information Theory to the Analysis of Population Distributions in Space." Economic Geography, Vol. 46, pp. 317-331.

Chapman, G.P. (1973). "The Spatial Organization of the Population of the United States and England and Wales." Economic Geography, Vol. 49, pp. 325-343.

Chapman, G.P. (1977). Human and Environmental Systems: A Geographer's Appraisal. London: Academic Press.

Charnes, A., W.M. Raike, and C.O. Bettinger (1972). "An Extremal and Information-Theoretic Characterization of Some Interzonal Transfer Models." Socio-Economic Planning Sciences, Vol. 6, pp. 531-537.

Charnes, A., K.E. Haynes and F.Y. Phillips (1976). "A 'Generalized Distance' Estimation Procedure for Intra-Urban Interaction." Geographical Analysis, Vol. 8, pp. 289-294.

Clark, C., F. Wilson, and J. Bradley (1969). "Industrial Location and Economic Potential in Western Europe." Regional Studies, Vol. 3, pp. 197-212.

Coffey, W. (1977). "A Macroscopic Analysis of Income Regions in Metropolitan Boston." Professional Geographer, Vol. 29, pp. 40-46.

Coffey, W. (1979). Geography: Towards a General Spatial Systems Approach. Doctoral Dissertation, Department of Geography, University of Western Ontario (in press).

Court, A. (1966). "Population Distributions and Self-Potentials." Unpublished manuscript, 47 pp. Presented at the Seminar on Properties of Spatial Series, University of Michigan, Ann Arbor, May, 1966.

Craig, J. (1972). "Population Potential and Population Density." Area, Vol. 4, pp. 10-12.

Curry, L. (1964). "The Random Spatial Economy: An Exploration in Settlement Theory." Annals of the Association of American Geographers, Vol. 54, pp. 138-146.

Curry, L. (1972). "Spatial Entropy." In International Geography 1972, edited by W.P. Adams and F.M. Helleiner, pp. 899-901. Toronto: University of Toronto Press.

Dalvi, M.Q., and K.M. Martin (1976). "The Measurement of Accessibility: Some Preliminary Results." Transportation, Vol. 5, pp. 17-42.

Dinkel, J.J., G.A. Kochenberger, and S-N Wong (1977). "Entropy Maximization and Geometric Programming." Environment and Planning A, Vol. 9, pp. 419-427.

Dutton, G.H. (1970). "Macroscopic Aspects of Metropolitan Evolution." Harvard Papers in Theoretical Geography, Paper No. 1, Laboratory for Computer Graphics and Spatial Analysis, Harvard University.

Echenique, M., D. Crowther and W. Lindsay (1969). "A Spatial Model of Urban Stock and Activity." Regional Studies, Vol. 3, pp. 281-312.

Fein, E. (1970). "Demography and Thermodynamics." American Journal of Physics, Vol. 38, pp. 1373-1379.

Fisch, O. (1977). "On the Utility of Entropy Maximization." Geographical Analysis, Vol. 9, pp. 79-84.

Garrison, C.B., and A.S. Paulson (1973). "An Entropy Measure of the Geographic Concentration of Economic Activity." Economic Geography, Vol. 49, pp. 319-324.

Getis, A., and B. Boots (1978). Models of Spatial Processes. London: Cambridge University Press.

Goodchild, M.F. (1979). "Commentary: Current Issues in Interaction." Ontario Geography, No. 13, pp. 85-89.

Goodchild, M.F., R.F. Milliff, and S.M. Davis (1981). "The Significance of Potential-Density Regressions." Professional Geographer, Vol. 33, pp. 341-349.

Gould, P.R. (1972). "Pedagogic Review: Entropy in Urban and Regional Modelling." Annals of the Association of American Geographers, Vol. 62, pp. 689-700.

Gurevich, B.L. (1969). "Geographical Differentiation and its Measures in a Discrete System." Soviet Geography, Vol. 10, pp. 387-413.

Hansen, S. (1972). "Utility, Accessibility, and Entropy in Spatial Modelling." Swedish Journal of Economics, Vol. 74, pp. 35-44.

Hansen, W.G. (1959). "How Accessibility Shapes Land Use." Journal of the American Institute of Planners, Vol. 25, pp. 73-76.

Harris, C.D. (1954). "The Market as a Factor in the Localization of Industry in the United States." Annals of the Association of American Geographers, Vol. 44, pp. 315-348.

Haynes, K.E., and W.J. Enders (1975). "Distance, Direction, and Entropy in the Evolution of a Settlement Pattern." Economic Geography, Vol. 51, pp. 357-365.

Haynes, K.E., F.Y. Phillips, and J.W. Mohrfeld (1980). "The Entropies: Some Roots of Ambiguity." Socio-Economic Planning Sciences, Vol. 14, pp. 137-145.

Hobson, A., and B. Cheng (1973). "A Comparison of the Shannon and Kullback Information Measures." Journal of Statistical Physics, Vol. 7, pp. 301-310.

Houston, C. (1969). "Market Potential and Potential Transportation Costs: An Evaluation of the Concepts and their Surface Patterns in the U.S.S.R." Canadian Geographer, Vol. 13, pp. 216-236.

Hyman, G.M. (1969). "The Calibration of Trip Distribution Models." Environment and Planning, Vol. 1, pp. 105-112.

Ingram, D.R. (1970). "The Concept of Accessibility: A Search for an Operational Form." Regional Studies, Vol. 5, pp. 101-107.

Inhaber, H. (1975). "Distribution of World Science." Geoforum, Vol. 6, pp. 231-236.

Inhaber, H., and K. Przednowek (1974). "Distribution of Canadian Science." Geoforum, Vol. 5, pp. 45-53.

Isard, W., et al (1960). Methods of Regional Analysis. Cambridge, Mass.: M.I.T. Press.

Jaynes, E.T. (1957). "Information Theory and Statistical Mechanics I." Physical Review, Vol. 106, pp. 620-630.

Jaynes, E.T. (1968). "Prior Probabilities." IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, pp. 227-248.

Johnston, R.J. (1979). Geography and Geographers: Anglo-American Human Geography Since 1945. London: Edward Arnold.

Kellogg, O.D. (1929). Foundations of Potential Theory. New York: Dover.

Klein, H.A. (1971). The New Gravitation. New York: Lippincott.

Kolars, J.F., and J.D. Nystuen (1974). Geography. McGraw-Hill: New York.

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

Lakshmanan, T.R., and W.G. Hansen (1965). "A Retail Market Potential Model." Journal of the American Institute of Planners, Vol. 31, pp. 134-143.

Leopold, L.B., and W.B. Langbein (1962). "The Concept of Entropy in Landscape Evolution." Theoretical Papers in the Hydrologic and Geomorphic Sciences. U.S. Geological Survey Professional Paper No. 500-A, pp. 1-20.

MacMillan, W.D. (1930). The Theory of the Potential. New York: Dover.

March, L., and M. Batty (1975). "Generalized Measures of Information, Bayes' Likelihood Ratio and Jaynes' Formalism." Environment and Planning B, Vol. 2, pp. 99-105.

Marchand, B. (1972). "Information Theory and Geography." Geographical Analysis, Vol. 4, pp. 234-257.

McAdams, A.K. (1970). Mathematical Analysis for Management Decisions. New York: Macmillan.

McCalden, G. (1972). A Review of Macrogeographical Concepts. Doctoral Dissertation, Department of Geography, Ohio State University.

McCalden, G. (1975). "Macrogeographic Functions: A Review and Extension." Geographical Analysis, Vol. 7, pp. 411-419.

Medvedkov, Y. (1967). "The Concept of Entropy in Settlement Pattern Analysis." Papers of the Regional Science Association, Vol. 18, pp. 165-168.

Medvedkov, Y. (1970). "Entropy: An Assessment of Potentialities in Geography." Economic Geography, Vol. 46, pp. 306-316.

Neft, D.S. (1966). Statistical Analysis for Areal Distributions. Monograph Series, No. 2, Regional Science Research Institute.

Olsson, G. (1965). Distance and Human Interaction: A Review and Bibliography. Bibliography Series, No. 2, Regional Science Research Institute.

Openshaw, S., and C.J. Connolly (1977). "Empirically Derived Deterrence Functions for Maximum Performance Spatial Interaction Models." Environment and Planning A, Vol. 9, pp. 1067-1079.

Perin, D. (1975). Spatial Income Inequalities in the United States, 1953-1972. Doctoral Dissertation, Department of Geography, Ohio State University.

Phillips, F., G.M. White, and K.E. Haynes (1976). "Extremal Approaches to Estimating Spatial Interaction." Geographical Analysis, Vol. 8, pp. 185-200.

Pooler, J., and J.F. de Abreu (1979). "Income Fronts and Migration Winds in Brazil: A Graphical Analysis." Ontario Geography, No. 13, pp. 25-39.

Ray, D.M. (1965). "Market Potential and Economic Shadow." Research Paper No. 101, Department of Geography, University of Chicago.

Rich, D.C. (1980). Potential Models in Human Geography. Concepts and Techniques in Modern Geography 26. University of East Anglia, Norwich: Geo Abstracts.

Semple, R.K. (1973). "Recent Trends in the Spatial Concentration of Corporate Headquarters." Economic Geography, Vol. 49, pp. 309-318.

Semple, R.K., and R.G. Golledge (1970). "An Analysis of Entropy Changes in a Settlement Pattern Over Time." Economic Geography, Vol. 46, pp. 157-160.

Semple, R.K., and J.M. Griffin (1971). "An Information Analysis of Trends in Urban Growth Inequality in Canada." Discussion Paper No. 19, Department of Geography, Ohio State University.

Semple, R.K., and H.L. Gauthier (1972). "Spatial-Temporal Trends in Income Inequalities in Brazil," Geographical Analysis, Vol. 4, pp. 169-180.

Semple, R.K., and G.J. Demko (1977). "An Information-theoretic Analysis: An Application to Soviet-COMECON Trade Flows." Geographical Analysis, Vol. 9, pp. 51-63.

Senior, M.L. (1979). "From Gravity Modelling to Entropy Maximizing: A Pedagogic Guide." Progress in Human Geography, Vol. 3, pp. 175-210.

Shannon, C.E., and W. Weaver (1949). The Mathematical Theory of Communication. Urbana: University of Illinois Press.

Sheppard, E.S. (1975). "Entropy in Geography: An Information Theoretic Approach to Bayesian Inference and Spatial Analysis." Discussion Paper No. 18, Department of Geography, University of Toronto.

Sheppard, E.S. (1976). "Entropy, Theory Construction and Spatial Analysis." Environment and Planning A, Vol. 8, pp. 741-752.

Sheppard, E.S. (1979a). "Geographic Potentials." Annals of the Association of American Geographers, Vol. 69, pp. 438-447.

Sheppard, E.S. (1979b). "Spatial Interaction and Geographic Theory." In Philosophy in Geography, edited by S. Gale and G. Olsson, pp. 361-378. Dordrecht, Holland: D. Reidel.

Smith, T.E. (1975). "A Choice Theory of Spatial Interaction." Regional Science and Urban Economics, Vol. 5, pp. 137-176.

Snickars, F., and J.W. Weibull (1977). "A Minimum Information Principle: Theory and Practice." Regional Science and Urban Economics, Vol. 7, pp. 137-168.

Stewart, J.Q. (1941). "An Inverse Distance Variation for Certain Social Influences." Science, Vol. 93, pp. 89-90.

Stewart, J.Q. (1942). "A Measure of the Influence of Population at a Distance." Sociometry, Vol. 5, pp. 63-71.

Stewart, J.Q. (1947). "Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population." Geographical Review, Vol. 37, pp. 461-485.

Stewart, J.Q. (1950). "The Development of Social Physics." American Journal of Physics, Vol. 18, pp. 239-253.

Stewart, J.Q. and W. Warntz (1958). "Physics of Population Distribution." Journal of Regional Science, Vol. 1, pp. 99-123. Reprinted in Spatial Analysis, edited by B.J. Berry and D.F. Marble, pp. 130-146. Englewood Cliffs, N.J.: Prentice-Hall.

Taylor, P.J. (1971). "Distance Transformation and Distance Decay Functions." Geographical Analysis, Vol. 3, pp. 221-238.

Taylor, P.J. (1977). Quantitative Methods in Geography. Houghton Mifflin: Boston.

Tegsjö, B., and S. Oberg (1966). "Concept of Potential Applied to Price Formation." Geografiska Annaler, Vol. 48, pp. 51-58.

Theil, H. (1967). Economics and Information Theory. Amsterdam: North Holland.

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

Tobler, W.R. (1979). "Estimation of Attractivities from Interactions." Environment and Planning A, Vol. 11, pp. 121-127.

Tribus, M. (1961). Thermostatics and Thermodynamics. Princeton: D. Van Nostrand.

Tribus, M. (1969). Rational Descriptions, Decisions and Designs. New York: Pergamon.

Vickerman, R.W. (1974). "Accessibility, Attraction, and Potential: A Review of Some Concepts and Their Use in Determining Mobility." Environment and Planning A, Vol. 6, pp. 675-691.

Wachs, M., and T.G. Kumagai (1973). "Physical Accessibility as a Social Indicator." Socio-Economic Planning Sciences, Vol. 7, pp. 437-456.

Walsh, J.A., and M.J. Webber (1977). "Information Theory: Some Concepts and Measures." Environment and Planning A, Vol. 9, pp. 395-417.

Warntz, W. (1959). Toward a Geography of Price: A Study in Geo-Econometrics. Philadelphia: University of Pennsylvania Press.

Warntz, W. (1965). Macrogeography and Income Fronts. Monograph Series, No. 3, Regional Science Research Institute.

Warntz, W. (1966). "The Topology of a Socio-Economic Terrain and Spatial Flows." Papers of the Regional Science Association, Vol. 17, pp. 47-61.

Warntz, W. (1973a). "First Variation on the Theme, Cartographics is to Geographical Science as Graphics is to Science Generally." In Applied Geography and the Human Environment, edited by R.E. Preston, pp. 54-85. Dept. of Geography Publications Series, No. 2, University of Waterloo.

Warntz, W. (1973b). "New Geography as General Spatial Systems Theory—Old Social Physics Writ Large?" In Directions in Geography, edited by R.J. Chorley, pp. 89-126. London: Methuen.

Warntz, W. (1979). "Places of Birth, Education, and Activity of American Leaders as Related to Flows." Ontario Geography, No. 13, pp. 3-23.

Warntz, W., and D. Neft (1960). "Contributions to a Statistical Methodology for Areal Distributions." Journal of Regional Science, Vol. 2, pp. 47-66.

Warntz, W., and M. Woldenberg (1967). "Concepts and Applications—Spatial Order." Harvard Papers in Theoretical Geography, Geography and the Properties of Surfaces Series, Paper No. 1, Laboratory for Computer Graphics and Spatial Analysis, Harvard University.

Warntz, W., and P. Wolff (1971). Breakthroughs in Geography. New York: Plume Books.

Warntz, W., et al (1971). Tabulations of Data. Harvard Papers in Theoretical Geography, Geography of Income Series, Paper No. 2, Laboratory for Computer Graphics and Spatial Analysis, Harvard University.

Webber, M.J. (1976). "The Meaning of Entropy Maximizing Models." In Mathematical Land Use Theory, edited by G.J. Papageorgiou, pp. 277-292. Lexington, Mass.: D.C. Heath.

Webber, M.J. (1977). "Pedagogy Again: What is Entropy?" Annals of the Association of American Geographers, Vol. 67, pp. 254-266.

Webber, M.J. (1979). Information Theory and Urban Spatial Structure. London: Croom Helm.

Weibull, J.W. (1976). "An Axiomatic Approach to the Measurement of Accessibility." Regional Science and Urban Economics, Vol. 6, pp. 357-379.

Weibull, J.W. (1980). "On the Numerical Measurement of Accessibility." Environment and Planning A, Vol. 12, pp. 53-67.

Wilson, A.G. (1967). "A Statistical Theory of Spatial Distribution Models." Transportation Research, Vol. 1, pp. 253-269.

Wilson, A.G. (1970). Entropy in Urban and Regional Modelling. London: Pion.

Wilson, A.G. (1974). Urban and Regional Models in Geography and Planning. New York: John Wiley.

Wilson, A.G., and M.L. Senior (1974). "Some Relationships Between Entropy Maximizing Models, Mathematical Programming Models, and their Duals." Journal of Regional Science, Vol. 14, pp. 207-215.

Wilson, A.G., and M.J. Kirkby (1975). Mathematics for Geographers and Planners. Oxford: Clarendon Press.

Yeates, M. (1974). An Introduction to Quantitative Analysis in Human Geography. New York: McGraw-Hill.

Zielinski, K. (1979). "Experimental Analysis of Eleven Models of Urban Population Density." Environment and Planning A, Vol. 11, pp. 629-641.

Zielinski, K. (1980). "The Modelling of Urban Population Density: A Survey." Environment and Planning A, Vol. 12, pp. 135-154.

END

21-09-82

FIN