



Det här verket har digitaliserats vid Göteborgs universitetsbibliotek och är fritt att använda. Alla tryckta texter är OCR-tolkade till maskinläsbar text. Det betyder att du kan söka och kopiera texten från dokumentet. Vissa äldre dokument med dåligt tryck kan vara svåra att OCR-tolka korrekt vilket medför att den OCR-tolkade texten kan innehålla fel och därför bör man visuellt jämföra med verkets bilder för att avgöra vad som är riktigt.

This work has been digitized at Gothenburg University Library and is free to use. All printed texts have been OCR-processed and converted to machine readable text. This means that you can search and copy text from the document. Some early printed books are hard to OCR-process correctly and the text may contain errors, so one should always visually compare it with the images to determine what is correct.



**Rapport**

**R16:1975**

# **Två modeller för rumsliga befolkningsfördelningar**

**Jörgen Weibull**



**Bygghforskningen**

# Två modeller för rumsliga befolkningsfördelningar

Jörgen Weibull

*Denna rapport är ett resultat av tio månaders vistelse vid Centre de Recherche d'Urbanisme (C.R.U.) i Paris 1972-73. Rapporten innehåller en kort presentation av C.R.U. och en tämligen utförlig redovisning av två matematiskt inriktade projekt. Båda projekten avser rumsliga befolkningsfördelningar; i det ena fallet gäller det befolkningstätheten i en stad, i det andra städernas rumsliga fördelning i en region eller ett land.*

## Några ord om C.R.U.

C.R.U. är ett av franska staten finansierat forskningscentrum. Forskningen är tämligen tvärvetenskaplig och omfattar projekt av mycket varierande slag, från sociologi och beskrivande statistik till trafikmodeller och matematiska modeller. Vid sidan av forskningen bedrivs viss kursverksamhet. Denna har till syfte att förbereda studenter med relevant akademisk grundutbildning för urbaniseringsforskning. Centret omfattar 30-40 personer varav ungefär hälften forskare.

## Den exponentiella modellen för befolkningstätheten i en stad

En vanlig modell för boendebefolkningens fördelning i en stad är den så kallade Clarks formel, enligt vilken befolkningstätheten avtar exponentiellt från centrum ("central business district"). Clark uppställde sambandet enbart på grundval av empiriska undersökningar av befolkningstätheten i några amerikanska städer. I en del senare arbeten har dock försök gjorts att ge modellen förankring i ekonomisk teori, dels deterministiskt, dels statistiskt genom resonemang av samma typ som inom termodynamiken. Det är det senare, statistiska synsättet, som utgör grunden i föreliggande arbete.

Eftersom de teoretiska grunderna för modellen har redovisats på andra ställen, ges endast en kort presentation av några av de grundläggande tankegångarna i ett inledande avsnitt. Därefter redovisas en grafisk metod att bestämma värdet på de två parametrarna i modellen. Metoden förutsätter endast kännedom om folkmängd och medelavstånd

till centrum samt papper och penna. Vidare presenteras en känslighetsanalys med syfte att ge en kvantitativ uppskattning av den osäkerhet i parametervärdena som beror på osäkerhet hos indata. Sist diskuteras ett mått på avvikelse mellan modell och data. Måttet kallas redundans och kommer ursprungligen från matematisk kommunikationsteori.

De teoretiska resonemangen illustreras i varje delavsnitt med tillämpningsresultat på olika städer. Tonvikten i framställningen är dock inte avsedd att ligga på det specifika för den exponentiella modellen utan snarare på mer allmänna metodologiska aspekter.

Som ett allmänt omdöme om modellen kan det konstateras att den visserligen är beräkningsmässigt lätthanterlig, men i många fall visar bristande överensstämmelse med observationsdata. I de flesta fall består denna diskrepans i att modellen överskattar boendestätheten i centrum. En mer realistisk modellansats skulle kunna ge bättre överensstämmelse med data, men förmodligen till priset av minskad överskådlighet och lätthanterlighet.

## Städer som punkter i planet

För att skapa sig någon överblick över den uppsjö av arbeten som återfinns i litteraturen om städernas rumsliga och storleksmässiga fördelning kan man göra följande grova indelning i tre "traditioner" eller "riktningar".

För det första har vi den så kallade centralortsteorin som bygger på ett arbete av W. Cristaller. Från vissa idealiserade geografiska och ekonomiska antaganden härledde Cristaller en hierarkisk struktur av städer på olika "funktionsnivå", rumsligt fördelade i ett hexagonalt mönster.

En annan riktning har sin upprinnelse i ett arbete av F. Auerbach. Till skillnad från i centralortsteorin, återfinns här inte någon egentlig teoretisk grund utan endast ett empiriskt funnet samband. Sambandet brukar kallas "rank-size rule" och tillskrivs ofta G. K. Zipf, "Zipfs lag". I sin enklaste form säger "lagen" att om alla städer i en region (eller ett land) ordnas efter folkmängd, så att den största staden får rang ett, den näst

# Byggnadsforskningen Sammanfattningar

R16:1975

Nyckelord:

urbaniseringsforskning, befolkningstäthet, ortssystem, matematisk modell

Rapport R16:1975 hänför sig till forskningsanslag Bs 471 från Statens råd för byggnadsforskning till Matematiska institutionen, KTH, Stockholm.

UDK 312.9 (44)  
711.132 (44)  
SfB A  
ISBN 91-540-2420-X

Sammanfattning av:

Weibull, J. 1975, *Två modeller för rumsliga befolkningsfördelningar*. (Statens råd för byggnadsforskning) Stockholm. Rapport R16:1975, 69 s., ill. 17 kr + moms.

Rapporten är skriven på svenska med svensk och engelsk sammanfattning.

Distribution:

Svensk Byggtjänst  
Box 1403, 111 84 Stockholm  
Telefon 08-24 28 60  
Grupp: samhällsplanering

största rang två osv, så kommer storleken av den n:te staden att vara en n:te del av storleken hos den största staden.

Slutligen har vi en tredje riktning, punktmönsteranalys, som använder sig av matematisk-statistiska metoder från kvantitativ ekologi. Det är två huvudtyper av metoden som där förekommer, nämligen "kvadrat analysis" (QA) och "k:th nearest neighbour analysis" (NA). I QA analyseras punktmönster genom indelning av studieområdet i ett rutnät och uppräknig av antalet punkter i varje ruta (eller i ett stickprov av rutor). I NA mäter man avståndet från en godtycklig punkt till den k:te närmsta punkten i punktmönstret ( $k = 1$ : allra närms-

ta,  $k = 2$ : näst närmsta osv).

Avsikten med den studie som redovisas i föreliggande avsnitt var att, utgående från "k:th nearest neighbour analysis", utarbeta redskap för analys av verkliga städers rumsliga fördelning.

Sålunda ges först en härledning av några statistiska egenskaper hos avstånd mellan punkter i ett slumpmässigt punktmönster. Därigenom möjliggörs bland annat en klassificering av verkliga punktmönster med avseende på agglomereringsgrad, alltifrån fullständig koncentration via slumpmässighet till absolut regelbundenhet enligt Cristallers centralortsteori. Efter att ha etablerat dessa teoretiska resultat beträffande den

rumsliga fördelningen görs sedan en ihopkoppling av dessa resultat med den storleksfördelning som uttrycks av rank-size rule. Preliminära tillämpningar på de franska städerna visar att de verkliga städerna tenderar att vara något mer regelbundet fördelade än punkterna i ett slumpmässigt punktmönster, dock inte så regelbundet som föreskrivs av Cristallers centralortsteori. Tendensen mot regelbundenhet visar sig vara starkare i en jordbruksregion än i en industriell region. Sist analyseras styrkan hos tre statistiska test av slumpmässighet hos punktmönster.

# Two models for the spatial distribution of population

Jörgen Weibull

*This report is the result of a ten-month stay at the Centre de Recherche d'Urbanisme (C.R.U.) in Paris during 1972-1973. The report contains a brief description of C. R. U. and a reasonably complete presentation of two mathematically oriented projects. Both projects concern the spatial distribution of population: urban population density in the first case and, in the second, the spatial distribution of cities in a region or in a country.*

## A Few Words About C. R. U.

C.R.U. is a government financed research center. The research is of a multi-disciplinary character and includes a great variety of different types of projects extending from studies of a sociological or descriptive statistical nature to others using mathematical models. In addition to research the center gives postgraduate courses annually. Some 30 persons are engaged in research at the center.

## The Exponential Model for Urban Population Density

A common model for urban population density is the so-called Clarks's formula, according to which population density diminishes exponentially from the central business district (CBD) outwards. Clark based his rule on the basis of empirical studies in some American cities. However, in some later work attempts have been made to provide the model with a foundation in economic theory — partly in a deterministic fashion, partly using statistical arguments of the same type as in thermodynamics. The latter, statistical viewpoint, forms the basis of the present work.

As the theoretical foundations of the model have been presented elsewhere only a brief description of a few of the basic ideas are presented in an introductory section of this report. Thereafter, a graphic method to determine the values of the two parameters in the model is presented. This method only assumes knowledge of total population and average distance to the CBD, plus paper and pencil. Moreover, a sensitivity analysis is described which provides a quantitative estimate of the error in parameter values that is due to lack of precision in the data used. Finally a measure of the discrepancy between a model and its observation data is dis-

cussed. This measure is called redundancy and comes originally from mathematical communication theory.

The theoretical reasoning is illustrated in each subsection with results from application on different cities. Emphasis in the description, however, is not on the specifics of the exponential model but rather on more general, methodological aspects.

As a general comment on the model, it can be said that it certainly is easy to use as far as calculation is concerned, but in many cases displays insufficient agreement with observation data. In most cases, the discrepancy consists in that the model overestimates population density in the CBD. A more sophisticated model could give better agreement with data, but probably at the price of greater complexity and difficulty of use.

## Cities as Points in a Plane

In order to obtain a general view of the mass of existing work on the spatial and size distribution of cities, one may make the following rough division into three major trends or approaches.

To begin with, there is the central-place theory based upon a work by W. Cristaller. Proceeding from certain idealized geographic and economic assumptions, Cristaller derived a hierarchic organization of urban settlements at different "functional" levels, spatially distributed in a hexagonal pattern.

Another approach originated with the work of F. Auerbach. Contrary to the central-place theory, this approach has (so far) no theoretical basis — it is merely based on an observed correlation. This correlation is usually called the rank-size rule, often ascribed to G. K. Zipf and known as Zipf's law. In its simplest form this law states that if all the cities in a region (or a country) are arranged in order of decreasing size of population so that the largest city is assigned rank one, the next largest rank two and so forth, then the size of the  $n$ :th city will be an  $n$ :th part of the size of the largest city.

Finally, we have a third approach, the so-called point pattern analysis, which is based on mathematical-statistical methods from quantitative ecology. There are two principal types of method which occur, namely quadrat analysis (QA) and  $k$ :th nearest neighbour analysis (NA). In QA a grid is placed over

## Swedish Building Research Summaries

R16:1975

Key words:

urban studies, population density central-place theory, mathematical model

Report R16:1975 refers to research grant Bs 471 from the Swedish Council for Building Research to the Department of Mathematics, Royal Institute of Technology, Stockholm.

UDC 312.9 (44)  
711.132 (44)  
SFB A  
ISBN 91-540-2420-X

Summary of:

Weibull, J., 1975, *Två modeller för rumsliga befolkningsfördelningar*. Two models for the spatial distribution of population. (Statens råd för byggnadsforskning) Stockholm. Report R16:1975, 69 p., ill. Kr 17 + moms.

The report is in Swedish with Swedish and English summaries.

Distribution:

Svensk Byggtjänst  
Box 1403, S-111 84 Stockholm  
Sweden

the area to be studied and the number of points in each square (or in a random sample of squares) is counted. With NA, one measures the distance from an arbitrary point to the  $k$ :th nearest point in the point pattern ( $k=1$ :nearest point,  $k=2$ :next nearest point, etc).

The purpose of the study described was to, on the basis of  $k$ :th nearest neighbour analysis, prepare instruments for an analysis of the spatial distribution of real cities.

Thus, a derivation is first given of a few statistical characteristics of distances between points in a random point pattern. This analysis makes it possible to classify real point patterns by their degree of agglomeration, from absolute concentration via random distribution to complete regularity according to Crisaller's central-place theory. After these theoretical results concerning spatial distribution have been established, they are put together with the size distri-

bution induced by the rank-size rule. Preliminary applications on French data reveal that real towns tend to be somewhat more regularly distributed than the points in a random point pattern, although not as regular as predicted by the Crisaller model. Agricultural regions display a stronger tendency towards regularity than industrial regions. Finally, the power of three statistical tests of randomness is analysed.

Rapport R16:1975

Två modeller för rumsliga  
befolkningsfördelningar

av Jörgen Weibull

Denna rapport hänför sig till forskningsprojekt Bs 471  
från Statens råd för byggnadsforskning till Matematiska  
institutionen, KTH, Stockholm.

Statens råd för byggnadsforskning, Stockholm  
ISBN 91-540-2420-X

LiberTryck Stockholm 1975



## INNEHÅLL

FÖRORD .....	4
TABELLFÖRTECKNING .....	5
FIGURFÖRTECKNING KAPITEL 2 .....	6
FIGURFÖRTECKNING KAPITEL 3.....	7
1. NÅGRA ORD OM C.R.U. ....	8
1.1 Allmänt .....	8
1.2 Forskningen.....	9
1.3 Några C.R.U.-publikationer.....	13
2. DEN EXPONENTIELLA MODELLEN FÖR BEFOLKNINGSTÄTHETEN I STÄDER .....	15
2.1 Inledning,.....	15
2.2 En härledning av modellen.....	16
2.3 Skattning av parametrarna.....	19
2.4 Känslighetsanalys för Å och ð.....	29
2.5 Modellens överensstämmelse med data.....	34
2.6 Sammanfattning.....	41
2.7 Referenser.....	41
3. STÄDER SOM PUNKTER I PLANET .....	43
3.1 Inledning .....	43
3.2 En allmän orientering .....	43
3.3 Några statistiska egenskaper .....	49
3.4 Ett samband mellan städers storlek och avstånd.....	53
3.5 Några statistiska test.....	59
3.6 Sammanfattning .....	67
3.7 Referenser.....	68

## Förord

Denna rapport är ett resultat av tio månaders vistelse vid Centre de Recherche d'Urbanisme (C.R.U.) i Paris 1972-73.

Arbetet kom att bestå av två skilda delar, den ena ägnad åt en matematisk modell för fördelningen av befolkningstätheten i en stad och den andra ägnad vissa matematiskt-statistiskt inriktade regionala studier. De båda delarna redovisas här i var sitt avsnitt. I avsnitt 2 behandlas sålunda vissa tekniska frågor kring den urbana modellen. Avsnitt 3 innehåller en mera översiktlig presentation av de regionala studierna. Men först av allt en kort presentation av C.R.U.

Tabellförteckning

TABELL 1. Tillämpningar av den exponentiella modellen

TABELL 2. Martin-Löfs normativa tabell för modellens redundansvärden

Figurförteckning Kapitel 2

- FIG. 1. Den exponentiella modellen
2. Den normerade maximum-likelihoodfunktionen  $t(s)$
  3. Den "normerade partitionsfunktionen"
  4. Schematisering av skattningsmetoden
  5. Del av grafen för  $t(s)$
  6. Del av grafen för  $w(t)$
  7. Parametern  $b$  som funktion av ytterradien  $R$  för Bordeaux 1962
  8. Parametern  $b$  som funktion av ytterradien  $R$  för Paris 1876-1962
  9. Schematisk bild av sambandet befolkningsdata - parametervärden
  10. Jämförelse mellan två modeller för London 1966
  11. Successiva modellreduktioner

Figurförteckning Kapitel 3

- FIG. 1. Ett centralortssystem à la Cristaller
2. Svenska städer rangordnade efter folkmängd
  3. Några typiska punktmönster
  4. Ett parvis agglomererat punktmönster
  5. Två frekvensfunktioner för avstånd till närmsta grannpunkt
  6. De franska städerna 1968
  7. Städerna rangordnade efter folkmängd. Frankrike 1962.
  8. Medelavstånd till närmsta grannstad som funktion av storleksklass. Frankrike 1962.
  9. Agglomereringsindexet  $D^*$  som funktion av storleksklass. Frankrike 1962.
  10. Frankrike uppdelat i en jordbruks- och en industridel
  11. Medelavstånd till närmsta grannstad, respektive agglomereringsindexet  $D^*$  som funktion av storleksklass för de två delarna av Frankrike 1962.
  12. De tre testernas styrkor som funktioner av agglomereringsgraden  $p$ . Stickprovsstorlek  $n=10$  resp.  $n=25$ .
  13. De tre testernas styrkor för ett regelbundet, kvadratisk punktmönster som funktioner av stickprovsstorleken.

## 1. NÅGRA ORD OM C.R.U.

### 1.1 Allmänt

C.R.U. (Centre de Recherche d'Urbanisme) bildades 1962. Bildandet följde på en resolution antagen av FN:s ekonomiska och sociala råd under den 32:a sessionen; resolutionen rekommenderade nämligen medlemsländernas regeringar att bilda organisationer på nationell nivå för forskning och studier av urbaniseringens olika aspekter. Centret är numera underordnat le Ministère de l'Amenagement du Territoire, de l'Equipement, du Logement et du Tourisme.

Man kan dela upp verksamheten på C.R.U. i tre delar: forskning, kursverksamhet och förlagsverksamhet.

Forskningen, som av naturliga skäl är tämligen tvärvetenskaplig, bedrevs under min vistelse där av ett tjugutal forskare av skilda nationaliteter (ungefär hälften fransmän) och dicipliner (arkitekter, sociologer, geografer, ingenjörer m fl). De flesta forskarna var fast anställda, men även konsultuppdrag förekom.

Kursverksamheten har till syfte att förbereda studenter för urbaniseringsforskning. För att bli antagen vid dessa kurser krävs relevant akademisk grundutbildning (examen i arkitektur, stadsplanering, geografi etc.). Kurserna består dels av lektioner (hållna av centrets forskare och inbjudna gästföreläsare), dels av praktiskt deltagande i de olika forskningsgruppernas verksamhet. Det ges för närvarande tre olika kurser varje år (av 2, 6 respektive 9 månaders längd).

Förlagsverksamheten är underordnad den övriga verksamheten såtillvida att endast arbeten utförda i C.R.U:s regi publiceras. De flesta publikationerna är på franska men

även en del engelskspråkiga utgåvor förekommer. En förteckning över några av titlarna ges i slutet av detta kapitel.

C.R.U. subventioneras av la Délégation Générale à la Recherche Scientifique och le Ministère de l'Aménagement du Territoire, de l'Équipement, du Logement et du Tourisme, men får även vissa inkomster från kurs- och förlagsverksamheten. 1973 års budget uppgick till drygt två miljoner francs. Sammanlagda antalet sålda och utdelade publikationer uppgick samma år till närmare 6.000 exemplar fördelade på ett trettiotal titlar.

## 1.2 Forskningen

För att ge en uppfattning av forskningens inriktning, ges här ett smakprov på några projekt. Som synes spänner verksamheten, trots de relativt begränsade personella resurserna, över ett vitt fält.

De franska städerna. En forskningsgrupp har under åtskilliga år ägnat sig åt statistiskt studium av de franska städerna. Resultatet har blivit fyra digra volymer, "Les villes françaises, I-IV" ([3]), den första publicerad redan 1966 och den fjärde och (förmodligen) sista så sent som 1974. Dessa böcker ger en för franska förhållanden ovanligt överskådlig och enhetlig statistisk beskrivning av de franska städernas utveckling under 1900-talet. De är rikt illustrerade med kartor och diagram, varför de är förhållandevis lättillgängliga även för en icke franskspråkig läsekrets.

I den första volymen ges en allmän demografisk redogörelse där särskilt städernas tillväxt under perioden 1911-62 studeras. Den andra volymen ägnas åt en socio-ekonomisk studie, huvudsakligen gällande år 1962. Med indelning i ekonomiska sektorer redovisas fördelningen av den aktiva befolkningen dels med utgångspunkt från bostadsort, dels utgående från

arbetsplatsort. Även de dagliga resorna studeras. I den tredje volymen ges (kultur)geografiska uppgifter: städernas yta, deras genomsnittliga befolkningstäthet, samband mellan urbaniseringsgrad och trafik, uppskattning av städernas omland m m. Den fjärde volymen slutligen, är avsedd som referensmaterial för franska stadsplanerare. Där finns de data som utgör underlag för statistiken i de tre första volymerna sammanställda stad för stad.

Orthopoles. Först några ord om bakgrunden. I en föregående studie ("fas 1"), slutförd 1967 ("Modèles de villes", [1]), var syftet att undersöka möjligheterna för en starkt utbyggd kollektivtrafiksektor (80% av resorna till och från arbetet) i en modellstad med 250.000, 500.000 och respektive 1 miljon invånare. Modellstaden bestod av ett hexagonalt mönster av stadskärnor med grönområden emellan. På grundval av resultaten från denna studie utfördes sedan en ny, tämligen detaljerad studie ("fas 2") av möjligheterna för ett kommunikationssystem denna gång baserat på privatbilism i samma modellstad ("Archipoles", [6]; engelsk upplaga [11]). Denna studie har beskrivits i "Några exempel på fransk urbaniseringsforskning"<sup>(1)</sup> och kommer därför inte att behandlas vidare här.

För närvarande befinner sig studierna i en tredje fas, "orthopoles". Denna fas utgör dels en syntes av de två tidigare faserna i det att man nu använder en geografiskt differentierad fördelning mellan kollektiv- och privat trafik. (Huvudsakligen kollektivtrafik och fotgängare i centrala delar, privatbilism i mer perifera delar). Dessutom försöker man "mjuka upp" modellen så att den lättare skall kunna användas vid planering av verkliga

-----  
 (1) Snickars, F., 1971, Stockholms Läns Landsting, Regionplanekontoret, Meddelande 1971:2.



städer. Syftet med studien är dels att utveckla ett instrument för utvärdering av verkliga trafikplaner, dels att kunna anvisa hur en viss stad skall kunna utvecklas mot modellstaden. Studien beräknas vara slutförd under 1974.

Fransk stadsplanering i teori och praktik. Under de två senaste åren har en forskningsgrupp ägnat sig åt studier av fransk stadsplaneringspraxis. Arbetet har skett i två etapper. Den första etappen, som avslutades 1973, resulterade i en bok i vilken det beskrivs hur fransk stadsplanering går till i allmänhet: vilka institutioner som är inkopplade, de olika planerna, vilka styrinstrument som står till planerarnas förfogande osv. ("Conception et instruments de la planification urbaine", [7]).

Den andra etappen, som visserligen är avslutad men ännu inte resulterat i någon publikation, har ägnats åt studium av några praktiska fall. Man har valt fem medelstora franska städer och i detalj jämfört de planer som fastställdes under en tjuguarperiod med den verkliga utvecklingen i städerna under samma tid. Med tanke på att resultaten har visat sig något kontroversiella, är det emellertid tveksamt huruvida någon fullständig publikation överhuvudtaget kommer att ske.

Fritid: från myt till verklighet. Under denna svepande rubrik har en forskningsgrupp publicerat resultat från studier av fritidsaktiviteter i Frankrike; deras omfattning mätt i tid och pengar m m ([4]).

Kostnaden för grönområden. Detta projekt påbörjades egentligen inte förrän under hösten 1973. Projektets syfte är att genom studium av några praktiska fall uppskatta den relativa samhällsekonomiska kostnaden för grönområden i jämförelse med kostnaden för bebyggelse av samma område.

Efter denna kortfattade presentation av några av de andra projekten på C.R.U. kommer slutligen turen till de två mer matematiskt inriktade projekt som vi sysslade med i den forskningsgrupp jag tillhörde. Nedan ges dock endast en kort redogörelse för bakgrunden till de två studierna eftersom en detaljerad redovisning återfinnes i de två följande avsnitten.

Den exponentiella modellen. Åtskilliga urbaniseringsforskare och kulturgeografer har uppställt mer eller mindre sofistikerade matematiska modeller för olika aktiviteters rumsliga fördelning i städer. En ofta förekommande modell för boendebefolkningens fördelning är den s k exponentiella modellen ("Clarks formel"), enligt vilken befolkningstätheten avtar exponentiellt med avståndet från centrum. Sedan några år tillbaka har denna modell varit utsatt för studier på C.R.U. Dessa tidigare studier finns dokumenterade i några C.R.U.-publikationer varav en är på engelska och en annan, "Modèle urbain de localisation résidentielle" (1972, [2]), publicerades inom ramen för ett avtal om vetenskapligt utbyte mellan Frankrike och Sovjetunionen och därför föreligger både i en franskspråkig och en ryskspråkig version.

Anledningen till att man valt just den exponentiella modellen för fördelningen av befolkningstätheten i en stad var att man ansåg att den (a) belyser en relevant aspekt av städers struktur, (b) har god överensstämmelse med verkligheten, (c) är relativt lätthanterlig ur matematisk

synvinkel och (d) har lättolkade parametrar. Dessa aspekter kommer att diskuteras i detalj i kapitel 2.

Den regionala studien. De senaste åren har man påbörjat vissa regionala studier på C.R.U. Forskningsgruppen fick sålunda i uppdrag att använda matematisk metodik för att belysa regioners rumsliga struktur.

Målsättningen kan sammanfattas i följande punkter:

- (a) utarbeta matematisk-statistisk metodik för studium av samband mellan å ena sidan städers storlek, tillväxt m m och å andra sidan deras rumsliga fördelning,
- (b) samla relevanta data för alla franska städer,
- (c) tillämpa metodiken på dessa data.

Studierna pågår fortfarande, varför jag endast kan presentera några preliminära resultat. Arbetet redovisas i avsnitt 3.

### 1.3 Några C.R.U.-publikationer

#### På franska:

- [1] Bussière, R., 1967, "Modèles de villes".
- [2] Bussière, R., 1972, "Modèle urbain de localisation résidentielle".
- [3] Charré, J.G., Coyaud, L.M., 1966-74, "Les villes françaises: I. Démographie générale, II. Socio-économie, III. Géo-urbanisation, IV. Bases documentaires".
- [4] Coronio, G., Muret, J.P., 1973 "Loisir: du mythe aux réalités".

- [5] Coyaud, L.M., 1974, "L'urbanisation des campagnes".
- [6] Magnan, R., Dryjski, 1970, "Archipoles", (eng.uppl. se nedan).
- [7] Magnan, R., Berthumé, G., Comby, J., 1973, "Conception et instruments de la planification urbaine".
- [8] Ostrowski, W., 1968-70, "L'urbanisme contemporain: I: Des origines à la Charte d'Athènes, II: Tendances actuelles".
- [9] Zeitoun, J., 1971, "Modèles en urbanisme".

På engelska:

- [10] Bussièrre, R., 1970, "The spatial distribution of urban populations".
- [11] Magnan, R., Dryjski, D., 1972, "Archipoles".
- [12] Ostrowski, W., 1970-73, "Contemporary town planning: I. From the origin to the Athens Charter, II. Present trends".

Publikationerna kan rekvireras genom att skriva till:

Centre de Recherche d'Urbanisme  
4, Avenue du Recteur Poincaré  
75016 PARIS  
Frankrike

På begäran erhålles en förteckning över publikationerna ("Catalogue des publications").

## 2. DEN EXPONENTIELLA MODELLEN FÖR BEFOLKNINGSTÄTHETEN I STÄDER

### 2.1 Inledning

En starkt aggregerad, matematiskt lätthanterlig modell för boendebefolkningens fördelning i en stad är den exponentiella modellen som ursprungligen lär ha uppställts av H. Bleicher år 1892 [2], men som ofta tillskrivs C. Clark (1951, [6]), "Clarks formel". Den brukar skrivas

$$D(r) = A \cdot e^{-b \cdot r}, \quad (1)$$

där  $D(r)$  är befolkningstätheten på avståndet  $r$  från stadens centrum (Central Business District).  $A$  och  $b$  är konstanter.  $A$  är tätheten i centrum och  $b$  ett mått på hur snabbt tätheten avtar med avståndet från centrum (se fig.1).

För det första förutsätter alltså modellen existensen av ett väldefinierat centrum. För det andra är den endimensionell så till vida att den endast avser befolkningstäthetens variation med avståndet från centrum och inte dess tangentiella variation.

Clark uppställde sambandet (1) på grundval av empiriska undersökningar. I en del senare arbeten har dock försök gjorts att ge modellen mer teoretisk förankring. Dessa arbeten kan indelas i två huvudgrupper. Den ena gruppen karaktäriseras av att man har velat förankra modellen i ekonomisk teori (se t ex Muth [10] och Amson [1]). I den andra gruppen har man i stället givit sig in på matematisk-statistiska resonemang av samma typ som förekommer i de fysikaliska teorierna för gasers uppträdande (statistisk mekanik). Prov på den senare inriktningen ges t ex av Bussière & Snickars [4] och March [8]. Gemensamt för båda grupperna är att den exponentiella modellen härleds som ett specialfall bland många olika möjliga fördelningar.

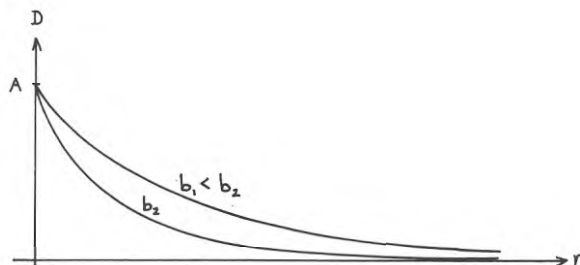


FIG. 1. Den exponentiella modellen

Det arbete som kommer att presenteras i resten av detta kapitel kan sägas vara en fortsättning på Bussièrè & Snickars arbete. Eftersom de teoretiska grunderna för modellen har redovisats på andra ställen (jfr Bussièrè [3], Bussièrè & Snickars [4], Bussièrè [5], Snickars [12] och [13]), kommer endast några grundläggande tankegångar att sammanfattas i ett inledande avsnitt (§ 2.2). Därefter redovisas en grafisk metod att bestämma värdet på parametrarna  $A$  och  $b$  (§ 2.3). Sedan presenteras en felkalkyl med syfte att ge en kvantitativ uppskattning av den osäkerhet i parametervärdena som beror på osäkerhet hos indata (§ 2.4). Sist diskuteras ett mått på avvikelse mellan modell och data (§ 2.5). Måttet kallas redundans och kommer ursprungligen från matematisk kommunikationsteori. De förda resonemangen illustreras i varje avsnitt med resultat från tillämpningar på olika städer. Tonvikten i den följande framställningen är inte avsedd att ligga på det specifika för den exponentiella modellen utan snarare på mer allmänna metodologiska aspekter.

## 2.2 En härledning av modellen

Bussièrè och Snickars [4] visar att den exponentiella modellen kan härledas med matematisk-statistiska metoder grundade på Wilsons entropi-maximeringsprinciper (jfr Wilson [14]). Härledningen utgår från ett explicit val av kostnadsfunktion  $c(r, \theta)$  som definieras som den generaliserade kostnaden att ta sig till stadens centrum för en person som befinner sig vid  $(r, \theta)$ , där  $r$  är avståndet från centrum och  $\theta$  vinkeln. I senare arbeten tolkas  $c(r, \theta)$

som olägenheten att bosätta sig vid  $(r, \theta)$  (se [12]).

Modellen avses vanligen gälla den boende (natt-) befolkningen, men kan naturligtvis även användas vid studium av t ex den arbetande (dag-) befolkningen, utgångspunkter för resor inom staden (för trafikmodeller), markpriser etc.

För förståelse av fortsättningen krävs en sammanfattning av Bussièrës och Snickars' härledning.

Antag att den urbaniserade region vi skall studera bebos av  $N$  personer och upptar ytan  $E$  i det tvådimensionella planet (t ex genom kartprojicering). Antag vidare att det enda vi för övrigt känner till om befolkningens rumsliga fördelning är ett medelvärde  $\bar{c}_x$  av den generaliserade kostnadsfunktionen  $c(r, \theta)$  från ett slumpmässigt urval av  $n$  invånare ( $n \leq N$ ):

$$\bar{c}_x = \frac{1}{n} \cdot \sum_{i=1}^n c(r_i, \theta_i) \quad (2)$$

$$x = \left\{ (r_i, \theta_i) \right\}_{i=1}^n \quad (3)$$

Uppgiften är nu att på grundval av denna bristfälliga kännedom göra en så god uppskattning som möjligt av boende-befolkningens rumsliga fördelning.

Bussière och Snickars definierar därför en kontinuerlig sannolikhetstäthet  $d(r, \theta)$  på  $E$ , där  $d(r, \theta) \cdot r \cdot dr \cdot d\theta$  är sannolikheten att en innevånare är bosatt i ytelementet  $r \cdot dr \cdot d\theta$  vid  $(r, \theta)$  <sup>(1)</sup>. Eftersom  $d$  är en sannolikhetstäthet, så gäller

$$d(r, \theta) \geq 0 \quad (4)$$

$$\int \int_E d(r, \theta) \cdot r dr d\theta = 1 \quad (5)$$

---

(1) Man kan tänka sig följande statistiska experiment: gör upp en förteckning över alla invånare i området och drag sedan en på måfå ur förteckningen. Notera koordinaterna för hans bostad.

Med den bristfälliga kännedom vi har är det enda ytterliga antagande vi kan göra om tätheten  $d$  att den skall vara (väntevärdes-) konsistent med det givna medelvärdet,  $d$  v s :

$$\int_E \int c(r, \theta) \cdot d(r, \theta) r dr d\theta = \bar{c}_x \quad (6)$$

Den mest sannolika fördelningen är den som maximerar entropin

$$H(d) = - \int_E \int d(r, \theta) \cdot \ln(d(r, \theta)) r dr d\theta \quad (7)$$

under bivillkoren (4), (5) och (6).<sup>(1)</sup>

Maximeringsproblemet kan lösas med klassisk variationskalkyl (Lagranges multiplikatormetod). Lösningen är

$$\hat{d}(r, \theta) = \frac{e^{-\hat{b} \cdot c(r, \theta)}}{Z(\hat{b})} \quad (8)$$

där  $\hat{b}$  är Lagrange-multiplikatorn associerad med bivillkoret (6),  $d$  v s  $\hat{b}$  skall väljas så att  $\hat{d}$  satisfierar ekvationen i (6). Partitionsfunktionen  $Z$  är till för att normalisera tätheten  $\hat{d}$  och definieras

$$Z(b) = \int_E \int e^{-b \cdot c(r, \theta)} r dr d\theta \quad (9)$$

-----  
(1) För motivering, se t ex [4].



Lagrange-ekvationen (6) kan alltså skrivas <sup>(1)</sup>

$$-\frac{d}{db} \left[ \ln(Z(b)) \right] = \bar{c}_x \quad (10)$$

Det ovan förda resonemanget ger således följande uppskattning av befolkningstätheten:

$$\hat{D}(r, \theta) = N \cdot \hat{d}(r, \theta) = \frac{N}{Z(\hat{b})} \cdot e^{-\hat{b} \cdot c(r, \theta)} \quad (11)$$

Om vi väljer den generaliserade kostnadsfunktionen

$c(r, \theta) = r$  så fås "Clarks formel":

$$\hat{D}(r, \theta) = \hat{D}(r) = A \cdot e^{-\hat{b} \cdot r} \quad (12)$$

där

$$A = N/Z(\hat{b}) \quad (13)$$

Observera att denna kostnadsfunktion kan tolkas som att olägenheten att bo vid  $(r, \theta)$  är proportionell mot avståndet  $r$  till stadens centrum.

### 2.3 Skattning av parametrarna

I fortsättningen kommer endast kostnadsfunktionen  $c(r, \theta) = r$  och den därav följande exponentiella modellen att användas. Det antages i fortsättningen att  $E$  är cirkulärt med mittpunkt i stadens centrum och har radien  $R$ . I detta specialfall gäller:

-----  
 (1) Antag att vi av andra skäl än ovan (inte nödvändigtvis på grundval av resonemang om entropimaximering) hade omsatt exponentiell form på täthetsfunktionen d. Antag vidare att vi liksom ovan hade ett stickprovs-medelvärde  $\bar{c}_x$ . Med "maximum-likelihood" metodik hade vi kommit fram till samma ekvation (10) för bestämning av parametern  $b$ 's värde.

$$(8) \Rightarrow \hat{d}(r) = \frac{e^{-\hat{b} \cdot r}}{Z(\hat{b})} \quad (14)$$

$$(9) \Rightarrow Z(b) = \frac{2\pi}{b^2} (1 - (1 + Rb) \cdot e^{-Rb}) \quad (15)$$

$$(10) \Rightarrow \frac{1}{b} \left[ 2 - \frac{R^2 b^2}{e^{Rb} - (1 + Rb)} \right] = \bar{r}_x \quad (16)$$

där  $\bar{r}_x$  alltså är det erhållna medelvärde

$$(2) \Rightarrow \bar{r}_x = \frac{1}{n} \cdot \sum_{i=1}^n r_i \quad (17)$$

$$\hat{D}(r) = \hat{A} \cdot e^{-\hat{b} \cdot r} \quad (12)$$

$$\hat{A} = N/Z(\hat{b}) \quad (13)$$

För att bestämma de båda parametervärdena i ett konkret fall, då medelavståndet  $\bar{r}_x$  och yttre radien  $R$  är givna, får vi alltså först försöka finna det  $b$ -värde  $\hat{b}$  som satisfierar ekvation (16) och sedan beräkna  $\hat{A}$  ur (13).

Tyvärr är  $b$  inte direkt lösbar ur ekvation (16).  $\hat{b}$  måste alltså sökas på numerisk eller grafisk väg. Eftersom ekvation (16) ser olika ut för varje kombination av värden på  $R$  och  $\bar{r}_x$  inför vi de två dimensionslösa variablerna

$$s = \bar{r}_x/R \quad (0 \leq s \leq 1) \quad (18)$$

och

$$t = R \cdot b \quad (19)$$

Ekvation (16) kan nu skrivas

$$\frac{2}{t} - \frac{t}{e^t - 1 - t} = s \quad (20)$$

Denna ekvation som varken innehåller  $R$  eller  $\bar{r}_x$  definierar  $t$  som funktion av  $s$  (för  $0 < s < 1$ ).  $t:s$  graf återfinns i figur 2.

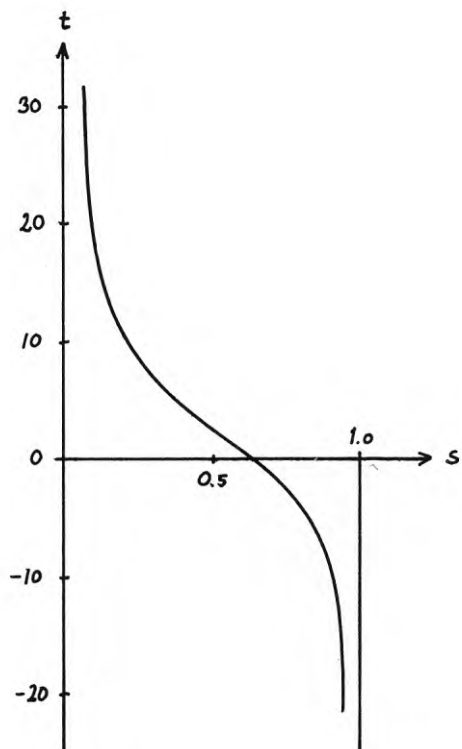


FIG. 2. Den normerede maximum-likelihoodfunktionen  $t(s)$

Som framgår av figuren är  $t$  en monotont avtagande funktion av  $s$ . Detta betyder att för givet värde på  $R$ , så är  $\hat{b}$  stort om  $\bar{r}_x$  är litet och vice versa. Om vi alltså låter  $\bar{r}_x$  öka från ett litet värde upp till  $\frac{2}{3} \cdot R$ , så kommer  $\hat{b}$  att minska ned till noll. Om  $\bar{r}_x$  sedan fortsätter att öka, så kommer  $\hat{b}$  att bli negativ. (1)

Vi kan också göra följande observation: för ett givet stickprov av invånare och fixt värde på ytterradien  $R$ , så beror värdet av  $\bar{r}_x$  naturligtvis på valet av centrum. Det (eller de) centrumval som ger minsta medelavstånd  $\bar{r}_x$  ger också högst värde på  $\hat{b}$ , dvs den spetsigaste fördelningen. Om "stickprovet" omfattar hela befolkningen ( $n=N$ ), så gäller alltså att om stadens "mest tillgängliga" punkt, dvs den punkt som minimerar medelavståndet  $\bar{r}$ , väljs som centrum i modellen, så fås det högsta möjliga värdet på  $\hat{b}$  (2), (3). Eftersom det för detta centrumval gäller  $d\bar{r}=0$ , så kommer även att gälla  $d\hat{b}=0$ , dvs  $\hat{b}$  är okänslig för små förflyttningar av centrumpunkter i närheten av denna punkt.

Med hjälp av den nya variabeln  $t$  kan partitionsfunktionen nu skrivas

$$Z(b) = \pi \cdot R^2 \cdot w(t) \quad (21)$$

där

$$w(t) = \frac{2}{t^2} \cdot \left[ 1 - (1+t) \cdot e^{-t} \right] \quad (22)$$

är en dimensionslös "normaliserad partitionsfunktion". Dess graf ges i figur 3.

- 
- (1) I fallet med ett oändligt område, dvs  $R = \infty$ , fås  $Z(b) = 2\pi/b$  och följaktligen  $\hat{b} = 2/\bar{r}_x$ .
  - (2) Ytterradien  $R$  måste naturligtvis vara så stor att samma befolkning kommer med för alla aktuella centrumpunkter.
  - (3) I vissa "triviala" fall kan det finnas mer än en punkt som ger  $\bar{r}_{\min}$ .

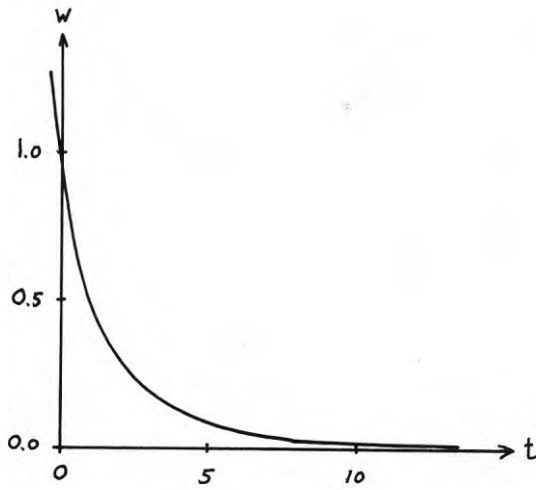


FIG. 3. Den "normerade partitionsfunktionen"

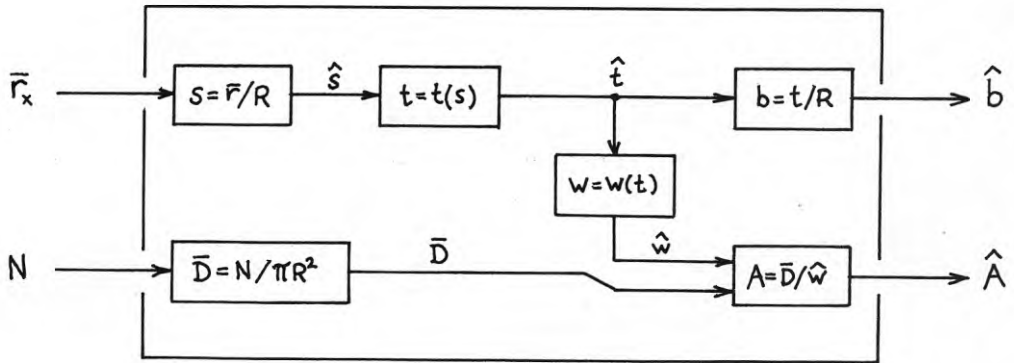


FIG. 4. Schematisering av skattningsmetoden

$\hat{A}$  kan nu skrivas som

$$\hat{A} = \frac{N}{\pi \cdot R^2} \cdot [w(\hat{t})]^{-1} = \bar{D} \cdot \hat{w}^{-1} \quad (23)$$

där  $\hat{t} = \hat{b}/R$ .  $\hat{A}$  kan alltså ses som produkten av medeltätheten  $\bar{D}$  och en viktfaktor  $\hat{w}^{-1}$ .  $\hat{A}$  är mindre än medeltätheten  $\bar{D}$  för negativa  $\hat{b}$  ( $\bar{r}_x > \frac{2}{3}R$ ), lika med  $\bar{D}$  för  $\hat{b} = 0$  ( $\bar{r}_x = \frac{2}{3}R$ ) och slutligen är  $\hat{A}$  större än  $\bar{D}$  för positiva  $\hat{b}$  ( $\bar{r}_x < \frac{2}{3}R$ ), vilket alltsammans är intuitivt högst rimligt.

Den föreslagna skattningsmetoden har schematiserats i figur 4. Varje par av "insignaler" ( $\bar{r}_x, N$ ) ger upphov till ett par av "utsignaler" ( $\hat{b}, \hat{A}$ ).

För att underlätta parameterskattning medelst grafisk metod har de i de flesta praktiska fall aktuella delarna av graferna för funktionerna  $t$  respektive  $w$  uppritats i figur 5 (sid. 25) och figur 6 (sid. 25).

Metoden kanske illustreras bäst av ett tillämpningsexempel: För boendebefolkningen i Bordeaux 1962 erhöles ett medelavstånd  $\bar{r}_x \approx 2,76$  km för  $R = 7$  km. Vi har således  $\hat{s} \approx 0,394$ . Figur 5 (sid. 25) ger  $\hat{t} \approx 4,45$ . Figur 6 (sid. 25) ger  $\hat{w} \approx 0,095$ . Eftersom folkmängden  $N$  inom 7 km:s avstånd från centrum är ungefär 432.000 personer, så är medelbefolkningstätheten  $\bar{D}$  i området ungefär 2.810 personer per kvadratkilometer. Vi får slutligen  $\hat{b} \approx 0,635$  (km<sup>-1</sup>) och  $\hat{A} \approx 29.600$  (personer/km<sup>2</sup>). (jfr resultaten från datorkörningar i tabell 1 sid. 26).

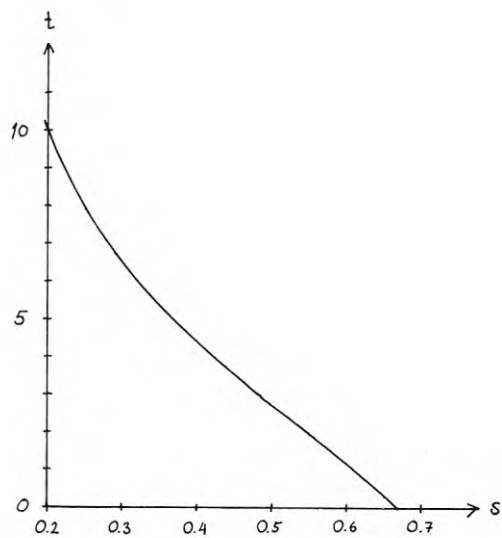


FIG. 5. Del av grafen för  $t(s)$

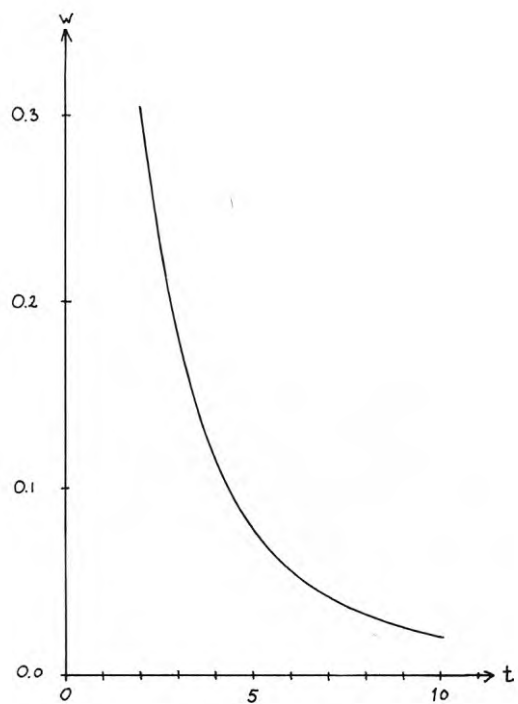


FIG. 6. Del av grafen för  $w(t)$

	R	N	$\hat{b}$	$e_{\hat{b}}$	$\hat{A}$	$e_{\hat{A}}$	redundans
	(km)	(personer)	( $\text{km}^{-1}$ )	( $\text{km}^{-1}$ )	(pers/ $\text{km}^2$ )	(pers/ $\text{km}^2$ )	
BORDEAUX	7,0	430 000	0,6335	0,0006	29 455	50	0,002
LONDON	35,0	9 420 000	0,1101	0,0002	20 252	50	0,008
MARSEILLE	13,0	800 000	0,5058	0,005	33 119	700	0,04
MONTREAL	12,5	1 930 000	0,2702	0,002	26 441	70	0,02
PARIS	25,0	7 520 000	0,2118	0,001	55 467	400	0,01
STOCKHOLM	15,0	1 080 000	0,2569	0,003	12 649	100	0,03
TORONTO	20,0	1 670 000	0,1921	0,001	10 971	10	0,01
TOULOUSE	6,0	310 000	0,8413	0,003	36 149	200	0,01
ZÜRICH	10,0	600 000	0,3544	0,003	13 738	100	0,02

TABELL 1. Tillämpningar av den exponentiella modellen (R=ytterradian, N=folkmängden,  $\hat{A}$  och  $\hat{b}$  de skattade parametervärdena,  $e_{\hat{b}}$  och  $e_{\hat{A}}$  den beräknade osäkerheten i parametervärdena, redundans = ett mått på modellens avvikelse från data).



Skattningsmetoden förutsätter endast ett medelavstånd baserat på ett stickprov. I de allra flesta praktiska fall har man dock tillgång till data för hela befolkningen. I dessa fall kan man alltså låta "stickprovet" omfatta hela befolkningen. "Stickprovsmedelvärdet"  $\bar{r}_x$  sammanfaller då med det "riktiga" medelvärdet  $\bar{r}$ . I samtliga tillämpningar som redovisas i fortsättningen, har vi haft tillgång till fullständiga befolkningsdata.

För en "ideal" stad, d v s en stad för vilken befolknings-  
tätheten avtar exponentiellt ända ut till oändligt avstånd  
från centrum, spelar det ingen roll vilken yttre radie  $R$   
man väljer; det skattade  $b$ -värdet kommer alltid vara  
lika med det "riktiga"  $b$ -värdet. För en verklig stad kommer  
 $\hat{b}$  dock att bero på valet av ytterradie.  $\hat{b}$ :s variationer  
med  $R$  kan alltså ge oss en uppfattning om hur verklighets-  
trogen den exponentiella modellen är.<sup>(1)</sup> Grafen för  
 $\hat{b} = \hat{b}(R)$  för boende-befolkningen i Bordeaux 1962 visas i  
figur 7 (sid. 28 ).

$\hat{b}$ :s beroende av ytterradien  $R$  har också studerats för Paris  
för ett antal år i tidsperioden 1876-1968. Graferna visas i  
figur 8 (sid. 28 ). Man ser att kurvan sjunker och blir  
jämnare för varje år. Under de tidigaste åren ( $\leq 1936$ ) fram-  
träder Versailles ( $r \approx 17$  km) tydligt som en ganska isolerad  
agglomeration ( $\hat{b}$  faller plötsligt). Under de senare åren  
"drunknar" Versailles i Stor-Parisagglomerationen (kurvan  
blir jämn).

---

(1) Genom att öka  $R$  tillräckligt mycket kan man dock få  $\hat{b}$   
att minska hur mycket som helst, vilken verklig stad man  
än väljer. Ty eftersom även en betydande folkmängd  
på stort avstånd från staden då kommer med, så kommer  
 $s = \bar{r}_x/R$  att öka, varav följer att  $t$  och därmed även  
 $\hat{b} = t/R$  minskar.

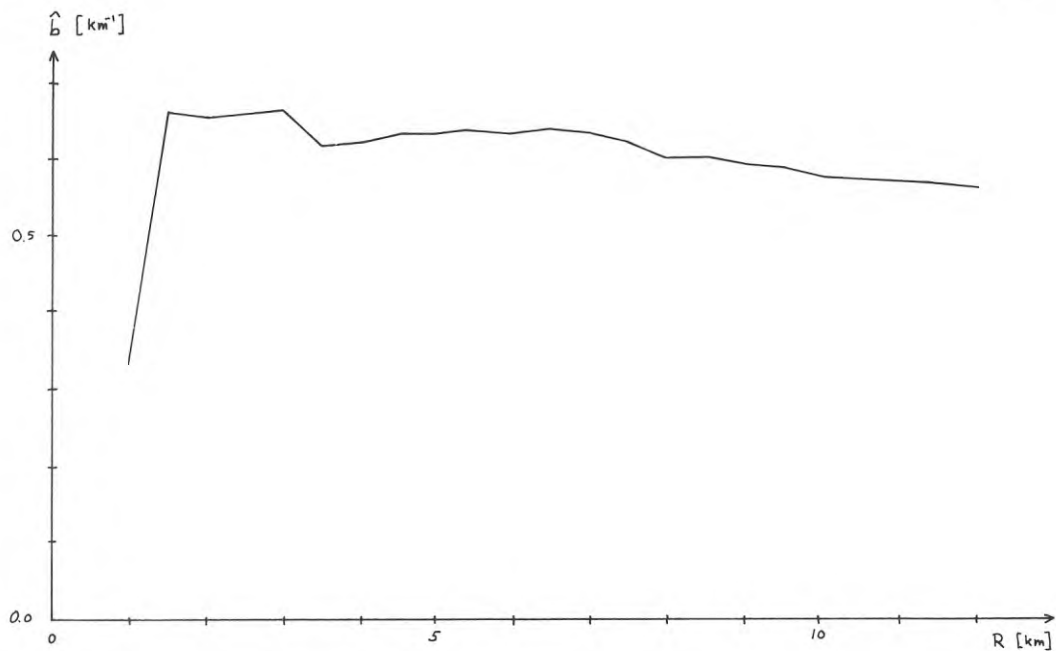


FIG. 7. Parametern  $b$  som funktion av ytterradien  $R$  för Bordeaux 1962

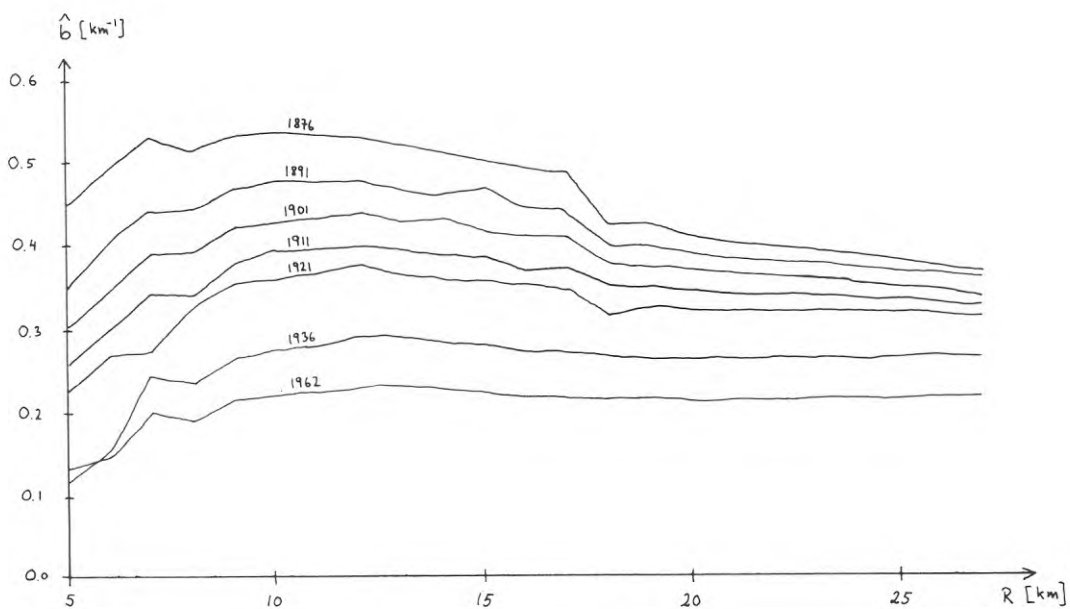


FIG. 8. Parametern  $b$  som funktion av ytterradien  $R$  för Paris 1876-1962

De båda parametrarna A och b har skattats för boendebefolkningen i ett antal städer. Skattningsvärdena ges i tabell 1 (sid. 26 ). Dessa värden, liksom underlaget för graferna i figur 7 och 8, har beräknats på dator. Newton-Raphsons iterationsmetod har fått ersätta den grafiska metoden för bestämning av parametervärdena.

#### 2.4 Känslighetsanalys för $\hat{A}$ och $\hat{b}$

De indata som används vid tillämpningar av modellen är mycket sällan felfria. Det förekommer felaktigheter i folkmängdsuppgifter och även i geografiska koordinater. Avsikten med de följande beräkningarna är att ge en kvantitativ uppfattning av hur stor den osäkerhet i parametervärdena  $\hat{A}$  och  $\hat{b}$  är som beror på denna osäkerhet i indata.

Situationen åskådliggörs i figur 9. Rutan "skattning" representerar den metod att skatta parametrarna som diskuterades i föregående avsnitt (jfr figur 4, sid. 23 ). Med terminologi från elektricitetsläran skulle man kunna uttrycka saken på följande vis: "störningar" eller "brus" i befolkningsdata ger upphov till "brus" överlagrat på "insignalerna"  $\bar{r}$  och N, vilket i sin tur ger upphov till "brus" på "utsignalerna"  $\hat{b}$  och  $\hat{A}$ . Hur stort detta senare brus blir, beror på "skattnings-lådans" överföringsfunktion.



FIG. 9. Schematisk bild av sambandet befolkningsdata - parametervärden

Med metoder från den numeriska felkalkylen har matematiska uttryck härletts för osäkerheten i parametervärdena som funktion av osäkerheten i befolkningsdata. Dessa samband har sedan använts för numeriska beräkningar av osäkerheten i de tidigare presenterade parametervärdena för några olika städer (jfr tabell 1, sid. 26 ).

Det bör kanske påpekas att denna känslighetsanalys inte säger något om överensstämmelsen mellan modell och data, utan endast något om tillförlitligheten hos parametervärdena.

I de tillämpningar som gjorts har befolkningsdata förekommit i form av uppgifter för hela bostadsområden; för varje bostadsområde har folkmängden angivits och med hjälp av en karta har de rumsliga koordinaterna för den befolkningsmässiga tyngdpunkten erhållits.

Låt oss anta att ytterradien är  $R$  och att det finns  $M$  bostadsområden inom detta avstånd från centrum. Vi inför följande beteckningar:

$n_i$  = folkmängden i bostadsområde  $i$   
 $(x_i, y_i)$  = koordinaterna för bostadsområde  $i$   
 $(x_c, y_c)$  = koordinaterna för stadens centrum

$$r_i = \left[ (x_i - x_c)^2 + (y_i - y_c)^2 \right]^{\frac{1}{2}} \quad (0 \leq r_i \leq R)$$

$$N = \sum_{i=1}^M n_i$$

$$\bar{r} = \frac{1}{N} \cdot \sum_{i=1}^M n_i \cdot r_i \quad (0 \leq \bar{r} \leq R)$$

d v s  $r_i$  är avståndet från bostadsområde  $i$  till stadens centrum,  $N$  den totala folkmängden inom avståndet  $R$  från centrum och  $\bar{r}$  medelavståndet till centrum.

Eftersom antalet bostadsområden i allmänhet är mycket stort ( $M \gg 10$ ), kommer medelfelen och inte maximalfelen att beräknas.

Det generella uttrycket för medelfelet  $e_F$  hos ett funktionsvärde  $F(x_1, \dots, x_n)$  är

$$e_F \approx \left[ \sum_{i=1}^n \left( \frac{\partial F}{\partial x_i} \right)^2 \cdot e_i^2 \right]^{1/2}$$

under förutsättning att felen  $\Delta x_1, \dots, \Delta x_n$  i  $x$ -värdena är oberoende stokastiska variabler med standardavvikelserna  $e_1, \dots, e_n$  <sup>(1)</sup>.

Från ovanstående uttryck följer genast

$$e_{\bar{r}} \approx \frac{1}{N} \cdot \left[ \sum_{i=1}^M n_i^2 \cdot e_{r_i}^2 + \sum_{i=1}^M (r_i - \bar{r})^2 \cdot e_{n_i}^2 \right]^{1/2} \quad (24)$$

$$e_{r_i} \approx \frac{1}{r_i} \cdot \left[ (x_i - x_c)^2 \cdot e_{x_i}^2 + (y_i - y_c)^2 \cdot e_{y_i}^2 \right]^{1/2} \quad (25)$$

där  $e_{x_i}$  och  $e_{y_i}$  är medelfelen hos koordinaterna för bostadsområde  $i$ ,  $e_{r_i}$  det resulterande medelfelet i  $r_i$  och slutligen  $e_{n_i}$  medelfelet i folkmängdsuppgiften för bostadsområde  $i$ .

Eftersom  $\hat{\delta}$  beräknas från  $\bar{r}$  via sambanden (18), (19) och (20) i föregående avsnitt, så följer det att felet  $\Delta \hat{\delta}$  i  $\hat{\delta}$  kan approximeras som

$$\Delta \hat{\delta} \approx R^{-2} \cdot \left. \frac{d\hat{\delta}}{ds} \right|_{s=\hat{s}} \cdot \Delta \bar{r} \quad (26)$$

För medelfelen  $e_{\hat{\delta}}$  och  $e_{\bar{r}}$  gäller alltså sambandet

$$e_{\hat{\delta}} \approx R^{-2} \cdot \left. \frac{d\hat{\delta}}{ds} \right|_{s=\hat{s}} \cdot e_{\bar{r}} \quad (27)$$

(1) Eftersom felen  $\Delta x_i$  är stokastiska variabler är  $\Delta F$  också en stokastisk variabel. Standarddeviationen hos  $\Delta F$  (och således också hos  $F$ ) kallas medelfelet i  $F$  och betecknas  $e_F$ .

På liknande sätt erhålls medelfelet  $e_A$  i parameter-  
värdet  $\hat{A}$ :

$$e_A \approx \frac{1}{Z(\hat{b})} \cdot \left[ \sum_{i=1}^M e_{n_i}^2 + \left( \frac{N}{Z} \cdot \left[ \frac{dZ}{db} \right]_{b=\hat{b}} \right)^2 \cdot e_{\hat{b}}^2 \right]^{\frac{1}{2}} \quad (28)$$

Med hjälp av sambanden (24)-(28) kan vi nu beräkna medelfelen i parametervärdena  $\hat{A}$  och  $\hat{b}$  som funktioner av medelfelen i indata. Dessa är dock sällan kända. Däremot brukar man kunna uppskatta dem approximativt. Vi gör följande för-  
enklande antaganden:

- (A1) Medelfelen  $e_{x_i}$  och  $e_{y_i}$  hos koordinaterna  $(x_i, y_i)$  för bostadsområde  $i$  kan approximeras av

$$e_{x_i} \approx e_{y_i} \approx \alpha \cdot \sqrt{a_i}, \quad i=1, \dots, M$$

där  $a_i$  är bostadsområdets yta och  $\alpha$  en (dimensionslös) icke-negativ konstant. (1)

---

(1) "Riktiga koordinater för bostadsområde  $i$  är sådana som gör  $r_i$  identisk med det "riktiga" medelavståndet från område  $i$ , d v s medelavståndet beräknat på basis av koordinaterna för varje invånare i området.

(A2) Det relativa felet i folkmängdsuppgifterna är approximativt lika stort för alla bostadsområden, d v s

$$e_{n_i} \approx \beta \cdot n_i, \quad i = 1, \dots, M$$

där  $\beta$  är en icke-negativ konstant.

Följande fiktiva exempel kan kanske illustrera innebörden av de två antagandena: För ett rektangulärt bostadsområde av formatet 400 x 500 m<sup>2</sup> och med en (officiell) folkmängd av 800 personer, ger (A1) och (A2)  $e_{x_i} \approx e_{y_i} \approx 110$  m och  $e_{n_i} \approx 8$  pers., under förutsättning att  $\alpha = 25\%$  och  $\beta = 1\%$ .

Med användning av antagandena (A1) och (A2) får vi slutligen

$$e_{\hat{b}} \approx \frac{1}{N \cdot R^2} \cdot \left| \frac{dt}{ds} \right|_{s=\hat{s}} \cdot \left[ \left( \sum_{i=1}^M n_i^2 \cdot a_i \right) \cdot \alpha^2 + \left( \sum_{i=1}^M (r_i - \bar{r})^2 \cdot n_i^2 \right) \cdot \beta^2 \right]^{\frac{1}{2}} \quad (29)$$

$$e_{\hat{A}} \approx \frac{1}{Z(\hat{b})} \cdot \left[ \left( \frac{N}{Z(\hat{b})} \cdot \left| \frac{dZ}{db} \right|_{b=\hat{b}} \right)^2 \cdot e_{\hat{b}}^2 + \left( \sum_{i=1}^M n_i^2 \right) \cdot \beta^2 \right]^{\frac{1}{2}} \quad (30)$$

Dessa båda samband möjliggör nu numerisk uppskattning av medelfelen hos parametervärdena  $\hat{A}$  och  $\hat{b}$ , under förutsättning att befolkningsdata är tillgängliga och att  $\alpha$  och  $\beta$  har givits numeriska värden. (1)

Numeriska beräkningar (på dator) av medelfelen i de tidigare presenterade parametervärdena för några olika städer visas i tabell 1 (sid. 26).  $\alpha$  och  $\beta$  har satts till 25% respektive 1% i samtliga fall.

-----

(1) Om bostadsområdenas ytor ( $a_i$ ) är okända, kan de teoretiskt uppskattas genom indelning av den cirkulära ytan innanför  $R$  i ett antal ringar (eller sektorer). Genom att räkna antalet bostadsområden i varje ring (eller sektor) kan ytorna sedan uppskattas.

## 2.5 Modellens överensstämmelse med data

Vanligen utvärderar man statistiska modeller genom att utföra ett signifikanstest. Om testet inte visar att modellen signifikativt avviker från data, så accepteras modellen (hypotesen) och allt är frid och fröjd. Om avvikelser emellertid är signifikativa, så ger testet i allmänhet inte någon direkt upplysning om storleken av avvikelserna.

Eftersom många modeller för samhällsliga fenomen visar statistiskt signifikativa avvikelser från data utan att för den skull vara ointressanta från praktisk synpunkt, så skulle ett mått på en modells avvikelse från data vara värdefullt bl a i dessa sammanhang. Martin-Löf [9] har nyligen föreslagit ett sådant mått, nämligen redundans, ett begrepp från matematisk kommunikationsteori (Shannon & Weaver [11]).

Martin-Löfs mått är dock bara tillämpligt för diskreta sannolikhetsfördelningar, och vi studerar en kontinuerlig. Vi måste därför göra vissa modifieringar. Låt oss emellertid först återge några definitioner som gäller det diskreta fallet.

Entropin hos en diskret sannolikhetsfördelning definieras som

$$H = -K \cdot \sum_{i=1}^n p_i \cdot \log p_i \quad (31)$$

där  $K$  är en positiv konstant (jfr [11]).

En statistisk modells redundans i förhållande till data definieras som

$$R = 1 - \frac{H_{\text{obs}}}{\hat{H}} \quad (32)$$

där  $H_{\text{obs}}$  är entropin hos den relativa frekvensfördelningen som erhålls från data och  $\hat{H}$  entropin hos den sannolikhetsfördelning som modellen ger. Det förutsätts härvid att modellen



är baserad på maximum-entropi- eller maximum-likelihood-teknik. Denna förutsättning garanterar att  $\hat{H}$  aldrig är mindre än  $H_{obs}$ . Av detta förhållande och faktumet att båda entropierna är icke-negativa följer att redundansen alltid ligger i intervallet  $[0,1]$ . Redundansen är noll om och endast om modellen överensstämmer fullständigt med data.

En sannolikhetsfördelningens entropi kan tolkas som ett mått på osäkerheten om utfallet. Ju "bredare" sannolikhetsfördelning, desto större entropi. Redundansen kan alltså tolkas som den ökning av osäkerhet om utfallet som modellen ger i jämförelse med data. Ju mindre denna ökning är, desto bättre överensstämmer modellen med data.

Redundansen är ett "absolut" mått i det avseendet att den möjliggör direkt jämförelse mellan helt olika statistiska modeller. Martin-Löf [9] föreslår följande normativa tabell:

<u>Redundans</u>	<u>Värdeomdöme</u>
1	meningslös
0,1	mycket dålig
0,01	dålig
0,001	bra
0,0001	mycket bra

TABELL 2. Martin-Löfs normativa tabell

Så långt det diskreta fallet.

Problemet med det kontinuerliga fallet är att man inte kan jämföra modellens entropi direkt med entropin för data. Modellen representeras ju av en kontinuerlig funktion, medan data endast består av en svit diskreta mätvärden.

För att lösa detta problem föreslår Snickars [13] uppdelning av det studerade urbaniserade området i  $M$  delområden av storlek  $s_1, \dots, s_M$  (ej nödvändigtvis samma indelning som i känslighetsanalysen). Ekvationerna (5), (6) och (7) kan då diskretiseras enligt

$$\begin{aligned}
 H &= - \sum_{i=1}^M s_i \cdot d_i \cdot \ln(d_i) & (33) \\
 \sum_{i=1}^M c_i \cdot d_i \cdot s_i &= \bar{c} \\
 \sum_{i=1}^M d_i \cdot s_i &= 1
 \end{aligned}$$

Här är  $d_i$  tätheter definierade enligt  $d_i = p_i/s_i$  där  $p_i = n_i/N$ ,  $n_i$  folkmängden i delområde  $i$  och  $s_i$  ytan. generaliserade kostnad som tillskrivs område  $i$  och  $\bar{c}$  medelvärdet av den diskretiserade kostnadsfunktionen.

På motsvarande sätt som i det kontinuerliga fallet kan man nu härleda en diskret exponentiell täthetsfördelning med hjälp av maximum-entropi-teknik. Lagrange-ekvationen för bestämning av värdet av parametern  $b$  blir formellt likadan; den enda skillnaden är att partitionsfunktionen nu är en summa i stället för en integral.

Vi kan nu beräkna redundansen för den sålunda diskretiserade exponentiella modellen genom att använda uttrycket (33) för entropierna i definitionen (32) av redundansen.

Det återstår emellertid ett problem, och det är att värdet av den sålunda beräknade redundansen kommer att bero på valet av ytenhet: antag att vi byter skala  $s'_i = \gamma \cdot s_i$  (t ex från  $\text{km}^2$  till  $\text{m}^2$ ), då kommer det nya redundansvärdet bli

$$R' = \frac{\hat{H}}{\hat{H} + \ln(\gamma)}, \quad R' \neq R \text{ om } \gamma \neq 1$$

För att undvika detta godtycke, normaliserar vi ytorna på följande vis. Om  $a_i$  betecknar ytan av område  $i$  uttryckt i någon viss ytenhet (t ex  $\text{km}^2$ ), så låt  $s_i$  definieras som

$$s_i = k \cdot \left( \sum_{i=1}^M a_i \right)^{-1} \cdot a_i$$

där  $k$  är en konstant. Vi väljer värdet på konstanten så att entropin definierad enligt (32) blir konsistent med entropin i (31) i det specialfall då alla delområden är lika stora. Det enda värde som satisfierar detta villkor är  $k = M$  ( $M =$  antalet delområden).

Detta konsistensvillkor är naturligt därför att om alla delområden vore lika stora, så kunde vi lika väl direkt använda det ursprungliga uttrycket (31) för entropin. För att redundansvärdet skall vara oberoende av vilken metod som används, så måste konsistensvillkoret ovan vara uppfyllt.

Den föreslagna "hybrida" redundansen är alltså identisk med Martin-Löfs redundans i specialfallet med lika stora delområden. Men vilka är den "hybrida" redundansens egenskaper i det allmänna fallet?<sup>(1)</sup>

För det första kan vi notera att den "hybrida" redundansen aldrig är negativ. Liksom i det "genuint diskreta" fallet är redundansen noll om och endast om den diskretiserade modellen överensstämmer med de likaledes diskretiserade data. Det gäller även i detta fall att ju lägre redundansvärde, desto bättre kan modellen sägas överensstämma med data.

-----  
 (1) Den "hybrida" redundansen kan ses som en generalisering av Martin-Löfs i det avseendet att entiteterna  $s_i$  utgör "vikter" som à priori tillordnas de olika möjliga utfallen. Specialfallet med lika "vikter" innebär att lika sannolikheter à priori tillskrivs varje möjligt utfall.

Det finns emellertid en - om inte praktiskt så åtminstone principiellt - viktig skillnad mellan den "hybrida" och den "genuint diskreta" redundansen. Nämligen att värdet av den "hybrida" redundansen i det allmänna fallet kan överskrida 1. Vid närmare eftertanke finner man att detta är intuitivt "riktigt" <sup>(1)</sup>.

Vi är nu äntligen rustade att utvärdera modellens överensstämmelse med data i de tidigare diskuterade tillämpningarna. Vid dessa tillämpningar har endast befolkningstäthetens beroende av avstånd till centrum och inte dess variationer i vinkelled studerats. Vid diskretiseringen har den cirkulära ytan därför delats upp i ett antal koncentriska ringar (av lika tjocklek).

Vid alla tillämpningar har modellen som väntat visat signifikativa avvikelser från data. Exempelvis ger ett signifikans-test på Bordeaux ett  $\chi^2$ -värde på 3.500 för 19 frihetsgrader, vilket kan jämföras med 0,1-percentilvärdet  $K_{99,9}^{-1}$  (19)  $\approx 44$ . (Vid betraktande av dessa resultat bör man dock komma ihåg att instrumentella fel beroende på datas grovhet i förhållande till ringvidden kan bidra till de höga  $\chi^2$ -värdena.)

För att mäta modellens avvikelse från data har ("hybrida") redundansvärden beräknats. De återfinns i tabell 1 (sid. 26 ). Som synes varierar de från "mycket dålig" (Marseille) till "bra" (Bordeaux). (I fallen Marseille, Stockholm och Zürich var data förhållandevis grova, varför man kanske inte bör fästa alltför stor tilltro till dessa redundansvärden).

Låt oss med hjälp av redundansen jämföra den exponentiella modellen med en annan modell! Ett - för den med den

- (1) Denna egenskap hos den "hybrida" redundansen beror på att
- $$\min_i \left\{ s_i \right\}_{i=1}^M \leq 1.$$

Om vid en tillämpning av den exponentiella modellen t ex hela befolkningen påträffades i ett enda bostadsområde med en yta mindre än genomsnittet, så skulle redundansen överskrida 1.

exponentiella modellen obekante - naturligt sätt att uppskatta befolkningstätheten i en stad vore att helt enkelt dividera stadens folkmängd med dess yta. I vårt studium av befolkningstäthetens variation med avståndet från centrum skulle detta tillvägagångssätt ge ansatsen  $D(r) = \bar{D}$  ( $0 \leq r \leq R$ ). I figur 10 kan vi i ett konkret fall (London 1966, jfr tabell 1, sid. 26) jämföra denna likformiga modell med den exponentiella. Som synes överensstämmer den exponentiella modellen trots sina brister bättre med data. Detta förhållande avspeglas också av redundansvärdena. Vi hade tidigare erhållit redundansvärdet 0,008 för den exponentiella modellen. Enkel handräkning ger ett redundansvärde på 0,16 för den likformiga modellen. Den exponentiella modellen kan alltså i detta fall närmast klassificeras som "dålig" medan den likformiga modellen är "mycket dålig" (jfr tabell 2, sid. 35).

Man kan se dessa modeller som två steg i en reduktionskedja från observationsdata ned till å prioriantagande om likformighet, där redundansen mäter den relativa ökningen i osäkerhet i varje steg. Detta sätt att betrakta en modell möjliggör bedömning av hur mycket "information" modellen ger, i jämförelse med ett rent å prioriantagnade.

I vårt exempel visar redundansvärdena att merparten av ökningen i osäkerhet sker i steget från exponentialmodellen till den likformiga modellen (se figur 11).

Detta förhållande visar att avståndet till centrum har betydelse för befolkningstäthetens rumsliga fördelning.

De erhållna redundansvärdena möjliggör också jämförelse av exponentialmodellen med helt andra statistiska modeller. Exempelvis erhöll Karlqvist & Marksjö [7] ett redundansvärde på 0.004 när de tillämpade en gravitationsmodell på resor i Stor-Stockholm.

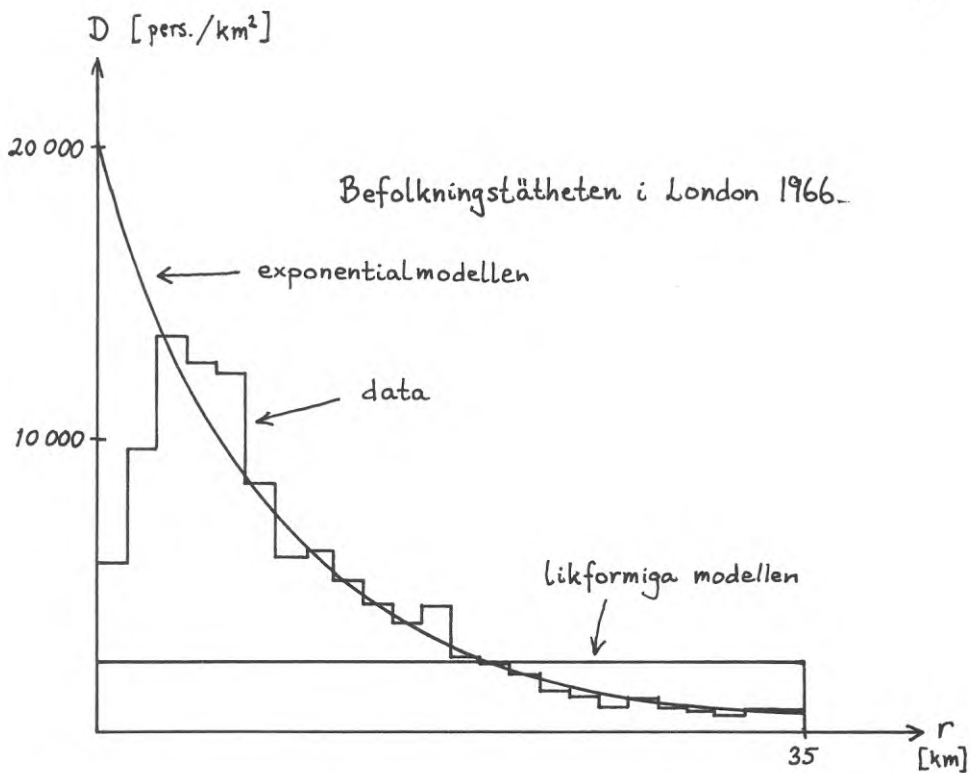


FIG. 10. Jämförelse mellan två modeller för London 1966

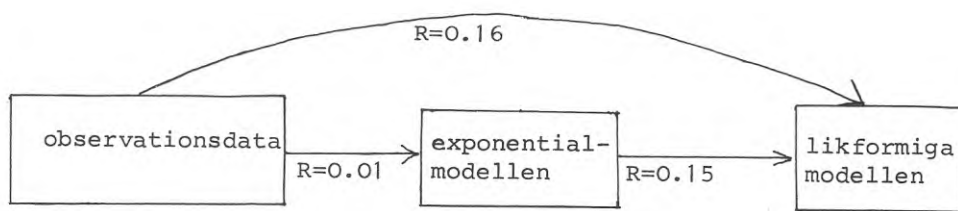


FIG. 11. Successiva modellreduktioner

## 2.6 Sammanfattning

Redogörelsen har huvudsakligen ägnats vissa tekniska frågor. Sålunda har dels en mättekniskt enkel, grafisk metod att skatta de båda parametrarna redovisats. Dessutom har en kalkyl för beräkning av osäkerheten i parametervärdena presenterats. Slutligen har ett mått på modellens avvikelse från observationsdata diskuterats.

Som ett allmänt omdöme om modellen kan det avslutningsvis konstateras att den visserligen är beräkningsmässigt lätthanterlig, men i många fall visar bristande överensstämmelse med observationsdata. I de flesta fall består denna diskrepans i att modellen överskattar boendetätheten i centrum. Genom att använda någon annan generaliserad kostnadsfunktion än det euklidiska avståndet till centrum skulle dock bättre överensstämmelse med data kunna erhållas, men förmodligen till priset av minskad beräkningsmässig lätthanterlighet.

## 2.7 Referenser

- [1] Amson, J.C., 1972, "The dependence of population distribution on location costs", *Environment and Planning*, 1972 vol.4.
- [2] Bleicher, H., 1892, "Statische Beschreibung der Stadt Frankfurt am Main und ihrer Bevölkerung",
- [3] Bussière, R., 1968, "Morphologie urbaine, repartition de la population", C.R.U.
- [4] Bussière, R., Snickars, F., 1970, "Derivation of the negative exponential model by an entropy maximizing method", *Environment and Planning*, 1970 vol.2.

- [5] Bussière, R., 1972, "Modèle urbain de localisation résidentielle", C.R.U.
- [6] Clark, C., 1951, "Urban population densities", Journal of the Royal Statistical Society, Series A, 114.
- [7] Karlqvist, A., Marksjö, B., 1971, "Statistical urban models", Environment and Planning, 1971 vol.3.
- [8] March, L., 1970, "A statistical theory of simple spatial distributions", Institute of Land Use and Built Forms Studies Working Paper Number 5, University of Cambridge.
- [9] Martin-Löf, P., 1970, "Statistiska Modeller", anteckningar från seminarier 1969-70 vid Stockholms Universitet, stencil.
- [10] Muth, R.F., 1961, "The spacial structure of the housing market", Papers and Proceedings of the Regional Science Association, 7.
- [11] Shannon, C.E., Weaver, W., 1949, "The mathematical theory of communication", University of Illinois Press (ny upplaga 1972).
- [12] Snickars, F., 1971, "Några exempel på fransk urbaniseringsforskning", Stockholms Läns Landsting, Regionalplane-kontoret, Meddelande 1971:2.
- [13] Snickars, F., 1972, "En modell för den rumsliga befolkningsfördelningen i vissa svenska städer", Närförläggningensutredningen, NÄR-185:1.
- [14] Wilson, A., 1968, "Notes on some concepts in social physics", Center for Environmental Studies, Working Paper 4.



### 3. STÄDER SOM PUNKTER I PLANET

#### 3.1 Inledning

I mitten på femtiotalet utvecklades inom kvantitativ biologi matematisk-statistiska metoder för analys av punktmönster för studium av hur individer tillhörande en viss art är fördelade över en yta. Dessa metoder har på senare år även börjat användas inom matematiskt orienterad kulturgeografi. Det är en speciell typ av sådan metodik, s k "k:th nearest neighbour analysis", som kommer att presenteras nedan, med vissa tillämpningar på de franska städernas fördelning.

Men först en kortfattad presentation av några klassiska arbeten inom området.

#### 3.2 En allmän orientering

För att skapa sig någon överblick över den uppsjö av arbeten som återfinns i litteraturen kan man kanske göra följande grova indelning i tre "traditioner" eller "riktningar":

##### Centralortsteori

Denna "klassiska" riktning bygger på W. Cristallers verk "Die zentralen Orte in Süddeutschland (publicerad 1933, [6]) och A. Löschs arbete "The Economics of Location" (engelsk översättning 1954, [13]).

Cristaller försökte förklara städers rumsliga och storleksmässiga fördelning utifrån ekonomisk teori. Från vissa idealiserande geografiska och ekonomiska antaganden härledde han en hierarkisk struktur av städer på olika "funktionsnivå", rumsligt fördelade i ett hexagonalt mönster (se figur 1).

Cristallers geografiska antaganden kan sammanfattas i tre punkter (Angel & Hyman [1]):

- (i) Området antas vara plant och sammanhängande
- (ii) Relevanta egenskaper hos området antas vara jämnt fördelade. (Sådana egenskaper kan t ex vara råvaruresurser, markens bördighet, "köpkraft", jordbrukarbefolkningens täthet.)
- (iii) Förflyttningar antas kunna ske lika lätt överallt, oberoende av läge och riktning för förflyttningen.

Dessa antaganden är som synes starkt idealiserade. I en nyligen gjord studie undersöker Angel och Hyman [1] möjligheterna att matematiskt transformera det verkliga geografiska rummet med dess varierande tätheter och transportkostnader till något abstrakt rum där dessa idealiserade antaganden skulle kunna gälla. Författarna visar att det finns transformationer som ger båda egenskaperna (i) och (ii), eller enbart egenskap (iii), men inga andra kombinationer. Det finns alltså inte någon transformation som ger alla tre egenskaperna samtidigt (se Angel & Hyman [1]).

En allmän översikt över centralortsteori ges i Robinson [15].

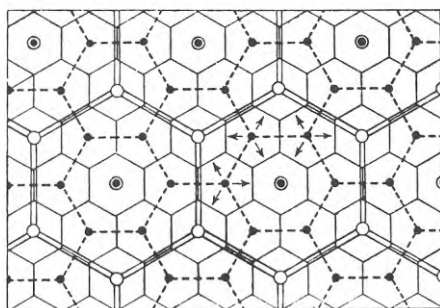


FIG. 1. Ett Centralortssystem à la Cristaller

### Rank-size rule

En annan riktning har sin upprinnelse i F. Auerbachs arbete "Das Gesetz der Bevölkerungskonzentration", vilket publicerades redan 1913, [2]. Till skillnad från i centralortsteori återfinns här inte någon egentlig teoretisk grund, utan endast ett empiriskt funnet samband. Sambandet brukar kallas rank-size rule och tillskrivs ofta G. K. Zipf, "Zipfs lag", (1941, [19]). Som hörs av benämningen gäller sambandet städernas rang och deras storlek. Däremot säger det ingenting om den rumsliga fördelningen av städerna.

I sin enklaste form säger "lagen" att om alla städer i en region (eller ett land) ordnas efter (befolknings-) storlek, så att den största staden får rang 1, den näst största 2 osv, så kommer storleken av den r:te staden att vara 1/r:te del av storleken hos den största staden. Mer generellt kan sambandet skrivas

$$P_r = \alpha \cdot r^{-\beta} \quad (1)$$

eller ekvivalent

$$\log(P_r) = \log(\alpha) - \beta \cdot \log(r) \quad (2)$$

där r står för rang, P för folkmängd och  $\alpha$  och  $\beta$  är konstanter (specialfallet ovan motsvaras alltså av  $\alpha = P_1$ ,  $\beta = 1$ ). Av den logaritmiska versionen framgår det att sambandet motsvaras av en rät linje med lutningen  $-\beta$  i ett log-log-diagram med rang på ena axeln och folkmängd på den andra (se figur 2).

Rank-size rule visar vid de flesta tillämpningar endast approximativ överensstämmelse med data<sup>(1)</sup>. Trots att "lagen" (ännu) saknar teoretisk förankring och endast har

-----  
 (1) Empiriska undersökningar är behäftade med vissa svårigheter. Resultatet påverkas t ex av hur städernas folkmängd beräknas och av hur det studerade området avgränsas. Dessutom är det tveksamt vilken signifikans sambandet kan tillskrivas.

viss empirisk täckning, så används den ibland som norm. Så nämner t ex Zipf avvikelser från rank-size rule som symptom på bristande socio-ekonomisk jämvikt (jfr Steward [17]).

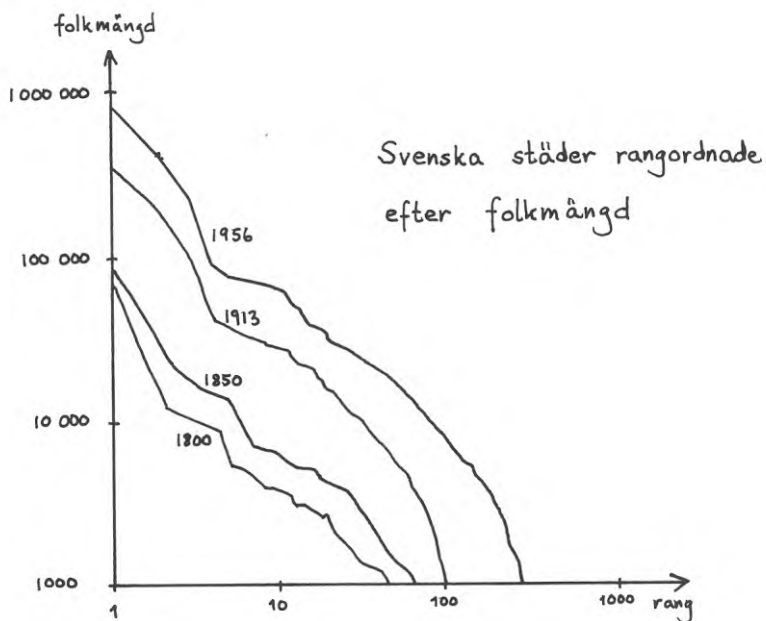


FIG. 2.

Trots att rank-size rule och centralortsteorin kan tyckas ha mycket lite med varandra att göra så har försök gjorts att förena de två. Ett exempel på ett sådant försök är M. Beckmanns artikel "City Hierarchies and the Distribution of City Size" 4 .

#### Punktmönster-analys

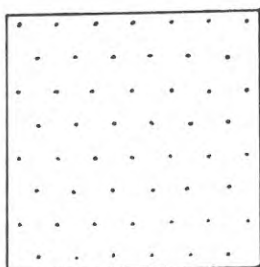
En modern riktning inom kvantitativ kulturgeografi bygger på statistisk metodik inom kvantitativ biologi för studium av arters geografiska spridning. Det är två huvudtyper av metoder som då kommit till användning: "quadrat analysis" (QA) och "k:th nearest neighbour analysis" (NA). I båda fallen är ett för oss relevant delsyfte att avgöra om punkterna (plantorna, städerna) är slumpmässigt fördelade eller om den rumsliga fördelningen är mer regelbunden eller agglomererad (jfr figur 3.

I QA försöker man lösa uppgiften genom att dela in studieområdet i ett rutnät av lika stora rutor, och antalet punkter i varje ruta (eller i ett stickprov av rutor) räknas. Med utgångspunkt i den sålunda erhållna frekvensfördelningen försöker man sedan göra omdömen om den rumsliga fördelningen.

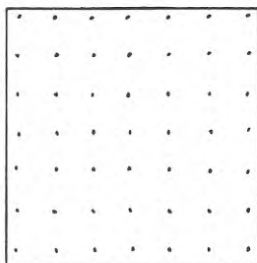
Åtskilliga problem är förknippade med QA. Bland annat kommer rutnätets placering och rutornas storlek att påverka resultaten.

En metodik som mäter ett observerat punktmönsters avvikelse från slumpmässighet men som inte lider av vissa av QA:s brister, är NA. I denna metodik mäter man avståndet från en godtycklig punkt till den k:te närmsta punkten i punktmönstret ( $k = 1$ : allra närmsta;  $k = 2$ : näst närmsta o s v). Ett medelvärde från sådana mätningar jämförs sedan med motsvarande förväntade värde för ett slumpmässigt punktmönster.

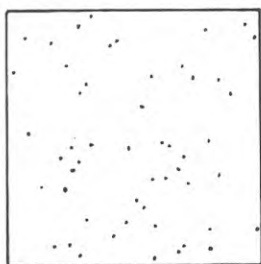
En brist hos NA är att den för ett fixt värde på  $k$  inte "ser" skillnad på mycket olika mönster (jfr t ex figur 3.d med figur 4, som båda ger samma resultat för  $k = 1$ ).



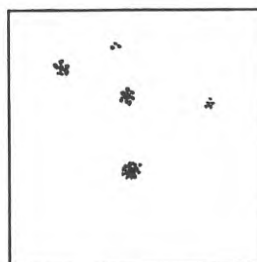
a. regelbundet tri-  
angulärt mönster.



b. regelbundet kva-  
dratiskt mönster.



c. slumpmässigt  
punktmönster.



d. agglomererat  
punktmönster.

FIG. 3. Några typiska punktmönster

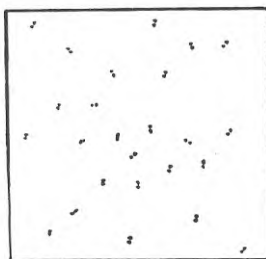


FIG. 4. Ett parvis agglomererat  
punktmönster

Exempel på QA tillämpad inom urbaniseringsforskning ges av Rogers [16]. Exempel på NA inom samma tillämpningsområde återfinns bl a hos: M. Dacey [7], [8] och [9], B. Barr m fl [3], P. Holgate [10] och [11].

### 3.3 Några statistiska egenskaper

I detta avsnitt kommer några statistiska egenskaper hos slumpmässiga punktmönster att härledas. Framför allt kommer intresset att riktas mot avståndet från en godtycklig punkt i planet till den  $k$ :te närmsta punkten ( $k = 1$ : allra närmsta,  $k = 2$ : näst närmsta o s v). Det mig veterligen tidigast publicerade arbetet är en artikel av Clark & Evans i tidskriften *Ecology* (1954, [5]). Författarna härledde där uttryck för fördelningsfunktion m m i fallet  $k = 1$ . Ett par år senare publicerades en artikel av Thompson [18] i samma tidskrift, denna gång med härledning för godtyckligt värde på  $k$ . Den härledning som presenteras nedan grundar sig inte direkt på någon av dessa båda artiklar, utan är delvis författarens egen skapelse. (Däremot ansluter framställningen delvis till Rogers [16])<sup>(1)</sup>.

Ett grundläggande begrepp i detta sammanhang är "slumpmässighet". Det är därför nödvändigt att ge det en entydig definition.

Låt oss föreställa oss ett statistiskt försök, vars utfall är en punkt  $x$  inom ett visst område  $A$  i planet<sup>(2)</sup>.

Detta försök kan kallas det enskilda försöket. Låt oss vidare kalla  $N$  oberoende upprepningar av detta försök för det stora försöket. Ett utfall  $X$  av detta stora försök

-----

- (1) Förståelse av alla tekniska detaljer i den följande härledningen är inte nödvändig för förståelse av fortsättningen. Läsare som inte nås av en alltför stark lidelse för matematisk statistik kan därför med gott samvete nöja sig med endast en flyktig granskning av de följande sidorna.
- (2) De studerade områdena antas vara "matematiskt snälla" områden i det Euklidiska planet.

består alltså av  $N$  stycken punkter inom  $A$ ,  $X = \{x_1, \dots, x_N; x_i \in A\}$ .

Om nu punkten i det enskilda försöket har lika stor chans att hamna var som helst inom  $A$ , eller mer formellt uttryckt om  $x$  är rektangulärfördelad på  $A$ , i så fall kallas utfallet från motsvarande stora försök för ett (rektangulärfördelat) slumpmässigt punktmönster. <sup>(1)</sup>

Det följer omedelbart från definitionen att antalet punkter ( $n$ ) inom ett godtyckligt delområde  $a \subseteq A$  är binomialfördelat,  $n \sim \text{Bin}(N, p)$ , där  $p = S_a/S_A$  ( $S_a$  och  $S_A$  betecknar respektive ytor).

Ett grundläggande resultat i sannolikhets teorin är att en binomialfördelning  $\text{Bin}(N, p)$  tenderar mot en Poisson-fördelning om  $N$  ökar och  $p$  minskar på sådant sätt att produkten  $N \cdot p$  är konstant. Om vi således antar att  $N$  är stort samt att  $a$  är mycket litet i jämförelse med  $A$ , så kommer  $n$  att vara approximativt Poisson-fördelat,  $\text{Poisson}(\lambda \cdot S_a)$ , där  $\lambda = N/S_A$ , d v s  $\lambda$  är medelantalet punkter per ytenhet.  $\lambda$  kommer att kallas punktmönstrets intensitet.

Vi kan nu härleda ett relativt enkelt matematiskt uttryck för sannolikhetsfördelningen för avståndet  $r_k$  från en godtycklig punkt  $y$  inom  $A$  (ej nödvändigtvis en punkt i det studerade punktmönstret) till den  $k$ :te närmsta punkten i det slumpmässiga punktmönstret om vi nämligen förutsätter att området  $A$  och antalet punkter  $N$  båda är så stora att de diskuterade avstånden är försumbart små i jämförelse med områdets dimensioner. Då är nämligen antalet punkter  $n_r$  inom ett avstånd  $r$  från  $y$  approximativt Poisson  $(\lambda \cdot \pi \cdot r^2)$ , d v s

$$P(n_r = m) \approx \frac{(\lambda \cdot \pi \cdot r^2)^m}{m!} \cdot e^{-\lambda \pi r^2} \quad m = 0, 1, \dots \quad (3)$$

-----  
 (1) Om inget annat påpekas, så avses i fortsättningen i detta avsnitt slumpmässiga punktmönster.



Men händelsen att det är åtminstone  $m$  punkter i denna cirkel är ekvivalent med händelsen att  $r_m$  inte överskrider avståndet  $r$ . Vi har således ekvationen

$$P(r_m \leq r) = P(n_r \geq m) \quad (4)$$

av vilken följer (för  $k = 1, 2, \dots$ ):

$$F_{r_k}(r) = 1 - \sum_{i=0}^{k-1} \frac{(\lambda \cdot \pi \cdot r^2)^i}{i!} \cdot e^{-\lambda \pi r^2} \quad (5)$$

$$f_{r_k}(r) = 2 \cdot (\lambda \cdot \pi)^k \cdot \frac{r^{2k-1}}{(k-1)!} \cdot e^{-\lambda \pi r^2} \quad (6)$$

$$E(r_k) = \frac{(2k)! \cdot k}{(2^k \cdot k!)^2} \cdot \lambda^{-1/2} \quad (7)$$

$$V(r_k) = \frac{k}{\lambda \cdot \pi} - [E(r_k)]^2 \quad (8)$$

där  $r$  står för avstånd och  $F$  är fördelningsfunktionen,  $f$  frekvensfunktion,  $E$  väntevärdet samt  $V$  variansen. <sup>(1)</sup>

Uttrycken (5) - (8) gäller avstånd uppmätta från en godtycklig punkt  $y$  i området, således även från en punkt i det studerade punktmönstret. I detta specialfall är alltså  $r_k$  avståndet till den  $k$ :te närmsta grannpunkten ("k:th nearest neighbour distance").

Genom att sätta  $k = 1$  i uttrycken (5) - (8) får vi sannolikhetsfördelningen för  $r_1$ , avståndet till den närmsta punkten i det slumpmässiga punktmönstret:

$$F_{r_1}(r) = 1 - e^{-\lambda \pi r^2} \quad (9)$$

---

(1) Transformationen  $t_k = 2\pi \lambda r_k$  ger en  $\chi^2$ -fördelning med  $2k$  frihetsgrader. Eftersom summan av oberoende  $\chi^2$ -fördelade variabler också är  $\chi^2$ -fördelad, kan denna transformation användas för exakta test av slumpmässighet (jfr avsnitt 3.5 samt Thompson [18]).

$$f_{r_1}(r) = 2\lambda \pi r \cdot e^{-\lambda \pi r^2} \quad (10)$$

$$E(r_1) = \frac{1}{2} \cdot \lambda^{-1/2} \quad (11)$$

$$V(r_1) = \left(\frac{1}{\pi} - \frac{1}{4}\right) \cdot \lambda^{-1} \quad (12)$$

Nedan har grafen för frekvensfunktionen för  $r_1$  skissats för två värden på intensiteten  $\lambda$  (se figur 5)

Med hjälp av de härledda uttrycken kan vi nu utvärdera funna punktmönster. Clark & Evans [5] föreslår användning av indexet  $D^* = \bar{r}_1 / r_E$  (13)

Medelvärde  $\bar{r}_1$  erhålls genom att avgränsa ett område i planet och mäta avståndet från varje punkt i mönstret till närmsta grannpunkt.  $r_E$  är väntevärdet för samma avstånd (ekv.(11)). Det följer av definitionen att  $E(D^*) = 1$  för ett slumpmässigt punktmönster. Man kan visa att  $D > 2.15$  för ett triangulärt punktmönster och  $D^* = 2$  för ett kvadratisk<sup>(1)</sup>.

För ett fullständigt agglomererat punktmönster gäller  $D^* \approx 0$  (jfr figur 3).

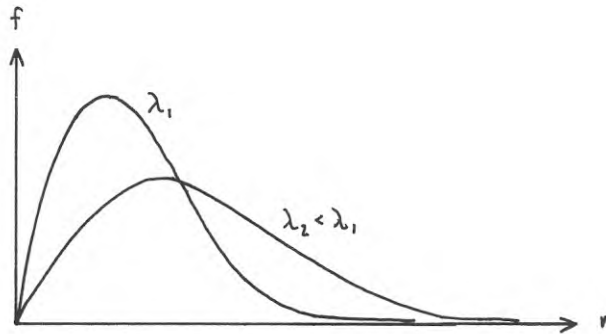


FIG. 5. Två frekvensfunktioner för avstånd till närmsta grannpunkt

(1) Observera att ett triangulärt punktmönster erhålles om man betraktar alla städer över en viss "funktionsnivå" i Cristallers modell (jfr figur 1, § 3.2).

$D^*$  kan alltså användas som ett mått på agglomeration.  $D^* > 1$  indikerar att punktmönstret är mer regelbundet än ett slumpmässigt;  $D^* < 1$  indikerar att punktmönstret i stället är mer agglomererat än slumpmässigt.  $D^*$ -indexet har emellertid begränsad användning eftersom det endast ger en fragmentarisk karaktäristik av det studerade punktmönstret (Ex:  $D^* \approx 0$  både för punktmönstret i figur 3.d och det i figur 4).

Statistiska test av slumpmässighet kommer att redovisas (§ 3.5), men först ägnas ett avsnitt åt tillämpningar av ovanstående avståndsanalys på franska städers rumsliga fördelning (§ 3.4).

### 3.4 Ett samband mellan städers storlek och avstånd

Nedan redovisas några tillämpningar på franska städers rumsliga fördelning (jfr fig.6, sid. 54) men först lite "teori".

Hittills har endast vissa egenskaper hos punktmönster i allmänhet diskuterats. Eftersom städer är olika stora, kunde det emellertid vara intressant att ta med även storleksaspekten i beräkningarna.

Låt oss alltså beskriva städernas storleksfördelning i ett visst område (t ex en region eller ett land) med (trappstegs-) funktionen  $N$ , där  $N(P)$  är antalet städer med mer än  $P$  invånare. Intensiteten  $\lambda(P)$  av städer med mer än  $P$  invånare blir således  $\lambda(P) = N(P)/S_A$ , där  $S_A$  är områdets yta.

Om städerna nu vore slumpmässigt fördelade i området (enligt § 3.3), så skulle det förväntade avståndet  $r_E(P)$  från en stad med mer än  $P$  invånare till närmsta grannstad med mer än  $P$  invånare vara:

$$r_E(P) = \frac{\sqrt{S_A}}{2} \cdot [N(P) - 1]^{-1/2} \approx \frac{\sqrt{S_A}}{2} \cdot [N(P)]^{-1/2} \quad (14)$$

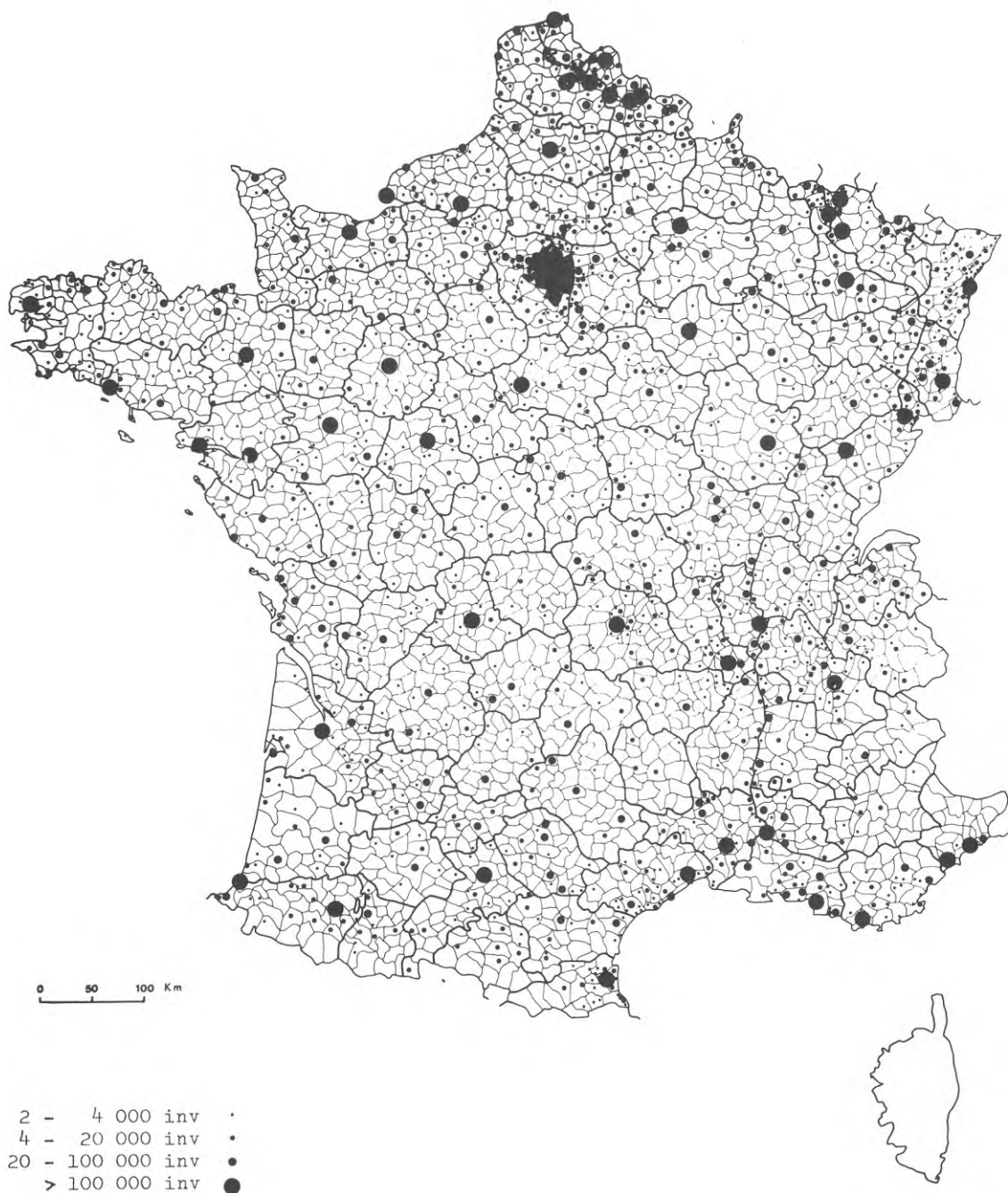


FIG. 6. De franska städerna 1968

(jfr ekv. (11), § 3.3). Med kännedom om ytan  $S_A$  och funktionen  $N$  för ett visst område kan man alltså beräkna det förväntade avståndet  $r_E$  för godtyckligt  $P$ .

Emellertid ger oss rank-size rule möjlighet att approximera funktionen  $N$ . Av den följer nämligen

$$N(p) \approx \mu \cdot P^{-\nu} \quad (15)$$

(jfr ekv. (1), § 3.2,  $\nu = 1/\beta$ ).

Under förutsättning av en slumpmässig rumslig fördelning och en storleksfördelning i enlighet med rank-size rule får vi alltså följande samband:

$$r_E(P) \approx \frac{\sqrt{S_A}}{2} \cdot \mu^{-1/2} \cdot P^{\nu/2} = \gamma \cdot P^{\nu/2} \quad (16)$$

eller i logaritmisk form:

$$\log(r_E(P)) = \log(\gamma) + \frac{\nu}{2} \cdot \log(P) \quad (17)$$

Av den logaritmiska versionen framgår det att sambandet motsvaras av en rät linje med lutningen  $\nu/2$  i ett log-log-diagram med folkmängd på ena axeln och avstånd på den andra. (1)

Vid empiriskt studium av städerna i Frankrike 1962 (totalt ca 1300 städer med mer än 2000 invånare) erhöles de resultat som visas i figur 7 och 8. (sid. 56). Som synes är "rank-size-kurvan" inte helt rätlinjig. Med stöd av figuren ansätter vi

-----  
(1) Detta att lutningen är  $\nu/2$  gäller inte bara för avståndet till närmsta grannstad, utan för godtyckligt värde på  $k$ . Däremot blir uttrycket för konstanten  $\gamma$  annorlunda. (jfr ekv. (7), § 3.3).

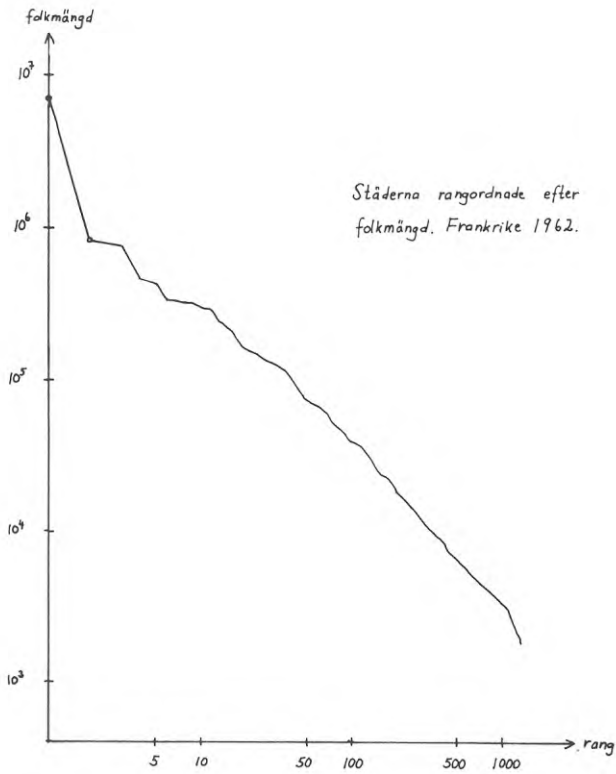


FIG. 7.

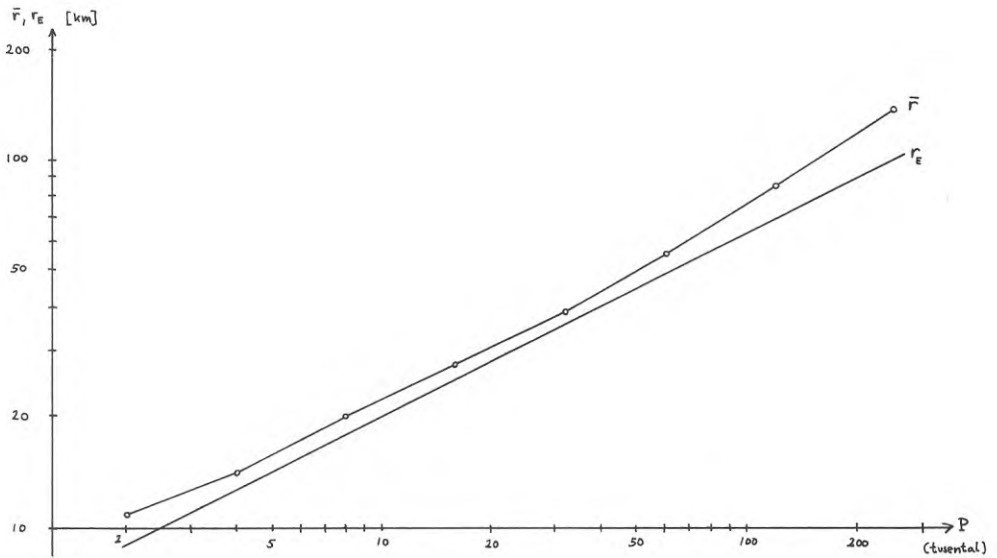


FIG. 8. Medelavstånd till närmsta grannstad som funktion av storleksklass. Frankrike 1962.

dock  $\mu \approx 3.3 \cdot 10^6$  och  $\nu = 1.0$ . Eftersom Frankrikes yta (utom Korsika och departementen bortom haven) är ungefär  $535\ 000\ \text{km}^2$  så erhåller vi

$$r_E(P) \approx 0.20 \cdot P^{0.5}$$

Både kurvan för  $\bar{r}(P)$  och för  $r_E(P)$  är inritade i figur 8 (sid. 56). Det framgår att  $r_E$  systematiskt överskattar avstånden med 10% - 35% men att lutningen på linjen är hyfsad. Vidare ser man hos kurvan för  $\bar{r}$  en svag tendens till krökning uppåt.

Kvoten  $D^*(P) = \bar{r}(P)/r_E(P)$  visas i figur 9 (sid. 58). Trots att  $D^*$  är större än 1, återstår ännu mycket till det "Cristallerska" värdet  $D^* = 2.15$  (jfr avsnitt 3.3).

Eftersom städers rumsliga och storleksmässiga fördelning kan tänkas ha med industrialiseringsgraden att göra, gjordes följande indelning av Frankrikes yta: en syd-västlig del (SV) där den andel av de manliga förvärvsarbetande som arbetar inom jordbruksnäringen överskrider 30% samt en nord-östlig del (ND) där samma andel underskrider 30% (se figur 10).<sup>(1)</sup>

---

(1) Uppgifterna får antas gälla år 1962.

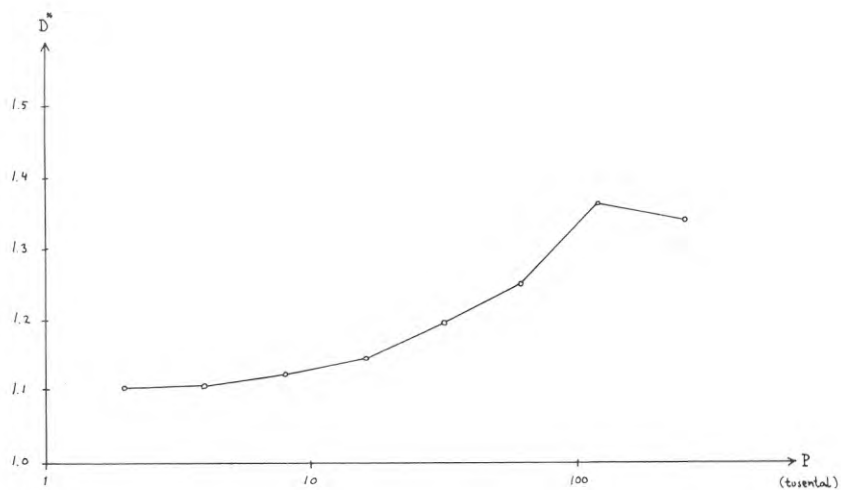


FIG. 9. Agglomereringsindexet  $D^*$  som funktion av storleksklass. Frankrike 1962.



FIG. 10. Frankrike uppdelat i en jordbruks- och en industridel



Medelavståndet  $\bar{r}(P)$  är plottat för båda regionerna i figur 11a (sid. 60). Som synes är medelavståndet i den agrara regionen för alla värden av  $P$  ungefär en och en halv gånger större än motsvarande medelavstånd i den industriella regionen. En del av denna skillnad kan förklaras av skillnaden i "intensitet" ( $\lambda \approx 0.002$  städer/km<sup>2</sup> i SV och  $\lambda \approx 0.003$  städer/km<sup>2</sup> i NO). Men även med hänsyn till denna skillnad i "intensitet", så kvarstår ändå en påtaglig skillnad mellan regionerna.  $D^*(P)$ , som ju är en av intensiteten oberoende indikator för agglomeration visar detta i figur 11b (sid. 60). Indexets värde är ju klart större i den agrara regionen (d v s punktmönstret är något mera "regelbundet").

### 3.5 Några statistiska test

En svårighet med avståndsbaserade test av slumpmässighet hos punktmönster är det statistiska beroendet mellan variablerna. Detta beroende förhindrar oss t ex att använda indexet  $D^*$  som testvariabel. Vi vet visserligen att förväntningsvärdet under nollhypotesen är 1, men det är också i stort sett allt. p g a denna svårighet använder man i stället vanligen testvariabler baserade på stickprov.

Låt oss titta närmare på några test som förekommer i litteraturen (A är det studerade området,  $N$  antalet punkter i det studerade punktmönstret i A och  $H_0$  står för nollhypotesen att punktmönstret är slumpmässigt).

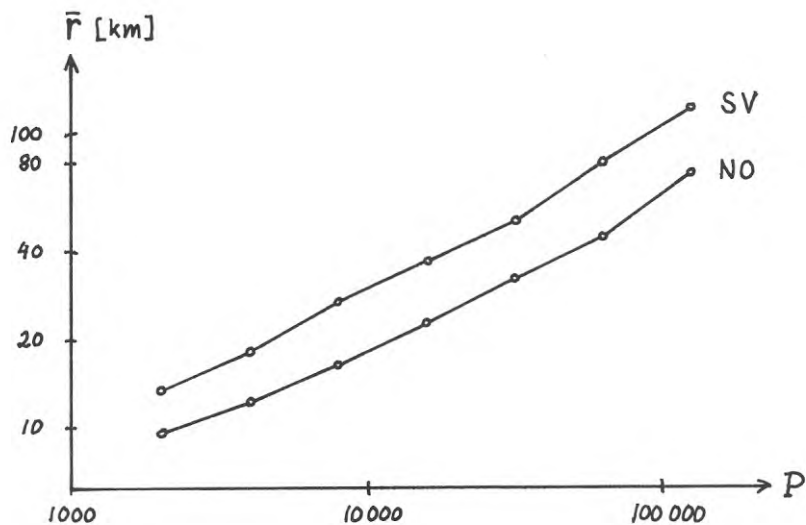


FIG. 11.a. fig. 11. a.

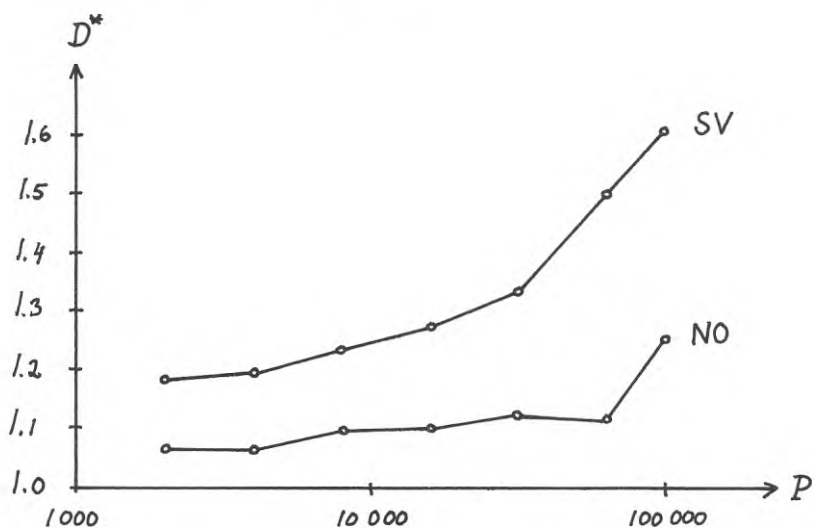


fig. 11. b.

FIG. 11.b.

Medelavstånd till närmsta grannstad, respektive agglomereringsindexet  $D^*$  som funktion av storleksklass för de två delarna av Frankrike 1962.

Test 1 (jfr Thompson [18]): Testvariabel är  $\bar{r}_k(n)$ , medelvståndet från  $n$  st. ( $n \ll N$ ) inom  $A$  slumpmässigt fördelade mätpunkter till den  $k$ :te närmaste punkten i det studerade punktmönstret. Av sambanden (7) och (8) i § 3.3 samt centrala gränsvärdesatsen följer att  $\bar{r}_k(n)$  under  $H_0$  är approximativt normalfördelad  $N(m_k, \sigma_k/\sqrt{n})$  där  $m_k = E(r_k)$  och  $\sigma_k = [V(r_k)]^{1/2}$ . Testet är således: förkasta  $H_0$  om och endast om

$$|\bar{r}_k(n) - m_k| > \frac{\sigma_k}{\sqrt{n}} \cdot \phi^{-1} \left(1 - \frac{\alpha}{2}\right), \quad (18)$$

där  $\phi$  är normalfördelningsfunktionen och  $\alpha$  risknivån.

Test 2 (jfr Holgate [10]): Testvariabel är  $\bar{z}_{ij}(n)$ , medelvärdet av  $n$  st. kvoter  $Z_{ij}$  ( $n \ll N$ ), beräknade för lika många i  $A$  slumpmässigt fördelade mätpunkter. Varje kvot beräknas enligt  $Z_{ij} = r_i^2/r_j^2$  ( $i < j$ ), dvs som kvoten mellan kvadraterna på avstånden till den  $i$ :te respektive  $j$ :te närmsta punkten i det studerade punktmönstret. Med hjälp av centrala gränsvärdesatsen kan man visa att  $\bar{z}_{ij}(n)$  under  $H_0$  är approximativt normalfördelad  $N(m_{ij}, \sigma_{ij}/\sqrt{n})$ , där

$$m_{ij} = i/j \quad (19)$$

$$\sigma_{ij} = \left[ \frac{i(j-i)}{j^2(j+1)} \right]^{1/2} \quad (20)$$

Testet blir alltså: förkasta  $H_0$  om och endast om

$$|\bar{z}_{ij}(n) - m_{ij}| > \frac{\sigma_{ij}}{\sqrt{n}} \cdot \phi^{-1} \left(1 - \frac{\alpha}{2}\right) \quad (21)$$

Test 3 (jfr Pielou [14]): Testvariabel är  $\bar{r}_k'(n)$ , medelvståndet från  $n$  st. ( $n \ll N$ ) slumpmässigt valda punkter i det studerade punktmönstret till den  $k$ :te närmsta grannpunkten i punktmönstret. Testet blir likadant som i test 1: förkasta  $H_0$  om och endast om

$$|\bar{r}_k'(n) - m_k| > \frac{\sigma_k}{\sqrt{n}} \cdot \phi^{-1} \left(1 - \frac{\alpha}{2}\right) \quad (22)$$

De två första testerna är alltså baserade på avstånd mätta från slumpmässigt fördelade mätpunkter medan det tredje testet utnyttjar avstånd från slumpmässigt valda punkter i det studerade punktmönstret.

En viktig skillnad mellan å ena sidan test 1 och 3 och å andra sidan test 2 är att i de två förstnämnda krävs kännedom om intensiteten  $\lambda = N/A$  vilket inte krävs i test 2. Denna skillnad har dock betydelse endast i sådana fall t ex i biologin, där  $N$  är så stort att  $\lambda$  inte kan beräknas direkt.

En annan skillnad, som gör sig gällande för stora  $N$ , denna gång mellan test 1 och 3, är att det i test 3 krävs någon form av identifiering av punkterna i punktmönstret för att man skall kunna göra ett slumpmässigt urval av de  $n$  punkterna bland alla  $N$  punkter.

I fallet  $k = i = 1$ ,  $j = 2$  kommer agglomerering att ge höga värden på testvariablerna i test 1 och 2 och ett lågt värde på testvariabeln i test 3. Det motsatta kommer att gälla för regelbunda mönster (jfr figur 3 a och b, § 3.2).

Ett mått på hur "bra" - eller "känsligt" - ett statistiskt test är, är dess styrka, d v s dess förmåga att upptäcka förefintliga avvikelser från nollhypotesen. De tre ovanstående testens styrka kommer att prövas dels för ett agglomererat punktmönster, dels för ett regelbundet punktmönster. För enkelhets skull behandlas endast fallet  $k = i = 1$ ,  $j = 2$ .

#### Ett aggregerat punktmönster:

Låt oss föreställa oss ett punktmönster där grupper av punkter är slumpmässigt fördelade. För att förenkla beräkningarna antas avståndet mellan punkterna i varje grupp vara försumbart litet

i jämförelse med avstånden mellan grupperna. Antag vidare att gruppernas storlek, d v s antalet punkter i dem, är approximativt geometriskt fördelade, d v s sannolikheten att en grupp innehåller  $m$  punkter ( $m = 1, 2, \dots$ ) är ungefär lika med  $q \cdot (1-q)^{m-1}$ , där  $q$  är ett tal mellan noll och ett. Om totala antalet punkter i området är  $N$ , där  $N$  antas vara ett stort tal, så blir under dessa förutsättningar totala antalet grupper - inklusive de "grupper" som består av bara en enda punkt - ungefär lika med  $q \cdot N$ . Intensiteten  $\mu$  av gruppen är således  $q \cdot N/S_A$ , d v s  $\mu = q \cdot \lambda$ . Sannolikheten att en grupp består av endast en punkt blir approximativt lika med  $q^2$ .  $p = 1-q$  är således ett mått på punktmönstrets grad av agglomerering:  $p=0$  motsvarar ett slumpmässigt punktmönster i enlighet med nollhypotesen och ju högre värde på  $p$  ( $p < 1$ ), desto större är graden av agglomerering.

Mot detta fiktiva, agglomererade punktmönster har vi nu möjlighet att pröva de tre testens styrka. Med ett tests styrka menar vi nu - något mera precist än ovan - sannolikheten att testet skall ge utslag, givet viss agglomereringsgrad. De tre styrkefunktionernas grafer återfinnes i figur 12 (sid. 64). I båda diagrammen är agglomereringsgraden abscissa ("x-axel") och styrkan ordinata ("y-axel"). I det övre diagrammet förutsätts stickprovets storlek vara  $n = 10$ , i det undre  $n = 25$ . I samtliga fall avses ensidigt test på risknivån  $\alpha = 5\%$ . Som synes hävdar sig test 3 bäst såväl i fallet med den mindre som i fallet med den större stickprovsstorleken.

I fallet med den större stickprovsstorleken,  $n = 25$ , ger test 3 utslag med ca 95 procents säkerhet redan vid agglomereringsgraden  $p = 0.3$ . Agglomereringsgraden måste öka till  $p = 0.4$  respektive  $p = 0.5$  för att test 2 respektive test 1 skall ge utslag med samma säkerhet. I det andra fallet,  $n = 10$ , kräver

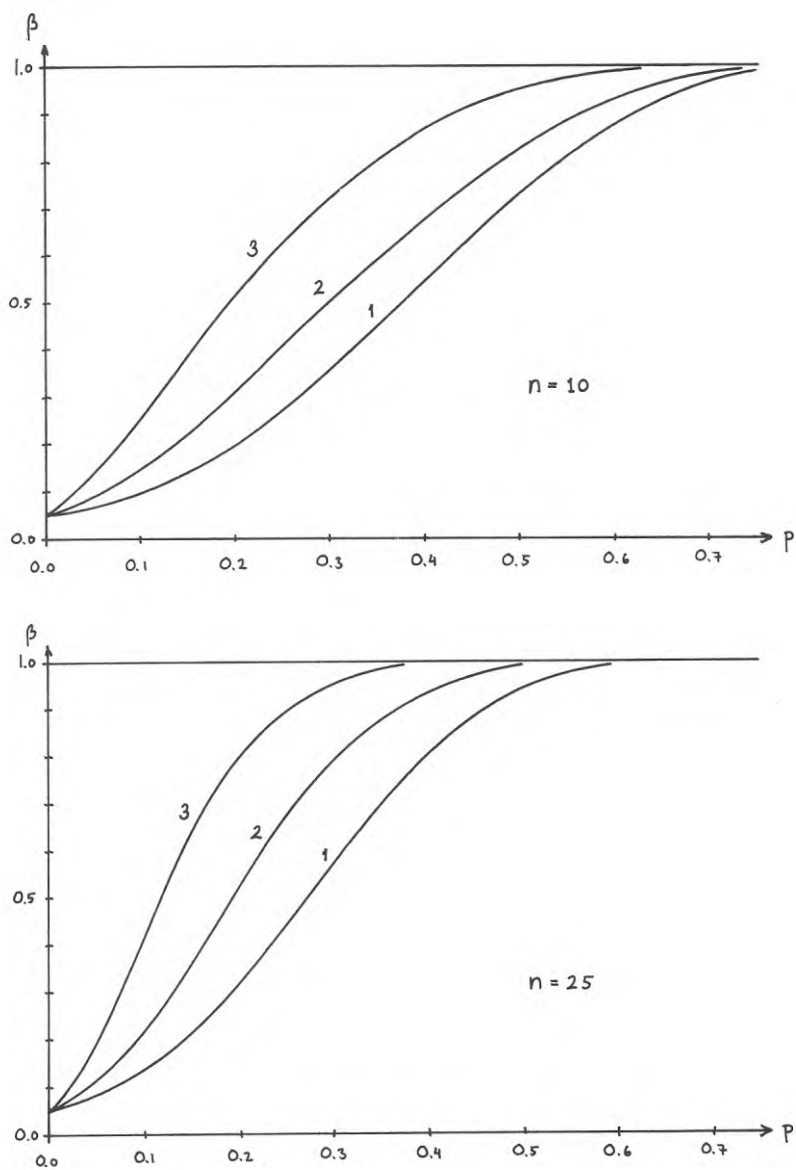


FIG. 12. De tre testernas styrkor som funktioner av agglomereringsgraden  $p$ . Stickprovsstorlek  $n=10$  resp.  $n=25$ .

däremot alla tre testerna en agglomereringsgrad överskridande 0.5 för utslag med 95 procents säkerhet.

Som en annan jämförelse mellan testerna kan vi konstatera att test 1 inte ens med den större stickprovsstorleken uppnår samma styrka som test 3 med den mindre stickprovsstorleken.

#### Ett regelbundet punktmönster:

För den matematiska enkelhetens skull behandlas endast ett fullständigt regelbundet, kvadratisk punktmönster (se figur 3b, § 3.2). Med ett tests styrka avses i detta fall alltså sannolikheten att testet skall ge utslag för ett helt kvadratisk punktmönster. Vi får nöja oss med att endast studera styrkans beroende av stickprovsstorleken. I figur 13 nedan återfinnes graferna för de tre testernas styrkefunktioner. Liksom i agglomereringsfallet är styrkan ordinata, medan stickprovsstorleken nu är abscissa.

Skillnaden mellan de tre testernas styrka är anmärkningsvärt stor. Liksom i fallet med det agglomererade punktmönstret har test 3 den största styrkan. Test 1 ger utslag med mer än 95 procents säkerhet om stickprovsstorleken överskrider ca 30 mätningar. Test 2 är påfallande "svagt". Det krävs mer än 100 uppmätta avstånd för att testet skall ge utslag med 95 procents säkerhet.

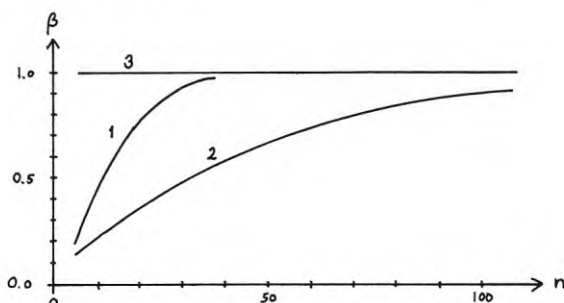


FIG. 13. De tre testernas styrkor för ett regelbundet, kvadratisk punktmönster som funktioner av stickprovsstorleken.

Den extrema styrkan hos test 3 kräver kanske en kommentar. Det är ju nämligen så att testvariabeln i detta test är medelavståndet från  $n$  st. punkter i punktmönstret till respektive närmsta grannpunkt. Eftersom punktmönstret antas vara kvadratisk regelbundet så är avståndet från varje punkt i punktmönstret till dess närmsta grannpunkt lika stort, nämligen lika med dubbla förväntade avståndet från ett slumpmässigt punktmönster (se ekv. (13) och fortsättningen, § 3.3). Därför kommer nollhypotesen att förkastas på varje rimlig risknivå.

Resultatet av våra undersökningar av de tre testerna kan sammanfattas i följande rekommendation: Vid test av båda slagen av avvikelser från slumpmässighet är test 3 att föredra framför de två andra testen. Vid test av agglomereringstendensen bör då stickprovsstorleken vara ganska stor ( $n \geq 25$ ) för att testet med rimlig säkerhet skall ge signifikant utslag.

Det bör dock kanske påpekas att dessa slutsatser om testen har begränsad räckvidd; vi har ju endast studerat en viss typ av agglomerering och ett visst slag av regelbundenhet. För andra typer av avvikelser från slumpmässighet kanske testen visar annan styrka.

En positiv egenskap hos de tre testerna är att de är tämligen enkla att hantera såväl beträffande själva mätförfarandet som beräkningarna. Dessutom är testvariablernas värden direkt tolkbara i termer av "agglomerering" respektive "regelbundenhet". Men det finns säkert andra statistiska test som har större styrka än de tre diskuterade. Eftersom de exakta fördelningarna av de i testerna grundläggande stokastiska variablerna är kända, skulle för det första motsvarande exakta test kunna utforskas. För det andra finns det i litteraturen andra relevanta statistiska test än de tre nämnda. Vid test av tendensen mot agglomerering bör särskilt Hopkins [12] nämnas. Han använder en testvariabel som i princip är kvoten mellan testvariablerna



i test 1 och test 3 ovan. När det gäller tendenser mot regelbundenhet kan Holgate [11] återigen nämnas. Som alternativ till testvariabeln  $\bar{Z}_{12}$  använder han den statistiska korrelationen mellan avståndet till närmsta respektive näst närmsta punkt.

Ytterligare ett fält att utforska vore slutligen indexet  $D^*$ :s statistiska fördelning, t ex medelst Monte-Carlo-metoden. Med kännedom om denna fördelning skulle man sedan kunna formulera ett möjligen starkt test direkt på  $D^*$ .

### 3.6 Sammanfattning

Efter en kortfattad översikt av litteraturen inom området, har en begränsning gjorts till studium av s k "k:th nearest neighbour analysis". Först har några statistiska egenskaper hos dessa avstånd presenterats. Sedan har ett samband mellan städers storlek och medelavstånd till närmsta grannstad härletts och tillämpats på de franska städerna. Slutligen har några statistiska test av slumpmässighet hos punktmönster dikuterats.

Tyvärr medgav den begränsade tiden på C.R.U. inte några mer omfattande tillämpningar. På sikt hoppas man dock kunna studera den rumsliga fördelningen av vissa dynamiska egenskaper, t ex samband mellan mindre städers tillväxt och avståndet till närmsta större stad (teorin om "utvecklingspoler"). Mer om detta förmäler dock ej denna historia.

### 3.7 Referenser

- [ 1 ] Angel, S., Human, G., 1971, "Transformations and Geographic Theory", Centre for Environmental Studies, Working Paper 72.
- [ 2 ] Auerbach, F., 1913, "Das Gesetz der Bevölkerungskonzentration", Petermann's Geographische Mitteilungen 59.
- [ 3 ] Barr, M., et al., 1971, "Patterns of urban spacing in the USSR: analysis of order neighbour statistics in twodimensional space", Journal of Regional Science, vol.11, nr.2.
- [ 4 ] Beckmann, M., 1958, "City hierarchies and the distribution of city size", Economic Development and Cultural Change, vol.6.
- [ 5 ] Clark, P., Evans, F., 1954, "Distance to nearest neighbour as a measure of spatial relationships in populations", Ecology, no.35.
- [ 6 ] Cristaller, W.,, 1933, "Die zentralen Orte in Süddeutschland", Jena.
- [ 7 ] Dacey, M.F., 1960, "A note on the derivation of nearest neighbour distances", Journal of Regional Science, vol.2, no.2.
- [ 8 ] Dacey, M.F., Tung, T., 1962, "The identification of randomness in point patterns, I", Journal of Regional Science, vol.4, no.1.
- [ 9 ] Dacey, M.F., 1963, "order neighbour statistics for a class of random patterns in multidimensional space", Annals of the Association of American Geographers, vol.53, no.4.

- [10] Holgate, P., 1965, "Some new tests of randomness", *Journal of Ecology*, 53.
- [11] Holgate, P., 1972, "The use of distance methods for the analysis of spatial distributions of points", Lewis (editor), *Stochastic Point Processes*, Wiley.
- [12] Hopkins, B., 1954, "A new method of determining the type of distribution of plant individuals", *Ann. Bot.*, no.18.
- [13] Lösch, A., 1954, "The economics of location", New Haven, (översättning).
- [14] Pielou, E.C., 1959, "The use of point-to-point distances in the study of the pattern of plant populations", *Journal of Ecology*, 47,
- [15] Robinson, K., "Central place theory I. A review", Centre for Environmental Studies, Working Paper 2.
- [16] Rogers, A., 1969, "Quadrat analysis of urban dispersion: 1. Theoretical techniques, *Environment and Planning*, vol.1, no.1, 1969.
- [17] Steward, C., 1958, "The size and spacing of cities", *Geographical Review*, vol.51, no.3.
- [18] Thompson, H.R., 1956, "Distribution of distance to nth neighbour in a population of randomly distributed individuals", *Ecology*, vol.37.
- [19] Zipf, G.K., 1941, "National unity and Disunity", Bloomington, Ind.,

















**R16: 1975**

**Denna rapport hänför sig till forskningsanslag Bs 471 från  
Statens råd för byggnadsforskning till Matematiska institutionen,  
KTH, Stockholm.**

**Distribution: Svensk Byggtjänst, Box 1403, 111 84 Stockholm**

**Grupp: samhällsplanering**

**Pris: 17 kronor + moms**