

UC Berkeley

IURD Working Paper Series

Title

On a Problem in Spatial Sampling

Permalink

<https://escholarship.org/uc/item/49g9p6qb>

Author

Falcke, Caj

Publication Date

1969-11-01

Peer reviewed

ON A PROBLEM IN SPATIAL SAMPLING

Caj Falcke

November 1969

Working Paper No. 107

Reprinted August 1971

All Rights Reserved

PREFACE

This paper is the sixth working paper of a group that has as their common concern the development of statistical methods for the analysis of the spatial structure of population and employment in urban areas and regions. Its principal concern is in application of quadrat analysis to spatial sampling.

This work is being conducted under the sponsorship of a grant from the Economic Development Administration of the U.S. Department of Commerce. A complete listing of the Working Papers under this grant appears on the back page of this paper. They deal with regional patterns, development, and policies.

Andrei Rogers
University of California,
Berkeley
November 1, 1969

Working Papers on Spatial Pattern Analysis

#86 Andrei Rogers and Norbert G. Gomar, STATISTICAL INFERENCE IN QUADRAT ANALYSIS, October, 1968.

#91* Schlomo Angel, POINT PATTERNS IN URBAN REGIONS: TWO COMPLEMENTARY TESTS, January, 1969.

#93* Andrei Rogers, QUADRAT ANALYSIS OF URBAN DISPERSION, March, 1969.

#96* Juan Martin, Norbert Gomar, Andrei Rogers, SOME COMPUTER PROGRAMS FOR QUADRAT ANALYSIS, May, 1969.

* Out of print

#97 · Andrei Rogers and Juan Martin, BIVARIATE QUADRAT ANALYSIS OF THE
DISPERSION OF RETAIL ESTABLISHMENTS AND POPULATION IN URBAN AREAS,
May, 1969.

#107 Caj Falcke, ON A PROBLEM IN SPATIAL SAMPLING, November, 1969.

1. Introduction

Let a random variable X have the density

$$(1.1) \quad f(x; \{\theta\}),$$

where $\{\theta\}$ is a set of parameters. There are a number of questions that can be asked with respect to (1.1):

1. What is the form of f ?
2. What are the elements of $\{\theta\}$?
3. What are the specific values of the elements of $\{\theta\}$ given that f is specified?
4. Does $\{\theta\}$ change over time and/or space?
5. Given $\{\theta\}$, what mechanism could have resulted in the specific $\{\theta\}$?
6. Given f , what is an effective sample design to obtain the (a) estimates of $\{\theta\}$?; (b) power to discriminate between alternative hypotheses?
7. Given $f(x; \{\theta\})$ and $Y = H(X)$, how can this information be utilized in obtaining information about the variable Y ?

The questions 1, ..., 4 and 6 are "classical" in statistical analysis. The fifth question is a much neglected, but an extremely important one. Andrei Rogers has in a recent paper [A. Rogers (1969) pp. 70...73] pointed out its relevance for statistical inference.

In this paper I shall be concerned with questions of type 7. I shall present some analytical tools for dealing with the problem, and an application. These tools have wide applicability, and will turn out to be

efficient in complex problems. Therefore I consider it important to present them here, even though the application I worked out turned out to be solvable by somewhat simpler methods.

The problems of sampling related to questions of type 7 should be worked out along the lines of the theory of ratio estimates. I shall not deal with these questions in this paper since the application I chose could be solved by analytical methods.

2. The General Nature of the Problem

Suppose a certain population (X) is spread over a two-dimensional space in a clustered way (see Fig. 1).

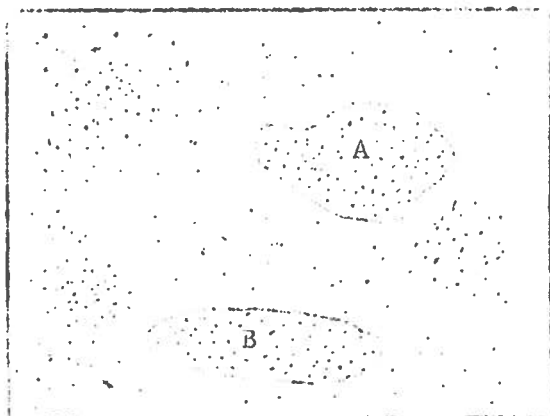


Figure 1

Suppose elements in A are similar w.r.t. some characteristic Y , and elements in B are similar w.r.t. the same characteristic Y , but elements in A are different from elements in B w.r.t. Y , i.e. if $y \in Y$, then $y_A \neq y_B$.

Now if we know the distribution of X we can take advantage of that knowledge when we are interested in a characteristic Y that is related to X , provided that we know the functional relationship of Y to X . The properties of Y can then be derived. Since the elements of the clusters are at least to some degree homogeneous w.r.t. Y , clustered sampling of Y would be much more efficient than random sampling, if we chose to sample Y directly instead of deriving its distribution from the density of X .

3. Functions of Random Variables. Derived Distributions.

Let the problem be to derive the density function of a random variable Y from the knowledge of the density of another random variable X when $Y = U(X)$. For example suppose that X has the density

$$(3.1) \quad f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and let

$$Y = 8X - 2$$

One way to derive the distribution of Y is the use of the cumulative distribution. In the next paragraph I shall present the more efficient method of characteristic functions.

Denote the density and the cumulative distribution of Y by $g(y)$ and $G(Y)$, respectively. From (3.1) we have

$$(3.2) \quad F(X) = \begin{cases} 0 & \text{for } x \leq 0 \\ \int_0^x f(x) dx = \int_0^x 1 dx = x & \text{for } 0 < x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

where $F(X) = P(X \leq x)$ is the distribution of X . From the definition of a distribution function we get

$$G(Y) = P(Y \leq y) = P(8X - 2 \leq y) = P(X \leq \frac{y+2}{8}).$$

From (3.2) we finally get

$$G(y) = \begin{cases} P(X \leq \frac{y+2}{8}) = 0 & \text{for } \frac{y+2}{8} \leq 0 & \text{i.e. for } y \leq -2 \\ P(X \leq \frac{y+2}{8}) = \frac{y+2}{8} & \text{for } 0 < \frac{y+2}{8} < 1 & \text{i.e. for } -2 < y < 6 \\ P(X \leq \frac{y+2}{8}) = 1 & \text{for } \frac{y+2}{8} \geq 1 & \text{i.e. for } 6 \leq y \end{cases}$$

To get the density $g(y)$ we differentiate $G(y)$ at points where the derivative exists:

$$G(y) = \frac{y+2}{8}; \quad \frac{d}{dy} G(y) = \frac{1}{8} \frac{d}{dy} G(y) = \frac{1}{8},$$

Thus we obtain the desired density

$$\begin{aligned} G(y) &= \frac{1}{8} & -2 < y < 6 \\ &= 0 & \text{elsewhere,} \end{aligned}$$

4. Characteristic Functions

Many problems in probability theory, particularly those connected with sums of independent random variables, can be tackled by means of

characteristic functions (the name "Fourier transformations" is used in analysis).¹

Definition: The expected value of the random variable e^{itX} is called the characteristic function of the random variable X ; X is a real parameter, i is the complex part. Denote the distribution function of X by $F(X)$. The characteristic function is then

$$(\text{Def}): \quad \zeta_X(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} dF(x) = \begin{cases} \int_{-\infty}^{\infty} e^{itx} f(x) dx & \text{in the continuous case} \\ \sum_{x=0}^{\infty} e^{itx} p(x) & \text{in the discrete case} \end{cases}$$

Where $f(x)$ and $p(x)$ denote the density of X (i.e., $[X = x]$), and the first integral is the Stieltjes integral.

The following properties hold for a characteristic function:

(i) A characteristic function exists for all distribution functions, i.e., for all random variables.

(ii) $\zeta(0) = 1$.

(iii) $|\zeta(t)| \leq 1$.

(iv) $\zeta(t)$ is uniformly continuous over the whole real line.

(v) Let X and Y be two random variables, and a and b two constants. Then, for the linear combination $Y = a + bX$ we have

$$\begin{aligned} \zeta_X(t) &= E(e^{itY}) = E(e^{it[a+bX]}) = E(e^{ita} e^{itbX}) = e^{ita} E(e^{itbX}) \\ &= e^{ita} \zeta_X(tb). \end{aligned}$$

(vi) The characteristic function of an independent random variable is equal to the product of their characteristic functions:

$$\zeta_{\sum_{j=1}^n X_j}(t) = E(e^{it \sum_{j=1}^n X_j}) = E(e^{itX_1} e^{itX_2} \cdots e^{itX_n}) = \prod_{j=1}^n \zeta_{X_j}(t).$$

¹Characteristic functions and moment generating functions are very similar in their essential features. By dropping the complex part of a characteristic function, one reduces it to a moment generating function. However, characteristic functions are more general since they exist for all random variables.

Note: This is only a necessary, but not a sufficient condition for independence.

(vii) If X has an absolute moment of n^{th} order, then its characteristic function has a k^{th} derivative ($k \leq n$):

$$\frac{d^{(k)}}{dt^{(k)}} \zeta_X(t) = \zeta_X^{(k)}(t) = \frac{d^{(k)}}{dt^{(k)}} \int e^{itx} dF(x) = i^k \int x^k e^{itx} dF(x).$$

From the last property we can derive useful expressions for the mean and the variance (and, as a matter of fact, also for higher order moments).

By definition

$$\int x^k dF(x) = E(X^k).$$

Then, if we evaluate the first derivative of the characteristic function at $t = 0$, we have that

$$\zeta_X'(t = 0) = i \int x e^{itx} dF(x) = i \int x dF(x) = iE(X).$$

Thus, $E(X) = \frac{1}{i} \zeta_X'(0).$

Similarly, consider the second derivative at $t = 0$:

$$\zeta_X''(0) = i^2 \int x^2 dF(x) = i^2 E(X^2) = -E(X^2).$$

Thus, $E(X^2) = -\zeta_X''(0)$

Hence

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = -\zeta_X''(0) + [\zeta_X'(0)]^2.$$

(viii) A distribution function is uniquely determined by its characteristic function.

Note 1: There exist characteristic functions whose values coincide on an interval as large as we please and which are not identically equal.

Note 2: It is possible to find independent random variables X_1 , X_2 , and X_3 such that the probability distribution of X_2 and X_3 are different but that the distribution functions of the sums $X_1 + X_2$ and $X_1 + X_3$ are

identical (Gnedenko, p. 300). This is another explanation to the result observed by Rogers (Rogers, p. 73) that the same distribution seems to fit data generated by different compound mechanisms.

Consider now an example. Calculate the characteristic function for the negative binomial distribution. Use the characteristic function to derive the mean and the variance of the distribution.

To derive the negative binomial distribution, let p denote the probability of success and q denote the probability of failure of a given event. Let $p(x)$ be the probability that exactly $x + r$ trials will be required to produce r successes. The last trial must be a success, and its probability is therefore p . Among the other trials there must be $r-1$ successes, and the probability of this is $\binom{x+r-1}{r-1} p^{r-1} q^x$.

Thus

$$(4.1) \quad p(x) = \binom{x+r-1}{r-1} p^{r-1} q^x p = \binom{x+r-1}{r-1} p^r q^x \quad [(\text{Mood and Graybill (1963), p. 72})].$$

This distribution is called the negative binomial distribution. Mood's derivation, which I have followed above, differs slightly from that of other authors. This is due to the fact that $\binom{x+r-1}{r-1} = \binom{x+r-1}{x}$.

The characteristic function for the negative binomial distribution is then given by

$$\begin{aligned} \zeta_X(t) &= \sum_{x=0}^{\infty} e^{itx} \binom{x+r-1}{r-1} p^r q^x = p^r \sum_{x=0}^{\infty} (qe^{it})^x \frac{(x+r-1)!}{(r-1)! (x+r-1-r+1)!} \\ &= p^r \sum_{x=0}^{\infty} \frac{(qe^{it})^x (x+r-1)!}{x! (r-1)!} = p^r (1 - qe^{it})^{-r} \end{aligned}$$

The first and the second derivative of $\zeta_X(t)$ are

$$\zeta'_X(t) = -rp^r (1 - qe^{it})^{-r-1} (-qe^{it})$$

$$X(0) = rp^r (1 - q)^{-r-1} (-qi) = rp^{-1} qi.$$

$$\begin{aligned} \zeta_X'(t) &= \frac{d}{dt} \left[\frac{rp^r (1 - qe^{it})^{-r-1}}{f(t)} \right] \frac{(-qe^{it} i)}{g(t)} = f(t) \frac{dg(t)}{dt} + g(t) \frac{df(t)}{dt} \\ &= -rp^r (1 - qe^{it})^{-r-1} (-qe^{it} i^2) + (-qe^{it} i) [(-r-1)(-rp^r)(1 - qe^{it})^{-r-2} \\ &\quad (-qe^{it} i)] \end{aligned}$$

$$\begin{aligned} \zeta_X''(0) &= -rp^{-1} q + (-qi)(-r^2 p^{-2} qi - rp^{-2} qi) = -rp^{-1} q + r^2 p^{-2} q^2 i^2 + rp^{-2} q^2 i^2 \\ &= -rp^{-1} q - rp^{-2} q^2 - r^2 p^{-2} q^2 = -r^2 p^{-2} q^2 - rp^{-1} q(1 + \frac{q}{p}). \end{aligned}$$

The mean and the variance are then

$$E(X) = \frac{1}{i} \zeta_X'(0) = \frac{1}{i} rp^{-1} qi = rp^{-1} q =$$

$$\begin{aligned} \text{Var}(X) &= -\zeta_X''(0) + [\zeta_X'(0)]^2 = r^2 p^{-2} q^2 + rp^{-1} q(1 + \frac{q}{p}) - r^2 p^{-2} q^2 \\ &= rp^{-1} q(1 + \frac{q}{p}) = \mu(1 + \frac{q}{r}). \end{aligned}$$

5. An Application

The State Board of Equalization receives during any one period N independent "contributions" from the retail establishments of our study area. Let $P(N = n) = p(n)$. If we denote the shipments of collected sales taxes by $y_1, y_2, \dots, y_N$ then the total tax revenue during the period is the sum $S_N = y_1 + y_2 + \dots + y_N$. Let the random variables y_i be functions of another random variable X = the number of stores in a given cell (block) of the area; specifically, let $y_i = a + bx_i$, for all i . Furthermore, let the distribution of the x 's be known; specifically let the x 's be identically distributed as the negative binomial distribution. Suppose we want to find the distribution, the mean and the variance of S_N . How can we benefit from our knowledge about the functional relationship between Y and X , and the distribution of X ? Note that N is also a random variable. By definition the distribution function of Y is

$$F(Y) = P(y_i \leq y), i = 1, 2, \dots$$

Let n be a fixed, arbitrary value for N . Then

$$(5.1) \quad P(S_N \leq y) = P(y_1 + y_2 + \dots + y_n \leq y)$$

which, using the assumption about independence, is

$$(5.2) \quad = F^{n*}(y), \text{ (see Rogers, p. 33)}$$

where n^* denotes "n fold convolution."

From Property (vi) of a characteristic function we then have that

$$(5.3) \quad \zeta_{S_N}(t) = [\zeta_y(t)]^n$$

From the definition of a characteristic function it follows that

$$(5.4) \quad \zeta_{S_N}(t) = E(e^{itS_N}) = E\left(\sum_n p(n) e^{itS_n}\right)$$

when N is a random variable.

Using $E p(n) = p(n)$ the expression (5.4) can be written as

$$(5.5) \quad \zeta_{S_N}(t) = \sum_n p(n) E(e^{itS_n})$$

We recognize the latter term of the right hand side as a characteristic function, namely $E(e^{itS_n}) = S_N(t)$

Thus, using (5.3)

$$(5.6) \quad S_N(t) = \sum_n p(n) [\zeta_y(t)]^n$$

We note that the first theorem of Helly states that any sequence of non-decreasing functions $F_1(Z), F_2(Z), \dots, F_n(Z), \dots$, which are uniformly bounded, contains at least one subsequence $F_{n_1}(Z), F_{n_2}(Z), \dots, F_{n_k}(Z) \dots$ which converges weakly ² to some non-decreasing function $F(Z)$. (Gnedenko, p. 280). Clearly the distribution functions of y fulfill the conditions of Helly's theorem. (A distribution function is always non-decreasing; the functions are uniformly bounded since $y = f(X)$, and the x 's are uniformly distributed.)

²We say that a sequence of non-decreasing functions $F_1(Z), F_2(Z), \dots, F_n(Z), \dots$ converges weakly to a non-decreasing function $F(Z)$ if for $n \rightarrow \infty$ it converges to $F(Z)$ at each of its points of continuity.

Furthermore, by referring to the direct limit theorem, which says that if the sequence of distribution functions converges weakly to $F(Z)$, then the sequence of corresponding characteristic functions converges weakly to the characteristic function $\zeta_Z(t)$, we can conclude that (5.6) converges, since $p(n) \rightarrow 1$. Hence, (5.6) is the stem function³ of the sequence $p(n)$. [See also Note 4 at the end of the paper]]. Denote a stem function by $B(Z)$. Then

$$(5.7) \quad \zeta_{S_N}(t) = B(\zeta_Y[t]).$$

Recall that N is a random variable. Now if we assume that N is for example Poisson distributed -- which is a plausible assumption of the period we are considering is a relatively short one -- we have to derive the stem function of the Poisson distribution with parameter $\zeta_Y(t)$. Denote this parameter by z for a moment. Then

$$\begin{aligned} B(p[z]) &= \sum_{n=0}^{\infty} p(n, \lambda) z^n \text{ where } \lambda \text{ is the parameter of the Poisson} \\ &\quad \text{distribution} \\ &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} z^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda z)^n}{n!} = e^{-\lambda} e^{\lambda z} \\ &= e^{-\lambda(1-z)}. \end{aligned}$$

By substituting $z = \zeta_Y(t)$ we get from (5.7)

$$(5.8) \quad B(p[\zeta_Y(t)]) = e^{-\lambda(1-\zeta_Y[t])} = \zeta_{S_N}[t]$$

Finally we shall make use of our knowledge that $Y = a + bX$. Recall Property (v) of a characteristic function:

$$\zeta_Y(t) = e^{ita\zeta_X(tb)}.$$

³

If $A(z) = \sum_{k=0}^{\infty} a_k z^k$ converges somewhere when z varies, then we say that $A(z)$ is the stem function of the sequence (a_k) . z is a real or complex parameter. $A(z)$ is also called the generating function of the sequence.

Insert this expression in (5.8):

$$(5.9) \quad \zeta_{S_N}(t) = e^{-\lambda(1-e^{ita})\zeta_X(tb)}$$

Now the characteristic function (and hence the distribution of the sum S_N) is uniquely determined by known quantities and the unknown parameter of the Poisson distribution.

To derive the mean and variance of S_N we need the first and second derivatives of $\zeta_{S_N}(0)$. Note that, in general,

$$\zeta_X(tb) = Ee^{itbx} = \sum_x e^{itbx} p(x).$$

The first derivative of $\zeta_{S_N}(t)$ is then

$$\zeta'_{S_N}(t) = \frac{d}{dt} e^{-\lambda(1-e^{ita})\zeta_X(tb)} = e^{-\lambda(1-e^{ita})\zeta_X(tb)} (\lambda \zeta'_X[ta] e^{ita} + \lambda \zeta_X[ta] e^{ita} ia)$$

$$S'_N(0) = e^{-\lambda(1-\zeta_X[0])} (\lambda \zeta'_X[0] + \lambda \zeta_X[0] ia) = \lambda \zeta'_X(0) + \lambda ia = \lambda(bE[X] + a)$$

Thus

$$E(S_N) = \frac{1}{i} \zeta'_{S_N}(0) = \frac{1}{i} i(bE[X] + a) = \lambda(bE[X] + a).$$

Now instead of using the second derivative (which gets quite messy, in this particular case) to derive the variance I shall illustrate the method of expected values:

$$E(S_N) = E(N)E(y_1) = \lambda E(y_1) = \lambda(a + bE[X])$$

$$\text{Var}(S_N) = \text{Var}\left(\sum_{i=1}^N y_i\right) = \text{Var}(y_1) + \text{Var}(y_2) + \dots + \text{Var}(y_N), \text{ by}$$

independence of the y 's. There are N terms in the sum, but since N is

random we have to use $E(N) = \lambda$. Note also that $\text{Var}(y_1) = \text{Var}(a + bX) =$

$\text{Var}(a) + \text{Var}(bX) = b^2 \text{Var}(X)$, for all i . We therefore have that

$$\text{Var}(S_N) = \sum \text{Var}(y_i) = \sum b^2 \text{Var}(X) = \lambda b^2 \text{Var}(X). \text{ Collecting the results:}$$

$$\begin{aligned}\zeta_{S_N}(t) &= e^{-\lambda(1-e^{ita})} \zeta_X[tb]) \\ E(S_N) &= \lambda(a + bE[X]) \\ \text{Var}(S_N) &= \lambda b^2 \text{Var}(X)\end{aligned}$$

The only additional information we need is an estimate of λ , the mean value of the number of "shipments" arriving at the Board during the period under consideration.

Note 1: The assumption of N being Poisson-distributed is not an essential one. Any realistic distribution would do equally well; only the algebra becomes then more troublesome.

Note 2: The fact that we were able to use expected values to derive $\text{Var}(S_N)$ must not lead the reader to the conclusion that characteristic functions are unnecessary, and lead only to messy calculations. In more complex situations the method of expected values will indeed lead to very cumbersome calculations, and the characteristic functions will prove to be more convenient.

Note 3: Since λ is totally independent of X , there is no way we can benefit from the knowledge of the distribution of X when we sample to get an estimate of λ .

Note 4: After writing up this paper and after having read through it I realized that equation (5.6) could directly have been established as a stem function by reference to property (iii) of a characteristic function, which states that $|\zeta_X(t)| \leq 1$. Then a comparison with the geometric series shows that (5.6) converges at least for $|\zeta_X(t)| \leq 1$.

REFERENCES

- Gnedenko, B.V. The Theory of Probability. New York: Chelsea Publishing Co., 1967.
- Mood, A.M. and Graybill, F.A. Introduction to the Theory of Statistics. McGraw-Hill, 1963.
- Rogers, Andrei. "Quadrat Analysis of Urban Dispersion." University of California, Berkeley: Center for Planning and Development Research, Working Paper No. 93, 1969.

IURD
Working Papers Sponsored by the
Economic Development Administration

- WP-65* Ronald Choy, "A Review of the Role of Demographic Variables in Macro-Econometric Models," August, 1967.
- WP-66* Andrei Rogers, "The Aggregation Problem in Demography," October, 1967.
- WP-67* Michael Teitz, "Toward a Theory of Urban Public Facility Location," November, 1967.
- WP-70* Michael Teitz, "Locational Strategies for Competitive Systems," December, 1967.
- WP-72* William Alonso, "The Quality of Data and the Choice and Design of Predictive Models," December, 1967.
- WP-74 William Alonso, "Industrial Location and Regional Policy in Economic Development," February, 1968.
- WP-78 William Alonso, "Equity and Its Relation to Efficiency in Urbanization," July, 1968.
- WP-81* Andrei Rogers and Susan McDougall, "An Analysis of Population Growth and Change in Slovenia and the Rest of Yugoslavia," June, 1968.
- WP-84* Andrei Rogers, "Activity Allocation Models in Transportation Planning," September, 1968.
- WP-86 Andrei Rogers and Norbert G. Gomar, "Statistical Inference in Quadratic Analysis," October, 1968.
- WP-87 William Alonso, "Aspects of Regional Planning and Theory in the United States," October, 1968.
- WP-89* David Darwent, "Growth Pole and Growth Center Concepts: A Review, Evaluation and Bibliography," October, 1968.
- WP-90 William Alonso, "Beyond the Inter-Disciplinary Approach to Planning," November, 1968.
- WP-91* Shlomo Angel, "Point Patterns in Urban Regions: Two Complementary Tests," January, 1969.

- WP-92* Andrei Rogers and Susan McDougall Choy, "Multisectoral Models of Regional Demographic and Economic Growth," February, 1969.
- WP-93* Andrei Rogers, "Quadrat Analysis of Urban Dispersion, March, 1969.
- WP-96* Juan Martin, Norbert Gomar, and Andrei Rogers, "Some Computer Programs for Quadrat Analysis," May, 1969.
- WP-97 Andrei Rogers and Juan Martin, "Bivariate Quadrat Analysis of the Dispersion of Retail Establishments and Population in Urban Areas," May, 1969.
- WP-99 Susan McDougall Choy, "The Role of Labor in Regional Multisectoral Growth Models," June, 1969.
- WP-107 Caj Falcke, "On a Problem in Spatial Sampling," November, 1969.