

O processo de difusão no sistema urbano brasileiro: análise do padrão de distribuição espacial de centros urbanos e seu ajustamento a distribuições de probabilidades

SPERIDIÃO FAISSOL
Geógrafo do IBGE

Participaram da presente pesquisa:

Analista Especializado
MARIA DAS GRAÇAS DE OLIVEIRA

Estatísticos
ANTONIO CARLOS FERNANDES DE MENEZES
MARÍLIA BIANGOLINO XAVIER

Geógrafo
EVANGELINA XAVIER GOUVEIA DE OLIVEIRA

Com a colaboração de:

Geógrafos
LANA LIMA MOREIRA
JANE FLORENTINA DE ABREU FERRO
DIVA DE QUINA ALMEIDA

Técnico de Estatística
NADJA LOUREIRO PERNES DA SILVA

1 — INTRODUÇÃO

Um dos objetivos tradicionais da pesquisa geográfica é o de descrever um padrão espacial relativo a determinado fenômeno ou evento geográfico e procurar explicar tal padrão através dos mecanismos que o geraram. O que isto implica é em uma maneira de deduzir a forma espacial a partir do conhecimento ou da análise do processo temporal — um procedimento dedutivo que poderíamos chamar de transformação tempo-espaço — um problema, a rigor, de natureza filosófica e quase metafísica, pois há sempre uma maneira de conceber espaço como o tempo entre um lugar e outro.

ARTIGOS

Uma das maneiras de se poder deduzir como uma determinada forma espacial pode estar associada a um determinado processo é a da utilização de certas distribuições de probabilidade, cujas funções geradoras são matematicamente definidas e especificadas. É o caso, por exemplo, das distribuições Poisson ou Binomial Negativa, que têm funções geradoras especificadas, nitidamente associadas a determinadas formas espaciais, e que ao mesmo tempo têm significado em termos de processo. Uma das dificuldades mais sérias a esta espécie de conexão espaço-tempo é a relativa à escala em que o processo (ou um determinado processo) funciona. Conhecido um processo e a escala em que ele funciona, seria possível prever a forma espacial que ele toma; inversamente, conhecida a forma espacial em que o processo funciona, é possível prever que tipo de processo está governando aquela forma.

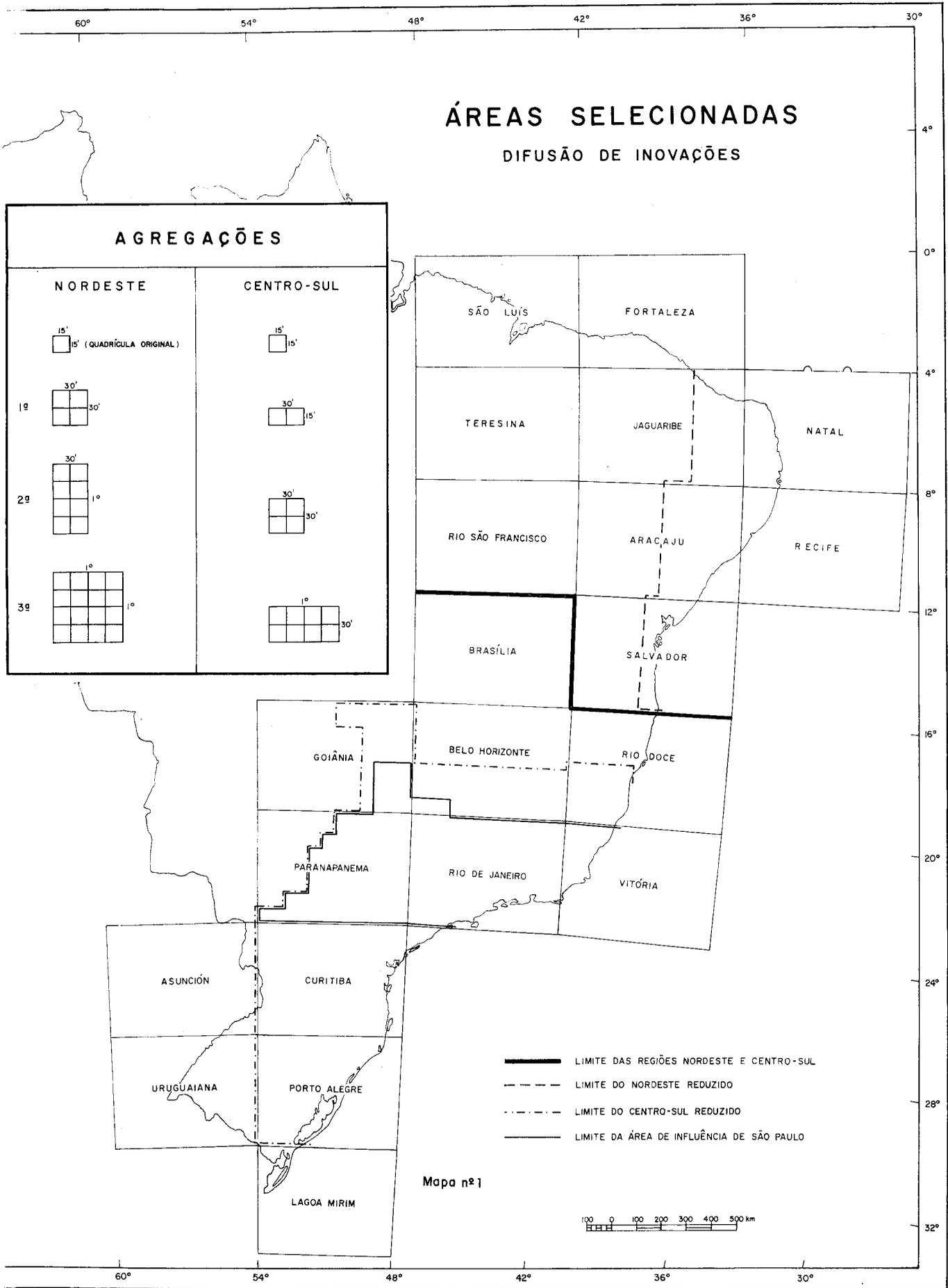
Pretendemos aqui explorar a natureza das relações entre o processo e forma espacial, utilizando algumas distribuições matemáticas e associando-as a um conjunto de informações sobre o sistema urbano brasileiro, resultantes de pesquisas realizadas no Departamento de Geografia do IBGE. Como a natureza dos estudos ainda pode ser considerada preliminar, um processo iterativo é utilizado, indo de processo a padrão espacial, e daí de volta ao processo, de forma a poder testar a validade da associação das duas coisas, em diferentes escalas e assim se ter melhor noção de processo e escala.

O presente estudo apresenta dois tipos de distribuição de cidades no Brasil, no Centro-Sul e no Nordeste, utilizando para isso as folhas da carta 1:1.000.000, conforme vai indicado mais adiante. As cidades selecionadas foram as de nível 4b, derivadas do sistema de regiões funcionais urbanas elaborado pelo Departamento de Geografia¹, considerando-se também as de nível acima (partindo-se do pressuposto de que em cada cidade de nível hierárquico superior existem funções do tipo 4b), para efeito de ter-se uma completa cobertura do sistema de cidades em um determinado nível de organização espacial. Em alguns casos sentiu-se necessidade de analisar-se também a distribuição de centros 3b, o que será indicado em cada caso. (ver mapa 1).

Cada folha foi dividida em um número de quadrículas de 15' por 15' (portanto de área igual) e para cada quadrícula foi indicado o número de centros urbanos existentes.

Adicionalmente foi feita uma outra análise, utilizando-se as mesmas folhas da carta 1: 1 000 000, dividida da mesma forma, mas ao invés de centros de nível 4b, foram lançadas todas as sedes de municípios. O objetivo desta análise adicional foi o de comparar as duas distribuições; uma resultante das forças de mercado, gerando centros de uma determinada categoria no sistema espacial; a outra resultante de uma evolução político-administrativa (como é a criação de municípios), com o propósito de se verificar em que medida um sistema de decisões afeta o outro. Na realidade uma indicação deste tipo pode contribuir inclusive para se ter uma idéia da capacidade de se afetar o processo econômico, por via de decisões políticas ao nível de organização da rede de municípios.

O objetivo do estudo é ainda o de comparar a distribuição nas duas áreas, de nível de desenvolvimento caracteristicamente diferentes, procurando associar a diferenciação de distribuição à diferenciação do nível de desenvolvimento, mapeando uma e outra em distribuições de probabilidades (Binomial Negativa e Poisson), que possam contribuir para a explicação da natureza do processo espacial que deu origem a tais distribuições. O estudo introduz, ao mesmo tempo, uma metodologia quantitativa probabilística, já amplamente utilizada² em numerosos trabalhos e para a qual utilizamos rotinas computacionais cedidas pelo Prof.



ÁREAS SELECIONADAS

DIFUSÃO DE INOVAÇÕES

AGREGAÇÕES

NORDESTE

CENTRO-SUL

15' (QUADRÍCULA ORIGINAL)

15'

1ª 30' 30'

30' 15'

2ª 30' 1º

30' 30'

3ª 1º 1º

1º 30'

SÃO LUÍS

FORTALEZA

TERESINA

JAGUARIBE

NATAL

RIO SÃO FRANCISCO

ARACAJU

RECIFE

BRASÍLIA

SALVADOR

GOIÂNIA

BELO HORIZONTE

RIO DOCE

PARANAPANEMA

RIO DE JANEIRO

VITÓRIA

ASUNCIÓN

CURITIBA

URUGUAIANA

PORTO ALEGRE

LAGOA MIRIM

- LIMITE DAS REGIÕES NORDESTE E CENTRO-SUL
- - - LIMITE DO NORDESTE REDUZIDO
- · - · LIMITE DO CENTRO-SUL REDUZIDO
- (thin) LIMITE DA ÁREA DE INFLUÊNCIA DE SÃO PAULO

Mapa nº 1

100 0 100 200 300 400 500 km

Mapa 1 — Áreas selecionadas para o estudo da distribuição dos Centros 4b e Cidades.

Duane Marble, da Northwestern University, e implantadas no sistema operacional do Centro de Informática, Fundação IBGE.

A natureza essencial do tipo de metodologia utilizada, probabilística, apoia-se no fato de que uma distribuição de pontos em um espaço geográfico pode ser o resultado de um processo estritamente estocástico, de distribuição aleatória, pode implicar em uma distribuição por contágio, ou ainda uma forma mista que resulte numa distribuição mais regular que aleatória. Os processos que gera diferentes tipos de distribuição fazem parte integral de teorias geográficas, conforme acentua Dacey em seu citado trabalho, exemplificando com a teoria de localidades centrais, que seria gerada por um processo de difusão estritamente estocástico, nos seus vários níveis que, por sua vez, se refletiria em uma distribuição hierárquica do tipo *Rank-Size*.

Apresentaremos no capítulo que se segue uma visão geral, de natureza conceitual e metodológica, das relações entre sistemas de localidade central e os processos que podem gerar tal distribuição. A seguir analisaremos o padrão espacial de cidades nas duas áreas indicadas, não só no nível 4b tomado como núcleo de tamanho mínimo, mas também no nível superior e inferior, com vistas a identificar, por via da metodologia utilizada, os desvios da distribuição adotada como ponto de referência teórico.

No capítulo 4 procuramos apresentar os resultados obtidos, mapeando as duas distribuições (Centro Sul e Nordeste), nos modelos conceituais adotados, comparando valores esperados com os observados, procurando assim testar a validade das hipóteses adotadas.

Finalmente uma ligeira síntese final, à guisa de conclusões, visa mais chamar a atenção para um diferente ângulo metodológico de análise, que ainda precisa sofrer o crivo de novas aplicações brasileiras, para dele se tirar conclusões mais válidas, tanto no que diz respeito à natureza do processo brasileiro, como a validade do modelo utilizado, em uma espécie de *feedback* teoria metodologia—teste empírico.

2 — O SISTEMA DE LOCALIDADES CENTRAIS ANALISADO SEGUNDO UM PROCESSO ESTOCÁSTICO. MODELOS DE DISTRIBUIÇÃO PUNCTIFORMES, AS DISTRIBUIÇÕES BINOMIAL NEGATIVA E POISSON E SUAS PREMISSAS BÁSICAS, APLICAÇÕES A PROBLEMAS GEOGRÁFICOS E DIFICULDADES METODOLÓGICO-TEÓRICAS

O sistema de localidades centrais pode ser examinado sob vários ângulos: de um lado pode ser estudado quase que como uma forma de organização espacial da teoria de localização — uma de suas premissas básicas é a clássica planície isomórfica — que associada à idéia de bens e serviços de diferentes alcances (*Range*) e limiares (*threshold*), gera uma hierarquia urbana e um espaçamento de cidades de diferentes tamanhos. De outro lado, pode constituir uma forma de análise de um padrão de distribuição, ao qual um modelo de distribuição teórica é aplicado, em busca de uma adequada descrição do mesmo e uma relação entre esta distribuição e a gênese do processo que leva à mesma.

No presente trabalho, como ficou esclarecido antes, procurou-se a segunda forma de análise, não apenas para fugir a um modelo determinístico, mas para, por via de modelos que têm uma função geradora conhecida, tentar identificar certos processos de difusão em uma estrutura espacial do sistema urbano já mais ou menos bem conhecida, através de uma série de estudos anteriores.³

2.1 — Localidades centrais e sua concepção em termos de um processo estocástico

A teoria de localidade central tem constituído, desde a sua formulação inicial por Christaller⁴ até suas reformulações mais recentes em termos de sua gênese estocástica⁵, o elemento mais importante na estruturação de uma teoria espacial de natureza essencialmente geográfica, mas com fortes implicações socioeconômicas, pois a cidade é um elemento fundamental para a organização econômica do espaço.

A teoria de localidade central baseou-se na demanda de bens de consumo (que foi tomada como estável e em equilíbrio), para derivar uma organização do sistema de localidades que atendesse essa demanda, partindo de princípios de minimização da distância percorrida, para o que a tecelagem hexagonal representava a forma otimizada. A uma demanda equilibrada correspondia uma determinada distribuição de localidades centrais, o que significa uma elevada proporção de isomorfismo entre a evolução dos processos espaciais que produzem uma forma particular de organização. É evidente o alto conteúdo mecanicista e geométrico Euclideano de tal concepção, tanto que numerosas experiências com sistemas eletromagnéticos (adotando a mesma premissa de uma planície isotrópica, sob a forma de uma superfície imantada por igual indicaram que a organização hexagonal foi gerada em quase 70% das experiências conduzidas.⁶

Christaller desenvolveu, de forma específica, a noção de centralidade como princípio de *Ordem*. “A cristalização de massa em torno de um núcleo é, no mundo orgânico como no inorgânico, uma forma elementar de ordem das coisas que existem juntas — uma ordem centralística. Esta ordem não é apenas um modo de pensar dos homens, existente em sua imaginação e desenvolvida porque o homem procura uma ordem; ela existe de fato, como um padrão inerente à própria matéria”.⁴

Assim a localidade central estabelece um sistema de ordem no espaço e muitas localidades centrais formam um sistema de localidades centrais. Aí se introduzem três noções:

1) Uma região complementar para a qual a localidade central é o centro, na qual se analisam as relações entre a cidade e o campo e vice-versa, implica em fluxo de/e para a cidade, fluxos de mercadorias e de pessoas que procuram e oferecem serviços.

2) Uma hierarquia de localidades centrais e, portanto, de regiões complementares, uma vez que certos bens e serviços necessitam de maior quantidade de consumidores para tornar possível sua produção, gerando assim uma relação entre produto e *threshold*, de natureza hierárquica.

3) Um espaçamento entre centros de diferentes tamanhos, descrito por uma função distância, que em síntese significa que a produção de um bem é limitada, também, por uma distância do consumidor ao local de produção; em outras palavras cada bem tem um alcance máximo, além do qual o seu consumo fica literalmente impossibilitado, sem afetar a demanda de outros.

Esta concepção estabelece uma relação entre a cidade e sua região complementar, em termos de um deficit de bens existentes na região e suprido pela cidade. E ao mesmo tempo significa que os dois conceitos gêmeos de *threshold* e *range* descrevem e explicam a teoria de localidade central.

Algumas formas numéricas podem indicar as aproximações ao modelo, que obviamente está muito ligado ao multiplicador urbano e sua evolução temporal. Assim o sistema urbano pode ser caracterizado também por uma série de relações de níveis de hierarquia, tanto para baixo como para cima, e tanto em termos de número de centros como de população servida.

O modelo hexagonal perfeito postula precisamente que, para cada centro de hierarquia $r' + 1$ deveriam existir seis centros de hierarquia r' , que somado ao próprio centro que é servido por ele mesmo, daria uma relação $k = 7$ (ou seja $6 + 1$). A título de exemplo derivado do próprio estudo das regiões funcionais urbanas teríamos N centros de hierarquia 2 (ou 2a apenas), para cada centro de hierarquia 1 (metropolitanos). O mesmo pode ser feito em relação à população (o que, de certa forma, constituiria uma dupla verificação); ao nível mais baixo da hierarquia, a população de um centro (pr') estaria relacionada à população que ela serve pela fórmula.

$$Mr' \text{ max} = Pr' + m$$

em que Mr' seria a população da área de mercado do centro Pr' e m seria a população rural servida pelo centro, no caso o de mais baixa hierarquia.

Para os centros de hierarquia superior, pelo menos dois caminhos podem ser seguidos e que podem oferecer indicações valiosas para a compreensão do poder de difusão dos centros. O primeiro seria a relação do centro com os de hierarquia imediatamente abaixo (acrescidos dos centros pequenos que lhe fossem diretamente subordinados), na qual o multiplicador urbano seria, por definição, o mais alto; o segundo seria a relação com a totalidade dos centros abaixo, ou mesmo só o primeiro e segundo níveis abaixo, onde o multiplicador urbano fosse, por definição, mais baixo. Esta hipótese somente é válida para uma rede urbana não perfeitamente ajustada, pois um modelo hexagonal e em *steady-state* teria um só multiplicador urbano.

A esta relação Pr'/Mr' (quer dizer a população de um centro de determinada hierarquia dividida pela população de toda a sua área de mercado, que incluiria a sua própria população), poderíamos chamar de u ; em um centro que servisse apenas sua própria população, u teria valor 1, o que significaria que aquele centro teria alcançado uma situação tal que teria absorvido nele mesmo toda a sua área de mercado e, portanto, teria um multiplicador urbano infinito. O multiplicador urbano pode ser descrito pela fórmula.

$$\text{Mult. Urb. } \frac{u}{1 - u}$$

e assim seria infinito se u fosse igual a 1.

O reverso seria um centro de população muito baixa relativamente à população rural que ele sirva (no caso de centro de mais baixa hierarquia), fazendo com que o multiplicador tenda para zero e indicando uma baixa capacidade do centro de servir à sua região complementar.

Estreitamente associado a este tipo de verificação do poder do centro como difusor de inovações está uma outra relação, ligada às distâncias que separam os centros de diversas hierarquias.

Pela fórmula $a(K + 1)$, a seria uma constante que descreveria o espaçamento dos centros, seja a partir de um subsistema para se saber qual deveria ser o espaçamento esperado nos outros subsistemas, seja

por simples comparação entre os espaçamentos observados (caso em que a deixaria de ser uma constante e seria apenas um elemento de comparação); a utilidade desta medida parece óbvia, não só pelo seu próprio mérito, mas particularmente à luz da metodologia principal que estamos adotando, e na qual a área da quadricula em que se divide a região estudada é um elemento essencial de validação do modelo. Pareceria óbvio presumir-se que o espaçamento entre os centros 2a em São Paulo seja muito menor que no Centro Sul como um todo ou no Nordeste. O que isto significa é que o tipo de bens que um centro (digamos como os de nível 4b) pode distribuir no Nordeste difere daquele que ele poderia distribuir em São Paulo, pois a principal implicação da idéia de “Range” é que, além de uma certa distância, um determinado bem não tem condições de ser consumido sem afetar o consumo de outros.

A maneira de associar as duas coisas é determinar o fator constante de crescimento pela fórmula:

$$\text{Fator Cresc.} = \frac{1 + k - u}{1 - u}, \text{ de tal forma que se, por exemplo, } u$$

fosse igual a 0,7 e k fosse igual a 6, então teríamos este fator igual a 21, pois $k + 1 - u$ seria igual a 6,3 e como $1 - u$ seria igual 0,3 o fator seria 21.

Uma análise do sistema urbano por via deste sistema de relações entre k (que a rigor indicaria o grau de centralização ou controle, pois representa o número de centros de ordem inferior, dependentes de um centro de ordem superior) e u (que é essencialmente o grau de urbanização, pois vimos que para uma cidade que sirva somente a ela mesma este valor de u seria 1 e em todos os outros casos seria inferior a 1) mostra o estado de equilíbrio no sistema. À medida que k aumenta ou u (ou eventualmente ambos) aumenta o fator crescimento e em consequência o grau de assimetria na distribuição dos tamanhos das cidades⁷, o que veremos na análise que faremos adiante, procurando comparar com a intensidade do processo de difusão. Veja-se, por exemplo, na tabela 2, que São Paulo tem 7 centros 2a subordinados, tem um valor 0,88 para u , portanto com um fator crescimento igual a 51. Já o Rio de Janeiro tem 3 + 1 centros e um valor de u igual 0,91, portanto com um fator crescimento igual a 34, bem inferior a São Paulo, o que mostra uma discrepância na hierarquia. É claro que valores deste tipo estarão associados ao poder de difusão de cada um destes centros metropolitanos, grau de urbanização da região e coesão no sistema.

É claro que tais discrepâncias são mais fáceis de examinar em um modelo probabilístico e por isso a forma hexagonal e o modelo determinístico do sistema de localidades centrais têm sido sujeitos a um conjunto de críticas, da mesma natureza daquelas que são feitas às premissas da teoria de localização e ligadas à planície isomórfica, competição perfeita, demanda igual em todos os sentidos etc.; destas críticas surgiram tentativas de colocar a teoria no contexto de um modelo probabilístico e inserida no princípio de satisfação mais que otimização.

Numerosos autores procuraram desenvolver a teoria de localidades centrais seguindo esta linha conceitual, com seus conseqüentes procedimentos metodológicos. Dacey², em seu artigo introduzindo a idéia de uma distribuição Poisson modificada e em numerosos outros estudos⁸, procurou mostrar que mesmo em áreas que se aproximam da planície homogênea teórica (que no caso dos Estados Unidos foi sempre partes do Meio-Oeste, especialmente o Estado de Iowa) padrões não ótimos e

não hexagonais podem emergir em função de condições particulares. Ele sugere, por exemplo, que o "Land Survey Act" americano afetou não somente a distribuição de terras e áreas administrativas, mas também a própria localização de centros de serviços. Dacey sugeria ainda que até mesmo os processos que governam a localização de sedes de condados podem ser diferentes daqueles que determinam a localização de outros centros. Procuramos investigar esta linha analítica no presente trabalho, examinando a distribuição de centros de serviços por via de distribuição de probabilidade, e da mesma forma, examinando a distribuição de cidades, sedes de município.

Assim, não considerando a hipótese de um modelo determinístico, na realidade partindo dos pressupostos de modelos probabilísticos que descrevam a distribuição dos centros pelas suas funções geradoras, as comparações entre os tipos de distribuições teóricas que mais se ajustem às distribuições observadas com o multiplicador urbano e com a constante a (por via das diferenças entre distanciamentos observados e esperados) podem ter o efeito de mútua validação ou levar à formulação de hipóteses diferentes. Na análise dos dois sistemas de cidades que apresentamos neste trabalho vamos encontrar exemplos destas comparações e sua validade interpretativa.

A linha metodológica do presente trabalho perseguirá, essencialmente, distribuições não ótimas, concentrando-se na análise do sistema urbano como uma distribuição de pontos, valendo-se do método de quadrículas como unidades básicas, nas quais o número de pontos é contado e usado como dado inicial básico.

2.2 — Modelos de distribuição punctiforme e o método das quadrículas

A metodologia aplicada baseia-se essencialmente no fato de que a distribuição de cidades constitui, a rigor, uma distribuição de pontos em uma superfície, tomada a cidade como uma unidade espacial básica. O tamanho da cidade ou número de indivíduos na cidade é uma consideração posterior. O segundo aspecto importante da metodologia é o de dividir-se a área em estudo em um número n de quadrículas que reduza a informação contida no mapa a um conjunto de números, para em seguida procurar-se uma expressão suscinta matemática que descreva tal conjunto de números. É óbvio que tal redução em dois níveis da informação cartográfica a um conjunto de números significa uma perda de detalhe na descrição da distribuição, mas se ela oferece a possibilidade de aplicar-se à mesma uma metodologia formal de análise e contribui para construção e verificação de teorias relativas à distribuição de povoamento, torna-se justificável.

Relacionada a esta metodologia existem problemas teóricos e práticos de vários tipos, desde os operacionais propriamente ditos (do tipo de tamanho da quadrícula ou tamanho mínimo da cidade a ser representada), até os mais complexos de interpretação do próprio resultado de ajustamento a uma determinada distribuição teórica.

Enquanto vamos descrevendo a metodologia que é aplicada, estes problemas vão sendo discutidos, o mesmo acontecendo quanto aos resultados analíticos obtidos.

O primeiro passo é o de dividir-se uma área em um número de quadrículas, a seguir contar-se o número de pontos em cada uma delas, computando-se a distribuição de freqüências ocorridas nas quadrículas.

Mencionamos o fato de que com este procedimento perde-se uma parcela (por vezes significativa) de detalhe referente à distribuição que queremos analisar. Considere-se, por exemplo, as duas distribuições indicadas no quadro abaixo e reproduzidas de um artigo de Dacey⁹, bem ilustrativas deste aspecto:

a) VALOR	FREQÜÊNCIA	DISTRIBUIÇÃO ESPACIAL				
0	6	0	0	1	2	2
1	3	0	0	1	2	2
2	6	0	0	1	2	2
b) 0	6	0	2	0	2	1
1	3	0	2	1	0	2
2	6	1	0	2	0	2

Como se vê, as duas distribuições de frequências são iguais, a função que pode gerar esta distribuição é igual, mas a distribuição espacial de pontos (transformados em conjunto de números) não é a mesma, o que pode levar a conclusões totalmente erradas. Isto faz do tamanho da quadrícula um problema extremamente importante, obviamente ligado ao problema da escala em que o fenômeno estudado é relevante.¹⁰

A escolha do tamanho da quadrícula é ditada por considerações de ordem prática e teórica. No primeiro caso, se o tamanho da quadrícula é grande, relativo à área em estudo, e temos assim um número pequeno de quadrículas, torna-se difícil obter estimativa precisa dos parâmetros. Por outro lado, se temos um número muito grande de quadrículas, um número muito pequeno de quadrados terá mais de 1 ou 2 pontos, a distribuição naturalmente tenderá para uma Distribuição Poisson, típica de um fenômeno de distribuição rara e aleatória; neste caso, pela escolha de uma quadrícula de tamanho muito pequeno, inapropriado ao fenômeno que estamos estudando, podemos descrever a distribuição como aleatória, quando na realidade ela pode não o ser. Harvey discute o fato de que a obtenção de um bom ajustamento a um modelo aleatório não justifica a inferência de que o padrão é aleatório em todas as escalas. Pode-se achar um bom ajustamento à Distribuição Poisson para um tamanho de quadrícula e um mau ajustamento para outro. Harvey acrescenta adiante, em seu trabalho¹⁰ que a representação de um processo de difusão por uma distribuição Binomial Negativa implica em definir um tamanho de quadrícula que seja aproximadamente igual ao “campo médio de informação” (mean information field). Dacey¹¹ mostra em estudo da distribuição de casas em Porto Rico, que uma distribuição de colônias (com populações diferentes) pode atender às especificações de uma função Poisson, mas uma distribuição Binomial Negativa pode ajustar-se bem para o número de indivíduos nas colônias, se este número de indivíduos tiver uma distribuição independente e com função logarítmica. Será o caso de obter-se uma distribuição Poisson para as cidades de um sistema que esteja em estado de quase equilíbrio (*Rank-Size*), mas cujos tamanhos (também típicos daquela distribuição *Rank-Size*) seguem uma função logarítmica. Voltaremos a estes problemas na análise empírica dos dados sobre o sistema de cidades, tanto em uma área como na outra.

Independência estatística entre as quadrículas, isto é, entre a ocorrência do fenômeno estudado em uma quadrícula e outra adjacente, é outro problema teórico importante. Em primeiro lugar, tratado de uma forma iterativa, o agrupamento de quadrículas com o objetivo de testar a independência acima indicada, pode mostrar a escala em que um determinado padrão mais se desviar de uma distribuição aleatória, e assim oferecer uma noção da estrutura do processo, e até mesmo das relações entre escala e processo. Em segundo lugar, pode nos levar a distinguir entre simples heterogeneidades entre populações no padrão de pontos, que pode gerar uma distribuição do tipo Binomial Negativa e uma distribuição realmente gerada por um processo de difusão do tipo contágio, também ajustado à mesma Binomial Negativa. A comparação entre os parâmetros p e k da Binomial Negativa entre quadrículas agregadas a diferentes tamanhos, pode oferecer indicações importantes, pois, segundo observa Dacey ¹¹, um ajustamento ao modelo binomial negativo implica em que para um quadrado de tamanho maior o parâmetro p permanece inalterado e o k varia linearmente de acordo com a área do quadrado. Adicionalmente medidas de contigüidade podem ser usadas para tais testes de independência. ¹²

Esta mais ou menos longa enumeração de problemas tem a finalidade apenas de chamar a atenção para as cautelas que se deve ter ao aplicar um modelo matemático abstrato a uma realidade empírica, na qual uma variada gama de fatores indeterminados pode perturbar a regularidade do modelo.

2.3 — A distribuição binomial negativa

A distribuição Binomial Negativa é geralmente associada a um tipo de processo de difusão do tipo contágio, no qual cada unidade espacial tenha uma oportunidade de receber um ponto, de forma independente embora não igual, formando agrupamentos.

O agrupamento de objetos (no caso do estudo básico que apresentamos), tais como cidades de nível 4b ou cidades sedes de município, refere-se a propriedade de uma distribuição aleatória concentrada ou uniforme. Uma distribuição aleatória é definida por N cidades, tratadas como pontos, uniforme e independentemente localizadas em uma determinada região tomada como amostra.

A distribuição binomial negativa é definida por dois parâmetros, p e k e a probabilidade de que um evento ocorra exatamente um número i de vezes, para $i = x$ é

$$P(x) = \binom{x+k-1}{k-1} p^k (1-p)^x \quad \begin{array}{l} 0 < P < 1 \\ k > 0 \\ x = 0, 1, 2, \dots \end{array}$$

O parâmetro p é obtido por

$$P = \frac{\bar{x}}{S^2}, \text{ em que } \bar{x} \text{ é a média da distribuição de pontos e } S^2 \text{ é}$$

a variância; o parâmetro k é obtido por

$$K = \frac{\bar{x}p}{(1-p)}$$

Cada um destes parâmetros tem uma significação particular, em termos de sua interpretação geográfica da difusão.

O parâmetro p descreve a relação entre a média da distribuição de pontos e a variância; considerando que na distribuição Poisson, que é aleatória, a média tende a ser igual a variância, este valor de p tenderia para 1 e, portanto, seria uma medida da tendência para aleatoriedade, pois se cada quadrícula tiver um ponto, a média será 1 e a variância zero e p será infinito. O outro parâmetro k é geralmente considerado a taxa de propagação¹³ ou taxa de difusão¹⁴ constituindo, de uma certa forma, uma medida de afinidade espacial entre objetos.¹⁵

Examinando-se, por exemplo, o comportamento deste parâmetro k em uma distribuição verdadeiramente Poisson, em que o parâmetro p é, por definição igual a 1, tem-se que k tende para infinito à medida que p tende para 1. O sentido genético de uma distribuição Poisson é o de que o processo estocástico funciona de forma irrestrita e em consequência a taxa de propagação ou difusão de um determinado fenômeno é infinita. É claro que uma interpretação deste tipo, para ser associada ao processo de desenvolvimento, precisaria estar relacionada com o nível ou estágio deste desenvolvimento. Da mesma forma que o estado de ordem (no sentido de teoria da informação) em uma região poderia ser zero (quer dizer a entropia seria maximizada) a um nível zero de desenvolvimento ou a um nível máximo, utópico, de desenvolvimento, esta taxa de difusão pode assumir o mesmo valor infinito nas duas extremidades do processo de difusão.

Olsson¹⁶ discute este problema, assinalando que é um fato longamente conhecido o de que “uma distribuição Binomial Negativa pode ser obtida para um Universo, se este Universo consiste de diferentes distribuições Poisson, com médias diferentes”. Olsson cita Greenwood e Yule como tendo demonstrado, já em 1920, que uma Distribuição Binomial Negativa pode ocorrer se a média (o parâmetro m) varia de acordo com uma distribuição Gama ou Pearson tipo III, com $2k$ graus de liberdade.

A observação de Olsson é de que, evidentemente, à medida que o parâmetro k tende para infinito, a distribuição Binomial Negativa tende para a Poisson, deixando de haver uma tendência para concentração espacial.

Analisando-se um outro exemplo teórico, pode-se ver que se a média de pontos por quadrado, \bar{x} , for a metade da variância, isto é, se o parâmetro p for igual a 0,55, o parâmetro k assume o valor da média; mas se o parâmetro p for inferior a 0,5 (por exemplo seja igual a 0,2), então o valor de k vai tendendo para zero, pois se p é uma relação entre a média e a variância e se esta relação for tendendo para zero, isto significa cada vez mais quadrados com valores mais baixos e menos quadrados com valores tanto mais altos quanto menor for o valor de p . Isto significaria, em termos de um processo de difusão por contágio, que o mecanismo de difusão praticamente inexistiria.

Estas considerações mostram que o problema do tamanho da quadrícula, conforme já assinalamos na descrição geral da metodologia adotada, é extremamente importante, pois ambos os parâmetros podem sofrer alterações substanciais pela simples modificação do tamanho da quadrícula (numa mesma medida, isto pode ocorrer com a definição da unidade tomada como ponto). Por isso mesmo, a interpretação do parâmetro k e comparações do referido parâmetro precisam estar rigorosamente controladas pela expressa observância dos critérios e testes apli-

cáveis ao modelo binomial negativo. Na análise dos resultados empíricos obtidos nas duas áreas, voltaremos ao problema, face aos próprios resultados obtidos.

É claro que o problema do comportamento do parâmetro k precisa ser examinado em sua componente temporal, pois é aí que temos a definição do processo. Getis¹⁷ sugeriu que se uma seqüência de mapas é examinada, para diferentes momentos de tempo, se o processo de difusão é do tipo contágio e desenvolvendo-se de uma forma exponencial logarítmica, então o valor de k crescerá como função linear do tempo. Entretanto, se o processo reflete apenas heterogeneidade espacial do universo analisado, o parâmetro k permanecerá inalterado, embora a densidade média de pontos possa aumentar.

O comportamento do parâmetro k , em um padrão espacial transversal, também deve obedecer a uma regularidade linear em função do tamanho do quadrado, caso o ajustamento à distribuição Binomial Negativa seja genuíno; aumentando-se a média de pontos por quadrado e se o valor do parâmetro p permanecer inalterado, o parâmetro k aumenta na mesma razão linear. Se k permanecer invariável ou variar de forma não linear, o problema é idêntico ao da série temporal, e então o ajustamento da Binomial Negativa pode refletir apenas não homogeneidade na população amostrada.

É claro que um problema fundamental está implícito neste comportamento; ou o fenômeno que estamos analisando é “scale free”, quer dizer tem soluções idênticas para diferentes níveis de resolução, ou torna-se necessário adotar premissas ou partir de conhecimentos “a priori” da escala (refletida no tamanho do quadrado) em que o processo opera, na maneira como formulamos seus mecanismos. O problema metodológico neste caso, segundo acentua Harvey,¹⁸ é o de identificar a escala em que uma determinada interpretação teórica é válida; pode ocorrer que em uma determinada escala a não homogeneidade local seja o fator mais importante, ao passo que em uma escala maior o processo de contágio pode ser efetivamente o mais significativo. Isto equivale a estabelecer, por interação e testes de contigüidade, o tamanho da quadricula que se enquadra no critério de independência, ou estabelecer “a priori” a escala em que o processo opera, conforme indicamos acima.

Nesta última instância, e considerando, por exemplo, o nível 4b de cidades, seria necessário estabelecer a sua área de influência, adotando-se uma quadricula de tamanho equivalente a esta, que seria algo parecido com o que Hagerstrand definiu como “mean information field” isto é, o seu campo médio de informação. Uma hipótese que se poderia formular a respeito seria a de que, no Brasil, foram identificados 718 centros¹, desde o nível 4b até as metrópoles, o que significa mais ou menos 12 mil quilômetros quadrados por centro, ou seja, aproximadamente quadrados de 1 grau (quando estamos utilizando quadrados de $15' \times 15'$).

Outro problema importante é o de que a função geradora do modelo binomial negativo pode provir de seis modelos diferentes, embora formalmente equivalentes, do tipo contágio, mas resultantes de processos geradores diferentes, todos mencionados no artigo de Dacey sobre a distribuição de casas em Porto Rico.¹¹

O primeiro é o denominado *Amostra Binomial Inversa*, segundo o qual a proporção de indivíduos a mais de k que precisam ser observados para se obter k indivíduos, com uma determinada propriedade, tem uma função de frequência Binomial Negativa.

O segundo é o da *distribuição aleatória de Colônias*, segundo o qual, se grupos de indivíduos são distribuídos aleatoriamente em uma área, de forma que o número de Colônias em amostras da área fixada tem função de frequência Poisson, e se o número de indivíduos nestas colônias é independentemente distribuído segundo uma função logarítmica, então também aplica-se a distribuição Binomial Negativa. Este caso particular tem especial significação na análise do sistema de cidades, pois se temos uma distribuição ao longo da hierarquia de números de indivíduos nas cidades (que de acordo com a regra Tamanho-Hierarquia obedece a uma distribuição lognormal), então a projeção espacial desta distribuição será do tipo Poisson. Brian Berry ⁷ no apêndice do artigo mencionado indica exatamente isso.

O terceiro refere-se a um simples modelo de crescimento populacional, em que se adote premissas de taxas constantes de natalidade, mortalidade e imigração. O número de habitantes, neste caso, segue uma distribuição Binomial Negativa.

O quarto modelo é do tipo aleatório misto, segundo o qual se a média de uma função de frequência Poisson varia de acordo com uma função tipo III ou Gama, então a função de frequência Poisson resultante tem uma distribuição Binomial Negativa. Olsson ¹⁶ chama a atenção para este problema em uma tentativa de estabelecer as bases de um modelo probabilístico para a teoria de localidades centrais.

O quinto tipo é, talvez, o mais comumente associado à distribuição Binomial Negativa, pois refere-se ao processo de nucleamento (*clustering*) amplamente utilizado na biologia e que é geralmente descrito como o processo de distribuição por contágio. Já mencionamos anteriormente o fato de que muitas vezes, por simples heterogeneidade na população, pode-se obter uma distribuição do tipo contágio que, na realidade, é o que Feller denomina de aparente contágio. ¹⁹ A diferença essencial é que na primeira a distribuição de eventos é independente uma da outra e, na segunda, esta independência não existe por falta de homogeneidade na população. Neste caso a gênese do processo é diferente e o ajustamento tem um caráter meramente descritivo, sem possibilidade de inferência estatística.

Dacey assinala em seu trabalho, que a forma limite de um processo de localização para um determinado número de indivíduos em uma área fixada pode ser descrita pela função Binomial Negativa se a probabilidade do *n*-ésimo indivíduo estar localizado numa área particular depender da proporção dos (*n*-1) indivíduos na referida área. Isto significa o próprio processo de contágio, em que o fato de existir um indivíduo com uma propriedade em um determinado lugar, aumenta a probabilidade de existir outro por efeito de proximidade (*neighbourhood effect*). ¹¹

Finalmente, o sexto modelo refere-se a uma distribuição Poisson composta, segundo a qual obtém-se uma distribuição do tipo Binomial Negativa se X_j , para $j = 1, 2, n$, forem variáveis independentes e distribuídas identicamente e de forma aleatória com função Poisson e n for uma função logarítmica; representando a soma $S_n = X_1 + \dots + X_n$, a distribuição Binomial Negativa.

Como se vê, o simples ajustamento de uma distribuição a uma função geradora da Binomial Negativa (na realidade para qualquer outro tipo de distribuição) não é suficiente para inferir uma relação entre o modelo e o exemplo empírico que estivermos analisando. Harvey, em seu extraordinário livro ²⁰, ressalta bastante o fato de que mapear uma realidade empírica em um modelo matemático não leva a nada se as duas coisas não estiverem geneticamente associadas.

2.4 — As distribuições Poisson e Poisson modificada

A distribuição Poisson tem importantes aplicações e descreve o funcionamento irrestrito de um processo estocástico, do tipo Markoviano, (porém apenas no sentido de que o estado do sistema no tempo T_0, T_1, \dots, T_n é independente da história prévia do sistema no estado imediatamente anterior a T_0 ou T_n conforme o caso); isto quer dizer que o número de eventos em um determinado momento de tempo é independente do número de eventos em qualquer outro intervalo. Outra premissa essencial é a de que $N(t_i) - N(t_j)$ é uma variável aleatória Poisson, com média $(t_i - t_j)$, o que significa que o número de eventos que ocorrem entre t_i e t_j têm média proporcional à diferença $t_i - t_j$. Por isso uma de suas principais aplicações é na medida de confiabilidade de sistemas físicos; uma das premissas de sistemas deste tipo e que podem ser mapeadas no modelo teórico é o de que o número de defeitos em um sistema de n componente, em um determinado intervalo de tempo, é distribuído como uma variável Poisson com média proporcional à extensão do intervalo.

A distribuição Poisson, entretanto, não se limita a aplicações em eventos raros a intervalos temporais; o intervalo pode ser o espaço, e aplicações em mecânica estatística indicam que o número de partículas em uma dada região, é também uma variável aleatória Poisson, com média proporcional ao volume da região.

A função de densidade da variável aleatória Poisson é dada por

$$P(x) = \frac{e^{-M} M^x}{x!} \quad \begin{array}{l} x = 0, 1, 2, 3 \dots \dots \dots \\ M > 0 \end{array}$$

Uma particularidade importante da distribuição Poisson é que ela é muitas vezes tomada como uma aproximação limite da distribuição Binomial, portanto um caso particular da mesma, em que n (o número de eventos) é grande e p (a probabilidade do evento ocorrer) é pequena em relação a n (n em geral superior a 50 e p inferior a 0.01), portanto com uma probabilidade muito elevada de o evento não ocorrer. Teoricamente, na distribuição Poisson, a média e a variância são iguais, do que se pode derivar um índice de dispersão $D = V/M$ (Variância/Média), que em casos práticos tenderia para 1², já que nesse caso a igualdade dos parâmetros é quase impossível.

O mesmo problema de tamanho de quadrado é significativo para a validade da interpretação de uma distribuição Poisson, como refletindo uma distribuição aleatória do fenômeno estudado e a média e a variância continuariam tendendo a igualdade proporcionalmente à diferença do volume (no caso espacial da área) entre as duas unidades consideradas.

A distribuição Poisson aplicada a uma distribuição de cidades descreveria a ação de um processo estocástico irrestrito (no caso muitas forças moldando o sistema urbano, sem que nenhuma fosse suficientemente forte para criar padrão próprio), responsável pela distribuição espacial de cidades e sua expansão entre um período t_i e t_j .

Dacey², partindo da premissa de que seria “altamente improvável que distribuições geográficas, particularmente padrões locais envolvendo decisões humanas fossem resultantes de eventos equiprováveis”, supõe que “padrões observados em um mapa refletem um certo sistema de ordem”. Dacey descreve mais adiante que “um padrão de

pontos pode ser descrito, para muitos propósitos de análise, por três tipos de processos espaciais:

- 1) O processo aleatório ou estocástico;
- 2) O processo de contágio ou concentrado; e
- 3) Um processo de distribuição regular ou geométrica.

As duas primeiras contêm exemplos abundantes na literatura e o próprio Dacey realizou numerosos estudos já citados em diferentes pontos deste trabalho sobre as mesmas. Quanto à terceira, Dacey procurou aplicar uma modificação da probabilidade Poisson, que descrevesse “um arranjo mais regular ou sistemático do ponto que o produzido por um processo estocástico”.

Para ilustrar e descrever tal processo Dacey usa o exemplo clássico da urna, contendo um número N de bolas, numeradas de 1 a N (a premissa necessária é a de que N seja grande e p pequeno, para poder usar-se a distribuição Poisson como limite da Binomial). Nesta urna forma-se uma amostra, retirando-se um número n de bolas, *sem reposição*; em uma segunda amostra (devolvidas todas as bolas à urna), um número m de bolas é retirado, com reposição. Considere-se que $P_1(1)$ seja a probabilidade de que uma determinada bola de número i seja retirada, uma vez na primeira amostra e $P_2(z)$ seja a probabilidade de que esta mesma bola seja retirada um número z de vezes na segunda amostra.

O problema é achar-se a probabilidade $p(x)$ de que uma bola de número i seja retirada um número x de vezes nas duas amostras (obviamente para $x = 1 + z$), vale dizer que no caso da segunda amostra (*com reposição*) o mecanismo é aleatório, enquanto que na primeira amostra um número é retirado apenas uma vez (*uma vez que não há reposição*) e o processo torna-se regular ou sistemático.

$$P_1(y) = \begin{cases} p & \text{para } y = 1 \\ q = 1-p & \text{para } y = 0 \end{cases}$$

Na segunda amostra (que segue a distribuição Binomial) para N grande a distribuição Poisson é tomada como aproximação, assim para o caso de na primeira amostra uma bola de número i ser retirada como probabilidade p , a probabilidade de a mesma ser retirada um número x de vezes nas duas amostras será $y = 1$ e $z = (x-1)$; para o caso de a mesma bola não ser retirada na primeira amostra a probabilidade será em função de $y = 0$ e $z = x$.

Para esta distribuição Dacey aplica a equação

$$\pi(x) = (1-p) \frac{e^{-M} M^x}{x!} + p \frac{e^{-M} M^{(x-1)}}{(x-1)!} \text{ para } 1 > p > 0$$

na qual M é a média da distribuição e p é $\sqrt{M_1 - M_2}$, sendo M_2 a variância. Observe-se que esta equação só é possível se a média for superior à variância, o que faz do parâmetro p , nesta equação, uma medida da tendência para regularidade na distribuição.

Já tínhamos assinalado que um dos problemas críticos em se avaliar a qualidade e validade do ajustamento da distribuição observada a uma distribuição teórica, refere-se ao tamanho da quadrícula utilizada e por via deste tamanho a própria idéia de independência da ocorrência de um evento, no caso um centro 4b, e outro evento espacialmente adjacente; isto é, existência de autocorrelação espacial.

De um lado podemos verificar isso pelo exame do comportamento dos diferentes parâmetros das diferentes distribuições, a níveis de agregação (*tamanho da quadrícula*) diferentes e já salientamos que numerosos autores tem chamado a atenção para o fato de que o parâmetro k , por exemplo, na Binomial Negativa, seja uma função linear do tamanho do quadrado, o que será utilizado neste estudo, no capítulo referente aos resultados empíricos obtidos.

2.5 — A análise de contigüidade e efeito de proximidade

De outro lado tanto Harvey como Dacey e Cliff chamam a atenção para a necessidade de análises complementares, do tipo contigüidade, por exemplo. Análises de contigüidades podem levar a diferenciar entre uma distribuição Poisson generalizada e uma Poisson composta (porque é resultante de um processo de contágio, mas no qual a heterogeneidade da amostra é evidente e constituída de diferentes sub-amostras que se ajustam a um dos modelos de contágio). Seria necessário que o tamanho da quadrícula fosse tal que a distribuição de pontos em torno de um núcleo fosse distinta de outra distribuição, vale dizer que a quadrícula deveria ser suficientemente grande para conter todo o conjunto de pontos que se aglutina em torno de um núcleo.

Para tal fim estamos utilizando um modelo de contigüidade desenvolvido por Cliff¹², para testar simultaneamente a independência de ocorrência do evento, mas também para examinar o possível direcionamento do processo de difusão.

A análise de contigüidade pode ser um instrumento auxiliar importante na interpretação da natureza do processo de difusão. Em primeiro lugar, porque ela contribui para avaliar o significado do efeito locacional sobre a estrutura de um lugar, permitindo ou facilitando a compreensão de uma variável tão significativa na Geografia como é a localização geográfica. Em segundo lugar, porque permite uma avaliação da escala em que um determinado fenômeno tem um significado em termos do processo espacial. Um mapa é preparado, para efeito desta forma de análise, indicando as propriedades de cada quadrícula (possui ou não a referida propriedade) e pode mostrar a distribuição de vários fenômenos. O teste de contigüidade aplicado a mapas assim, pode ser usado para dois propósitos distintos:

1) A constatação de concentração espacial de um determinado fenômeno, que pode estar associada a um determinado tipo de processo de difusão. Na realidade pode-se, por esta via analítica, constatar se a concentração existe ou não: mas, simultaneamente, se o fenômeno é conhecido — quer dizer existe uma distribuição concentrada — o uso de diferentes modelos de contigüidade pode oferecer ao pesquisador a possibilidade de determinar se o processo de difusão obedece a uma orientação própria, ou se ele se processa concêntricamente a partir de um centro em que ele ocorra.

2) Como os modelos de contigüidade calculam (por via dos algoritmos usuais) a distribuição aleatória do fenômeno e assim se pode ter para determinados tipos de distribuição o valor esperado e o valor observado, há um resíduo: a diferença entre o valor esperado e observado. Uma das premissas básicas da análise de correlação é a de que tais resíduos são variáveis mutuamente independentes. Em uma distribuição espacial, esta premissa implica em independên-

cia da distribuição dos resíduos, vale dizer que o fato de haver uma determinada quadrícula na área em estudo e em que o fenômeno ocorra, não afeta a ocorrência do mesmo fenômeno em outra quadrícula; isto a rigor significa que a área da quadrícula é igual à área sobre a qual o fenômeno estudado exerce sua influência, e para que o mesmo ocorra na quadrícula adjacente é preciso que a natureza geral do fenômeno seja tal que se espere a ocorrência do mesmo naquela quadrícula, independentemente da existência da quadrícula anterior adjacente.

O modelo de contigüidade poderia ser bem utilizado para testar o espaçamento entre centros de nível não só 4b, mas também os outros níveis acima. A teoria de localidade central postula precisamente que centros de níveis hierárquicos diferentes têm espaçamentos também diferentes, proporcionais ao tamanho do centro e por via de consequência de sua área de influência. Usando-se a distribuição espacial de centros dos diferentes níveis, com quadrículas de áreas aproximadamente iguais à área de influência dos centros, o modelo de contigüidade poderia ser aplicado e ele testaria a validade da distribuição, partindo-se da premissa de que o tamanho da área de influência seria corretamente dada. A não validação implicaria em se tentar identificar a causa, se ligada ao próprio tamanho da quadrícula ou se a não aleatoriedade da distribuição dos centros. O ajustamento ao modelo BN ou Poisson, em termos de distribuição de pontos, poderia ser assim testado, pois não aleatoriedade no modelo de contigüidade e ajustamento ao modelo BN na distribuição de pontos seriam evidências coincidentes, de um processo contágio; principalmente se, por exemplo, o modelo BN se ajustasse melhor com agregação de quadrículas na mesma direção em que fosse indicado no modelo de contigüidade.

No presente estudo estamos usando quadrículas agregadas como foi utilizado para o ajustamento à Poisson e Binomial Negativa. Cada quadrícula é assinalada com um B se contém um centro 4b, com um W se não contém um centro 4b. A seguir são formados pares de quadrículas (*juntas*), que contém um centro 4b em cada uma, que uma contém um centro 4b e a outra não, e que nenhuma contém; quer dizer formamos pares BB, BW e WW de quadrículas, com o objetivo de aplicar o modelo.

Com o propósito de determinar se a presença de um centro 4b em quadrículas vizinhas é ou não um fenômeno independente, a distribuição do número de quadrículas contíguas, ambas B (quer dizer contém um centro 4b), é examinada, usando-se o método da taxa de contigüidade. O método permite analisar esta distribuição, determinando o grau de aleatoriedade (e inversamente o grau de concentração) de pares de quadrículas que tenham sido caracterizadas como BB, BW ou WW; se um efeito de vizinhança existe, a presença de um centro 4b (ou um adotador da inovação no caso genérico) em uma determinada quadrícula deverá tornar a presença de um outro adotador ou centro 4b mais provável, naturalmente em uma quadrícula adjacente. Quer dizer, se tal efeito existe e opera no sentido de organizar uma distribuição espacial de centros 4b, pares de quadrículas BB seriam significativamente mais numerosos, as BW seriam menos numerosas quase que por definição.

Simultaneamente, partindo-se ou não de hipóteses sobre direcionamento no efeito de vizinhança (veremos mais adiante que a distribuição de pontos segue um padrão Binomial Negativo no Nordeste, quando as quadrículas são agrupadas na direção Norte-Sul), procura-se agrupar as quadrículas com diferentes sentidos direcionais, exata-

mente à procura de uma tendência do processo de difusão que o efeito de vizinhança descreve. É óbvio que o primeiro estudo parte da hipótese inicial de que não há uma espécie de condicionamento direcional ao processo de difusão, e que ele obedeceria simplesmente a uma função distância, a partir dos centros de difusão.

Assim a hipótese inicial a testar-se é a de que o número de pares de quadrículas, ambas contendo um centro 4b, não seria significativamente maior que o esperado por uma distribuição aleatória, ao mesmo tempo que o número de pares BW não seria significativamente menor que o esperado.

A fórmula que descreve esta aleatoriedade na distribuição é:

$$K(j) = \frac{J - E(J)}{T(J)}$$

na qual $K(J)$ é tratado como o desvio normal estandardizado. J é o número de pares de um certo tipo (cor) observado, $E(J)$ é o valor esperado e $T(J)$ é o desvio padrão, respectivamente, para uma cor J , em uma distribuição aleatória. Não aleatoriedade no padrão de pares de uma determinada cor é indicada por maior ou menor diferenciação, positiva ou negativa, do valor aleatório esperado, portanto de $K(J)$.

Um problema precisa ser esclarecido, em termos operacionais e conceituais, e que é de fundamental importância: como definir a contigüidade? Esta definição precisa estar previamente associada a uma noção sobre o direcionamento do processo de difusão e pode ser testado segundo cinco linhas diferentes:

- 1) Um processo de difusão leste-oeste, em que se considera duas quadrículas contíguas apenas se elas são vizinhas nesta direção;
- 2) Um processo de difusão norte-sul, segundo a mesma forma indicada acima;
- 3) Um processo de difusão noroeste-nordeste e sudoeste-sudeste;
- 4) Idem noroeste-sudoeste e nordeste-sudeste;
- 5) Não direcionado (o chamado Queen's case), em que a contigüidade seria definida nos oito sentidos, dos lados e dos vértices da quadrícula.

O teste de contigüidade foi aplicado para os dois primeiros tipos citados acima, tomando-se um quadrado composto de quatro quadrículas de $15' \times 15'$, com o objetivo de testar a hipótese de um direcionamento do processo de difusão.

Finalmente uma análise em diferentes momentos de tempo daria uma indicação da evolução temporal do processo, realmente em termos de saber-se se a presença de um centro 4b em um período t induz à existência de outro centro 4b no tempo $t + 1$, intervalo este obviamente suficiente para tal tipo de difusão. Este tipo de análise não foi feito, entretanto, uma vez que não dispomos dos dados necessários.

Uma das alternativas de investigação sobre velocidade do tempo de difusão, naturalmente sujeita a restrições compreensíveis, é a de fazer-se a análise de contigüidade para diferentes tipos de bens duráveis e para os mesmos na população rural e urbana, e separadamente para as duas regiões (Nordeste e Centro-Sul).

3 — O EXEMPLO EMPÍRICO UTILIZADO: DOIS SUBSISTEMAS DE CIDADES NO BRASIL — NORDESTE E CENTRO SUL, E SUAS CARACTERÍSTICAS PARTICULARES

Tratar o processo de desenvolvimento como um processo de difusão de inovações já não constitui mais tema controverso. Tanto economistas clássicos como Kuznets²¹, ou economistas — geógrafos como Friedmann²² ou geógrafos propriamente ditos, como Brian Berry²³ definem o processo de desenvolvimento da mesma forma. Friedmann define o desenvolvimento como “processo cumulativo, descontínuo, que ocorre como uma série de inovações elementares, que se organizam em núcleos (*clusters*) de inovações e finalmente sistemas, em larga escala, de inovações. As inovações podem ser técnicas ou institucionais, caso em que podem ser subdivididas nas costumeiras categorias de social, econômica, política e cultural”, pp. 86. O efeito cumulativo, acrescenta Friedmann, de sucessivas inovações, é o de “transformar a estrutura estabelecida da sociedade, atraindo personalidades inovativas e criadoras aos núcleos de mudança acelerada, encorajando a formação de novos valores, atitudes e comportamentos consistentes com a inovação, fomentando um ambiente social favorável à atividade inovativa e gerando, continuamente, novas inovações”, pp. 87.

Friedmann deixa claro que a inovação apresenta um padrão espacial, localizado em um ou em um pequeno número de centros de mudança, dos quais ela tende a se difundir para baixo e para os lados (*downward and outward*), constituindo, em determinado momento, o que ele mesmo denomina região núcleo (*core region*), em contraste com o resto do sistema espacial definido por ele como periférico. O fato mais importante do mecanismo é que estas regiões núcleo constituem subsistemas da sociedade, territorialmente organizados, e que possuem uma elevada capacidade de mudança inovativa; as regiões periféricas constituem, de outro lado, subsistemas cujo padrão de desenvolvimento é “determinado principalmente pelas instituições da região (ões) núcleo, em relação às quais guardam uma relação de dependência”, pp. 93.

Berry, a partir desta concepção e procurando incorporar conceitos desenvolvidos por numerosos outros autores (especialmente Hagerstrand, Hudson e Pedersen, todos mencionados no artigo de Berry), acrescenta novos argumentos “focalizando modelos que envolvem premissas de maximização de probabilidades, e associando a difusão hierárquica e espacial das oportunidades de crescimento — a nível nacional — de um centro urbano para outro aos efeitos de difusão espacial no âmbito de cada centro urbano, resultantes do uso que as pessoas (ou domicílios, como conjunto de pessoas em unidade invisível) fazem destas oportunidades”, pp. 109²³. Numa visão ampla trata-se de distinguir entre o que Pedersen²⁴ define como inovação empresarial — considerada aquele tipo de inovação que, uma vez adotada, tem consequências diretas para pessoas outras que não as que adotaram a inovação, de um lado; de outro lado o que ele denomina de inovações domiciliares, cujo impacto imediato se faz somente na unidade de adoção (o indivíduo, a família ou os moradores de um domicílio). Berry acrescenta que o crescimento regional depende e abrange as duas coisas, pois é tanto função da difusão de inovações empresariais, quanto da utilização que as pessoas possam ou venham a fazer da oportunidade da inovação.

A identificação de centros (tais como os centros 4b e superiores) que possam comandar a distribuição hierárquica e espacial das inovações, constitui, obviamente, um elemento de importância essencial para

a análise do processo espacial de difusão. Ao mesmo tempo, a distribuição destes centros, a área que cada um deles comanda (vale dizer a densidade de ocorrência dos mesmos) constitui elemento de igual importância.

É claro que tal mecanismo pode ser analisado em seu aspecto agregado, por via de um índice de urbanização que contenha, intrinsecamente, um conjunto de inovações (uma vez que, em geral, os mesmos domicílios que têm uma, têm também numerosas outras) *, da mesma forma que se pode — na realidade é desejável que se faça — analisar o padrão de difusão de cada elemento em particular, em busca de conjuntos que tenham padrões de difusão semelhantes, talvez associados a mecanismos diferentes ou níveis de renda.

No presente trabalho estamos restritos ao padrão agregado, pois que ele pode ser comparado com a análise do padrão de distribuição de pontos, que representa o centro de difusão de menor tamanho (naturalmente estabelecido em termos de um “*threshold*” mais ou menos arbitrário).

3.1 — As dimensões básicas do sistema urbano — hierarquia e grau de urbanização — e seus reflexos no processo de difusão de inovações

Em numerosas análises feitas do sistema urbano brasileiro, duas dimensões básicas apareceram, de forma nítida, (entre outras não relevantes para o propósito da presente análise) e que são o *Tamanho Funcional* e a *Urbanização e Desenvolvimento*, obviamente associadas ao mecanismo conjugado de difusão hierárquica e espacial³. Na última destas análises, abrangendo um conjunto de 644 cidades e 30 variáveis estas duas dimensões apareceram, também, de forma bem definida, a primeira servindo de base para uma hierarquia e a segunda para a definição do nível de desenvolvimento (em suma delimitando a região Núcleo e a Periferia). Como tem sido postulado sempre em estudos do processo de difusão, ele é mais rápido nas regiões mais desenvolvidas e de maior acessibilidade que nas regiões subdesenvolvidas e de menor acessibilidade.

A tabela 1 mostra os índices de tamanho funcional para um certo número de cidades brasileiras (na região sub e na região desenvolvida), que refletem bem o fato de que na primeira os índices de urbanização e desenvolvimento são sempre mais baixos e, em geral, muito baixos nas cidades pequenas, enquanto que na região desenvolvida, cidades grandes, inclusive as metropolitanas, têm níveis de urbanização e desenvolvimento mais elevados, mas pouco diferenciados dos níveis de cidades menores.

A implicação destas constatações é a de que na região desenvolvida a difusão de inovações já está percorrendo a hierarquia e disseminando-se pelos centros subordinados. Compare-se na tabela mencionada Campinas com Fortaleza, por exemplo, ambas com tamanhos funcionais quase iguais, mas com índices de urbanização e desenvolvimento bem diferentes, Campinas com um “*score*” quase duas vezes

* Em análise fatorial do conjunto das 360 microrregiões brasileiras, um fator desenvolvimento emergiu da correlação de numerosas variáveis deste tipo, como população servida por água, esgoto, rádio, geladeira, televisão etc...

TABELA 1

TAMANHO FUNCIONAL E DESENVOLVIMENTO URBANO

Cidades	Tamanho Funcional	Desenvolvimento Urbano
Fortaleza	25,5	6,6
Recife	33,9	11,2
Salvador	30,5	9,9
Natal	12,2	4,7
João Pessoa	12,3	5,9
São Luís	12,8	4,4
Teresina	9,2	2,9
Ihéus	3,9	1,5
Campina Grande	11,4	4,7
Crato	3,6	3,3
Juazeiro do Norte	2,1	— 2,9
Parnaíba	1,9	— 3,0
Moreno	— 5,2	— 7,4
Santarém	— 0,4	— 3,8
Olinda	1,7	— 4,7
Araçatuba	7,6	7,7
Bauru	13,5	9,5
Botucatu	4,2	8,5
Jundiá	11,5	8,1
Piracicaba	11,9	9,3
Santos	24,0	14,5
Campinas	23,0	11,6
Ribeirão Preto	17,9	11,3
São José do Rio Preto	11,8	11,8
São Bernardo	11,7	6,5
São Caetano	11,3	8,4
Guarulhos	8,8	— 0,9
São Paulo	60,6	18,8
Mauá	— 0,8	— 4,0
Rio	55,9	18,7
Duque de Caxias	12,7	1,7
Nova Iguaçu	14,2	— 0,9
Vitória	14,2	9,6
Vila Velha	2,1	— 0,9
Belo Horizonte	35,5	13,3
Barbacena	6,5	7,1
Governador Valadares	7,6	3,7
Ituiutaba	2,5	5,5
Juiz de Fora	15,6	8,6
Uberlândia	8,5	6,7
Goiânia	17,3	7,7
Curitiba	30,4	12,0
Londrina	13,1	9,2
Maringá	8,8	7,2
Campo Mourão	— 0,2	1,3
Florianópolis	9,7	7,6
Blumenau	11,2	6,9
Joinville	9,8	4,8
Porto Alegre	37,3	15,2
Pelotas	14,3	8,7
Santa Maria	8,7	6,2
Caxias do Sul	10,6	7,2
Passo Fundo	5,9	5,6
Bagé	4,3	5,2
Sapuçaia do Sul	— 2,6	— 3,0
Alvorada	— 3,6	— 8,2
Cachoeira do Sul	3,3	4,6
Cruz Alta	3,6	5,0
Erechim	4,0	6,3
Dom Pedrito	— 2,6	— 0,1
Caçapava do Sul	— 3,0	— 0,3

maior. Observe-se que a população de Campinas é menos da metade da população de Fortaleza. Ao nível das metrópoles, a diferença não é tão sensível, exatamente porque o processo de difusão está ocorrendo — embora numa aceleração reduzida — na hierarquia metropolitana, mas não (no caso da região subdesenvolvida) pela hierarquia abaixo do nível das metrópoles e capitais de Estado no Nordeste. Porto Alegre (apesar de ter população bem menor que Recife) tem um tamanho funcional maior, e um nível de urbanização e desenvolvimento superior, mas não muito ao de Recife (15,2 para 11,2). Já uma cidade do interior de São Paulo, como Ribeirão Preto, tem um nível de urbanização igual ao de Recife, embora tamanho funcional consideravelmente menor; São José do Rio Preto tem tamanho funcional igual ao de Natal (11,8 e 12,2) e urbanização muito superior (11,8 e 4,7), o que ilustra bem o que afirmamos acima. Numerosas outras comparações podem ser feitas, observando-se os valores da tabela mencionada.

Um processo de difusão assim caracterizado, bem diferente no Centro Sul comparado com o Nordeste, tem um poder de geração de centros de tamanho menor também diferenciado. Não é por outra razão que cerca de 70% dos centros de nível 4b e mais estão na região Centro Sul (506 centros), contra apenas 215 no Nordeste.

Uma distribuição deste tipo, em cada uma destas duas regiões, pode ser expressa em termos da probabilidade de cada área (cada quadrado seguindo a metodologia que estamos utilizando na análise do padrão de distribuição de pontos) conter um ou mais centros 4b ou superiores. Esta distribuição, por sua vez, ajustada a uma distribuição teórica, pode revelar, ao mesmo tempo, a natureza do processo de difusão e, por via dela, uma compreensão do estágio de urbanização e desenvolvimento. O que estamos procurando fazer no presente trabalho é comparar o tipo de distribuição que mais se ajusta a cada caso, com os resultados das análises do sistema urbano, utilizando um número de variáveis que descreva tal nível de urbanização e desenvolvimento.

3.2 — A hierarquia funcional e o multiplicador urbano em suas relações com o processo de difusão de inovações

Um dos postulados básicos da “*Field Theory*” de Berry²⁵ é o de que os atributos dos lugares (como podem estar medidos nas duas dimensões básicas do sistema urbano que apresentamos) estão funcionalmente associados às relações entre estes mesmos lugares; portanto, embora constituindo duas coisas distintas, são interrelacionadas e interdependentes. Assim o multiplicador urbano pode ser calculado de forma teoricamente correta, usando-se de um lado uma hierarquia funcional (derivada do sistema de relações) e de outro um atributo dos lugares (por exemplo, a sua população e a população servida pelos centros); o multiplicador urbano e o fator crescimento, como foram descritos na parte metodológica, podem constituir importantes elementos de comparação com resultados que pretendem descrever o processo de difusão, desde que se aceite a premissa básica de que um sistema em equilíbrio geraria um fator crescimento igual para todos os níveis hierárquicos. Diferenças entre os diferentes fatores crescimento seriam indicadoras do desequilíbrio no sistema e em seus diferentes níveis hierárquicos.

Estas diferenças podem existir — e em diferentes graus — tanto nas relações entre número de centros servidos, como em termos de população dos referidos centros, como veremos mais adiante.

O padrão de distribuição de cidades nos dois subsistemas analisados foi obtido, conforme indicação inicial, identificando-se os centros de nível 4b e acima, à base das folhas da carta do Brasil ao milionésimo¹ dividindo-se cada folha em quadrículas de 15' × 15' e contando o número de centros 4b em cada uma delas, obtendo-se assim uma distribuição de pontos sobre o mapa. O mesmo procedimento foi repetido, substituindo-se os centros 4b pelas sedes de municípios (que naturalmente constituiriam a área de influência de cada centro 4b ou de hierarquia superior) (ver mapas 2, 3 e 4).

Mais de uma vez mencionamos no presente trabalho a importância do tamanho da quadrícula para uma adequada interpretação do padrão observado em termos de uma distribuição probabilística, seja Poisson, Binomial Negativa ou outra. É que cada uma delas parte de premissas precisas sobre as características da distribuição gerada e se o processo que estivermos analisando (no caso os mecanismos de mercado que geram a distribuição de centros 4b e superiores) não puder ser mapeado neste modelo particular, estaremos sendo levados a interpretações incorretas.

Uma análise superficial da distribuição destes centros já pode oferecer algumas hipóteses iniciais importantes. A divisão em regiões funcionais urbanas a que aludimos partiu inicialmente das dez metrópoles (incluindo-se a área de influência de Goiânia pela sua significação territorial e pelo número de centros em sua área), que foram classificadas como de nível 1, desprezando-se diferenças intermetropolitanas, de hierarquia ou subordinação.*

A divisão em dois subsistemas — na realidade o Núcleo e a Periferia — constitui já um aspecto indiscutido da organização espacial brasileira, tratado por numerosos autores de diversas maneiras. Na própria análise global do sistema de regiões funcionais urbanas, Elza Keller¹ chama a atenção para o fato de que “como decorrência destas desigualdades regionais pode-se considerar o espaço nacional dividido em duas áreas absolutamente distintas sob o ponto de vista de organização urbana: de um lado o Centro-Sul, onde se localizam as metrópoles nacionais e mais as metrópoles regionais de Belo Horizonte, Porto Alegre e Curitiba; de outro lado, a Amazônia e o Nordeste dominados pelas metrópoles regionais de Recife, Salvador, Fortaleza e Belém”.

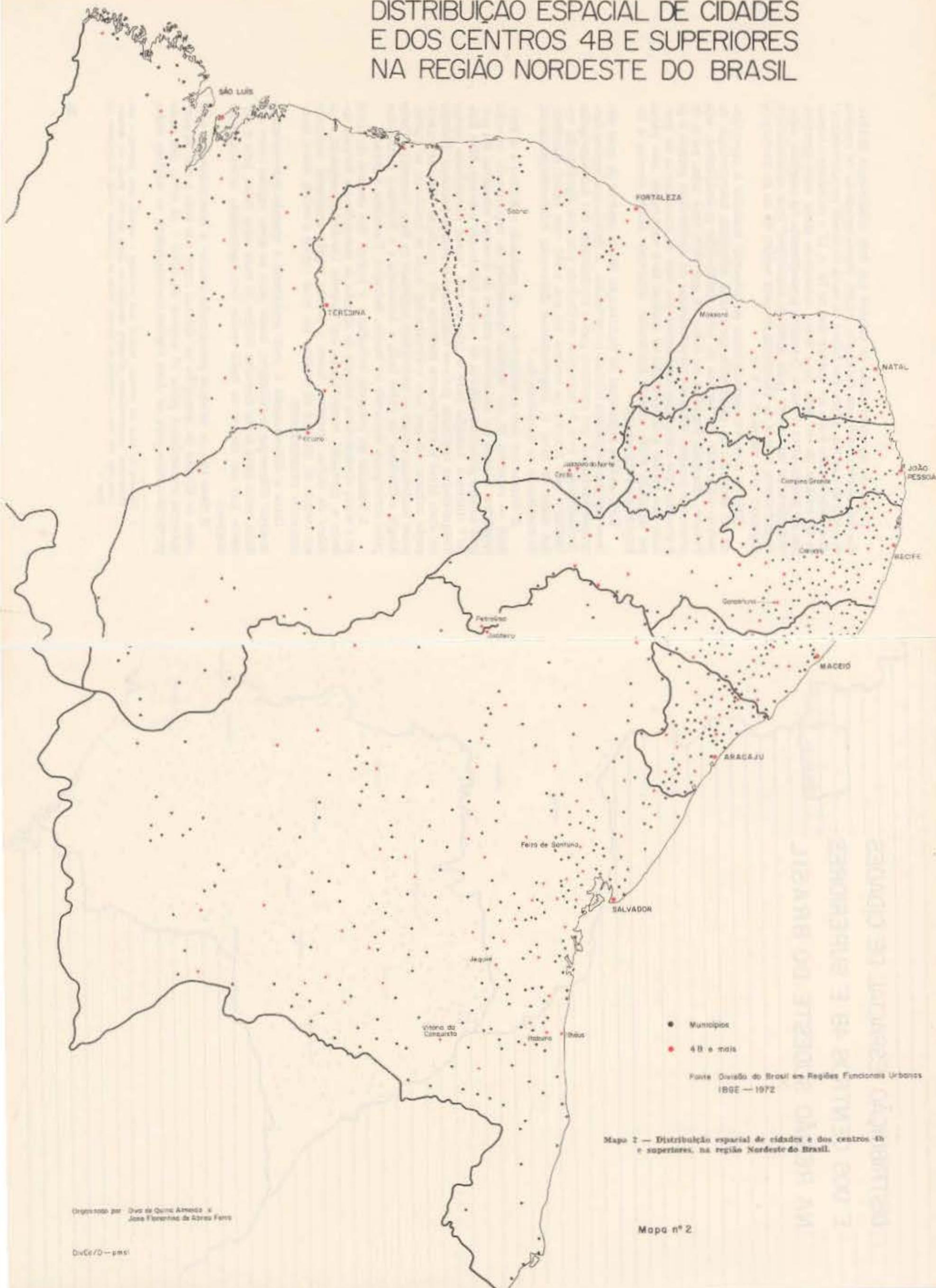
A essência do presente trabalho é a análise dos centros de nível 4b e mais, considerados núcleos de difusão de inovações no sistema urbano e, portanto, constituindo um sistema hierárquico-espacial, que difere de várias maneiras numa e noutra região analisada, inclusive em termos de número de centros em uma região e outra, como mencionamos anteriormente.

Onde a diferenciação é mais acentuada é quando começamos a analisar diferenças particulares, tanto para as duas regiões como um todo, como para áreas menores dentro de cada uma delas, pois é evidente que em ambos os casos existe heterogeneidade intra-regional também.

É claro que, se estamos procurando estudar a intensidade do processo de difusão de inovações no sistema como um todo e nos 10 subsistemas comandados pelos 10 centros de nível 1 (é óbvio que mais particularmente para as duas regiões consideradas), as relações entre

* Embora no texto do trabalho aludido seja chamada a atenção para diferenças nas hierarquias metropolitanas, inclusive relativas ao papel de São Paulo no sistema, como este não era o propósito da divisão, todas as metrópoles foram classificadas no nível 1.

DISTRIBUIÇÃO ESPACIAL DE CIDADES E DOS CENTROS 4B E SUPERIORES NA REGIÃO NORDESTE DO BRASIL



Fonte: Divisão do Brasil em Regiões Funcionais Urbanas IBGE — 1972

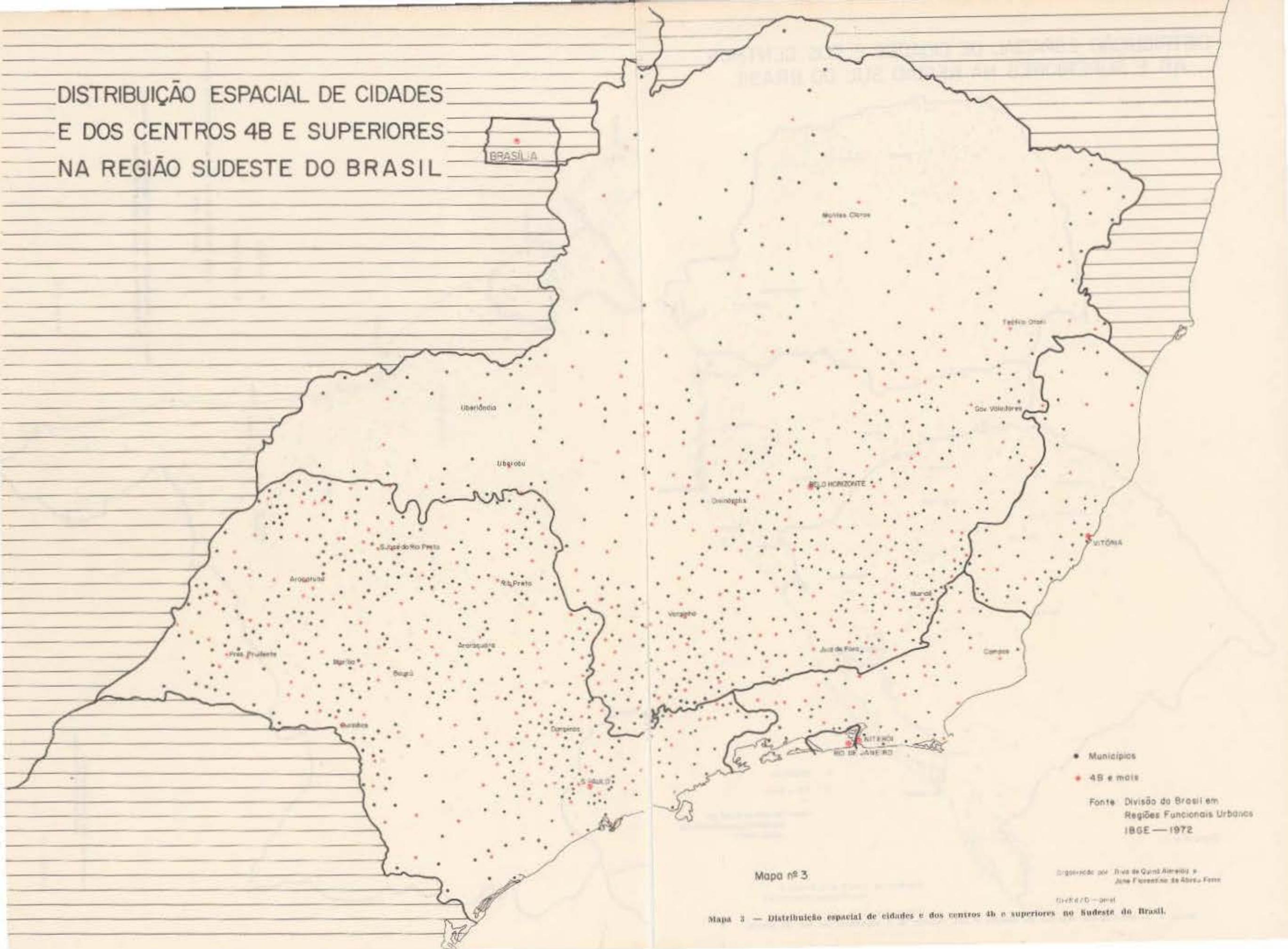
Mapa 2 — Distribuição espacial de cidades e dos centros 4B e superiores, na região Nordeste do Brasil.

Organizado por: Divo de Queiroz Almeida e José Florentino de Aguiar Ferraz

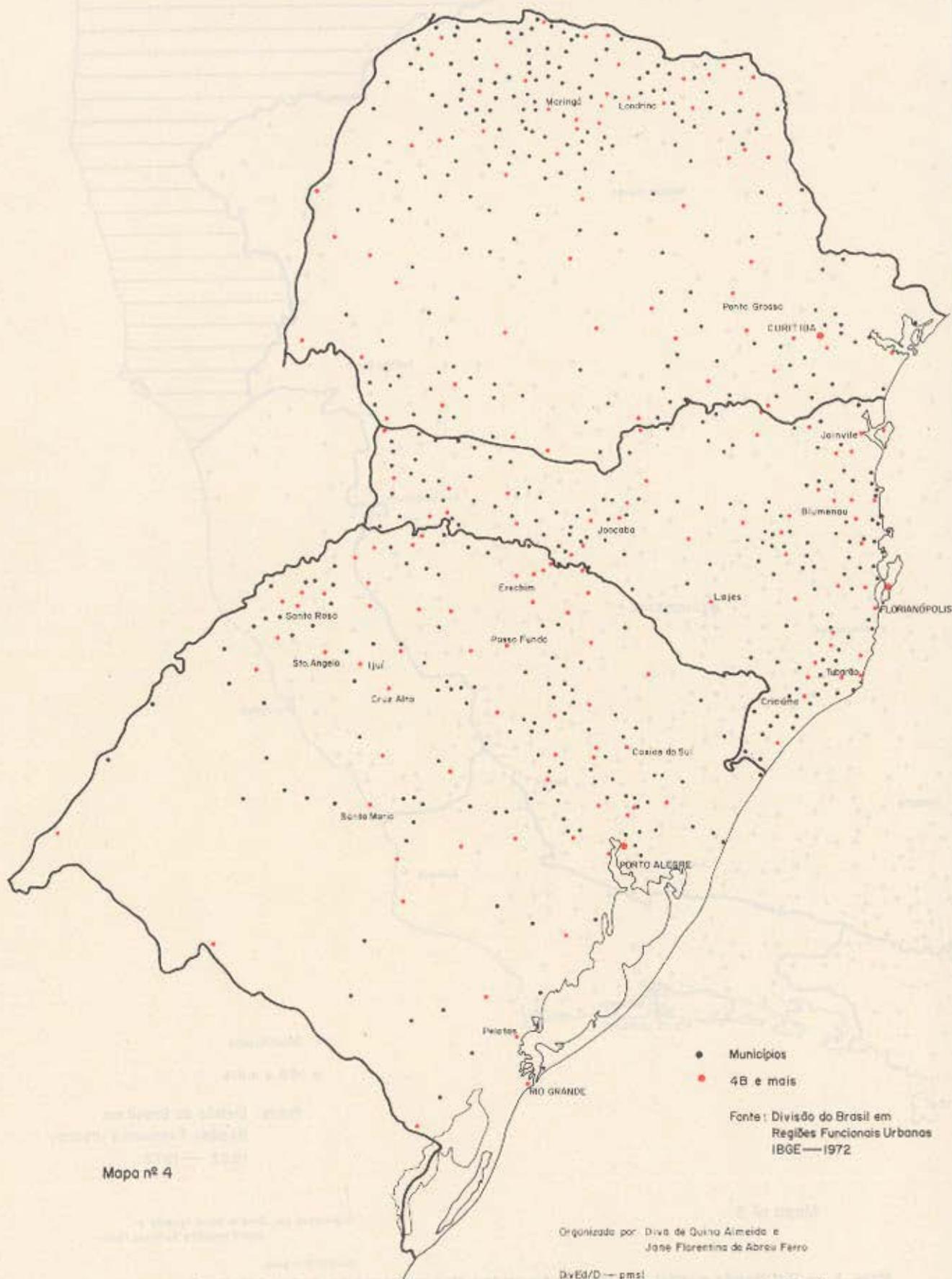
DivGe/D — pms

Mapa nº 2

DISTRIBUIÇÃO ESPACIAL DE CIDADES
E DOS CENTROS 4B E SUPERIORES
NA REGIÃO SUDESTE DO BRASIL



DISTRIBUIÇÃO ESPACIAL DE CIDADES E DOS CENTROS 4B E SUPERIORES NA REGIÃO SUL DO BRASIL



Mapa 4 — Distribuição espacial de cidades e dos centros 4b e superiores no Sul do Brasil.

os centros de nível 1 e 2 são as de interesse mais imediato. Se cada centro metropolitano constitui o principal elemento no processo de difusão, será na sua área mais imediata que o efeito se fará sentir de forma mais intensa. Isto significa medir ao mesmo tempo o espaçamento (ou a área de influência de cada um) e o multiplicador urbano.*

A análise da tabela 2 permite obter-se uma visão global do número de centros de cada categoria em cada subsistema, e de cidades (estas abaixo do nível 4b), com evidentes disparidades entre as mesmas, às quais estão associados numerosos fatores de ordem econômica.

TABELA 2
ORGANIZAÇÃO DA HIERARQUIA URBANA POR REGIÃO METROPOLITANA

Centros de Nível 1*	Nível									Muni- cípios
	2			3			4			
	Total	a	b	Total	a	b	Total	a	b	
Fortaleza	4	2	2	10	5	5	40	9	31	304
Recife	8	5	3	18	7	11	67	10	57	618
Salvador	7	4	3	11	4	7	55	8	47	344
Belo Horizonte	4	1	3	18	10	8	57	22	35	398
Rio de Janeiro	5	3	2	13	10	3	36	7	29	215
São Paulo	17	7	10	60	20	40	153	38	115	897
Curitiba	7	3	4	11	5	6	42	13	29	202
Porto Alegre	12	3	9	11	2	9	53	23	30	225
Goiânia	2	1	1	8	5	3	31	11	20	200
Total	66	29	37	160	68	92	534	141	393	3.403

* Belém não foi incluído por não fazer parte da área selecionada para este estudo.

A tabela 3 mostra bem as diferenças entre a área média de influência dos centros 4b e superiores para as dez cabeças de sistemas, o que já em si mesmo constitui importante elemento para compreensão não só das diferenças regionais, como dos efeitos reais do multiplicador urbano (complementando a análise do fator crescimento). É bastante evidente, por exemplo, que sistemas urbanos como o de São Paulo e Rio de Janeiro com áreas médias inferiores a 3 mil km² (associado a desvio padrão inferior à média) são muito mais homogêneos que o de Belo Horizonte, por exemplo, com áreas média bem maior e forte desvio padrão. Por outro lado, a hierarquia dos três sistemas do Nordeste aparece bem nítida pela área média; até mesmo o fato de que a posição de comando regional de Recife faria numerosas cidades distantes estarem diretamente subordinadas, com isso aumentando um pouco a média, e muito mais o desvio padrão.

* Existem algumas dificuldades a superar no uso do multiplicador urbano em regiões tão diferentes, pois parece claro que só o espaçamento entre centros não será suficiente para ser comparado com o multiplicador urbano. Existem diferenças entre capacidade aquisitiva de uma e outra região, não estritamente mensuráveis ou refletidas por uma função distância entre um centro e outro.

TABELA 3

ÁREA MÉDIA, VARIÂNCIA E DESVIO PADRÃO DOS CENTROS
4b E MAIS(Em 1.000 km²)

Centros de Nível 1*	Média	Variância	Desvio Padrão
Fortaleza	9,577	110.481,942	10,511
Recife	4,361	71.159,821	8,435
Salvador	7,830	133.947,976	11,573
Belo Horizonte	6,438	56.715,764	7,530
Rio de Janeiro	2,854	6.294,544	2,508
São Paulo	2,487	5.639,889	2,374
Curitiba	3,459	9.118,118	3,019
Porto Alegre	3,911	22.041,874	4,694
Goiânia	23,355	665.298,775	25,793

* Belém não foi incluído por não fazer parte da área selecionada para este estudo.

Considerando apenas os centros 2a (ver tabela 2), observamos que apenas São Paulo tem seis centros (tem em verdade 7, mas excluindo-se Campo Grande ficam seis), o que adicionado ao centro São Paulo, daria uma relação $k = 7$, típica de um sistema urbano com trama espacial e hierarquia organizada de forma completa. Por outro lado, se analisarmos a distribuição espacial dos centros 2a em São Paulo (Mapa 5) (incluindo-se Sul de Minas e Norte do Paraná — parte da área de influência de São Paulo — mas excluindo-se Mato Grosso), veremos que Bauru, Campinas, Londrina, Ribeirão Preto, Uberlândia e São José do Rio Preto, indicam uma distribuição nitidamente direcionada dos centros 2a ao longo do eixo principal São Paulo—Nordeste, com quatro dos seis centros ao longo deste eixo, e apenas um para oeste (Bauru) e o outro para Sudoeste (Londrina); enquanto isso os centros 2b seguem uma distribuição mais dispersa, indicando uma ação menos direcionada dos que os de nível 2a.

Depois de São Paulo, a metrópole que mais possui centros 2a é Recife, com cinco centros e com uma distribuição muito mais direcionada; na realidade três destes centros são capitais (Natal, Maceió e João Pessoa). Rio de Janeiro tem uma situação semelhante, pois dos três centros 2a, apenas Juiz de Fora não é capital de Estado (os outros dois são Niterói e Vitória).

Os centros 2b são muito poucos para apresentar qualquer padrão discernível.

Ainda a nível nacional, algumas comparações são interessantes, quando se analisa a estrutura da rede de São Paulo e de Recife que, embora muito diferentes, têm posições regionais que podem ser comparadas.

São Paulo tem, como vimos, seis centros de nível 2a, (excluindo-se Campo Grande) enquanto Recife tem cinco (incluindo-se Caruaru que, embora seja um núcleo urbano de tamanho bem menor que os outros, está próximo a Recife), o que mostra bem, em primeiro lugar, a não comparabilidade inter-regional dos centros e, em segundo lugar, o desnível enorme entre os centros metropolitanos e os de nível imediatamente abaixo. Por outro lado, a simples análise da posição geográfica e posição político-administrativa de tais centros indica que os centros

2a da área de Recife são capitais de Estado, litorâneas, sendo Campina Grande de um lado e Caruaru de outro as únicas que são interiores (assim mesmo Campina Grande muito próxima e com características próprias); enquanto isso, a distribuição espacial dos centros 2a da área de São Paulo dá bem uma idéia de sua projeção interior, conforme se vê no Mapa 5. É claro que quando mencionamos a não comparabilidade inter-regional é porque dificilmente se poderia conceber similaridade de funções de centros 2a do tipo João Pessoa, Natal etc., com outros do tipo Campinas, Ribeirão Preto etc., dadas as funções administrativas das primeiras; mas também não se poderia comparar Campinas ou Ribeirão Preto com Caruaru, Campina Grande, Feira de Santana, Ilhéus-Itabuna, os únicos centros 2a em todo o Nordeste que não são capitais de Estado. Mas esta não comparabilidade inter-regional já constitui, em si mesma, uma forma de medir os desequilíbrios existentes nos dois sistemas urbanos (Nordeste e Centro-Sul).

Todos os outros núcleos metropolitanos têm 1 ou 2 centros 2a, não capitais de Estado, com exceção de Fortaleza e Belém, que só têm capitais de Estado como centros 2a. Entretanto, Porto Alegre possui três centros 2a, (Caxias do Sul, Passo Fundo e Santa Maria) além do Rio Grande—Pelotas (que é 2b, mas praticamente poderia ser 2a se fluxos externos fossem considerados), formando um verdadeiro semicírculo em torno de Porto Alegre. Na verdade, depois de São Paulo, é a metrópole que tem maior número de centros 2a sem que sejam capitais de Estado.

Passando aos centros de nível 3, de forma geral, observa-se que 60 em 171 encontram-se em São Paulo; se considerarmos a relação centros de nível 3/2, para o Brasil, teremos um $k = 3,4$ (enquanto que para 10 centros metropolitanos e 71 centros 2, ou seja, a relação 2/1, teríamos $k = 8$ a nível nacional e $k = 18$, considerando somente a área de São Paulo). É claro que isto coloca o problema de saber-se se devemos usar os centros 2a e 2b somados, ou somente os centros 2a, em cujo caso teríamos $k = 7$ para São Paulo, como indicamos anteriormente (se excluirmos Campo Grande).

Ainda analisando a tabela 2 podemos observar que Fortaleza tem 4 centros de nível 2 e 10 centros de nível 3, o que indica $k = 5$ para o nível 1 e $k = 3,5$ para o nível abaixo, $k = 5$ na relação centros de nível 4 como 3 e, finalmente, $k = 8,6$ para a relação cidades/centros de nível 4. Estas comparações mostram bem o desequilíbrio ao nível de centros intermediários, pois se considerarmos que dois centros 2 são capitais de Estado, então teríamos uma sucessão 3, 3,5, 5 e 8,6 nas quatro hierarquias.

Recife, embora tendo oito centros de nível 2, destes apenas 5 não são capitais de Estado, e um deles é Campina Grande que, como dissemos, tem uma situação especial e, portanto, k seria 5 ou 6 conforme o critério que usássemos. Descendo a hierarquia, na área de Recife, vemos 18 centros 3 para 8 centros 2, o que indica k aproximadamente 3,3, quase como Fortaleza; a mesma relação k aproximadamente = 5, também se verifica na relação níveis 3 e 4 em Recife, mas já na relação cidades/centros de nível 4, o valor k para Recife passa a ser mais de dez, o que, de alguma forma, indica um maior desequilíbrio na hierarquia na rede de Recife que na de Fortaleza.

Analisando-se dois outros núcleos metropolitanos, São Paulo e Rio de Janeiro, observam-se diferenças marcantes entre um e outro. São Paulo tem 17 centros 2 para a metrópole, o que pode ser uma indicação de que o verdadeiro nível 2 é somente o 2a, pois aí existem seis

centros (eliminando-se Campo Grande), enquanto Rio de Janeiro tem 3 núcleos 2a e dois 2b, o que daria (se adotássemos a hipótese 2a somente para São Paulo) um valor de $k = 6$ ($5 + 1$) para o Rio de Janeiro. Nesse caso teríamos uma relação 13 centros 3 no Rio, para cinco de nível 2 (k sendo portanto pouco acima de 3), enquanto que para São Paulo teríamos uma relação 60 centros (mais os 10 2b, portanto 70), para 7 centros 2a, o que daria um valor de $k = 11$, enormemente superior ao do Rio de Janeiro.

A hipótese subjacente é a de que a rede urbana de São Paulo apresenta uma tendência para “*steady state*” muito maior, com a ação do principal centro metropolitano penetrando muito mais hierarquia abaixo, a ponto de manter e mesmo superar a relação nível 2 com nível 1. Já ao nível de centros 4 sobre os de nível 3, temos um valor k pouco inferior a 4 no Rio (36 para 13) e um valor oscilando entre 3 e 3,5 para São Paulo (se somarmos os 10 2b), portanto não muito diferente entre as duas, como a indicar o limite de penetração intensa tanto de uma como de outra, na geração de centros de nível 4. É claro que poderia estar havendo uma supervalorização dos centros 5 em São Paulo, o que transformaria numerosas cidades em centros 4b e não núcleos subordinados como atualmente acontece; tanto que a relação cidade-centro 4 em São Paulo dá um k quase igual ao do Rio de Janeiro, como se pode ver na mencionada tabela 2.

Um dos subsistemas mais peculiares é o de Porto Alegre que, tendo 12 centros de nível 2, tem o valor mais alto de todo o Brasil depois de São Paulo, (que contados os de nível 2 como um todo, teria 17). Mas a relação centros de nível 3 com os de nível 2 é praticamente 1, pois para os 12 de nível 2, existem 11 de nível 3. A relação centros de nível 4 para os de nível 3 já passa a ser quase 5 (11 para 53), o mesmo ocorrendo cidades/centros de nível 4 (53 para 225). Se considerássemos apenas as cidades de nível 2a, teríamos uma relação 3 na cabeça do sistema; passando-se as nove de nível 2b para a categoria 3, teríamos aí então 20 centros de nível 3 para de nível 2, o que daria uma relação aproximada de 7. Mas, nesta hipótese, a relação 4 com 3 passaria a ser 2,5 e não quase 5, como indicamos acima. Em qualquer caso, o problema crítico seria identificar o “*breaking-point*” na rede urbana comandada por Porto Alegre, se ao nível dos centros 2 para o 3, ou se ao nível dos centros de nível 3 para os de nível 4. Convém assinalar, a esta altura, que o problema surgiu na própria definição da rede urbana de Porto Alegre, no estudo referente às regiões funcionais urbanas. Elza Keller chama a atenção para o fato de que “a principal característica da rede de Porto Alegre é o grande número de pequenos centros regionais com áreas de influência reduzida. Dividindo-se os relacionamentos entre esses numerosos pequenos centros, não há possibilidade de se estruturarem centros de atuação regional poderosa”.¹ Elza Keller atribuiu tal fato à economia colonial de grande parte de sua área, o que evidentemente reforça a idéia de que o “*breaking point*” está ao nível dos centros 2a/metrópole e não no nível de centros 3/2, ocorrendo também ao nível de centros 4/3, que passaria a ser pouco mais de 2,5 e não cinco, como a caracterização de regiões funcionais urbanas sugere.

De um lado, a posição excêntrica da metrópole gaúcha explicaria a baixa relação 2/1, e de outro lado, a economia colonial explicaria a baixa relação 4/3. A nova relação 3/2, de quase 7, se explicaria pelo forte influxo que Porto Alegre já vem exercendo no sistema, logo abaixo dos núcleos regionais 2a do tipo Caxias, Santa Maria e Passo Fundo (e neles também, evidentemente).

3.3 — O multiplicador urbano e o fator crescimento para as regiões metropolitanas (cabeças de subsistemas)

Como assinalamos antes, pode-se analisar os efeitos dos centros no processo de difusão — e talvez mais adequadamente — por via da análise das relações entre população dos centros de hierarquia diferentes, a partir das idéias de um multiplicador urbano e o correspondente fator crescimento. A tabela 4 indica apenas o que se poderia chamar de grau de urbanização, e que é definido como a relação entre a população que serve e a população que é servida. É claro que existem problemas operacionais (daí termos apresentado duas hipóteses) de um lado; de outro lado especulações sobre uma perfeita adaptabilidade dos níveis hierárquicos produzidos em regiões funcionais urbanas podem ser igualmente feitas, como já o fizemos na análise do número de centros, no início deste capítulo. A análise dos valores obtidos com a primeira e segunda hipóteses pode alimentar numerosas conjecturas a respeito, o que não é o propósito da presente análise, pelo que faremos somente poucas indicações a título de sugestão para estudos posteriores.

Individualizamos para as metrópoles, não só porque obviamente cada uma é a cabeça de um subsistema, como porque pareceria lógico supor que aí obteríamos o multiplicador mais elevado em cada subsistema, podendo assim identificar bem as diferenciações regionais existentes.

O cálculo do valor de u foi feito segundo duas hipóteses (o que nos pareceu razoável, dado o caráter preliminar e metodológico do trabalho) que passaremos a discutir e cuja fórmula pode ser vista na própria tabela 4.

TABELA 4
GRAU DE URBANIZAÇÃO DAS ÁREAS METROPOLITANAS

Centros de Nível 1*	População Área Metropolitana	População da Área Servida		Grau de Urbanização	
		1.ª Hipótese	2.ª Hipótese	1.ª Hipótese	2.ª Hipótese
Fortaleza	876.642	1.760.953	1.262.961	0,49	0,69
Recife	1.651.260	3.437.454	2.743.957	0,48	0,60
Salvador	1.078.033	2.101.244	1.550.867	0,51	0,69
Belo Horizonte	1.501.828	2.015.834	1.631.475	0,74	0,92
Rio de Janeiro	6.674.410	8.340.990	7.319.083	0,80	0,91
São Paulo	7.863.528	10.239.024	8.901.571	0,76	0,88
Curitiba	655.855	1.334.831	976.691	0,49	0,67
Porto Alegre	1.408.402	2.499.390	1.717.030	0,56	0,82

1.ª Hipótese: $\frac{\text{População Área Metropolitana}}{\text{Pop. Área Met.} + \text{Pop. Urb. Centros 2} + \text{Pop. Urb. Cidades Diretamente ligadas}}$

2.ª Hipótese: $\frac{\text{População Área Metropolitana}}{\text{População Área Metropolitana} + \text{População Urbana Centros 2a}}$

* Belém não foi incluído por não fazer parte da área selecionada para este estudo.

A primeira é a que contém o denominador mais alto, constituído pela população da área metropolitana, somada à da população de todos os centros de nível imediatamente abaixo (portanto os de nível 2 e as cidades diretamente ligadas à metrópole); a segunda contém, no denominador além da população das áreas metropolitanas, as cidades de nível 2a, consideradas assim as que seriam servidas mais diretamente pela área metropolitana.

Algumas considerações preliminares nos parecem importantes, à luz de alguns exemplos particulares, que poderão nos levar à conclusão da necessidade de calcular uma população servida de modo diferente. Consideremos um caso específico, particularmente significativo sob este aspecto, o de Recife. A tabela 4 indica uma população da área metropolitana de 1,6 milhão de habitantes e 3,4 e 2,7 para área servida, segundo as duas hipóteses. Na primeira hipótese estão incluídos entre os centros 2, além das cidades de Natal João Pessoa, Maceió e Campina Grande, (que são ou capitais ou uma cidade do tipo Campina Grande), apenas cidades como Caruaru, Moçoró, Garanhuns e Florianópolis; além disso compõem aquele total de cidades diretamente subordinadas, localidades bastante distantes de Recife, sobre as quais o efeito multiplicador de Recife deve ser virtualmente nulo; por outro lado, na segunda hipótese, estão apenas as cidades de nível 2a, acrescidas à população da área metropolitana de Recife e poder-se-ia indagar, em que medida as principais capitais do Nordeste, do tipo Natal, João Pessoa e Maceió poderiam ser consideradas população servida, para efeitos do cálculo do valor de u . Feitas estas considerações, os resultados 0,48 e 0,60 como os dois valores obtidos para Recife parecem não refletir uma realidade objetiva.

Embora a análise do padrão de distribuição de pontos e de seus parâmetros conseqüentes não esteja sendo feita a nível de cada centro de nível 1, e portanto a relação entre os valores de u (grau de urbanização) e o fator crescimento de cada centro não possa ser estritamente comparável com o corpo da análise, parece-nos que as variações intra-regionais, embora sejam menores, de qualquer maneira são importantes de serem detectadas para comparação.

Vemos o próprio exemplo de Recife, já mencionado antes, que tem 5 centros de nível 2a (mesmo considerando que as três capitais sejam por igual dependentes), o que daria um fator crescimento 13,5, ou seja para 1 + 5 centros, menos o valor de u que é de 0,60 na segunda hipótese; aplicando-se a fórmula

Fat. Cresc. = $\frac{1 + k - u}{1 - u}$ teríamos o acima mencionado valor de 13,5. *

Fazendo-se idêntico cálculo para Fortaleza, ainda seguindo a mesma segunda hipótese, teríamos um valor igual a 7,3, bem menor que Recife, o que parece consistente com o conhecimento empírico que se tem e alguns resultados de estudos anteriores, dando a Recife uma posição proeminente no sistema urbano do Nordeste, embora não diferente de Salvador, que tem valor ligeiramente superior a 13,5.

Todas as outras metrópoles regionais têm valores pouco superiores a 10, pois Curitiba tem valor 10, Porto Alegre tem valor 12,2 e São Paulo e Rio têm valores elevados, (51 e 34,3), resultados que são

* A diferença entre o valor do Fator Crescimento mencionado e o que aparece na tabela 5 é causada pelas diversas maneiras experimentais utilizadas para calcular este índice, como foi observado anteriormente.

consistentes com as expectativas. A única metrópole regional que parece em desacordo é Porto Alegre (Belo Horizonte também em certa medida), sendo que no caso de Porto Alegre isto se relaciona à sua posição excêntrica de um lado, e de outro à economia colonial de grande parte da área. Tanto que se usarmos para Porto Alegre e Belo Horizonte a primeira hipótese, isto é, todos os centros 2 e população de centros diretamente subordinados, o fator crescimento passa a 28 no caso de Porto Alegre e 16 no caso de Belo Horizonte. Mas, adotada esta hipótese, Recife passaria a um valor 16 também, Salvador passaria a um valor pouco acima de 23 e Fortaleza para quase 11.

De qualquer maneira, o que ressalta da análise destes valores é que, de um lado, se destaca o sistema São Paulo-Rio, com grande diferença entre São Paulo e Rio de Janeiro, e Porto Alegre assumindo uma posição singular dada a sua excentricidade, embora menos distanciada do Rio do que este de São Paulo. Belo Horizonte, Recife, Salvador e Curitiba não se diferenciam demais ao passo que Fortaleza apresenta valor bem inferior.

Evidentemente não é o propósito deste trabalho fazer uma análise mais aprofundada da hierarquia urbana, nem do fator crescimento dos diferentes níveis hierárquicos, mas apenas procurar associar os efeitos dos diferentes graus de urbanização e de centralização, em cada subsistema, ao processo de difusão, examinando a compatibilização entre ambos.

A tabela 5 mostra os valores de u e do fator crescimento para os centros de diferentes níveis, subordinados às várias metrópoles, com o mesmo propósito de identificar ruturas (*breaking points*) na hierarquia, e verificar até onde elas correspondem às medidas que procuraremos obter na aplicação dos modelos de probabilidades, por via de seus principais parâmetros.

Convém assinalar que a tabela 5 foi elaborada de forma talvez mais ortodoxa (desde que assumiu que todos os centros 2 são servidos pelo centro 1 correspondente, os 3 pelos de nível 2 etc.), com apenas uma hipótese diferente: usando-se os centros diretamente subordinados ou desprezando tais centros, no cálculo de u e dos outros valores derivados.

O fator crescimento e mesmo o grau de urbanização apresentam valores bem diferentes, o que mostra bem a importância de se considerar a hipótese ou não, para melhor compreensão da força propulsora de cada núcleo metropolitano. Compare-se apenas o valor do multiplicador urbano das três metrópoles do Nordeste com São Paulo, para verificar que Fortaleza, embora tenha um multiplicador superior ao de São Paulo, tem um fator crescimento quase cinco vezes menor; mas o multiplicador cai bruscamente ao se considerar os municípios diretamente subordinados.

Um aspecto muito importante (e para o qual já tínhamos chamado a atenção antes) é o relativo à medida de desequilíbrio no sistema por via de fatores crescimento diferentes nos vários níveis da hierarquia. Observe-se que, em todos os casos, o maior fator crescimento é verificado entre os níveis 3/4, indicando uma elevada expansão do sistema ao nível de centros 3/4, muito maior que o dos níveis hierárquicos superiores. Mas mesmo esta defasagem é diferenciada; compare-se, por exemplo, o fator crescimento de Goiânia no nível 1/2 (e que é de 4,17) com o nível 3/4 (e que é de 42,89, portanto pouco mais de dez vezes maior), com o fator crescimento em São Paulo e que é de 58,69 e 277,36, respectivamente. Em primeiro lugar ele revela um poder de penetração e (conseqüentemente de difusão) da metrópole nível

TABELA 5
ÍNDICES DE URBANIZAÇÃO

Metrópole	Nível dos Centros	U	Multiplicador Urbano	Fator Crescimento
Recife	1/2	0,45	0,82	15,55
	1/2+m.d.s	0,25	0,33	11,67
	2/3	0,64	1,78	51,00
	3/4	0,34	0,52	99,48
Fortaleza	1/2	0,76	3,17	13,50
	1/2+m.d.s	0,53	1,13	7,38
	2/3	0,31	0,45	16,94
	3/4	0,29	0,41	55,93
Salvador	1/2	0,73	2,70	30,63
	1/2+m.d.s	0,52	1,08	17,67
	2/3	0,66	1,94	33,35
	3/4	0,26	0,35	75,32
Belo Horizonte	1/2	0,72	2,57	15,29
	1/2+m.d.s	0,51	1,04	9,16
	2/3	0,35	0,54	28,69
	3/4	0,38	0,61	92,94
Rio de Janeiro	1/2	0,82	4,56	28,78
	1/2+m.d.s	0,67	2,03	16,15
	2/3	0,55	1,22	34,33
	3/4	0,32	0,47	53,94
São Paulo	1/2	0,74	2,85	58,69
	1/2+m.d.s	0,63	1,70	41,54
	2/3	0,37	0,55	96,24
	3/4	0,45	0,82	277,36
Curitiba	1/2	0,47	0,89	14,21
	1/2+m.d.s	0,34	0,52	11,61
	2/3	0,50	1,00	27,00
	3/4	0,35	0,54	65,62
Porto Alegre	1/2	0,42	0,72	23,41
	1/2+m.d.s	0,30	0,43	19,57
	2/3	0,70	2,33	37,67
	3/4	0,25	0,33	71,67
Goiânia	1/2	0,37	0,59	4,17
	1/2+m.d.s	0,29	0,41	3,82
	2/3	0,73	2,70	30,63
	3/4	0,26	0,35	42,89

NOTA: u = Grau de Urbanização

$$\text{Multiplicador Urbano} = \frac{u}{1-u}$$

$$\text{Fator de Crescimento} = \frac{1 + k - u}{1 - u}$$

1 em relação aos centros de nível 2, logo abaixo, muito maior do que o de Goiânia; por outro lado ao indicar que o fator crescimento é muito maior em todos os outros níveis, em São Paulo está reiterando a mesma idéia, com a diferença que o hiato entre 1/2 e o 2/3 é muito grande em Goiânia (embora seja pequeno entre o 2/3 e o 3/4); enquanto isso a diferença em São Paulo é maior entre 2/3 e o 3/4; no primeiro caso encontraríamos uma quebra no sistema, logo entre o nível 1 e 2, e mais articulação no sistema aos níveis abaixo, como se ele fosse se constituindo de baixo para cima; no caso de São Paulo (embora isso possa ter ocorrido inicialmente) processo já se encontra em fase posterior e ainda que haja maior articulação entre o nível 3/4, o desnível é muitíssimo menor que no caso de Goiânia.

Muitas outras comparações podem ser feitas, inclusive sobre o efeito da hipótese de inclusão dos municípios diretamente subordinados, que destaca as metrópoles sem rede urbana melhor hierarquizadas, até mesmo no Nordeste; veja-se por exemplo o que ocorre com o fator crescimento de Recife, Fortaleza e Salvador, que diminui muito mais nas duas últimas que em Recife, o que evidentemente estaria ligado a esta menor hierarquização. Outras considerações podem ser ainda feitas, procurando avaliar a propriedade da classificação de determinados centros como 2 ou mesmo 3 e 4, do tipo das que fizemos pela simples análise do número de centros. Considerações deste tipo parecem indicar o caminho para uma revisão, em profundidade, da classificação dos centros à luz dos conceitos de localidade central, e tendo em vista seu poder como centros de difusão de inovações.

O que pretendemos fazer com estas indicações mais ou menos sumárias é mostrar suas relações com o processo de difusão, usando-as como elementos de comparação com os modelos de distribuição de pontos que estamos testando, reservando a análise de mais detalhe para a oportunidade de uma eventual revisão da Divisão do Brasil em Regiões Funcionais Urbanas.

Os dados que apresentamos, tanto os que relacionam os tamanhos funcionais das cidades com seu grau de urbanização, como os que mostram o poder multiplicador de cada centro metropolitano, visam oferecer algumas indicações de extensão do processo de difusão no sistema e em suas partes componentes. Eles são apresentados com o propósito de permitir uma comparação com resultados da aplicação de modelos de distribuição de pontos, que descrevem a função geradora da distribuição, por sua vez associada a diferentes estágios do processo de desenvolvimento — vale dizer da difusão de inovações no sistema.

Por isso mesmo tais dados são apresentados sem análise em profundidade, feita em trabalhos anteriores ou paralelos (como é o caso do multiplicador urbano e fator crescimento).

Parece importante salientar, antes de entrarmos na análise dos resultados obtidos na aplicação do modelo, que a premissa básica adotada é a de que o centro 4b é um núcleo de difusão de inovações e que seu campo de ação é o definido por sua área de influência. Mesmo levando em conta que esta área foi tomada como sendo diferente (duas vezes maior no Centro Sul que no Nordeste), é óbvio que ainda assim o poder de penetração de tais centros em São Paulo (e no Centro Sul em geral) é maior que no Nordeste, tanto que o fator crescimento dos centros 3/4 em São Paulo é quase três vezes superior ao de Recife. Daí ser necessário interpretar com cuidado os resultados obtidos, que apresentaremos a seguir, por isso mesmo, a diferentes níveis de agregação.

4 — OS RESULTADOS ANALÍTICOS OBTIDOS: COMPARAÇÃO DAS DUAS DISTRIBUIÇÕES COM OS MODELOS TEÓRICOS ADOPTADOS, NAS DUAS REGIÕES, EM DIFERENTES NÍVEIS DE AGREGAÇÃO, PARA CENTROS 4b, CIDADES E CENTROS 3b

Salientamos que o objetivo do presente estudo era de natureza exploratória, em busca de ajustamentos entre uma distribuição observada e modelos probabilísticos de distribuições teóricas, que tenham expressão genética. Com isso se procuraria explorar a gênese do processo de distribuição de centros de vários níveis, portanto o seu processo de difusão, para poder compará-la com outros indicadores do mesmo obtidos em outros estudos. No caso, as duas distribuições teóricas utilizadas foram a Poisson e a Binomial Negativa, dadas as suas conotações com construções teóricas relativas a localidades centrais.

Veremos assim a região Nordeste e depois a Zona da Mata, como uma subárea de natureza diferente; a seguir veremos o Centro-Sul, e áreas cada vez mais reduzidas da mesma, até a área de influência de São Paulo.

4.1 — A região Nordeste e a distribuição de centros 4b

O primeiro exemplo utilizado foi o Nordeste, e foram testadas ambas as distribuições, a diferentes níveis de agregação.

A linha metodológica seguida foi a de primeiro fazer uma análise praticamente do Universo (Nordeste), no qual foram contados 215 centros 4b, em nove folhas da carta 1:1.000.000, conforme se pode observar na tabela 6. A seguir foi feita uma amostra, tomada de forma subjetiva, embora levando em conta a idéia de diminuir a heterogeneidade da área, por via de escolha de uma amostra que represente a área mais desenvolvida do Nordeste, como indicaremos na análise de tal amostra.

Uma simples inspeção dos valores de centros, municípios e quadrículas, em cada uma das folhas, na tabela 6, mostra bem esta heterogeneidade. A folha Teresina tem, praticamente, o mesmo número de centros que a folha Natal, menos municípios, mas por outro lado contém 384 quadrículas comparado com 53 na folha Natal. Mesmo uma folha como a de Jaguaribe, que tem número de quadrículas igual, tem 58 centros e número de municípios mais ou menos proporcional, portanto 3 vezes mais que a folha Teresina.

Esta distribuição espacial foi, inicialmente, submetida à análise a nível de centros 4b, tanto pela distribuição Binomial Negativa como pela Poisson (Tabela 7). O teste qui-quadrado aceitou ambos os ajustamentos e algumas observações sobre os parâmetros obtidos podem trazer alguma luz ao problema. Observando-se a tabela relativa à distribuição Poisson, verifica-se que embora se aceite o ajuste, a média e variância não são rigorosamente iguais, embora a diferença seja mínima. Por isso mesmo, o parâmetro P na Binomial Negativa tem um valor 0,98, e K é igual a 5,21 indicando uma elevada taxa de propagação. Considerando o tamanho da quadrícula de $15' \times 15'$, portanto, de 800 km² (muito pequena para ser interpretada como o campo médio de informação do centro 4b), o ajustamento à Binomial Negativa e a Poisson podem estar indicando uma distribuição composta; Harvey¹⁸ acentua em seu trabalho que “mesmo que o processo operante seja genuinamente contágio, a versão composta da Binomial

TABELA 6
NORDESTE
215 Centros 4b
Quadrícula Original

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A	C/B
São Luís	240	13	94	0,05	7,23
Fortaleza	66	7	47	0,10	6,71
Teresina	384	21	109	0,05	5,19
Jaguaribe	384	58	354	0,15	6,10
Natal	53	22	162	0,41	7,36
Rio São Francisco	384	8	47	0,02	5,87
Aracaju	356	45	285	0,12	6,33
Recife	32	11	68	0,34	6,18
Salvador	220	30	185	0,13	6,16
	2.119	215	1.351		

TABELA 7
AJUSTAMENTOS — 2119 QUADRÍCULAS

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	1.916	1.914,54	1.916	1.916,42
1	192	194,25	192	190,73
2	10	9,85	10	11,31
3	1	0,33	1	0,51
	2.119		2.119	

Poisson

M = 0,101
X² obs. = 0,125
Aceita-se o ajuste

$\alpha = 0,05$
g.l. = 2
X² tab. = 5,991
Var. = 0,103
D = 1,02

Binomial Negativa

p = 0,98
k = 5,21
X² obs. = 0,066
Aceita-se o ajuste

Negativa pode dar um bom ajustamento, simplesmente reduzindo o tamanho do quadrado. É geralmente verdade que se Q (tamanho da quadrícula) for tornado suficientemente pequeno, a distribuição de frequências observadas corresponderá a um modelo Poisson simples. Aumentando o tamanho da quadrícula, pode gerar uma distribuição Binomial Negativa, porque a amostra é heterogênea". pp. 90/91.

Na primeira agregação (Tabelas 8 e 9), resultante de quatro quadrículas de $15' \times 15'$, o ajustamento a Poisson, evidentemente, é rejeitado, porque a média de centros por quadrícula passou a 0,39, deixando de caracterizar-se como um evento raro. A variância passou a 0,50, deixando de ser uma distribuição aleatória irrestrita.

Quase que pela mesma razão, a distribuição Binomial Negativa é aceita. Já neste caso o parâmetro K assume um valor igual a 1,47.

TABELA 8
NORDESTE
215 Centros 4b
Primeira Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
São Luís	64	13	94	0,20
Fortaleza	20	7	47	0,35
Teresina	96	21	109	0,21
Jaguaribe	96	58	354	0,60
Natal	16	22	162	1,37
Rio São Francisco	96	8	47	0,08
Aracaju	90	45	285	0,50
Recife	10	11	68	1,10
Salvador	58	30	185	0,51
	546	215	1.351	

TABELA 9
AJUSTAMENTOS — 546 QUADRÍCULAS

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Freqüências Observadas	Freqüências Esperadas	Freqüências Observadas	Freqüências Esperadas
0	387	368,30	387	385,67
1	115	145,00	115	119,39
2	35	28,54	35	31,05
3	6	3,74	6	7,56
4	3	0,36	3	1,78
	546		546	

Poisson
M = 0,394
g. l. = 2
X²obs. = 11,07
X²tab. = 5,991
Rejeita-se o ajuste

$\alpha = 0,05$
Var. = 0,499
D = 1,27

Binomial Negativa
p = 0,79
k = 1,47
g. l. = 3
X²obs. = 0,680
X²tab. = 7,815
Aceita-se o ajuste

A segunda agregação (Tabelas 10 e 11), passou a conter dois quadrados anteriores agrupados, observando-se um direcionamento Norte-Sul, a fim de testar a hipótese de maior efeito de difusão nesta direção, isto é, oito quadrículas originais e uma área da ordem de 6 mil

TABELA 10
NORDESTE
215 Centros 4b
Segunda Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
São Luís	42	13	94	0,30
Fortaleza	19	7	47	0,36
Teresina	48	21	109	0,43
Jaguaribe	48	58	354	1,20
Natal	11	22	162	2,00
Rio São Francisco	48	8	47	0,16
Aracaju	45	45	285	1,00
Recife	5	11	68	2,20
Salvador	29	30	185	1,03
	295	215	1.351	

TABELA 11
AJUSTAMENTOS — 295 QUADRÍCULAS

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	169	142,33	169	172,07
1	78	103,73	78	70,51
2	28	37,80	28	29,88
3	7	9,18	7	12,80
4	7	1,67	7	5,51
5	5	0,24	5	2,38
6	0	0,02	0	1,03
7	1	0,00	1	0,44
	295		295	

Poisson

M = 0,729
g. l. = 3
X²obs. = 20,97
X²tab. = 7,815
Rejeita-se o ajuste

α = 0,05
Var. = 1,296
D = 1,78

Binomial Negativa

p = 0,56
k = 0,94
g. l. = 4
X²obs. = 5,012
X²tab. = 9,488
Aceita-se o ajuste

km², aproximadamente igual a área média de influência de centros 4b no Nordeste. Isto satisfaria a condição de tamanho da quadrícula correspondente ao tamanho do Núcleo, aí entendido como o centro 4b e sua área de influência, com seus centros urbanos menores dependentes. Uma média de 0,73 centros por quadrícula, obviamente eliminaria a hipótese Poisson. Ao mesmo tempo, se já na primeira agregação, a variância passava a ser um pouco maior que no caso das quadriculas (na sua relação com a média), nesta segunda, a relação V/M passa a ser bem superior a 1, caracterizando uma distribuição já do tipo contágio. O teste qui-quadrado aceita a distribuição Binomial Negativa que tem, entretanto, anomalias no ajustamento em algumas frequências, que são bem ilustrativas do sistema urbano do Nordeste. As frequências zero, um e dois são bastante aproximadas, mas já na frequência 3 centros por quadriculas, a diferença é bem grande, diminuindo na frequência 4, tornando a aumentar na frequência 5. Estas discrepâncias maiores parecem indicar ainda um certo grau de heterogeneidade, pois enquanto se esperaria ter 13 quadriculas com 3 centros, obtém-se 7; enquanto se espera 2,3 quadriculas com 5 centros, obtém-se 5, o que estaria indicando áreas de concentração menor que a esperada, e outras com concentrações maiores, o que parece representativo do fenômeno urbano do Nordeste, ajustado ao modelo Binomial Negativo, mas ainda com alguma inadaptação frequência por frequência, que refletem uma rede urbana desconectada, fenômeno conhecido e também observável pela relação centros de hierarquia $r'/r + 1$ já mencionada.

A terceira agregação, (tabela 12 e 13) de novo rejeitada a distribuição Poisson, pelo simples fato de ter média e variância 1,4 e 3,8 aproximadamente. Esta terceira agregação contém de novo duas quadriculas anteriores, isto é, 16 quadriculas originais, cerca de 12 mil km², constituindo assim uma quadrícula de 1 grau quadrado. O teste do qui-quadrado aceita o ajuste à Binomial Negativa, mas a análise das distribuições para cada frequência indica um ajuste pior que o anterior, pois, desde a frequência zero, as diferenças entre valores esperados e observados é proporcionalmente bem maior.

O ajustamento aí foi processado por agrupamento de quadriculas no sentido Norte/Sul, com o objetivo de testar a hipótese de que o processo de expansão destes centros 4b obedecesse a um efeito de proximidade nesta direção, o que foi apenas parcialmente indicado por um ajuste mais ou menos precário. As tabelas 12 e 13 mostram os valores correspondentes.

Uma análise do conjunto dos quatro ajustamentos oferece algumas sugestões interessantes e permite algumas comparações tanto no que diz respeito ao problema tamanho das quadriculas como no que diz respeito ao comportamento dos diferentes parâmetros.

É importante assinalar que, desde a primeira agregação, o ajuste na probabilidade Poisson passou a ser rejeitado e aceito — neste nível e nos níveis de agregação seguintes — o modelo da Binomial Negativa, no qual o parâmetro k foi diminuindo progressivamente. Na realidade, na primeira agregação, este valor de k é de 1,47, na segunda ela passa a 0,94, o que já é bem diferente e mostra o processo de nucleação bem típico do sistema de cidades no Nordeste; já na terceira agregação tal parâmetro assume um valor 0,80.

A tabela 78 mostra a variação dos diversos parâmetros com o aumento do tamanho das quadriculas.

TABELA 12
 NORDESTE
 215 Centros 4b
 Terceira Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
São Luís	22	13	94	0,59
Fortaleza	10	7	47	0,70
Teresina	24	21	109	0,87
Jaguaribe	24	58	354	2,41
Natal	7	22	162	3,14
Rio São Francisco	24	8	47	0,33
Aracaju	24	45	285	1,87
Recife	3	11	68	3,66
Salvador	16	30	185	1,87
	154	215	1.351	

TABELA 13
 AJUSTAMENTOS — 154 QUADRÍCULAS

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	64	38,12	64	68,58
1	42	53,22	42	34,95
2	23	37,15	23	20,00
3	8	17,29	8	11,86
4	4	6,03	4	7,16
5	4	1,68	4	4,36
6	3	0,39	3	2,68
7	3	0,07	3	1,65
8	1	0,01	1	1,02
9	1	0,00	0	0,63
10	2	0,00	2	0,39
	154		154	

Poisson

M = 1,396
 g. l. = 4
 $X^2_{obs.} = 37,24$
 $X^2_{tab.} = 9,488$

Rejeita-se o ajuste

$\alpha = 0,05$
 Var. = 3,824
 D = 2,74

Binomial Negativa

p = 0,37
 k = 0,80
 g. l. = 5
 $X^2_{obs.} = 5,309$
 $X^2_{tab.} = 11,070$
 Aceita-se o ajuste

Verifica-se que, enquanto o parâmetro p vai se reduzindo de forma mais ou menos contínua, sem grandes desníveis, e de certa forma seguindo a variância, o parâmetro k se comporta de maneira diferente: da quadrícula original para a primeira agregação a diferença é bastante grande (5,21 para 1,47), portanto mais de três vezes maior na quadrícula original que na primeira agregação. Já daí para as agregações seguintes e mais particularmente entre a segunda e a terceira a diferença passa a ser quase nenhuma, refletindo assim uma certa estabilização da taxa de difusão daí para cima.

A invariabilidade do parâmetro p sendo uma das premissas básicas da distribuição Binomial Negativa genuína, e considerando que p , para as quatro análises feitas, varia de forma mais ou menos regular, ficaria caracterizada a não independência do evento (centro 4b, no caso), ao mesmo tempo que a variação não linear de k em relação ao tamanho da quadrícula estaria sugerindo heterogeneidade na área estudada, ajustamento descritivo e não genético.

No primeiro caso — o da não independência do evento — procuramos um teste da mesma por via da análise de contigüidade, pois esta não independência poderia estar ligada ao próprio mecanismo do processo de difusão do tipo contágio. No segundo caso, a não homogeneidade era desde logo admitida, razão pela qual uma das etapas subseqüentes do trabalho foi a de dividir a região em duas partes e analisá-las separadamente.

A análise de contigüidade, descrita na parte metodológica do presente estudo, foi feita para o Nordeste como um todo, utilizando-se apenas a agregação de quatro quadrículas (primeira agregação) e consideradas contíguas apenas quadrículas colocadas no sentido Norte-Sul. Segundo este método, um efeito de proximidade seria detectado se o número de quadrículas que contém um centro 4b ao lado de outra igual (também com um centro 4b), portanto formando uma junta 4b (junta BB), fosse significativamente maior do que o que ocorreria em uma distribuição aleatória.

O modelo de contigüidade descrito por Cliff adota duas formas analíticas, uma com reposição e outra sem reposição, que aplicadas ao Nordeste não apresentaram grande diferença, o que seria de esperar quando N é grande; apenas a variância e o conseqüente Desvio Padrão são diferentes.

As tabelas 14 e 15 mostram os valores para os dois casos.

TABELA 14
COM REPOSIÇÃO

Juntas	N.º Calculado	N.º Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	213,52	172,0	159,55	12,63	-3,29
BB	44,3	69,0	57,92	7,61	3,24
WW	257,16	274,0	229,78	15,16	1,11

TABELA 15
SEM REPOSIÇÃO

Juntas	N.º Calculado	N.º Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	213,9	172,0	90,1	9,5	- 4,41
BB	44,1	69,0	22,9	4,8	5,18
WW	256,9	274,0	27,2	5,2	3,28

Analisando-se os dados verifica-se que o número esperado de juntas BB é significativamente menor e os de juntas BW é significativamente maior que o observado, portanto rejeitando a hipótese nula de aleatoriedade na distribuição das mesmas; conseqüentemente aceita-se a hipótese da existência de um efeito de proximidade, segundo o qual a probabilidade de existência de um centro 4b é aumentada pela existência de outro em quadrícula adjacente.

Mesmo considerando que no modelo com reposição a variância seria (como se pode observar) bem maior, a diferença entre o valor observado e calculado foi suficientemente grande para não tornar a taxa de contigüidade baixa, e assim indicar aleatoriedade na distribuição.

4.1.1 — A distribuição de centros 4b na Zona da Mata

A seguir foi feita uma amostra (na realidade foi escolhido um novo Universo que, pelas características gerais da região — a Zona da Mata —, apresentasse um grau de homogeneidade maior), com a qual se pudesse testar novamente os mesmos tipos de distribuição.

Esta área — que chamaremos daqui por diante de Nordeste reduzido ou Zona da Mata — foi definida a partir de cinco folhas da carta ao milionésimo (ver tabelas 16 e 17). Algumas comparações com a

TABELA 16
NORDESTE REDUZIDO
107 Centros 4b
Quadrícula Original

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A	C/B
Natal	64	22	159	0,34	7,22
Recife	32	11	65	0,34	5,90
Jaguaribe	112	27	165	0,24	6,11
Salvador	44	11	58	0,25	5,27
Aracaju	168	36	215	0,21	5,97
	420	107	662		

TABELA 17
AJUSTAMENTO
420 Quadrículas
DISTRIBUIÇÃO DE POISSON

Número de Centros	Frequência Observada	Frequência Esperada
0	321	325,54
1	92	82,93
2	6	10,56
3	1	0,89
420		

M = 0,255
 VAR = 0,232
 $X^2_{obs.} = 2,789$
 $X^2_{tab.} = 5,991$
 $\alpha = 0,05$
 g. l. = 2
 D = 0,91
 Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Distribuição Binomial Negativa.

tabela 6 indicam a redução feita (houve simples eliminação de folhas inteiras como as de São Luís, Fortaleza, Teresina e Rio São Francisco). Ao lado disso, outras folhas foram parcialmente reduzidas, de forma a reduzir a área aos limites aproximados da Zona da Mata, o que redundou em reduzir o número de quadrículas de 2.119 para 420 (portanto a um quinto do original); entretanto, o número de centros 4b e mais reduziu-se de 215 para 107, o que representa apenas 50% do original, fato que em si já é um indício da diferença entre uma parte e outra da região originalmente analisada.

Obviamente o primeiro efeito de tal redução foi aumentar a média e a variância que passaram de 0,101 e 0,103 na primeira, para 0,255 e 0,232 na segunda. A consequência de ter-se média inferior à variância na área reduzida foi a impossibilidade de calcular-se o parâmetro p na Binomial Negativa e, quase que por via de consequência, obter-se um ajuste à distribuição Poisson. Este resultado mostra ao mesmo tempo que a heterogeneidade existente no Nordeste como um todo (no primeiro resultado houve ajuste à Binomial Negativa na quadrícula), como também que o não ajustamento à Binomial Negativa na área reduzida indica um maior grau de dispersão de centros 4b na Zona da Mata, o que seria naturalmente lícito esperar. É interessante observar, a título de uma primeira comparação, que na análise do Centro Sul reduzido também não houve possibilidade de ajuste à BN, por impossibilidade de cálculo do parâmetro p , pois a média e variância são 0,261 e 0,248, conforme se pode ver na tabela 50, mais adiante, indicando um comportamento semelhante.

Quando passamos à primeira agregação, (ver tabelas 18 e 19) a média já passa a ser inferior à variância, embora esta diferença deva ser avaliada à luz do tamanho da quadrícula. É óbvio que, quando to-

amos a área reduzida, a área de influência dos centros 4b torna-se menor (420 quadrículas e 107 centros significam quatro quadrículas por centro, enquanto que na região toda tínhamos praticamente 10 quadrículas por centro).

TABELA 18
NORDESTE REDUZIDO
107 Centros 4b
Primeira Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
Natal	16	22	159	1,37
Recife	8	11	65	1,37
Jaguaribe	28	27	165	0,96
Salvador	11	11	58	1,00
Aracaju	42	36	215	0,85
	105	107	662	

TABELA 19
AJUSTAMENTOS
105 Quadrículas

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	45	37,89	45	40,77
1	26	38,62	26	35,86
2	24	19,67	24	18,22
3	7	6,68	7	7,00
4	3	1,70	3	2,25
	105		105	

Poisson

M = 1,019
X²obs. = 6,712
Aceita-se o ajuste

$\alpha = 0,05$
g. l. = 3
Var. = 1,181
X²tab. = 7,815
D = 1,15

Binomial Negativa

p = 0,86
k = 6,43
X²obs. = 5,036
Aceita-se o ajuste

Ao mesmo tempo, aceita-se tanto o ajuste à BN como à Poisson, embora já aí o ajuste à BN (avaliando-se apenas porque as aproximações frequência por frequência são maiores), seja melhor. Entretanto, comparando-se com a região como um todo, observa-se que o parâmetro k (taxa de difusão), que teve um valor 1,47, passa nesta área a ter um valor de 6,43, conforme as expectativas. Para o Centro-Sul reduzido não houve sequer possibilidade de ajuste à BN.

Em termos de comparações preliminares, vale observar que esta mesma taxa de difusão é encontrada para a região Centro-Sul como um todo, a nível de quadrícula original (tabela 44), embora obviamente a quadrícula seja muito pequena para ser tomada como “*mean information field*”, mesmo no Centro-Sul.

A segunda agregação (tabelas 20 e 21) foi feita apenas para se testar a continuidade da aceitação do ajuste à BN, pois era óbvio que não mais se aceitaria o ajuste à Poisson, desde que na segunda agregação a média e variância são, respectivamente, de 1,91 e 3,12 e o parâmetro k assume um valor 3,03, já bem inferior ao da primeira agregação. Também não é lícito supor que este tamanho de quadrícula tenha validade para a Zona da Mata, pois a sua área, sendo de cerca de 6 mil km², ultrapassa amplamente o tamanho da área de influência dos centros 4b. Mas é interessante comparar estes valores com sua contra-partida no Centro Sul reduzido, em que na segunda agregação (tabelas 48 e 49), a média e a variância são 1,007 e 1,176, com a consequência de que o parâmetro k toma um valor de 5,98, bem maior que o da Zona da Mata, o que já indica a primeira diferença importante. A única dúvida seria se esta mesma diferença é válida quando se trata da Zona da Mata, considerando-se, sistematicamente, que as agregações no sul se fazem a nível de quadrículas menores (a segunda agregação tem quatro quadrículas originais no Centro Sul e oito no Nordeste). Neste caso, a primeira agregação da Zona da Mata seria equivalente à segunda no Centro Sul, mas voltaremos a estes problemas mais adiante.

TABELA 20
NORDESTE REDUZIDO
107 Centros 4b
Segunda Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
Natal	8	22	159	2,75
Recife	5	11	65	2,20
Jaguaribe	15	27	165	1,80
Salvador	6	11	58	1,83
Aracaju	22	36	215	1,63
	56	107	662	

TABELA 21
AJUSTAMENTOS
56 Quadriculas

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Freqüências Observadas	Freqüências Esperadas	Freqüências Observadas	Freqüências Esperadas
0	15	8,28	15	12,73
1	13	15,83	13	14,91
2	8	15,12	8	11,61
3	11	9,63	11	7,53
4	4	4,60	4	4,39
5	2	1,75	2	2,38
6	2	0,56	2	1,23
7	1	0,15	1	0,61
	56		56	

Poisson

M = 1,911
X²obs. = 10,021
Rejeita-se o ajuste

$\alpha = 0,05$
g. l. = 4
Var. = 3,117
X²tab. = 9,488
D = 1,63

Binomial Negativa

p = 0,61
k = 3,03
X²obs. = 3,384
Aceita-se o ajuste

Considerando que na análise da distribuição de pontos para a Zona da Mata o mesmo tipo de ajustamento à distribuição BN se observou, naturalmente com parâmetros *K* com valores bem mais elevados, associados a maior intensidade do processo de difusão nesta área, repetimos a análise de contigüidade para esta área, bem como para o Sertão, com o mesmo propósito com que ela foi feita para a região como um todo.

As tabelas 22 e 23 mostram os valores para as duas áreas, inicialmente para o modelo sem reposição, notando-se que na Zona da Mata foram contadas 114 quadriculas, 63 das quais pretas e 51 brancas; já no Sertão para 428 quadriculas, 96 são pretas e 332 são brancas, valores estes que mostram bem a diferenciação entre estas duas áreas.

TABELA 22
ZONA DA MATA

Juntas	N.º Calculado	N.º Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	51,88	44,0	25,46	5,04	— 1,56
BB	31,53	39,0	7,61	2,76	2,71
WW	20,59	21,0	7,17	2,67	0,15

TABELA 23

SERTÃO

Juntas	N.º Calculado	N.º Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	142,31	125,0	51,14	7,15	-2,42
BB	20,36	30,0	12,61	3,55	2,71
WW	245,33	253,0	16,01	4,01	1,91

Uma simples comparação do número de juntas e quadrículas nas duas áreas já permite algumas especulações sobre os mecanismos do processo de difusão na Zona da Mata e no Sertão.

Enquanto na Zona da Mata se contém 25% das quadrículas (portanto 75% no Sertão) a quantidade de juntas brancas na Zona da Mata é 12 vezes menor que no Sertão (253 e 21), que resulta em uma distribuição aleatória nos dois casos, mas a diferentes níveis de intensidade. Por outro lado, das 69 juntas pretas, 39 estão na Zona da Mata e 30 no Sertão, em ambos os casos com elevada taxa de contigüidade, também em diferentes níveis de intensidade. O sertão contém 22% de quadrículas pretas (96 em 428), mas 44% das juntas pretas, (somadas as 2 regiões) o que mostra bem a concentração de juntas pretas, separadas por extensos espaços vazios, estes com uma distribuição aleatória de juntas brancas.

Já na Zona da Mata, 55% das quadrículas são pretas (63 em 114), e 56% das juntas pretas de toda a região estão nela localizadas (39 em 69), mas ainda aí esta distribuição não é aleatória, observando-se um efeito de proximidade, menos intenso que para o Nordeste como um todo, mas igual ao do Sertão. Esta igualdade é a indicação mais precisa da identidade do processo, mas com intensidade diferente. Esta intensidade diferente e maior na Zona da Mata corrobora a indicação anterior através de um parâmetro k da BN mais elevado na Zona da Mata que no Nordeste como um todo. Por outro lado o fato de que o padrão de distribuição de centros mais importantes no Sertão está muito associado à distribuição de Serras e vales úmidos, também é um indício de que a distribuição de juntas BB no Sertão formaria agrupamentos, separados por extensos espaços vazios, entre tais serras e vales.

Esta diferença entre Zona da Mata e Sertão aparece também, usando-se o modelo *com reposição*, no qual a probabilidade de encontrar-se um par preto é fixa (no caso do modelo sem reposição usa-se a distribuição hipergeométrica), conforme se pode ver nas tabelas 24 e 25.

Aí se observa que a distribuição de juntas BB no Sertão quase forma um padrão concentrado (observe-se que taxas de contigüidade oscilando entre 2 e -2 indicariam aleatoriedade), enquanto que na Zona da Mata a distribuição é nitidamente aleatória, com taxa de contigüidade bem baixa.

Este tipo de distribuição baseou-se, conforme foi observado inicialmente, em uma definição de contigüidade ao longo do sentido nortesul, seguindo a linha adotada na agregação de quadrículas para o ajust-

TABELA 24
ZONA DA MATA

Juntas	N.º Calculado	N.º Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	51,42	44,0	26,51	5,15	- 1,44
BB	31,76	39,0	36,26	6,02	1,20
WW	20,81	21,0	25,95	5,09	0,04

TABELA 25
SERTÃO

Juntas	N.º Calculado	N.º Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	141,97	125,0	133,62	11,56	- 1,47
BB	20,53	30,0	26,29	5,12	1,85
WW	245,50	253,0	179,02	13,38	0,56

tamento às distribuições Binomial Negativa e Poisson. Entretanto, para realmente validar a hipótese de um efeito de proximidade no sentido norte-sul, ele precisaria ser negado em seu direcionamento leste-oeste, pois em caso contrário o direcionamento nos dois sentidos teria significado diferente.

Com este objetivo a análise de contigüidade foi repetida, para Zona da Mata e Sertão, aplicando-se a definição de junta às quadrículas que fossem contíguas na direção leste-oeste apenas. Os resultados aparecem nas tabelas 26 a 29 para a Zona da Mata e para o Sertão.

A primeira diferenciação importante é a de que, embora o efeito proximidade seja evidenciado no modelo sem reposição para a Zona da Mata, ele não o é para o Sertão, ao contrário do que ocorreu em relação ao agrupamento de quadrículas na direção norte-sul. Este fato parece estar demonstrando que o processo já atingiu uma etapa de maior dispersão de núcleos na Zona da Mata a ponto de que o efeito de proximidade existe, porém tanto na direção norte-sul como na leste-oeste; já em relação ao Sertão o processo de difusão aparece direcionado apenas em sua componente norte-sul, ao contrário do que se poderia supor baseado na idéia de que a fonte do processo de difusão fosse a região litorânea, e a direção da expansão fosse leste-oeste. Por trás desta diferenciação está o fato simples de que juntas formadas por quadrículas contendo centros 4b no sentido norte-sul são em número de 30 no Sertão, enquanto que quando se adota o direcionamento leste-oeste, o número de juntas BB passa a 22, substancialmente menor e,

já em si, indicativo de menor efeito de proximidade nesta direção que na direção norte-sul. Já na Zona da Mata o número de juntas BB foi de 39 e 35 nos dois casos, portanto quase idêntico. Já o mesmo não ocorre em relação às juntas BW, em número de 130 quando as juntas são formadas por quadrículas juntadas com contigüidade leste-oeste, enquanto são 125 no sentido norte-sul, em ambos os casos indicando distribuição aleatória.

TABELA 26
ZONA DA MATA — LESTE-OESTE
COM REPOSIÇÃO

Juntas	Número Calculado	Número Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	44,50	37,0	22,86	4,78	-1,57
BB	27,49	35,0	29,21	5,40	1,39
WW	18,01	18,0	21,04	4,59	-0,003

TABELA 27
SEM REPOSIÇÃO

Juntas	Número Calculado	Número Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	44,90	37,0	22,10	4,70	-1,68
BB	27,29	35,0	7,78	2,79	2,76
WW	17,82	18,0	6,98	2,64	0,07

TABELA 28
SERTÃO — LESTE-OESTE
COM REPOSIÇÃO

Juntas	Número Calculado	Número Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	138,15	130,0	128,80	11,35	-0,72
BB	19,97	22,0	25,38	5,04	0,40
WW	238,88	245,0	171,78	13,11	0,47

TABELA 29
SEM REPOSIÇÃO

Juntas	Número Calculado	Número Observado	Variância	Desvio Padrão	Taxa de Contigüidade
BW	138,47	130,0	50,72	7,12	-1,19
BB	19,81	22,0	12,43	3,53	0,62
WW	238,72	245,0	17,54	4,19	1,50

Com o propósito de verificar-se o processo de difusão descrito pela distribuição de centros 4b, segundo o modelo Binomial Negativo, portanto seguindo a linha do contágio, mas ainda em uma fase aglomerada, seria válido, por igual, no que diz respeito aos centros de hierarquia superior (3b por exemplo), fez-se um teste da distribuição dos mesmos, porém somente na Zona da Mata. É o que veremos a seguir.

4.1.2 — A distribuição dos centros 3b para a Zona da Mata

A análise da distribuição de pontos (Centros 3b) para o Nordeste foi feita apenas para a Zona da Mata, não só porque eles aí são mais numerosos, mas também por terem maior densidade de ocorrência. Para as mesmas 420 quadriculas originais foram contados 29 centros 3b (observe-se a título de comparação que a área de influência de São Paulo tem 94 centros 3b para 664 quadriculas), enquanto para todo o Nordeste o número de centros 3b é de 59.

É claro que sendo a área de influência do centro 3b bem maior que a do centro 4b, a distribuição limite não precisaria ser a relativa à quadricula original e sim possivelmente a primeira agregação (quatro quadriculas). Assim passamos logo à análise dos resultados da primeira agregação, na qual temos os 29 centros acima mencionados, para 106 quadriculas, conforme se pode observar nas tabelas 30 e 31.

TABELA 30
NORDESTE REDUZIDO
29 Centros 3b

Primeira Agregação

Folhas	Número de Quadriculas -A-	Número de Centros 3b -B-	B/A
Jaguaribe	28	6	0,21
Natal	16	5	0,31
Aracaju	43	11	0,21
Recife	8	4	0,50
Salvador	11	3	0,27
	106	29	

Uma média de 0,27 e variância 0,24 indicam logo a impossibilidade de se calcular os parâmetros da BN e a aceitação da Poisson, com qui-quadrado 0,18, enquanto que o tabulado foi de 3,8, para um nível de significância de 0,05. Além disso o ajuste, frequência a frequência, foi bastante bom, conforme se pode ver na tabela 31.

TABELA 31
AJUSTAMENTO
106 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	79	80,62
1	25	22,05
2	2	3,01
	106	

Poisson

M = 0,274
VAR. = 0,237
g. l. = 1
 α = 0,05
 $X^2_{obs.}$ = 0,180
 $X^2_{tab.}$ = 3,841
D = 0,86
Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Distribuição Binomial Negativa.

Duas outras agregações foram feitas, com oito e com dezesseis quadrículas (2a. e 3a. agregações), correspondentes a um retângulo de 30 minutos por um grau e um grau quadrado, respectivamente. As tabelas 32 e 33 mostram os valores obtidos na segunda agregação e as tabelas 34 e 35 indicam os resultados da terceira agregação. Logo à primeira inspeção verifica-se que na folha de Recife o número de centros por quadrícula é bem maior e na de Jaguaribe bem menor, o que irá se refletir na variância e conseqüentemente no ajuste.

Em ambos os casos não só aceita-se o ajustamento à distribuição Poisson como dados os valores dos parâmetros (média e variância), nem sequer pode-se calcular o ajuste à distribuição Binomial Negativa. Considerando que na área estudada (cerca de 350 mil km² para 29 centros) a área média dos centros 3b e mais é de aproximadamente um grau quadrado, (o que se obtém na terceira agregação), pode-se analisar a distribuição na segunda agregação como uma distribuição (no caso Poisson); na terceira agregação, entretanto, ainda se aceita a mesma distribuição Poisson, embora a análise a nível de cada frequência indique um ajuste mais precário que na segunda agregação. Tanto que o próprio índice *D* de dispersão (que na Poisson tenderia para 1), já passa de um valor 0,89 na segunda agregação para 0,52 na terceira, que reflete uma distribuição tendendo para uniforme (o que aliás se pode supor com 29 centros e 30 quadrículas).

Esta diferença entre o comportamento espacial dos centros 4b (que seguem um padrão descrito pela Binomial Negativa e os 3b que seguem uma distribuição que oscila entre aleatória e tendendo para uniforme), pode estar ligada ao multiplicador urbano de cada um dos dois níveis hierárquicos. Se voltarmos à tabela 4 podemos verificar que o fator crescimento, nas três áreas do Nordeste, oscila entre quase o dobro e quase o triplo, da relação centro 2/3 e 3/4. Isto indicaria uma ação significativa dos centros 3, gerando muito mais centros 4b, do que a ação dos centros 2, gerando outros de nível imediatamente abaixo; entretanto, é difícil determinar precisamente, de um lado, o próprio efeito de longo prazo que centros metropolitanos tiveram em ge-

TABELA 32
NORDESTE REDUZIDO
29 Centros 3b
Segunda Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 3b e mais -B-	B/A
Jaguaribe	15	6	0,40
Aracaju	22	11	0,50
Natal	8	5	0,62
Recife	5	4	0,80
Salvador	6	3	0,50
Total	56	29	

TABELA 33
AJUSTAMENTO
56 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	33	33,36
1	17	17,27
2	6	4,47
	56	

M = 0,518
 VAR. = 0,464
 $\alpha = 0,05$
 g. l. = 1
 $\chi^2_{obs.} = 0,075$
 $\chi^2_{tab.} = 3,841$
 D = 0,89
 Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Distribuição Binomial Negativa.

TABELA 34
 NORDESTE REDUZIDO
 29 Centros 3b
 Terceira Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 3b -B-	B/A
Jaguaribe	8	6	0,75
Aracaju	11	11	1,00
Natal	5	5	1,00
Recife	3	4	1,33
Salvador	3	3	1,00
Total	30	29	

TABELA 35
 AJUSTAMENTO
 30 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	8	11,41
1	15	11,03
2	7	5,33
	30	

M = 0,967
 VAR. = 0,499
 α = 0,05
 g. l. = 2
 $X^2_{obs.}$ = 2,970
 $X^2_{tab.}$ = 5,991
 D = 0,52
 Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Distribuição Binomial Negativa.

rar, diretamente, centros 4b dispersos pelo interior, que aos poucos foram crescendo e se transformando em centros maiores; de outro lado, o efeito que buscamos é o de centros que se elevaram à posição de 3b e estão produzindo crescimento para baixo (através do processo de difusão de inovações), portanto dentro da linha metodológica do multiplicador urbano.

Tornam-se necessárias mais pesquisas sobre tal mecanismo no Nordeste e Zona da Mata, que possam responder tais questões de forma

mais completa; vale esclarecer que o significado prático de tais pesquisas é enorme, pois de tal compreensão depende a possibilidade de intervir no sistema de forma adequada.

4.1.3 — A distribuição de cidades no Nordeste e Zona da Mata

A análise seguinte, ainda do sistema urbano do Nordeste, foi feita seguindo a distribuição de cidades (sedes de municípios) e não mais de centros 4b, como a princípio. O objetivo essencial, como foi salientado, era o de testar dois processos espaciais de distribuição de núcleos urbanos, afetados pelas forças de mercado (centro 4b) e por mecanismos político-administrativos (distribuição de municípios). Como no caso anterior, a simples observação dos valores pelas várias folhas da *Carta ao Milionésimo* mostra a heterogeneidade da amostra (tabela 36) ao comparar-se as mesmas folhas de Teresina e Natal, com praticamente o mesmo número de centros 4b; embora com mais de 7 vezes o número de quadriculas, o número de cidades é apenas 1,6 vezes superior, indicando mais concentração de cidades que de centros 4b. Para uma área de 40 mil km² na folha Natal, 22 centros indicam área média de cerca de 1,8 mil km², enquanto que na de Teresina esta área média é de quase 15 mil km²; já no que diz respeito à área média em torno de cada sede de município, é de 3 mil km² na folha de Teresina e de 250 mil km² na área de Natal. Outras comparações indicam heterogeneidade igual. Basta constatar que a média de municípios é de quase 0,7 municípios por quadricula na primeira análise, quase 2,5 na segunda, 4,5 na terceira e 9 na última análise (terceira agregação).

TABELA 36
NORDESTE
1.351 Cidades
Quadricula Original

Folhas	Número de Quadriculas -A-	Número de Cidades -B-	B/A
São Luís	240	94	0,39
Fortaleza	66	47	0,71
Teresina	384	109	0,28
Jaguaribe	384	354	0,92
Natal	53	162	3,05
Rio São Francisco	384	47	0,12
Aracaju	356	285	0,80
Recife	32	68	2,12
Salvador	220	185	0,84
Total	2.119	1.351	

As comparações de médias de centros 4b e de cidades permitem logo a constatação de maior concentração de cidades que de centros 4b, o que revela que o processo político, que afeta a criação de municípios, não segue a linha dos centros 4b, isto é, organização espacial da demanda.

Nesta análise seguiu-se o mesmo procedimento utilizado na análise dos centros 4b (ver tabela 37 a 39).

A primeira observação a fazer-se é que a quadrícula original não foi suficientemente pequena para forçar a distribuição binomial típica do processo nordestino à sua distribuição limite Poisson. Comparando-se a média e variância da distribuição de centros 4b com as das cidades, na análise a nível de quadrícula original, verifica-se uma enorme discrepância, pois se a média é cerca de cinco vezes maior, a variância é dez vezes maior, o que indica que a concentração de cidades é muito maior que a de centros 4b, fato que pode ser certamente atribuído ao processo político-administrativo regional ao mesmo tempo, a maior densidade de população; possivelmente o maior poder político da Zona da Mata tende a multiplicar mais o número de cidades — sedes de municípios — comparado com o Sertão. Tanto isto parece verdade que, na primeira agregação, o parâmetro K que apresentou um valor muito baixo na primeira análise (0,797), continua baixo (0,694), mas a média e variância passaram dos valores de 0,638 e 1,15 na quadrícula original, para 2,47 e 11,29 na primeira agregação. O resultado foi um ajuste muito melhor à BN que na quadrícula original, reiterando a noção acima indicada de uma forte concentração de municípios, que obviamente seria na Zona da Mata. A própria distribuição de frequências observadas indica isto, quando vemos que uma quadrícula original, cuja área aproximada é de 800 km², contém números elevados de mu-

TABELA 37
AJUSTAMENTO
2.119 Quadrículas

Número de Centros	Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas
0	1.316	1.326,43
1	506	469,85
2	155	187,61
3	75	77,73
4	40	32,79
5	17	13,98
6	6	6,00
7	3	2,59
8	1	1,12
2.119		

Binomial Negativa

p = 0,56
k = 0,79
g. 1 = 6
M = 0,638
 α = 0,05
VAR. = 1,148
D = 1,79
X²obs. = 10,871
X²tab. = 12,592
Aceita-se o ajuste

Obs.: O ajuste pela distribuição *Poisson* foi de má qualidade. Por isso a tabela correspondente não foi mencionada.

TABELA 38
NORDESTE
1.351 Cidades
Primeira Agregação

Folhas	Número de Quadrículas -A-	Número de Cidades -B-	B/A
São Luís	64	94	1,46
Fortaleza	20	47	2,35
Teresina	96	109	1,13
Jaguaribe	96	354	3,68
Natal	16	162	10,12
Rio São Francisco	96	47	0,48
Aracaju	90	285	3,16
Recife	10	68	6,80
Salvador	58	185	3,18
Total	546	1.351	

nicípios: 40 quadrículas contêm 4 municípios, quer dizer, municípios com cerca de 200 km² de área, enquanto uma quadrícula contêm oito municípios, portanto com área de 100 km² apenas. Este aspecto pode ser constatado, também, quando se compara a frequência zero na quadrícula original e na primeira agregação. Na quadrícula original, 1.316 quadrículas têm zero centros, enquanto que na primeira agregação, constituída de quatro quadrículas originais, apenas 178 quadrículas têm zero centros; da mesma forma, 506 quadrículas contêm um centro na quadrícula original, contra 111 na primeira agregação, mostrando que se na frequência zero a relação é 7 vezes maior, na frequência 1 esta relação passa a ser apenas pouco mais de 4,5 vezes, portanto em linha com o tamanho da quadrícula quatro vezes maior.

É preciso atentar, entretanto, para o fato de que a área examinada contêm 1.351 cidades, distribuídas em 2.119 quadrículas (daí a média 1,5 quadrículas por cidade), o que faria com que se pudesse supor que a área de influência de cidades seria da ordem de duas quadrículas, ou seja uma etapa intermediária entre a quadrícula original e a primeira agregação. O ajuste à distribuição BN foi bastante bom em ambos os casos (na realidade talvez um pouco melhor no caso da primeira agregação, se considerarmos os ajustes frequência a frequência), como já havíamos salientado, melhor mesmo que o ajuste obtido para os centros 4b e mais. Na realidade, em nenhum ajustamento de centros 4b se obteve tal diferença entre a média e a variância, gerando índices D de dispersão (e em consequência, parâmetros P da BN) tão baixos. O mais baixo índice de dispersão (ou, se preferirmos, o mais elevado índice de concentração obtido em todos os ajustes feitos foi o relativo à primeira agregação de cidades para o Nordeste, de 4,56, seguido do obtido para a primeira agregação de cidades da Zona da Mata, que foi de 3,4 (veja-se a tabela 78). Comparando-se estes índices com os da área de influência de São Paulo, verifica-se que a diferença é enorme (tabela 75 e 77), pois os índices de São Paulo são 0,98 para a quadrícula original e 1,3 para a primeira agregação.

TABELA 39
AJUSTAMENTO
546 Quadrículas

Número de Cidades	Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas
0	178	190,26
1	111	103,16
2	79	68,25
3	49	47,86
4	35	34,52
5	25	25,31
6	17	18,75
7	9	14,00
8	8	10,52
9	7	7,93
10	3	6,00
11	4	4,56
12	7	3,47
13	3	2,64
14	1	2,02
15	3	1,54
16	3	1,18
17	2	0,90
18	1	0,69
19	0	0,53
20	0	0,41
21	0	0,31
22	1	0,24
	546	

Binomial Negativa

$p = 0,22$
 $k = 0,69$
 $g. 1 = 11$
 $M = 2,474$
 $\alpha = 0,05$
 $VAR. = 11,291$
 $D = 4,56$
 $X^2_{obs.} = 9,532$
 $X^2_{tab.} = 19,675$
 Aceita-se o ajuste

Obs.: O ajuste pela distribuição *Poisson* foi de má qualidade. Por isso a tabela correspondente não foi mencionada.

Passando à análise da distribuição de cidades na Zona da Mata (ver tabelas 40 a 43), a primeira constatação é ainda a de que o tamanho da quadrícula não foi suficientemente pequeno para gerar uma distribuição *Poisson* limite, embora já apresente uma diferença essencial. Observe-se que quando o ajuste à distribuição *Poisson* para a região toda foi feito (embora nem indicássemos os valores), o qui-quadrado observado foi de tal ordem que praticamente teria sido desnecessário tentar o ajuste pela simples observação dos valores dados. No caso, embora se rejeite o ajuste à distribuição *Poisson*, a diferença entre média e variância não é tão grande (os parâmetros D são 1,52 e 3,44, enquanto a taxa de difusão (valor de K) passa de 0,79 para 3,07 da região como um todo para a Zona da Mata, mostrando que, mesmo

TABELA 40
 NORDESTE REDUZIDO
 670 Cidades
 Quadrícula Original

Folhas	Número de Quadrículas -A-	Número de Cidades -B-	B/A
Jaguaribe	112	165	1,47
Natal	64	159	2,48
Aracaju	168	223	1,32
Recife	32	65	2,03
Salvador	44	58	1,31
Total	420	670	

TABELA 41
 AJUSTAMENTOS
 420 Quadrículas

Número de Cidades	Distribuição de Poisson		Distribuição Binomial Negativa	
	Freqüências Observadas	Freqüências Esperadas	Freqüências Observadas	Freqüências Esperadas
0	118	85,20	118	116,25
1	125	135,91	125	122,00
2	73	108,40	73	84,88
3	51	57,64	51	49,05
4	32	22,98	32	25,45
5	13	7,33	13	12,31
6	5	1,95	5	5,66
7	2	0,44	2	2,51
8	0	0,08	0	1,08
9	1	0,01	1	0,45
	420		420	

Poisson

M = 1,595
 g. 1 = 5
 X²obs. = 42.133
 X²tab. = 11,070
 Rejeita-se o ajuste

α = 0,05
 Var. = 2,425
 D = 1,52

Binomial Negativa

p = 0,66
 k = 3,07
 g. 1 = 6
 X²obs. = 3,864
 X²tab. = 12,592
 Aceita-se o ajuste

TABELA 42
NORDESTE REDUZIDO
670 Cidades

Primeira Agregação

Folhas	Número de Quadrículas -A-	Número de Cidades -B-	B/A
Jaguaribe	28	165	5,89
Natal	16	159	9,93
Aracaju	42	223	5,30
Recife	8	65	8,12
Salvador	11	58	5,27
	105	670	

TABELA 43
AJUSTAMENTOS
105 Quadrículas

Número de Cidades	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	9	0,17	9	4,14
1	7	1,13	7	7,69
2	7	3,62	7	9,86
3	7	7,69	7	10,76
4	10	12,28	10	10,72
5	17	15,67	17	10,06
6	5	16,67	5	9,05
7	7	15,19	7	7,90
8	6	12,12	6	6,74
9	5	8,59	5	5,63
10	5	5,48	5	4,64
11	3	3,18	3	3,77
12	4	1,69	4	3,04
13	5	0,83	5	2,42
14	0	0,37	0	1,91
15	1	0,16	1	1,50
16	3	0,06	3	1,17
17	3	0,02	3	0,91
18	0	0,008	0	0,70
19	0	0,002	0	0,54
20	0	0,0009	0	0,41
21	1	0,0003	1	0,31
	105		105	

Poisson
M = 6,381
g. 1 = 8
X²obs. = 70,933
X²tab. = 15,507
Rejeita-se o ajuste

$\alpha = 0,05$
Var. = 21,949
D = 3,44

Binomial Negativa
p = 0,291
k = 2,615
g. 1 = 11
X²obs. = 11,988
X²tab. = 19,675
Aceita-se o ajuste

ainda ajustado a uma distribuição do tipo BN, a taxa de difusão, como seria de se esperar, é bastante mais elevada. O próprio ajuste à BN é melhor na Zona da Mata que para a região como um todo, principalmente nas frequências 1 e 2 centros por quadrícula, que é onde se contém a maior parte das cidades (816 das 1.351 na região inteira); esta concentração não existe na Zona da Mata, onde 570 dentre 670 estão quase que distribuídos por igual entre as frequências 1 a 4 inclusive (ver tabela 41).

O padrão de concentração aparece mais evidente ainda na primeira agregação, em que o parâmetro D (Variância/Média) é de 3,44, portanto indicando forte tendência para agrupamento (por isso mesmo o parâmetro P da BN é de 0,29 e K tem valor 2,61 também menor que em relação à primeira análise, relativa à quadrícula original).

Mesmo considerando somente a Zona da Mata, e comparando-se o que ocorre com a distribuição de centros 4b e cidades, verifica-se que no caso dos centros 4b (na segunda agregação temos 56 quadrículas e 107 centros, portanto dois centros por quadrícula (1,91) a variância é de 3,12 e o parâmetro K tem valor 3,03; (tabelas 20 e 21) em relação às cidades, na quadrícula original, a média é de 1,595 (portanto não muito diferente) e a variância é de 2,425 relativamente consistente com o que ocorre com centros 4b.

4.2 — A região Centro-Sul e a distribuição de centros de nível 4b, a diferentes tamanhos de quadrículas

O segundo exemplo utilizado foi o da região Centro-Sul, fazendo-se numerosos testes e subdivisões espaciais, embora não tivessem sido feitos testes de contigüidade.

O primeiro conjunto de testes foi feito para a região como um todo, somente a nível de centros 4b usando-se sucessivamente o tamanho das quadrículas originais, a primeira agregação, formada por duas quadrículas agregadas em sentido leste-oeste, e a segunda agregação, constituída de duas quadrículas da primeira, formando um quadrado.

O segundo teste foi feito já para a região Centro-Sul reduzida, isto é, a parte mais desenvolvida da região, com o objetivo de que ele tivesse um significado semelhante ao que a Zona da Mata teria para o Nordeste. Já aí foram feitos testes relativos às mesmas agregações, não somente para centros 4b, como também para os de nível 3b e para cidades, sedes de municípios. No caso dos centros 3b foi feita mais uma agregação de quadrículas (neste caso agrupando no sentido leste-oeste), pois sendo o centro 3b de hierarquia superior, era óbvio que sua área de influência seria maior. No caso das cidades usou-se apenas a quadrícula original e a primeira agregação pois, sendo seu número bastante elevado, o tamanho adequado seria menor.

Um terceiro tipo de teste foi ainda realizado para a área de influência de São Paulo, somente para centros 3b e cidades, estas apenas a nível de quadrícula original e primeira agregação, ao passo que o exame dos centros 3b foi feito até a terceira agregação.

O sistema urbano do Centro Sul, embora apresente um padrão de organização espacial e nível de desenvolvimento econômico diferente do sistema Nordestino, ainda apresenta muitas das características do

Universo brasileiro, tomando-se a região como um todo. Na realidade, um Centro e uma Periferia são tão típicos no Centro-Sul brasileiro como no Brasil como um todo.

Utilizada a mesma metodologia de divisão em quadrículas bem pequenas, inicialmente, procedeu-se a análise da distribuição de pontos representativos dos centros 4b, com as quadrículas iniciais constituídas por quadrados de $15' \times 15'$, portanto com 800 km^2 de área.

As tabelas 44 a 49 mostram os resultados das várias análises, indicando os valores esperados e observados, bem como os diferentes parâmetros.

A primeira constatação que se faz é que na primeira análise (quadrícula original) aceita-se tanto o ajustamento à distribuição Poisson como à Binomial Negativa, caracterizando-se assim o processo de forçar-se o tamanho da quadrícula para baixo, obtendo-se uma distribuição Poisson, limite da Binomial Negativa. Uma média 0,171 e uma variância de 0,17, resultante do fato de que de um total de 2.954 quadrículas, 2.496 têm frequência zero, indicam bem a inadequação do tamanho da quadrícula ao objeto da análise, caracterizando o fenômeno raro, ao mesmo tempo que indicando uma taxa de difusão elevada ($K = 6,49$) segundo o modelo BN. Logo na primeira agregação, que para a região Centro-Sul foi de dois quadrados e não de 4 como na Nordeste, a distribuição Poisson é rejeitada.

Nesta primeira agregação (constituída por retângulos de 1.600 km^2 , agregados no sentido leste-oeste, o ajuste à BN é aceito tendo sido o ajuste à Poisson rejeitado, como vimos). O parâmetro K baixou de 6,49 para 2,09 e P baixa de 0,974 para 0,861; ao mesmo tempo, a média de pontos por quadrícula passa de 0,171 para 0,336 e a variância para 0,391. Com estes valores fica bem evidenciado o sentido de mais concentração, corroborado pelo fato de que o parâmetro K baixa também.

Na segunda agregação, com quadrados formados por quatro quadrículas originais (portanto com área de 3.200 km^2), o ajuste à distribuição Poisson é de novo rejeitado e, mais uma vez, aceita a distribuição Binomial Negativa. O número de quadrículas aí ficou reduzido a 763 (a cada ajustamento há necessidade de aumentar ligeiramente o número de quadrículas para torná-las múltiplo de 2). A média de pontos por quadrículas e sua variância, que eram de 0,336 e 0,391, passam agora a 0,663 e 0,931, o que em si mesmo já indica quase que previamente a não aceitação da distribuição Poisson. O efeito do aumento do tamanho da quadrícula no aumento da variância é uma boa indicação do grau de concentração urbana em certas áreas, revelado a este nível de forma ainda mais clara.

Analisando-se o ajuste à Binomial Negativa observa-se que, nesta segunda agregação, o parâmetro K reduz-se (de 2,09 para 1,65), o mesmo ocorrendo com P . Entretanto, a comparação das diferentes frequências, na primeira e na segunda agregação, mostra alguns aspectos desta concentração; no caso da frequência zero as duas agregações ajustam-se quase que por igual, pois as frequências esperadas e observadas quase não diferem uma da outra (a diferença é da ordem de 2%). Já na frequência 1 centro por quadrícula na primeira agregação observa-se uma subestimação do número de quadrículas e, na segunda, uma superestimação, pequena em ambos os casos, mas indicadora de leve tendência para maior regularidade ao nível de duas que no nível de quatro quadrículas originais.

Mas na frequência dois centros por quadrícula, na primeira agregação há superestimação (proporcionalmente maior que a subestima-

TABELA 44
CENTRO-SUL
506 Centros 4b
Quadrícula Original

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A	C/B
Brasília	384	15	90	0,04	6,00
Goiânia	384	24	136	0,06	5,67
Belo Horizonte	384	48	241	0,13	5,02
Rio Doce	172	14	95	0,08	6,79
Paranapanema	384	109	488	0,28	4,48
Rio de Janeiro	341	142	617	0,42	4,35
Vitória	57	11	47	0,19	4,27
Assunción	29	7	24	0,24	3,43
Curitiba	368	85	354	0,23	4,16
Uruguaiana	150	7	29	0,05	4,14
Porto Alegre	267	42	171	0,16	4,07
Lagoa Mirim	34	2	6	0,06	3,00
	2.954	506	2.298		

TABELA 45
AJUSTAMENTOS
2.954 Quadrículas

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	2.496	2.488,97	2.496	2.494,73
1	412	425,33	412	416,12
2	44	36,51	44	40,05
3	2	2,08	2	2,91
	2.954		2.954	

Poisson

M = 0,171
X²obs. = 2,616
Aceita-se o ajuste

α = 0,05
g. l = 2
Var. = 0,176
X²tab. = 5,991
D = 1,03

Binomial Negativa

p = 0,97
k = 6,49
X²obs. = 0,256
Aceita-se o ajuste

TABELA 46
CENTRO-SUL
506 Centros 4b
Primeira Agregação

Folhas	Número de Quadriculas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
Brasília	192	15	90	0,08
Goiânia	192	24	136	0,13
Belo Horizonte	192	48	241	0,25
Rio Doce	88	14	95	0,16
Paranapanema	192	109	488	0,57
Rio de Janeiro	172	142	617	0,83
Vitória	33	11	47	0,33
Assunção	17	7	24	0,41
Curitiba	186	85	354	0,46
Uruguaiana	81	7	29	0,09
Porto Alegre	137	42	171	0,31
Lagoa Mirim	21	2	6	0,09
1.503		506	2.298	

TABELA 47
AJUSTAMENTOS
1.503 Quadriculas

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	1.098	1.073,43	1.098	1.100,08
1	326	361,31	326	318,91
2	60	60,80	60	68,34
3	16	6,82	16	12,92
4	3	0,57	3	2,28
1.503			1.503	

Poisson
M = 0,336
 $X^2_{obs.} = 23,25$
Rejeita-se o ajuste

$\alpha = 0,05$
g. l = 3
Var. = 0,391
 $X^2_{tab.} = 7,815$
D = 1,16

Binomial Negativa
p = 0,86
k = 2,09
 $X^2_{obs.} = 2,129$
Aceita-se o ajuste

TABELA 48
CENTRO-SUL
506 Centros 4b
Segunda Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
Brasília	96	15	90	0,16
Goiânia	96	24	136	0,25
Belo Horizonte	96	48	241	0,50
Rio Doce	45	14	95	0,31
Paranapanema	96	109	488	1,14
Rio de Janeiro	88	142	617	1,61
Vitória	16	11	47	0,69
Assunción	11	7	24	0,64
Curitiba	94	85	354	0,90
Uruguaiana	42	7	29	0,17
Porto Alegre	70	42	171	0,60
Lagoa Mirim	13	2	6	0,15
	763	506	2.298	

TABELA 49
AJUSTAMENTOS
763 Quadrículas

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	446	393,13	446	436,32
1	190	260,68	190	206,76
2	82	86,43	82	78,75
3	30	19,10	30	27,55
4	13	3,16	13	9,21
5	2	0,42	2	2,99
	763		763	

Poisson

M = 0,663
g. 1 = 3
X²obs. = 48,45
X²tab. = 7,815
Rejeita-se o ajuste

$\alpha = 0,05$
Var. = 0,931
D = 1,404

Binomial Negativa

p = 0,71
k = 1,65
g. 1 = 4
X²obs. = 2,574
X²tab. = 9,487
Aceita-se o ajuste

ção na frequência 1), enquanto que na segunda há subestimação (e já agora, proporcionalmente menor que na frequência 1), indicando que no primeiro caso temos os mesmos sinais de maior regularidade — quer dizer menos centros observados que esperados; enquanto isso, no segundo caso, temos de novo tendência a maior concentração, isto é, mais centros observados que esperados, o que passa a ocorrer em todas as outras frequências seguintes, de dois, três e quatro, apenas, na segunda agregação (quatro quadrículas originais); acrescenta-se ainda que a frequência cinco centros por quadrícula ocorre nesta segunda agregação, o que não ocorreu na primeira.

Convém assinalar que a área utilizada para esta análise resulta da soma de áreas de influência de duas metrópoles nacionais, com suas três metrópoles regionais correspondentes, e é constituída por partes de 12 folhas da *Carta do Brasil ao Milionésimo*, desde um pouco ao norte de Brasília até o Rio Grande do Sul (naturalmente excluído Mato Grosso, que embora parte da área de influência de São Paulo distorceria muito o resultado da análise pela sua grande extensão territorial e baixa densidade de ocupação).

Estas discrepâncias podem ser vistas sob vários ângulos; de um lado, a simples análise do número de centros contidos em cada área de influência metropolitana pode dar algumas indicações: Rio de Janeiro tem 52 centros distribuídos por mais de 120 mil km², o que daria uma área média da ordem de 2,9 mil km², por centro; Curitiba e Porto Alegre têm área média por centro, de 3,5 e 3,9 respectivamente, e Belo Horizonte tem 6,4 (veja-se tabela 3).

Enquanto isso, São Paulo, que tem 164 dos 430 centros de toda a região Centro-Sul, tem área média de 8 mil km², mas excluindo-se os 15 centros de Mato Grosso, a área média passa 2,5 mil km², o que mostra bem a conveniência de ter sido excluída aquela parte da análise.

De outro lado, como apresentamos nas tabelas respectivas, o número de quadrículas, por folha, foi feito também o cálculo de centros 4b por quadrícula e número de municípios por centro 4b.

O número de centros 4b por quadrícula constitui elemento relevante não só para análise, como também para indicar o grau de coesão interna da região estudada, comparando-se as diferentes folhas. Bastaria mencionar que a folha Rio de Janeiro tem 0,4 centros por quadrícula de 15' × 15' e 4,3 municípios por centro 4b; enquanto isso, na periferia da região, as folhas Goiânia e Rio Doce, respectivamente, têm 0,06 e 0,08 centros por quadrícula, mas têm 5,7 e 6,8 municípios por centro 4b. A conclusão lógica que se pode tirar daí é que, na folha Rio de Janeiro, o sistema urbano apresenta-se bem articulado, e que um centro 4b, que distribui uma quantidade de bens limitada e tem um poder de difusão também limitado, age sobre uma área pequena e um número de municípios pequeno, também, nesta parte da região; enquanto nas duas outras, para uma mesma dimensão espacial, o número de centros 4b é 5 vezes menor, mas o número de cidades é cerca de 50% maior, indicando o baixo grau da taxa de difusão numa rede pouco articulada, como o grande espaçamento entre cidades e centros entre si, conseqüência ou causa daquela baixa taxa de difusão.

É interessante comparar Curitiba e Porto Alegre (embora a comparação das duas folhas não reflita, necessariamente, comparação das áreas de influência das duas metrópoles); observe-se que, na folha Porto Alegre, temos 267 quadrículas, 42 centros e 171 municípios, do que resulta 0,16 centros por quadrícula e 4,1 municípios por centro 4b;

enquanto isso Curitiba tem 368 quadrículas (daí a impropriedade de uma comparação rigorosa), 86 centros e 354 municípios, do que resulta 0,23 centros por quadrícula e 4,12 municípios por centro. A diferença relativa não é grande, mas reforça o conjunto de observações que se faz, ao longo deste trabalho, a respeito das peculiaridades do sistema urbano comandado por Porto Alegre, em função das características econômicas da área. Tanto a economia essencialmente colonial (no sentido de área de colonização agrícola e pequenas propriedades) como área de pecuária extensiva estão tendendo a gerar um número de centros intermediários menor e maior quantidade de municípios.

Esta elevada diferenciação interna (observe-se a média de centros e de municípios para a área como um todo, comparado com o desvio padrão (ver tabela 3), levou-nos a refazer a análise para uma área menor (na realidade fizemos uma terceira apenas com a área de influência de São Paulo), com o propósito de investigar a natureza do processo em uma área reduzida, na qual além de continuar excluída a parte de Mato Grosso, também foram excluídas algumas folhas, diminuiu-se de 12 folhas para 8, conforme se pode ver nas tabelas 50 a 55), ao mesmo tempo que se reduzia o número de quadrículas em algumas folhas. Por exemplo, as folhas de Belo Horizonte e Goiânia, que continham 384 quadrículas cada, passaram a 192 e 143 na amostra reduzida, excluindo-se as partes menos densamente povoadas das mesmas. O mesmo ocorreu com outras, como se pode ver na tabela.

4.2.1 — *A distribuição de centros de nível 4b na região Centro-Sul reduzida*

A fim de obter um grau menor de heterogeneidade macrorregional, procurou-se ajustar o contorno da amostra reduzida aos limites da área mais desenvolvida no Centro-Sul — que em estudos do sistema urbano brasileiro temos denominado o Núcleo e a periferia imediata ao Núcleo, tanto que reduziu-se o número de quadrículas originais de 2.954 para 1.754 (enquanto os centros 4b eram reduzidos de 506 para 457).

Esta redução evidentemente produziu resultados diferentes, e que precisam ser entendidos em suas conotações próprias. A tabela 50 mostra inicialmente a distribuição de centros e número de quadrículas por folha, podendo ser observada, por exemplo, a diferença que ainda se constata entre a folha Goiânia (143 quadrículas e 17 centros) e a folha Rio de Janeiro (341 quadrículas e 142 centros), portanto cerca de 8,5 e 2,5 quadrículas por centro, o que equivaleria a dizer uma área média de 2 mil km² por centro no primeiro caso, e quase 7 mil no segundo caso, bastante heterogêneo. Outros exemplos podem ser vistos na tabela.

Submetidos os dados a ajuste nas duas distribuições, verifica-se que não há sequer possibilidade de ajuste à Binomial Negativa, pois, como vimos, o parâmetro K da BN é calculado pela fórmula (média $\times P/1-P$). Como P é média/variação e, no caso desta área reduzida, a média é 0,251 e a variância é 0,249, P é superior a 1, o que torna o denominador da fração, que calcula K , um número negativo. Entretanto, se recordarmos que a quadrícula original tem cerca de 800 km² (e temos 1.754 quadrículas e 457 centros), veremos que a área média do centro nesta região seria da ordem de 3 mil km² e, assim, teríamos que ter 4 quadrículas originais para compor o “*mean information field*”, o que seria alcançado na segunda agregação.

No primeiro ajustamento (quadrícula original), aceita-se o modelo Poisson, com um ajuste bastante bom, a nível de cada freqüência. Como salientamos acima, neste tamanho de quadrícula a média é superior à variância, o que torna o índice D de dispersão inferior a 1, um indicador de uma distribuição mais sistemática, menos dispersa que aleatória (ao contrário das distribuições do tipo contágio, que são mais concentradas que aleatórias), portanto tendendo a formar núcleos de concentração. Este índice inferior permitiu o teste de distribuição Poisson modificada, embora o mesmo tenha sido rejeitado pelo teste qui-quadrado.

A primeira agregação na análise desta região Sudeste reduzida, quer dizer, com duas quadrículas originais e 1.600 km² de área, submetida à análise, indicou, de novo, aceitação da distribuição Poisson, mas já agora não foi possível tentar o ajuste para a Poisson dispersa, e se tornou possível fazê-lo para a BN (ver tabelas 52 e 53).

A média e variância calculadas foram, respectivamente, 0,512 e 0,529 que, embora não muito diferentes, inverteu formalmente as situações, permitindo-se calcular *P* para a BN, mas não mais *P* para a Poisson dispersa. Neste nível de agregação foi aceito o ajuste à distribuição Poisson, indicando uma taxa de difusão bastante elevada, associada assim a um estágio bem adiantado do processo de difusão de inovações e, com ele, do próprio processo de desenvolvimento. (Recorde-se que estamos interpretando a existência de um centro 4b como um núcleo de difusão de inovações).

A agregação seguinte, com quatro quadrículas originais, somou um total de 454 quadrículas para os mesmos 457 centros, portanto com mais de um centro por quadrícula, mas, praticamente, com diferença irrelevante. Portanto aí estaríamos na situação de um centro por quadrícula e, assim, teoricamente, cada centro comandaria a área de uma quadrícula, ou seja, 3,2 mil km², correspondendo, deste modo, ao "mean information field" (ver tabelas 54 e 55).

Com uma média de 1,007 e variância 1,176 o ajuste à distribuição Poisson foi aceito de novo. Aceita-se também o ajuste à distribuição

TABELA 50
CENTRO-SUL REDUZIDO
457 Centros 4b

Quadrícula Original

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A	C/B
Rio de Janeiro	341	142	617	0,42	4,35
Vitória	57	11	47	0,19	4,27
Curitiba	368	86	354	0,23	4,12
Porto Alegre	267	42	171	0,16	4,07
Rio Doce	73	11	53	0,15	4,82
Belo Horizonte	192	39	194	0,20	4,97
Paranapanema	313	109	478	0,35	4,39
Goiânia	143	17	98	0,12	5,76
	1.754	457	2.012		

TABELA 51
AJUSTAMENTO
1.754 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	1.344	1.351,68
1	365	352,17
2	43	45,87
3	2	3,98
1.754		

Poisson

M = 0,261

X²obs. = 0,985

α = 0,05

g. 1 = 2

X²tab. = 5,991

Var. = 0,249

D = 0,95

Accepta-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Binomial Negativa.

TABELA 52
CENTRO-SUL REDUZIDO
457 Centros 4b
Primeira Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
Rio de Janeiro	173	142	617	0,82
Vitória	33	11	47	0,33
Curitiba	186	86	354	0,46
Porto Alegre	136	42	171	0,31
Rio Doce	38	11	53	0,29
Belo Horizonte	96	39	194	0,41
Paranapanema	158	109	478	0,69
Goiânia	73	17	98	0,23
893		457	2.012	

TABELA 53
AJUSTAMENTO
893 Quadrículas

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	536	535,30	536	677,71
1	279	237,94	279	183,75
2	59	70,09	59	28,04
3	16	11,96	16	3,17
4	3	1,53	3	0,29
	893		893	

Poisson
M = 0,512
X²obs. = 4,105
X²tab. = 7,815
g. 1 = 3
Aceita-se o ajuste

$\alpha = 0,05$
VAR. = 0,529
D = 1,03

Binomial Negativa
p = 0,97
k = 7,96
X²obs. = 147,60
X²tab. = 5,991
g. 1 = 2
Rejeita-se o ajuste

TABELA 54
CENTRO-SUL REDUZIDO
457 Centros 4b
Segunda Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 4b -B-	Número de Cidades -C-	B/A
Rio de Janeiro	88	142	617	1,61
Vitória	16	11	47	0,69
Curitiba	92	86	354	0,93
Porto Alegre	68	42	171	0,62
Rio Doce	25	11	53	0,44
Belo Horizonte	48	39	194	0,81
Paranapanema	79	109	478	1,38
Goiania	38	17	98	0,45
	454	457	2.012	

TABELA 55
AJUSTAMENTO
454 Quadrículas

Número de Centros	Distribuição de Poisson		Distribuição Binomial Negativa		
	Freqüências Observadas	Freqüências Esperadas	Freqüências Observadas	Freqüências Esperadas	
0	183	165,92	183	179,06	
1	146	167,01	146	154,29	
2	81	84,06	81	77,58	
3	29	28,20	29	29,73	
4	13	7,09	13	9,62	
5	2	1,43	2	2,76	
		454			
			454		

Poisson

M = 1,007
X²obs. = 9,451
Aceita-se o ajuste

$\alpha = 0,05$
g. 1 = 4
X²tab. = 9,488
VAR. = 1,176
D = 1,17

Binomial Negativa

p = 0,86
k = 5,98
X²obs. = 1,2544
Aceita-se o ajuste

BN, embora com K elevado 5,982. O ajuste em cada freqüência, como se pode verificar, foi bastante bom (ver tabela 55), embora com as mesmas características básicas do ajuste na região inteira, isto é, mais freqüências zero observadas que esperadas e mais freqüências 1 esperadas que observadas, que mostram uma ligeira tendência à maior dispersão, confirmada pelo parâmetro K elevado, muito mais que para a região como um todo (1,6 para a região toda e quase 6 para a região reduzida).

Ao contrário do que ocorreu não só para a região como um todo, como para o Nordeste e mesmo Zona da Mata, obteve-se um ajuste à distribuição Poisson nos três níveis, como médias proporcionais ao aumento dos quadrados e variância quase que exatamente proporcional, o que constitui significativa indicação de que a distribuição de pontos é realmente gerado por um processo Poisson.

Esta indicação sugere a necessidade de fazer-se uma análise dos centros de hierarquia logo acima (os centros de nível 3b e mais), com o propósito de se verificar se o processo de difusão segue linha semelhante; pois o fato de que na 2.^a agregação aceita-se o ajuste à BN poderia estar sugerindo concentração espacial a este nível de agregação.

4.2.2 — A distribuição de centros 3b e mais na região Centro-Sul reduzida

Considerando a região Centro-Sul reduzida, fez-se também um ajuste às duas distribuições para os centros 3b e superiores, cujos resultados aparecem nas tabelas 56 a 61, correspondente a quadrículas resultantes da primeira, segunda e terceira agregação (esta última não foi utilizada para centros 4b no Centro-Sul).

As tabelas 56 e 57 mostram os resultados da primeira agregação, cujo tamanho ainda é bastante pequeno para funcionar como “*mean information field*” do centro 3b e que isso funcionaria como distribuição limite. O simples fato de termos 150 centros e quase 900 qua-

TABELA 56
CENTRO-SUL REDUZIDO
152 Centros 3b
Primeira Agregação

Folhas	Número de Quadriculas -A-	Número de Centros 3b -B-	B/A
Rio de Janeiro	173	49	0,28
Vitória	33	4	0,12
Curitiba	186	24	0,13
Porto Alegre	136	15	0,11
Rio Doce	38	4	0,11
Belo Horizonte	96	10	0,10
Paranapanema	158	40	0,25
Goiânia	73	6	0,08
	893	152	

TABELA 57
AJUSTAMENTO
893 Quadriculas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	751	755,77
1	132	126,10
2	10	10,52
	893	

M = 0,170
VAR. = 0,168
X²obs. = 0,151
X²tab. = 5,991
 α = 0,05
g. 1 = 2
D = 0,98
Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste à Binomial Negativa.

TABELA 58
CENTRO-SUL REDUZIDO
152 Centros 3b
Segunda Agregação

Folhas	Número de Centros 3b -A-	Número de Quadrículas -B-	A/B
Rio de Janeiro	49	88	0,56
Vitória	4	16	0,25
Curitiba	24	92	0,26
Belo Horizonte	10	48	0,21
Paranapanema	40	79	0,51
Porto Alegre	15	68	0,22
Goiânia	6	38	0,16
Rio Doce	4	25	0,16
Total	152	454	

TABELA 59
AJUSTAMENTO
454 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	319	324,83
1	119	108,75
2	15	18,21
3	1	2,03
	454	

M = 0,335
VAR. = 0,302
 $\alpha = 0,05$
g. 1 = 2
 $X^{\circ}obs. = 1,957$
 $X^{\circ}tab. = 5,991$
D = 0,902
Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Distribuição Binomial Negativa.

TABELA 60
CENTRO-SUL REDUZIDO
152 Centros 3b
Terceira Agregação

Folhas	Número de Centros 3b -A-	Número de Quadrículas -B-	A/B
Rio de Janeiro	49	44	1,11
Vitória	4	9	0,44
Curitiba	24	48	0,50
Belo Horizonte	10	24	0,42
Paranapanema	40	40	1,00
Porto Alegre	15	36	0,42
Goiânia	6	18	0,33
Rio Doce	4	15	0,27
Total	152	234	

TABELA 61
AJUSTAMENTO
234 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	121	122,21
1	81	79,38
2	25	25,78
3	7	5,58
	234	

M = 0,649
VAR. = 0,621
 $\alpha = 0,05$
g. 1 = 3
X²obs. = 0,429
X²fab. = 7,815
D = 0,96

Aceita-se o ajuste pela Distribuição Poisson

Obs.: Não houve possibilidade de ajuste pela Distribuição Binomial Negativa.

drículas, com média 0,17 e variância 0,16, já indica uma certa tendência para uma distribuição mais uniforme. Quase que por via do mesmo raciocínio o mesmo ocorre com as tabelas 58 e 59, em que média e variância praticamente dobraram com o aumento, em dobro, do tamanho das quadrículas; observe-se que a média passou de 0,170 para 0,335 e a variância passou de 0,168 para 0,302, em ambos os casos praticamente o dobro, o que, de certa forma, indica a persistência do mesmo processo a estes dois níveis de agregação. As tabelas 60 e 61, em que já aparecem os mesmos centros e 234 quadrículas e ainda aí a média e a variância são proporcionais ao tamanho das quadrículas, com valores 0,649 e 0,621, o que indica persistência do mesmo tipo de distribuição, ao longo de três níveis de agregação. O comportamento dos centros 3b mostra-se, assim, essencialmente semelhante aos centros 4b na região e corrobora a diferença do Centro-Sul com o Nordeste.

Comparando-se a distribuição de centros 4b e mais com a de centros 3b e mais, aceitou-se o ajuste à distribuição Poisson, sem muita diferença entre uma e outra, e verificou-se que o parâmetro D na distribuição dos centros 3b é menor (o que quer dizer apenas que a relação (variância/média) é menor no caso dos centros 3b, e indica uma leve tendência à nucleação de centros 4b, embora não suficiente para rejeitar o ajuste à distribuição Poisson). A hipótese formulada inicialmente, quando se supôs que a aceitação do ajuste à BN, na segunda agregação dos centros 4b (correspondendo à quatro quadrículas originais), se devia apenas ao fato de estarmos ultrapassando a área média de influência do centro (e, portanto, seu "*mean information field*"), é comprovada, pelo menos, de uma maneira preliminar, como de resto o é toda esta aplicação metodológica a este nível de generalização. O fato de que, ao analisarmos a distribuição dos centros 3b, tendo ao nível da primeira, como da segunda e terceira agregações, estarmos obtendo uma distribuição ajustada ao modelo Poisson, tanto quanto se obteve o mesmo resultado em relação aos centros 4b, constitui uma forte indicação da regularidade da rede urbana nesta área, que apresenta de um lado uma distribuição espacial tendendo para sistemática (quer dizer ajustada à Poisson, mas permitindo cálculo da Poisson modificada, embora não ajustada à mesma) e, de outro lado, uma hierarquia nitidamente definida em outros estudos (veja-se também a tabela 1), ambas ligadas a um estado de equilíbrio no sistema urbano.

4.2.3 — *A distribuição de centros 3b na área de influência de São Paulo*

Ainda a título de ilustração procuramos reduzir a área ainda mais, desta vez para a área de influência de São Paulo (tabelas 62 a 69), e tanto na primeira agregação como na quadrícula original obteve-se uma média bastante baixa e variância menor que a média, portanto reiterando os resultados anteriores relativos à região Centro-Sul reduzida.

Não só média e variância aumentaram proporcionalmente ao tamanho dos quadrados, como em nenhum caso se aceitou sequer o ajuste à BN.

Evidentemente só analisamos a distribuição a partir da primeira agregação, vez que a quadrícula original seria muito pequena para análise de centros 3b; com 94 centros para 664 quadrículas o ajuste óbvio seria a Poisson. Na primeira agregação, para 332 quadrículas e os mesmos 94 centros, obtém-se uma média 0,28 e variância 0,25, o que indicaria uma aceitação da distribuição Poisson. A análise dos

ajustes nas duas frequências 1 e 2 indica uma maior quantidade de centros na frequência 1 que o esperado, enquanto na frequência 2 o esperado é superior, o que constitui uma leve tendência à maior uniformidade, (de tal forma que sendo a média maior que a variância, seria possível calcular o ajuste à Poisson modificada).

Na segunda agregação a média passou a 0,56 e a variância passou a 0,42, mostrando assim que a média dobrou (como seria óbvio esperar), mas a variância não aumentou da mesma forma, tanto que o índice D

TABELA 62
ÁREA DE INFLUÊNCIA DE SÃO PAULO
94 Centros 3b
Quadrícula Original

Folhas	Número de Quadrículas -A-	Número de 3b e Mais -B-	B/A
Parapanema	264	39	0,15
Belo Horizonte	12	2	0,17
Goiânia	42	4	0,10
Rio de Janeiro	346	49	0,14
Total	664	94	

TABELA 63
AJUSTAMENTO
664 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	573	576,35
1	88	81,59
2	3	5,77
	664	

Poisson

M = 0,142

VAR. = 0,130

g. 1 = 2

α = 0,05

$X^2_{obs.}$ = 1,830

$X^2_{tab.}$ = 5,991

D = 0,92

Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Distribuição Binomial Negativa.

TABELA 64
ÁREA DE INFLUÊNCIA DE SÃO PAULO
94 Centros 3b
Primeira Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 3b -B-	B/A
Paranapanema	132	39	0,29
Belo Horizonte	6	2	0,33
Goiânia	21	4	0,19
Rio de Janeiro	173	49	0,28
Total	332	94	

TABELA 65
AJUSTAMENTO
332 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	246	250,13
1	78	70,82
2	8	10,02
	332	

Poisson

$M = 0,283$

$VAR. = 0,251$

$g. 1 = 2$

$\alpha = 0,05$

$X^2_{obs.} = 1,180$

$X^2_{tab.} = 5,991$

$D = 0,89$

Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Distribuição Binomial Negativa.

TABELA 66
 ÁREA DE INFLUÊNCIA DE SÃO PAULO
 94 Centros 3b
 Segunda Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 3b -B-	B/A
Parapanema	67	39	0,58
Belo Horizonte	3	2	0,67
Goiânia	12	4	0,33
Rio de Janeiro	87	49	0,56
Total	169	94	

TABELA 67
 AJUSTAMENTO
 169 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	89	96,90
1	67	53,89
2	12	14,99
3	1	2,78
	169	

M = 0,556

VAR. = 0,424

α = 0,05

g. 1 = 2

X²obs. = 5,109

X²tab. = 5,991

D = 0,76

Accepta-se o ajuste

Obs.: Não houve possibilidade de ajuste pela Binomial Negativa.

TABELA 68
 ÁREA DE INFLUÊNCIA DE SÃO PAULO
 94 Centros 3b
 Terceira Agregação

Folhas	Número de Quadrículas -A-	Número de Centros 3b -B-	B/A
Paranapanema	33	39	1,18
Belo Horizonte	2	2	1,00
Goiânia	8	4	0,50
Rio de Janeiro	45	49	1,09
Total	88	94	

TABELA 69
 AJUSTAMENTO
 88 Quadrículas

Número de Centros	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	25	30,24
1	38	32,30
2	19	17,25
3	6	6,14
	88	

M = 1,068

VAR. = 0,768

D = 0,72

α = 0,05

g. 1 = 3

X²obs. = 2,003

X²tab. = 7,814

Accepta-se o ajuste

Obs.: Não houve possibilidade de ajuste pela distribuição Binomial Negativa.

passou de 0,89 para 0,76, aumentando assim a tendência para uma distribuição dispersa. Observe-se que o número de quadrículas com 1 centro calculado é de 54, enquanto o observado foi de 67, reiterando aquela característica já indicada na primeira agregação.

Finalmente, na terceira agregação, que é a que corresponderia mais claramente à área de influência do centro 3b, (8 quadrículas originais), a análise da tabela 68 mostra precisamente isso: quase todas as folhas têm 1 centro 3b por quadrícula, com exceção de Goiânia, portanto confirmando a média da área de influência. Ao lado disso a média é de 1,06 e a variância é de 0,768 (índice D é de 0,72). Um aspecto significativo deste ajuste é o de que, frequência a frequência, ele aparece bem melhor que o anterior, conforme se vê na tabela. Ainda nesta terceira agregação observa-se que o número de quadrículas com frequência 1 esperada é menor que o observado, constituindo-se em mais um elemento de confirmação da tendência da distribuição para o modelo Poisson modificada (mais uniforme que aleatória).

Um ponto importante é que a distribuição dos centros 3b segue o modelo Poisson nas várias agregações, ao mesmo tempo que tanto a média como a variância vão se modificando proporcionalmente, esta última até um pouco mais rapidamente; uma diminuição maior da variância que a média é indicadora da tendência para uma distribuição uniforme. O que estamos pretendendo dizer com isso é que, seguindo o modelo Poisson em todas as escalas da análise, pode-se supor que o modelo, neste caso, contém uma interpretação genética do processo, e não apenas descritivo. Este dado é tanto mais importante, quando se considera que a distribuição de centros 4b também seguiu o mesmo modelo (o que ocorrerá também com as cidades).

Simetricamente com o que fizemos em relação ao Nordeste, veremos a seguir o padrão de distribuição de cidades, para compará-lo com os centros com atuação regional, e da mesma maneira verificar se o processo de criação de municípios é simétrico ou não com o surgimento de centros regionais.

4.2.4 — *A distribuição de cidades para a região Centro-Sul reduzida e áreas de influência de São Paulo*

Da mesma forma que fizemos em relação ao Nordeste, fizemos também uma análise da distribuição das cidades na região Centro-Sul (apenas a área reduzida) e para a região de influência de São Paulo, cujos resultados aparecem nas tabelas 70 a 73 (área reduzida) e 74 a 77 para a área de São Paulo.

A análise da distribuição de cidades para a região Centro-Sul reduzida indica que a distribuição de cidades parece obedecer a um padrão de concentração um pouco maior, pois já na quadrícula original rejeita-se o ajuste à distribuição Poisson, aceitando-se o ajuste à BN, embora a diferença entre média e variância não seja grande (1,147 e 1,403), com um parâmetro K bem elevado (5,14). Enquanto isso, para os centros 4b, como se verificou na segunda agregação, quando o tamanho da quadrícula aproxima-se do tamanho da área de influência do centro, aceita-se tanto o ajuste à Poisson como à BN (esta última com parâmetro K quase igual a 5,98). Já na primeira agregação o parâmetro K diminui bastante, passando a 3,1 (ver tabela 73) e a diferença entre a média e variância é bastante grande (2,25 e 3,9).

Considerando-se, entretanto, somente a área de influência de São Paulo, o processo torna-se diferente, obtendo-se um ajuste à distribuição Poisson tanto na quadrícula original quanto na primeira agrega-

ção, embora nesta última aceite-se também ajuste à BN. Observe-se que na quadricula original, para 664 quadriculas temos 1.070 municípios, o que poderia dar margem à concentração; o fato de não ter-se observado maior concentração (média 1,61 e variância 1,59) e pouco

TABELA 70
CENTRO-SUL REDUZIDO
2.012 Cidades
Quadricula Original

Folhas	Número de Quadriculas -A-	Número de Cidades -B-	B/A
Rio de Janeiro	341	617	1,81
Vitória	57	47	0,82
Curitiba	368	354	0,96
Porto Alegre	267	171	0,64
Rio Doce	73	53	0,73
Belo Horizonte	192	194	1,01
Paranapanema	313	478	1,53
Goiânia	143	98	0,69
1.754		2.012	

TABELA 71
AJUSTAMENTOS
1.754 Quadriculas

Número de Cidades	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	622	556,99	622	622,79
1	599	638,92	599	584,07
2	298	366,45	298	327,15
3	151	140,12	151	142,06
4	62	40,18	62	52,74
5	19	9,22	19	17,59
6	2	1,76	2	5,42
7	0	0,26	0	1,57
8	0	0,04	0	0,44
9	1	0,01	1	0,12
1.754		1.754		

Poisson

M = 1,147
g. 1 = 5
X²obs. = 45,722
X²tab. = 11,070
Rejeita-se o ajuste

$\alpha = 0,05$
VAR. = 1,403
D = 1,22

Binomial Negativa

p = 0,82
k = 5,14
g. 1 = 6
X²obs. = 8,022
X²tab. = 12,591
Aceita-se o ajuste

TABELA 72
CENTRO-SUL REDUZIDO
2.012 Cidades
Primeira Agregação

Folhas	Número de Quadrículas -A-	Número de Cidades -B-	B/A
Rio de Janeiro	173	617	3,57
Vitória	133	47	1,42
Curitiba	186	354	1,90
Porto Alegre	136	171	1,26
Rio Doce	38	53	1,39
Belo Horizonte	96	194	2,02
Paranapanema	158	478	3,03
Goiânia	73	98	1,34
	893	2.012	

TABELA 73
AJUSTAMENTOS
893 Quadrículas

Número de Cidades	Distribuição de Poisson		Distribuição Binomial Negativa	
	Freqüências Observadas	Freqüências Esperadas	Freqüências Observadas	Freqüências Esperadas
0	168	93,83	168	164,75
1	220	211,41	220	214,19
2	176	238,16	176	184,53
3	117	178,86	117	132,00
4	92	100,75	92	84,78
5	56	45,39	56	50,73
6	34	17,04	34	28,87
7	16	5,48	16	15,83
8	9	1,54	9	8,43
9	2	0,38	2	4,38
10	2	0,08	2	2,24
11	0	0,01	0	1,12
12	0	0,003	0	0,55
13	0	0,0006	0	0,27
14	0	0,0001	0	0,13
15	1	0,0000	1	0,06
	893		893	

Poisson

M = 2,253
g. 1 = 7
X²obs. = 183,766
X²tab. = 14,067
Rejeita-se o ajuste

$\alpha = 0,05$
VAR. = 3,905
D = 1,733

Binomial Negativa

p = 0,58
k = 3,07
g. 1 = 9
X²obs. = 6,063
X²tab. = 15,507
Aceita-se o ajuste

TABELA 74
ÁREA DE INFLUÊNCIA DE SÃO PAULO
1.070 Cidades

Quadrícula Original

Folhas	Número de Quadrículas -A-	Número de Cidades -B-	B/A
Paranapanema	264	433	1,64
Belo Horizonte	12	5	0,42
Goiânia	42	17	0,40
Rio de Janeiro	346	615	1,78
Total	664	1.070	

TABELA 75
AJUSTAMENTO
664 Quadrículas

Número de Cidades	Distribuição de Poisson	
	Frequências Observadas	Frequências Esperadas
0	126	132,53
1	226	213,57
2	164	172,08
3	97	92,43
4	35	37,24
5	14	12,00
6	1	3,22
7	0	0,74
8	0	0,15
9	1	0,03
	664	

Poisson

$M = 1,611$

$Var. = 1,593$

$g. 1 = 5$

$\alpha = 0,05$

$X^2_{obs.} = 1,76$

$X^2_{tab.} = 11,070$

$D = 0,99$

Aceita-se o ajuste

Obs.: Não houve possibilidade de ajuste pela distribuição Binomial Negativa.

TABELA 76
ÁREA DE INFLUÊNCIA DE SÃO PAULO
1.070 Cidades

Primeira Agregação

Folhas	Número de Quadrículas -A-	Número de Cidades -B-	B/A
Paranapanema	132	433	3,28
Belo Horizonte	6	5	0,83
Goiânia	21	17	0,81
Rio de Janeiro	173	615	3,55
Total	332	1.070	

TABELA 77
AJUSTAMENTOS
332 Quadrículas

Número de Cidades	Distribuição de Poisson		Distribuição Binomial Negativa	
	Frequências Observadas	Frequências Esperadas	Frequências Observadas	Frequências Esperadas
0	21	13,22	21	20,07
1	49	42,62	49	49,33
2	65	68,69	65	66,47
3	59	73,79	59	64,97
4	54	59,46	54	51,49
5	44	38,32	44	35,09
6	23	20,58	23	21,31
7	9	9,47	9	11,82
8	3	3,81	3	6,09
9	2	1,36	2	2,94
10	2	0,44	2	1,35
11	0	0,12	0	0,59
12	0	0,03	0	0,25
13	0	0,008	0	0,10
14	0	0,002	0	0,04
15	1	0,0004	1	0,01
	332		332	

Poisson

$M = 3,223$
 $g. 1 = 8$
 $X^2_{obs.} = 11,14$
 $X^2_{tab.} = 15,507$
 Aceita-se o ajuste

$\alpha = 0,05$
 $VAR. = 4,227$
 $D = 1,31$

Binomial Negativa

$p = 0,76$
 $k = 10,34$
 $g. 1 = 9$
 $X^2_{obs.} = 5,460$
 $X^2_{tab.} = 16,919$
 Aceita-se o ajuste

mais na primeira agregação (média 3,2 e variância 4,2) é indicação de que a nível de área de influência de São Paulo o processo permanece constante em diferentes agregações, o que já havia ocorrido em relação a centros 4b e 3b também.

Esta diferenciação da área de influência de São Paulo é importante de ser assinalada, pois corresponde aos diferentes estágios de desenvolvimento identificados nas análises do sistema urbano, isto é, um núcleo mais desenvolvido na área de São Paulo, uma periferia imediata próxima e uma periferia mais ampla, correspondendo às áreas mais remotas do próprio Centro-Sul. A estas distinções estruturais estão correspondendo diferenças nas funções geradoras das distribuições de pontos, tomados como centros de difusão. É importante salientar que, na área de São Paulo, até mesmo a distribuição de cidades, que é gerada por fatores de ordem político-administrativa, segue a mesma linha dos centros 4b e 3b, portanto com o mesmo nível de racionalidade e relações com as forças de mercado que os centros de hierarquia superior.

Apenas observe-se que na primeira agregação (332 quadrículas e 1.070 cidades), aceita-se também a distribuição BN, embora com o parâmetro K muito elevado (10,34) e sem grande diferença entre média e variância (3,22 e 4,23). Neste caso, que a rigor não caberia analisar, pois cada quadrícula contém mais de três cidades, quando deveria conter apenas uma para se presumir o critério de relação entre tamanho da quadrícula e evento estudado, o ajuste à BN parece melhor que o Poisson, quando se compara frequências observadas e esperadas.

4.3 — Análise comparativa das várias distribuições nas duas regiões e suas subdivisões

Uma síntese dos ajustamentos testados pode ser feita nos dois tipos de objetivos procurados:

1 — A distribuição de centros 4b e, em alguns casos, 3b, tanto no Nordeste como Centro-Sul, tanto na região como um todo como em partes mais ou menos desenvolvidas: o objetivo era o de procurar relacionar o tipo de distribuição que melhor se ajustasse aos diferentes estágios do processo de desenvolvimento, definido no presente estudo tanto pelas duas dimensões básicas apresentadas na tabela 1 como pelos diferentes índices de urbanização apresentados, tais como o multiplicador urbano e o fator crescimento.

2 — A distribuição de cidades nas duas regiões: o objetivo era o de verificar se os processos que governam a distribuição e portanto a formação de cidades, são semelhantes aos que se relacionam à formação de centros de mercado, tanto os 4b como 3b.

No primeiro caso as várias distribuições foram testadas a vários níveis de agregação de quadrículas, tanto com o objetivo de testar a validade das mesmas aos níveis de agregação de quadrículas sugeridos pela extensão da área média de influência dos centros, como partindo do sentido inverso (isto é, da suposição de que a distribuição de centros nas duas regiões como um todo obedecesse ao modelo BN) e a área média pudesse ser definida por via do melhor ajustamento.

Como os valores obtidos para os multiplicadores urbanos e fatores de crescimento foram sempre maiores na relação centros 3/4 que na centros 2/3, procurou-se testar a distribuição de centros 3b também,

com o propósito de verificar, em ambos os casos, se o processo gerador era semelhante. Obviamente, diante dos diferenciais de multiplicador urbano e fator crescimento, a hipótese nula é a de que os processos sejam semelhantes, talvez menos na região Centro Sul que no Nordeste ou Zona da Mata.

Por fim, alguns testes de independência da ocorrência de eventos (no caso os centros 4b apenas) foram feitos através de análises de contigüidade e descritos na análise dos resultados obtidos.

Numerosas comparações foram feitas entre os vários resultados, mas o que procuraremos fazer neste capítulo é uma comparação geral, inter e intra-regional, associando estas com comparações com alguns resultados de estudos do sistema urbano feitos em outros lugares. É claro que seria prematuro ainda tentar conclusões definitivas, mas algumas idéias podem ser fixadas que, pela forma de reiteração de resultados anteriores, podem constituir mais um elemento de comprovação daqueles estudos.

A primeira é a de que, quando estabelecemos o tamanho da quadrícula correspondente à área de influência dos centros 4b (oito quadrículas para o Nordeste e quatro para o Centro-Sul), obtivemos, em ambos os casos, um bom ajustamento à distribuição Binomial Negativa. É claro que o sentido deste ajustamento não é genético, porque seria desnecessário enfatizar a heterogeneidade do próprio processo de distribuição de pontos constituídos de centros 4b nas duas regiões; mas não se pode fugir à interpretação — mesmo por via de um ajustamento que tem caráter descritivo do padrão atual — de que os estágios no processo de difusão são diferentes nas duas regiões (o que é também um fato conhecido); observe-se que o parâmetro K toma um valor 0,94 para o Nordeste e 1,65 para o Centro-Sul, indicando uma taxa de difusão bem superior nesta última região, reiterando os resultados da tabela 1, de maiores índices de urbanização ao longo da hierarquia de cidades do Centro-Sul comparados com os da região Nordeste; da mesma forma reitera os resultados obtidos com os diferentes índices de urbanização e fator crescimento (embora se possa, neste caso, argüir uma certa circularidade nos resultados, provindos da mesma classificação em regiões funcionais urbanas).

Esta comparação é feita, já admitida a diferenciação regional, pois como se indicou estamos tomando os centros como tendo áreas de influência diferentes, menores no Centro-Sul; comparando-se os parâmetros K para o mesmo tamanho de quadrículas, teríamos valores mais próximos uns dos outros, em relação ao parâmetro K , porém com médias de pontos por quadrículas substancialmente menores.

Observa-se no Centro-Sul que da quadrícula original para a segunda agregação (constituída de quatro quadrículas originais e equivalentes, em área, à primeira agregação no Nordeste) a variância aumenta de 0,176 para 0,931, portanto mais de cinco vezes, enquanto a média aumenta quatro vezes, conforme a variação do tamanho da quadrícula. Assim a variância no Centro-Sul, para o mesmo grau de agregação, é maior que a do Nordeste como um todo, indicando um maior grau de concentração espacial. Mas isto não quer dizer que a distribuição de centros 4b seja mais concentrada no Centro-Sul e mais dispersa no Nordeste, pois que o equivalente em termos de área coberta por um centro 4b no Centro-Sul comparado com Nordeste é duas quadrículas no Centro-Sul para quatro no Nordeste e o índice D para o Nordeste é superior ao do Centro-Sul. Da mesma forma o parâmetro K é bem maior, pois sendo 0,94 para a segunda agregação no Nordeste, passa a 2,09 no Centro-Sul na mesma agregação.

Em todas as comparações feitas (na realidade na própria hipótese inicialmente formulada sobre tamanho adequado da quadrícula), estamos considerando uma quadrícula de área duas vezes maior no Nordeste, em relação ao Centro-Sul. A hipótese foi baseada no fato de que as medidas de extensão da área de influência de centros 4b indicam que elas são cerca de duas vezes maior no Nordeste (tanto as que partem da soma das áreas dos municípios abrangidos, como as que resultam da divisão do número de centros pelo número de km² correspondentes às quadrículas em que os centros são, em média, observados).

Isto significa que as diferenças na intensidade do processo de difusão não são expressas unicamente pelos valores de K mas que, na própria área da quadrícula, já esta implícita uma função distância diferenciadora. É evidente, portanto, que, nos casos das distribuições que se ajustam ao modelo Binomial Negativo, a intensidade do processo de difusão deve ser entendida, considerando, simultaneamente, as variações do parâmetro K e do tamanho da quadrícula, embora a única medida objetiva seja a do referido parâmetro.

Comparando-se a Zona da Mata com o Centro-Sul reduzido, o parâmetro K, que é de 3,0 na segunda agregação, adquire um valor 5,98 no Centro-Sul, cerca de duas vezes maior, como já havia sido observado para a região como um todo. Da mesma forma, analisando-se a variância e a média, verifica-se que na Zona da Mata ela é de 3,12 e 1,91 respectivamente, enquanto no Centro-Sul reduzido ela é de 1,17 e 1,0; no primeiro caso rejeita-se a distribuição Poisson e aceita-se a BN e no segundo caso aceita-se a Poisson e BN ao mesmo tempo, com média pouco inferior à variância; já na Zona da Mata a variância é bem superior à média, conforme se vê na tabela.

Esta seqüência de valores do parâmetro K = 0,94 para o Nordeste, 1,65 para o Centro-Sul e 5,99 para o Centro-Sul reduzido, são medidas da intensidade do processo de difusão nas três áreas, mostrando quase que diferentes estágios do processo de desenvolvimento, também bem compatíveis com os resultados já verificados em diferentes análises do sistema urbano brasileiro.

Entretanto, o simples ajustamento a uma determinada distribuição (ou a mais de uma como foi o caso na análise a nível da quadrícula original nas duas regiões) pode refletir — conforme salientamos mais de uma vez — apenas heterogeneidade regional. Não fosse isso, como explicar que para a Zona da Mata aceita-se uma distribuição Poisson na primeira agregação (correspondente à área de influência do centro 4b nesta área), diferentemente do que ocorre para a região Nordeste. Isto é tanto mais válido quando estamos operando em uma escala macrorregional e em regiões que sabemos conter um maior ou menor grau de heterogeneidade. Não foi por outra razão que seguimos a sugestão de Harvey¹⁰ em usar um método iterativo, “pois ele pode tornar possível discriminar entre um modelo de difusão por contágio e sua aceitação por heterogeneidade, comparando-se a escala em que uma distribuição Binomial Negativa produz um melhor ajustamento, com a escala do campo médio de informação”.

A segunda observação a fazer é que, quando as duas regiões são reduzidas às suas partes mais desenvolvidas, (neste caso a primeira agregação no Nordeste Reduzido com quatro quadrículas originais corresponde a segunda no Centro-Sul reduzido e em ambos os casos temos média de uma quadrícula por centro aproximadamente). Observa-se que os dois parâmetros K são muito semelhantes, até levemente mais elevado na Zona da Mata do Nordeste. Para esta área reduzida das

duas regiões as duas distribuições são aceitas, o que mostra o caráter genético do ajuste, refletindo, mesmo na área reduzida, heterogeneidade espacial resultante do padrão concentrado de conglomerados de pontos formando um padrão geral aleatório. Observe-se que não estamos tratando, neste caso, da aproximação limite, dada com o ajuste ao modelo Poisson, válido apenas para o caso de médias de pontos muito baixas, como foi o caso da quadrícula original. Trata-se de um ajuste à distribuição Poisson, com média próxima ou pouco acima de 1 ponto por quadrícula, como é o caso tanto da primeira agregação no Nordeste reduzido como da segunda agregação no Centro-Sul reduzido.

A tabela 78 indica os valores dos parâmetros para as várias distribuições no Nordeste. Verifica-se que apesar da média de pontos estar variando (necessariamente) com o tamanho das quadrículas, a variância está se acentuando a cada agregação, tanto que é de cinco vezes mais quando se agregam quatro quadrículas, subindo para 38 vezes quando, na terceira agregação, temos 16 quadrículas. Tanto que o índice D de dispersão (ou concentração) passa de 1,02 na quadrícula original para 2,7 na última agregação. Esta significativa ampliação da variância com o aumento do tamanho da quadrícula é bem indicativa do processo de concentração do Nordeste, que reflete claramente a enorme discrepância no próprio tamanho das cidades metrópoles em relação ao nível hierárquico logo abaixo. O padrão Binomial Negativo descreve um padrão espacial de cidades, que é reflexo do processo urbano da região; portanto, mesmo dada a dualidade de ajustamento que normalmente levaria à rejeição da hipótese de validade genética o modelo pode ser interpretado como tal, apoiado nos estudos anteriores e nos próprios valores dos multiplicadores urbanos e fator crescimento (em Recife passa de 15 na relação centros 1 para 2 a 51 na relação centros 2 para 3).

Isto está significando bem que a ação da metrópole sobre o nível imediatamente abaixo é fraco, situação característica de rede urbana não articulada.

Outro aspecto importante a considerar, como um terceiro ponto a fixar, é que, quando analisamos as variações intra-regionais, elas são igualmente importantes. No Nordeste, por exemplo, verificamos que na Zona da Mata, para a primeira agregação, que corresponde à área de influência dos centros 4b, temos uma taxa de difusão de 6,4, conforme assinalamos, ligeiramente superior ao Centro-Sul reduzido, comparado com um valor 0,94 indicado para o Nordeste como um todo. Com esta taxa quase 7 vezes superior ao conjunto da região, associada a um valor da variância pouco superior à média (o que levou a um ajuste das duas distribuições, fato que não havia ocorrido para a região com um todo), pode-se perceber o elevado grau de heterogeneidade existente na região. Esta observação reitera, de forma bem nítida, os resultados de estudos e valores indicados na tabela 1, de forte diferenciação no sistema urbano do Nordeste.

Para a região Nordeste, como já indicamos no capítulo de análise dos resultados para a região, foi feito um teste de independência da distribuição dos centros 4b, para testar a hipótese de haver um efeito de contágio (de certa forma sugerido pelo ajuste à distribuição Binomial Negativa), mas com um direcionamento norte-sul. O teste de contigüidade, feito para testar esta hipótese, aceita não só a própria existência do efeito de proximidade, mas também seu direcionamento norte-sul; mas repetido o teste para as duas componentes regionais (Sertão e Mata) verifica-se que as hipóteses sobre contágio mais adiantado do processo de difusão na Zona da Mata são confirmadas; é que na

Zona da Mata se detecta o efeito proximidade também no sentido leste-oeste, quer dizer ele se aproxima do efeito multidirecionado, enquanto no Sertão o efeito fica restrito ao direcionamento norte-sul. A distribuição de juntas formadas por quadrículas contendo, ambas, um centro 4b, tem caráter aleatório no Sertão, quando as juntas são formadas por quadrículas agregadas no sentido leste-oeste.

As variações intra-regionais são importantes no Centro-Sul também. A tabela 79 mostra os mesmos parâmetros das duas distribuições, tanto para a região como um todo, como para as suas áreas mais reduzidas.

O primeiro fato a assinalar é que, ao passar da região como um todo para sua área reduzida, na quadrícula correspondente a um centro por quadrícula (segunda agregação) aceita-se as duas distribuições, o que não tinha acontecido para a região inteira. Neste ponto verifica-se uma simetria com o que ocorreu no Nordeste, guardada a diferença de tamanho da quadrícula, sempre duas vezes menor para o Centro-Sul.

Ao mesmo tempo, esta diferenciação regional significativa pode ser medida pelos valores diferentes do parâmetro K , de 1,6 para a região e de quase 6 para a região reduzida.

Passando a uma comparação do padrão espacial dos centros 3b, verifica-se, desde logo, um comportamento bem diferente: em nenhum caso a média de pontos por quadrícula é inferior à variância, o que mostra o caráter bem mais disperso de centros desta hierarquia, inclusive para a área de influência de São Paulo (embora aí não se possa comparar com os centros 4b que não foram analisados). Entretanto, é curioso observar que somente na terceira agregação se atinge o valor um centro em São Paulo, quase 1 na Zona da Mata e 0,65 na Centro-Sul reduzido; mas ao lado disso a variância mantém-se proporcional ou tende a diminuir (como é o caso da Zona da Mata), o que parece estar indicando uma tendência para regularidade muito maior no que diz respeito aos centros 3b que nos centros 4b.

Já as cidades não se comportam nem como os centros 4b nem como os centros 3b. A distribuição Poisson só é aceita na área de influência de São Paulo. Considerando-se o Nordeste como um todo, ao passar da quadrícula original para a primeira agregação a média aumenta 4 vezes, enquanto a variância aumenta mais de 10 vezes, enquanto na Zona da Mata ela aumenta nove vezes (índices D 3,4 e 4,5 respectivamente), ao passo que no Centro-Sul reduzido o índice D é de 1,7, portanto indicando um grau de concentração muito maior no Nordeste que no Centro-Sul.

Em síntese pode-se dizer que os centros apresentam os seguintes padrões de distribuição:

1) Tomando-se a região por inteiro, tanto a região Nordeste como o Centro-Sul apresentam bom ajustamento à distribuição Binomial Negativa, nos vários níveis de agregação. No caso particular da quadrícula original, em ambos os casos, aceita-se a distribuição Poisson como forma limite.

2) A região reduzida (parte mais desenvolvida), tanto Nordeste como Centro-Sul, já o tipo de distribuição aceito é o Poisson. Mas neste caso para a Zona da Mata aceita-se também a Binomial Negativa na primeira e segunda agregação, enquanto a Poisson somente foi aceita na quadrícula original e primeira agregação, esta última já com área

TABELA 78

PARÂMETROS DAS DISTRIBUIÇÕES DE PONTOS
SEGUNDO AS ÁREAS

Áreas	Categorias	Estágios	Parâmetros					Distribuições	
			M	VAR.	D	P	K	Poisson	Binomial Negativa
Nordeste	Centros 4b	Quad. Original	0,101	0,103	1,02	0,98	5,21	Aceita	Aceita
		1. ^a Agregação	0,394	0,499	1,27	0,79	1,47	Rejeitada	Aceita
		2. ^a Agregação	0,729	1,296	1,78	0,56	0,94	Rejeitada	Aceita
		3. ^a Agregação	1,396	3,824	2,74	0,37	0,80	Rejeitada	Aceita
	Cidades	Quad. Original	0,638	1,148	1,79	0,56	0,79	Rejeitada	Aceita
		1. ^a Agregação	2,474	11,291	4,56	0,22	0,69	Rejeitada	Aceita
Nordeste Reduzido	Centros 4b	Quad. Original	0,255	0,232	0,91	—	—	Aceita	—
		1. ^a Agregação	1,019	1,181	1,16	0,86	6,43	Aceita	Aceita
		2. ^a Agregação	1,911	3,117	1,63	0,61	3,03	Rejeitada	Aceita
	Centros 3b	1. ^a Agregação	0,274	0,237	0,86	—	—	Aceita	—
		2. ^a Agregação	0,518	0,464	0,89	—	—	Aceita	—
		3. ^a Agregação	0,967	0,499	0,52	—	—	Aceita	—
	Cidades	Quad. Original	1,595	2,425	1,52	0,66	3,07	Rejeitada	Aceita
		1. ^a Agregação	6,381	21,949	3,44	0,29	2,62	Rejeitada	Aceita

TABELA 79

PARÂMETROS DAS DISTRIBUIÇÕES DE PONTOS
SEGUNDO AS ÁREAS

Áreas	Categorias	Estágios	Parâmetros					Distribuições	
			M	VAR.	D	P	K	Poisson	Binomial Negativa
Centro Sul	Centros 4b	Quad. Original	0,171	0,176	1,03	0,97	6,49	Aceita	Aceita
		1. ^a Agregação	0,336	0,391	1,16	0,86	2,09	Rejeitada	Aceita
		2. ^a Agregação	0,663	0,931	1,40	0,71	1,65	Rejeitada	Aceita
Centro Sul Reduzido	Centros 4b	Quad. Original	0,261	0,249	0,95	—	—	Aceita	—
		1. ^a Agregação	0,512	0,529	1,04	0,97	7,96	Aceita	Rejeitada
		2. ^a Agregação	1,007	1,176	1,17	0,86	5,98	Aceita	Aceita
	Centros 3b	1. ^a Agregação	0,170	0,168	0,98	—	—	Aceita	—
		2. ^a Agregação	0,335	0,302	0,90	—	—	Aceita	—
		3. ^a Agregação	0,649	0,621	0,96	—	—	Aceita	—
	Cidades	Quad. Original	1,147	1,403	1,22	0,82	5,14	Rejeitada	Aceita
1. ^a Agregação		2,253	3,905	1,73	0,58	3,07	Rejeitada	Aceita	
Área de Influência São Paulo	Centros 3b	Quad. Original	0,142	1,130	0,92	—	—	Aceita	—
		1. ^a Agregação	0,283	0,251	0,89	—	—	Aceita	—
		2. ^a Agregação	0,556	0,424	0,76	—	—	Aceita	—
		3. ^a Agregação	1,068	0,768	0,72	—	—	Aceita	—
	Cidades	Quad. Original	1,611	1,593	0,99	—	—	Aceita	—
		1. ^a Agregação	3,223	4,227	1,31	0,76	10,34	Aceita	Aceita

$$D = \frac{\text{VAR.}}{M}$$

$$P = \frac{M}{\text{VAR.}}$$

$$K = \frac{M \cdot P}{1-p}$$

superior à área de influência do centro 4b. No Centro-Sul aceita-se a Poisson nos três níveis; para a Binomial Negativa, no primeiro nível nem pôde ser calculado o parâmetro P , porque a média foi superior à variância; no segundo nível a distribuição foi rejeitada e finalmente aceita no terceiro nível. Para esta região reduzida (nas duas áreas e estudadas), a distribuição dos centros de nível acima (3b) obedece, em ambos os casos, o padrão Poisson. Na realidade o fato de que em todos os níveis de agregação as médias e variâncias aumentaram de forma regular com o tamanho das quadriculas, indica bem que a natureza do processo gerador é também Poisson, o que torna o modelo mais explicativo que descritivo.

3) Analisada apenas a área de influência de São Paulo, inclusive em níveis de agregação maiores (como a segunda e terceira) a distribuição de centros 3b, em todos os níveis, segue o padrão Poisson, reforçando a interpretação de que se trata de processo Poisson. O fato de que, ao fazermos a análise da distribuição de cidades, também para a área de influência de São Paulo, a distribuição ajustada foi a de Poisson, reforça mais o argumento. Este fato é significativo, pois a única área em que a distribuição de cidades seguiu o modelo Poisson foi na área de influência de São Paulo, pois em todas as outras o padrão seguido foi o da Binomial Negativa. A dedução que se pode derivar é a de que o mecanismo que governa a distribuição de cidades somente segue a linha dos centros regionais em São Paulo (quer dizer obedece aos mecanismos de mercado), o que pode significar elevado grau de subestimação de cidades (numerosas das quais seriam centros 4b), ou que, em São Paulo, o mecanismo de criação de municípios segue realmente as mesmas linhas de racionalidade econômica que seguem os desenvolvimentos de centros sub-regionais e regionais.

Como a distribuição para o Centro-Sul reduzido é do tipo Binomial Negativo, o que inclui as cidades da área de influência de São Paulo, parece que a hipótese de maior racionalidade econômica é a mais viável; inclusive porque ela segue a tônica de outros estudos que identificam no Centro-Sul como um todo o Núcleo básico brasileiro, com uma área mais contraída constituindo o que chamamos o Núcleo propriamente dito, sendo o restante a sua periferia imediata. Finalmente a área de influência de São Paulo constituiria o "score" deste Núcleo, em três aproximações sucessivas, marcadas por três etapas do processo de desenvolvimento.

Simetricamente, para a Zona da Mata no Nordeste, o fato de se obter uma distribuição ajustada, tanto Poisson como Binomial Negativa, de centros 4b e Poisson para os centros 3b, também indica que aí a subdivisão evidencia dois estágios do processo de desenvolvimento. É o que, em outros estudos, temos chamado de Núcleo Secundário. Observe-se que ele é simétrico, de certa forma, em relação ao Centro-Sul reduzido, pois em ambos aceita-se a distribuição Poisson e a Binomial Negativa quando a média de pontos por quadricula é próxima de 1, em ambos os casos aceita-se a distribuição Poisson para os centros 3b e em ambos os casos rejeita-se a Poisson e aceita-se a Binomial Negativa para cidades. Só não é simétrico em relação a área de influência de São Paulo, consistente com o que temos observado em outros estudos de que a periferia imediata de São Paulo corresponde ao Núcleo Secundário do Nordeste, em termos de padrão de urbanização, mas o Núcleo básico não tem correspondente no Nordeste.

5 — CONCLUSÕES

O que vamos apresentar à guisa de Conclusões tem sobretudo três sentidos diferentes, embora complementares:

1) Alguns resultados obtidos, embora ainda de natureza preliminar, podem constituir uma contribuição ao estudo do sistema urbano brasileiro, principalmente num aspecto pouco explorado que é o mecanismo de difusão de inovações.

2) Algumas sugestões de pesquisas complementares, que podem dar novos “*insights*” em relação a diferenciações regionais no Brasil, especialmente naquilo que representem ângulos de análise diferentes dos até agora utilizados.

3) A clássica palavra de cautela em relação a modelos matemáticos, fascinantes pela sua própria natureza, mas susceptíveis de gerar mecanismos circulares de explicação (o modelo explicado pelo modelo), que afinal constitui, também, uma constatação da necessidade de se desenvolver teoria geográfica, único caminho que poderá colocar a Geografia no plano das outras Ciências Sociais já estabelecidas. Quando mencionamos este aspecto, não se contém nele nenhuma espécie de narcisismo profissional no sentido de que um determinado ramo do conhecimento humano adquira *status* científico, mas apenas um sentido objetivo de tornar este ramo do conhecimento um instrumento útil ao progresso social. E esta nos parece ser a tendência atual da Geografia.

Em relação ao item 1 (resultados obtidos), parece-nos que algumas idéias esboçadas em diferentes ocasiões sobre o sistema urbano brasileiro continuam obtendo confirmação por via de novos ângulos de análise, no caso presente, a difusão de inovações.

O processo de difusão de inovações — visto através do conteúdo implícito de inovações contido na distribuição de centros de nível 4b como o menor núcleo de comando regional — pode ser estudado de forma adequada, mapeando-se sua distribuição espacial em uma função geradora de uma distribuição de propabilidades. As distribuições Poisson e Binomial Negativa são muito apropriadas a este tipo de estudos, pois descrevem fenômenos que têm nítidas aplicações no campo da análise espacial, particularmente em termos do mecanismo da difusão de inovações.

Algumas das premissas adotadas (mais de forma implícita que explícita) precisariam ser testadas de forma mais concreta ainda; uma delas é a de que a área média do campo de ação de cada centro 4b seja da ordem de 6 mil km² no Nordeste e 3 mil no Centro-Sul e que tal área possa ser interpretada como o campo médio de informação de cada centro.

Entretanto o simples fato de que para o Nordeste, por exemplo, uma distribuição do tipo Binomial Negativa é aceita para quadrículas com estas áreas e superiores e a distribuição Poisson não é aceita, pode bem constituir evidência significativa de que o processo de localização espacial de centros 4b siga um modelo do tipo contágio no Nordeste, típico de uma região com rede urbana desconectada, portanto com maior grau de concentração.

Observe-se que, nas quadrículas originais, em ambas as regiões se obteve um ajustamento às duas distribuições (Binomial Negativa e Poisson), forçada esta última como distribuição limite, por via de um tamanho de quadrícula bastante pequeno. Entretanto, ao aumentar-se

o tamanho das quadriculas (duas para o Centro-Sul e quatro para o Nordeste), e daí por diante, rejeita-se a Poisson e aceita-se a Binomial Negativa, com ajustes mais ou menos precários, conforme o caso, como descrevemos nos capítulos próprios. Apenas o parâmetro K para a região Centro-Sul foi bem superior, indicando uma taxa de difusão mais elevada na mesma, reiterando resultados de estudos até agora realizados, seguindo linhas metodológicas diferentes.

Esta diferença entre o Centro-Sul e Nordeste diminui ao analisarmos apenas a região reduzida, nos dois casos. Tomando-se o Centro-Sul reduzido verifica-se a aceitação do modelo Poisson nas três agregações feitas, com índices de dispersão muito próximos de 1 e com média e variância proporcionais aos diferentes tamanhos de quadriculas. É importante salientar que, no Centro-Sul reduzido, somente na segunda agregação são aceitas as duas distribuições, e que, na quadricula original, nem ao menos foi possível o cálculo dos parâmetros da Binomial Negativa, o que indicaria uma certa tendência para um modelo Poisson, embora com algumas características de uma distribuição composta, que evidenciam um certo grau de heterogeneidade intra-regional atuando no processo de difusão. Este mesmo fenômeno se observa na Zona da Mata, pois na quadricula original ocorreu a mesma impossibilidade de se calcular o parâmetro para a Binomial Negativa; ao passo que na primeira agregação ambas as distribuições são aceitas, com índice de dispersão próximo de 1, como ocorreu na segunda agregação do Centro-Sul reduzido. Esta conformidade com o modelo Poisson para a Zona da Mata, diferentemente do que ocorreu com a região como um todo, é corroborada com a análise de contigüidade feita para o Nordeste. Conforme foi assinalado, observa-se um efeito de proximidade em ambas as áreas do Nordeste (Mata e Sertão), mas no Sertão este efeito é direcionado norte-sul (embora não conforme com as expectativas), enquanto na Zona da Mata ele é direcionado tanto no sentido norte-sul como no leste-oeste; aí temos uma indicação da maior intensidade do processo de difusão que tende a uma forma concêntrica característica de um estágio mais adiantado.

Não foi sequer necessário fazer o teste do modelo a nível de centros 4b para a área de influência de São Paulo, que seria obviamente Poisson (inclusive porque o teste foi feito para cidades e o modelo ajustado foi Poisson).

Seguindo ainda a linha de diferenciações regionais, por via da distribuição de centros de hierarquias diferentes, foi examinada a distribuição dos Centros 3b, tanto para a Zona da Mata, como para o Centro-Sul reduzido e, ainda, para a área de influência de São Paulo. É importante salientar, desde logo, que na Zona da Mata, somente na terceira agregação (1 grau quadrado) alcançou-se a média de 1 ponto por quadricula, com variância bastante menor (tanto que o índice de dispersão é pouco superior a 0,5). O modelo Poisson, entretanto, foi aceito para todos os níveis de agregação, embora o ajuste no último nível tenha sido menos nítido que no segundo. Convém apenas ressaltar que tanto o número de centros como de quadriculas é relativamente pequeno, o que pode ter acarretado distorções na medida.

Tomando-se a região Centro-Sul reduzida, verifica-se que o padrão é semelhante, pois aceita-se o modelo Poisson em todos os níveis. A variância, no entanto, é bem maior, tanto que o índice de dispersão é pouco superior a 0,9 nas três agregações. Fenômeno semelhante ocorre com a área de influência de São Paulo, com a diferença que a média 1 ponto por quadricula é alcançada na terceira agregação que corresponde a meio grau (duas vezes inferior ao da Zona da Mata) e cerca

de 1,5 vezes inferior ao do Centro-Sul reduzido. Não é sem razão que o fator crescimento de São Paulo é bem superior ao de todos os centros metropolitanos, em todos os níveis hierárquicos.

Um outro aspecto significativo a ressaltar e mencionado com um dos objetivos iniciais do estudo que apresentamos, foi o de verificar em que medida os processos que governam a distribuição de centros 4b é diferente, numa e noutra região dos que governam a distribuição de cidades.

Ainda aí a diferença entre Centro-Sul reduzido e Zona da Mata aparece muito pequena, uma vez que em ambos os casos o ajuste aceito foi o da Binomial Negativa, porém com um número de cidades por quadrícula bem maior no Nordeste, indicando fracionamento de municípios muito mais acentuado (a média na quadrícula original é de 1,6 contra 1,1 na Zona da Mata e Centro-Sul reduzido respectivamente). A variação do parâmetro K nas duas áreas, a nível de quadrícula original, também reitera a maior intensidade do processo de difusão no Centro-Sul.

Entretanto é na região de influência de São Paulo que se observa a diferenciação do processo, em relação às outras áreas; ao lado disso revela a consistência do processo gerador na área de São Paulo, pois tanto os centros 3b como cidades obedecem ao mesmo padrão espacial, descrito por uma distribuição Poisson. O que esta coincidência de padrões de distribuição espacial indica é uma certa uniformidade do processo de difusão na área de São Paulo, que segue todos os níveis hierárquicos (não foi testado o nível 2 pois o número de centros é muito pequeno para que a análise seja significativa), mas o fato de que fator crescimento do centro metropolitano paulista para os de nível 2 ser também o maior dentre todos, ilustra bem esta diferença entre o subsistema São Paulo e os outros.

Quando procuramos investigar esta diferença de padrões entre os centros de nível 4b e outros e o padrão de cidades, estávamos tentando identificar diferenças essenciais entre as duas, as primeiras geradas pelos mecanismos de mercado e a outra por um conjunto de fatores de ordem político-administrativa. A análise da tabela 3, por sua vez, mostra bem o descompasso que existe ao nível das cidades centros 3/4 com as de nível 1/2, revelando um enorme hiato entre o nível metropolitano e o de centros regionais e subregionais. Este padrão é típico de áreas subdesenvolvidas de ocupação relativamente antiga, em que uma metrópole com função polarizadora (no sentido que Berry emprega esta palavra), liga-se diretamente a todo um *hinterland*, sem o surgimento de centros intermediários (ou com poucos centros) entre uma ponta e outra do sistema. Observe-se que Recife, por exemplo, tem mais de 60 cidades diretamente subordinadas (entre as quais até dois centros 3a), embora quase todas dividam esta subordinação com outro centro. O exemplo mais típico é destes dois centros 3a indicados: um deles é Propriá, subordinado também a Aracaju, na área de Salvador; o outro é Picos, subordinado também a Teresina, na área de Fortaleza. Enquanto isso São Paulo tem apenas doze cidades diretamente subordinadas, todas elas pertencentes à região metropolitana de São Paulo, portanto num raio de menos de 50 quilômetros de distância.

Não é por razão diferente que o fator crescimento na área de São Paulo passa de quase 59 para 96 (menos do dobro) do nível 1/2 para o nível 2/3, enquanto Recife passa de quase 15 para 51 (mais de três vezes superior), reiterando este descompasso muito menor na primeira, entre o nível metropolitano e o imediatamente abaixo. O fato de ter o multiplicador urbano aumentado, no caso de Recife, do nível 1/2 para

o 2/3, muito mais que o de São Paulo, está obviamente associado à classificação de capitais regionais/estaduais da área de Recife, do tipo Maceió, Natal, João Pessoa e mesmo Campina Grande no nível 2 e dependentes de Recife, quando seu grau de autonomia é muito maior, principalmente no caso das capitais estaduais.

Quanto ao item 2, uma das linhas sugeridas de pesquisa, que se pode talvez analisar, implica em uma utilização heterodoxa do modelo; tem sido ressaltado por diversos autores (e muitas vezes o problema e os autores foram citados ao longo do presente estudo), que uma determinada distribuição como a Binomial Negativa ou mesmo a Poisson, pode resultar de diferentes distribuições agregadas, que invalidam a aplicação do modelo, o que é obviamente correto e metodologicamente ortodoxo. Entretanto, parece-nos que quando tomamos, por exemplo, a análise feita para todo o Centro-Sul temos uma distribuição ajustada ao modelo Binomial Negativo, com parâmetros K decrescentes; quando reduzimos a região, aumentando-lhe a homogeneidade, conforme assinalamos na análise feita no capítulo próprio, o ajuste ao nível da quadrícula original nem sequer permite o cálculo dos parâmetros da Binomial Negativa, mas permite até o cálculo de uma Poisson modificada, que descreve uma distribuição mais sistemática que aleatória; estes dois resultados podem ser comparados entre si, pois as duas áreas podem constituir dois níveis de heterogeneidade diferentes, um maior e outro menor, identificados pelo tipo de ajuste a que melhor se adaptam e por parâmetros que se comportam de maneira diferente.

O que estamos tentando demonstrar com isso é que a medida de heterogeneidade no sistema é, a rigor, mais importante que a definição de homogeneidade, pois é a primeira que estamos procurando medir. A utilização pouco ortodoxa do modelo consiste em aproveitar seus parâmetros não só para medir a validade genética do ajuste, mas descrever diferentes níveis de diferenciação regional, em cada um dos quais o processo de difusão de inovações teria intensidade diferentes. É claro que tal linha metodológica precisa ser utilizada com cuidado, associado a outros estudos do sistema urbano, realizada por via de numerosos testes em tamanhos de quadrículas diferentes, e utilizando limites espaciais para a análise da distribuição de pontos que tenham um mínimo de validade previamente testada. É o caso de tomarmos a região Centro-Sul desenvolvida (somente a periferia imediata ao Núcleo associada ao Núcleo), ou somente a parte mais próxima ao litoral do Nordeste, e Nordeste como um todo, comparando situações intra e inter-regionais.

As análises aqui feitas, de cidades, centros 4b e 3b, partiram do pressuposto de que os mecanismos que geram a distribuição de núcleos de uma determinada hierarquia podem ser diferentes quando se analisa uma outra hierarquia. Na realidade esta seria uma constatação fundamental, pois qual seria o significado de uma distribuição mais sistemática que aleatória em São Paulo (na realidade a distribuição aceita foi a Poisson, mas tendendo para mais sistemática que aleatória, como indicamos), tanto de cidades como centros 4 e 3b. E o fato de que somente a nível de centros 4b ela é Binomial Negativa para a região como um todo e Poisson para a área reduzida e área de influência de São Paulo, com parâmetros K crescentes pode ser interpretado como três estágios, como indicamos?

Como estamos utilizando os modelos aqui apresentados de forma um tanto heterodoxa, pois estamos aplicando os mesmos a áreas bastante grandes e obviamente heterogêneas, violando um preceito básico de homogeneidade da amostra, as interpretações precisam ser feitas

com cautela. O seu conteúdo descritivo é óbvio, mas ele ganha foros de explanatório quando é associado a estudos outros, não só já divulgados anteriormente e indicados nos valores da tabela 1, como os referentes ao multiplicador urbano e fator crescimento, que oferecem um importante subsídio para a interpretação e mensuração dos desequilíbrios na rede urbana.

O objetivo da divulgação de resultados, que ainda comportariam pesquisas subseqüentes para sua maior avaliação, visa principalmente receber a crítica de outros pesquisadores na mesma linha; é claro que visa também, como numerosos outros estudos do sistema urbano que temos divulgado, introduzir metodologias talvez pouco divulgadas no contexto de análises espaciais, particularmente no Brasil.

Por fim uma palavra de cautela.

Muitas das dificuldades inferenciais com a utilização de modelos matemáticos do tipo distribuições teóricas de probabilidades de ocorrência de eventos (como centros 4b), relacionam-se ao fato de que teorias de localização urbana ainda estão incompletamente especificadas. Harvey chama a atenção para este aspecto ao indicar que “a lição a se tirar desta situação é a de que nossa habilidade em usar modelos matemáticos está ultrapassando a construção de teoria geográfica. Há um mundo de diferença entre a simples utilização (ou “consumo”) do modelo e a construção de poderosas teorias na pesquisa geográfica. O uso de modelos matemáticos é indubitavelmente um instrumento poderoso, desde que os critérios que orientam sua aplicação possam ser estabelecidos e postos em operação”¹⁸

A utilização de tais modelos, cercada dos cuidados necessários à identificação do problema que estamos analisando, com as premissas do modelo especificadas de forma clara, constitui hoje o elemento de maior vigor científico na atual fase do desenvolvimento da Geografia, em particular, e das Ciências Sociais em geral.

BIBLIOGRAFIA

1. IBGE. Departamento de Geografia — *Divisão do Brasil em Regiões Funcionais Urbanas*, Rio de Janeiro, 1972.
2. DACEY, M. F. — “Modified Poisson probability law for point pattern more regular than random”, in *An. Ass. Am. Geogr.* 54, 559/565.
3. FAISSOL, S. — “Tipologia de cidades e regionalização do desenvolvimento econômico”, *Boletim Geográfico*, n.º 223 e outros do mesmo autor.
4. CHRISTALLER, W. — *Central Places in Southern Germany*, Trad. C. W. Baskin, Ed. Prentice-Hall. 1966.
5. THOMAS, E. N. — “Toward an Expanded Central Place Model”, in *Geogr. Rev.*, jul. 1961, além de numerosos outros como Morrill, Olsson etc., que estudaram localidades centrais como associadas a um processo estocástico.
6. CHORLEY, R. J., e Haggett, P. (eds.) — *Models in Geography*, Londres, Methuen & Co., 1967.
7. BERRY, B. J. L. — “City Size and Economic Development”, in *Urbanization and National Development*, Vol. I, ed. Jakobson, L. e Prakash, J., Sage Publications, 1971.
8. DACEY, M. F. — “Imperfections of the Uniform Plane”, in *Discussion Paper*, Michigan Inter-University Community of Mathematical Geographers, n.º 4.
9. DACEY, M. F. — “A County-seat model for the areal pattern of an urban system”, *Geogr. Rev.*, 56, 527-542.
10. HARVEY, D. — “Pattern, Process and the Scale Problem in Geographical Research”, in *Trans. Inst. Br. Geogr.*, 45, 71-78.
11. DACEY, M. F. — “An Empirical Study of the Areal Distribution of Houses in Puerto Rico”, in *Trans. Inst. Br. Geogr.*, 45, 51-69.
12. CLIFF, A. D. — “The neighbourhood effect in the diffusion of innovations”, in *Trans. Inst. Br. Geogr.*, 44, 75-84.
13. ANSCOMBE, F. J. — “The statistical analysis of insect counts on the negative binomial distribution”, in *Biometrics*, 5, 165-173.
14. HARVEY, D. — “Geographical processes and the analysis of point patterns: testing models of diffusion by quadrat sampling”, *Trans. Inst. Br. Geogr.*, 1966, 81-95.
15. ROGERS, A. — “A stochastic analysis of the spatial clustering of retail establishments”, in *Journ. An. Stat. Ass.* 60, 1094-1103.
16. OLSSON, G. — “Central Place Systems, Spatial Interaction and Stochastic Processes”, in *Urban Economics, Theory, Development and Planning*, The Free Press, 1970.

17. GETIS, A. — "Occupancy Theory and map pattern analysis", Seminar Paper Series A, Dep. Geogr., University of Bristol.
18. HARVEY, D. — "Some methodological problems in the use of the Neyman type A and the Negative Binomial Probability Distribution for the analysis of spatial point patterns" in *Trans. Inst. Br. Geogr.*, 44, 85-95.
19. FELLER, W. — "On a general class of contagious distributions", in *Annals of Mathematical-Statistics*, 14, 389-400.
20. HARVEY, D. — *Explanation in Geography*, Londres, Edward Arnold (Pub.), 1969.
21. KUZNETS, S. — *Modern Economic Growth*, New Haven, Yale Univ. Press, 1966, cap. 1.
22. FRIEDMANN, J. — "A generalized theory of polarized development", in *Growth Centers in Regional Economic Development*, (Hansen, ed.), The Free Press, 1972, 82-107.
23. BERRY, B. J. L. — "Hierarchical Diffusion: The Basis of Development Filtering and Spread in a System of Growth Centers", in *Growth Centers in Regional Economic Development*. (Hansen, ed.). The Free Press, 1972, 108-138.
24. PEDERSEN, P. O. — "Innovation Diffusion within and between National Urban Systems", *Geographical Analysis*, jul. 1970, 203-254.
25. BERRY, B. J. L. — "A Synthesis of Formal and Functional Regions Using a General Field Theory of Spatial Behavior", in *Spatial Analysis — A Reader in Statistical Geography* (Berry e Marble, eds.), New Jersey, Prentice-Hall, 1968.

SUMMARY

The author, in his work, aims at demonstrating the utilization of the technique of grids, connected with the probabilistic distribution models of Poisson and the Negative Binomial, adopting for application the study of the spatial distribution patterns of the urban centers. This technique originates with the division of the field under study in grids covering this same area, which engenders a distribution of frequencies of the urban centers per grid. This distribution is then adjusted to the theoretic model, in this case, Poisson or the Negative Binomial.

The adjustment to a Poisson model signifies that there is no factor operating with sufficient force to engender a definite spatial pattern. An adjustment to the Negative Binomial, on the other hand, reveals a spatial concentration of the urban centers. Bearing in mind that what is to be arrived at through application of these models is the determination of what process is at work in the emergence of the centers, a contiguity analysis was effected, in addition, for the purpose of verifying whether this implies the existence of a process of diffusion and whether this latter possesses any specific directionality.

For practical application of these techniques, two areas were chosen, both covering approximately the whole of the Northeast and the Center-South. These two initial areas were later on subdivided, the Northeast into "Sertão" (drought-ridden hinterland) and "Zona da Mata" (vegetation-rich zone) — also called the Reduced Northeast — and the Center-South into two more developed areas, one of which with farther reach, which was called the Reduced Center-South, and the other corresponding more closely to the boundaries of the São Paulo Area of Influence.

The analysis was realized at two levels: for the whole of the urban system, taking into account the totality of the seats of the municípios (cities), and besides, only for the centers grown by influence of the market forces, just as they were classified in the "Division of Brazil into Functional Urban Regions" IBGE Foundation.

It became apparent that, as far as the Northeast is concerned, the model which describes the spatial pattern is that of the Negative Binomial, just as much applying to the centers as to the cities. Taking into consideration the Reduced Northeast only, the distribution of cities and 4b level centers continues to follow the same model. The 3b level centers which are hierarchically superior, however, offer a pattern described by the Poisson distribution.

In the Centre-South, the 4b centers adjust themselves to a Negative Binomial distribution, but in the reduced area this pattern is substituted by a Poisson model. The cities tend to follow this model in the more developed area, that is to say, in the São Paulo Area of Influence. The 3b level centers were only studied in the two reduced areas and there, the pattern they show is typically Poisson.

The results of these analyses once again confirm that, the more a region is developed, the less subject to splitting into nuclei is the distribution of its urban centers and the better this latter is balanced.

Versão de Patrice Charles F. X. Wullaume

RESUMÉ

Dans ce travail, l'auteur se propose de démontrer l'utilisation de la technique des quadrillages, associée aux modèles probabilistes de distribution Poisson et le Binôme Négatif, prenant comme application les patrons de distribution spatiale des centres urbains. Cette technique part de la division du champ d'études en quadrillages de cette même région, ce qui engendre une distribution de fréquences des centres urbains par quadrillage. La dite distribution est alors ajustée à un modèle théorique, dans le cas, celui de Poisson ou le Binôme Négatif.

L'ajustement à un modèle Poisson signifie qu'aucun facteur agit avec une force suffisante pour engendrer un patron spatial définitif. Un ajustement à la distribution Binôme Négatif, d'un autre côté, révèle une concentration spatiale des centres urbains. En prenant en considération le fait que, ce qu'on cherche obtenir par l'application de ces modèles est de déterminer quel est le processus en cours agissant pour la création des centres, une analyse de contiguïté fut accomplie, par surcroît, afin de vérifier si elle était conditionnée à l'existence du processus de diffusion et si ce dernier possède un sens directionnel spécifique.

Pour procéder à une application pratique de ces techniques, deux régions furent choisies, comprenant plus ou moins tout le Nord-Est et le Centre-Sud. Ces deux zones initiales furent, par la suite, sous-divisées, celle du Nord-Est en "Sertão" (arrière-pays souvent atteint par la sécheresse) et en "Zona da Mata" (zone riche en végétation) — connue comme le Nord-Est Réduit — et puis le Centre-Sud en deux régions plus développées, l'une la plus absorbante que fut appelée le Centre-Sud Réduit, et l'autre correspondant de plus près aux limites de la zone d'influence de São Paulo.

L'analyse fut fait à deux niveaux: pour tout le système urbain, tous les sièges des "municipes" (les villes) dûment considérés et, en plus, seulement pour les centres produits par les forces du marché, tels qu'ils furent classifiés dans "La Division du Brésil en Régions Fonctionnelles Urbaines" — Fondation IBGE.

On constata que, pour le Nord-Est, le modèle qui décrit le patron spatial est celui du Binôme Négatif aussi bien pour les centres que pour les villes — En prenant seulement en considération le Nord-Est Réduit, la distribution des villes et des centres du niveau 4b continue à obéir au même modèle. Les centres niveau 3b, hiérarchiquement supérieurs, cependant, présentent un patron décrit par la distribution Poisson.

Dans le Centre-Sud, les centres 4b s'ajustent à une distribution Binôme Négatif mais, dans la région réduite, ce patron est substitué par un modèle Poisson. Les villes penchent vers ce modèle seulement dans la région plus développée, c'est à dire, dans la Zone d'Influence de São Paulo. Les centres du niveau 3b furent seulement étudiés dans les deux régions réduites et là leur patron est nettement du type Poisson.

Le résultat de ces analyses confirme, encore une fois, que plus une région est développée, moins nucléée et plus équilibrée est la distribution de ses centres urbains.

Versão de Patrice Charles F. X. Guillaume