

2910



# Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS

## ANALISIS DE DOS MODELOS ESTADISTICOS ESPACIALES

### T E S I S

Que para obtener el Grado de  
A C T U A R I O  
P r e s e n t a

MARIO JOSE FRANCISCO CORTINA BORJA

México, D. F.

1984



Universidad Nacional  
Autónoma de México



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## INDICE

Introducción	1
I.1) Estadísticas de distribuciones espaciales	5
I.2) Interpretaciones de la aleatoriedad	10
I.3) Estadísticas espaciales en arqueología	13
II.1) Modelos de vecino más cercano	18
II.2) Modelo original de Clark y Evans	25
II.3) Generalización para n vecinos más cercanos	38
II.4) Generalización para k dimensiones del método para el primer vecino más cercano	46
II.5) Generalización para k dimensiones del método para n vecinos más cercanos	49
II.6) Método de la mediana	92
III.1) Índice A	95
III.2) Aproximaciones a la distribución de A bajo algunas hipótesis	108
Apéndice I	117
Apéndice II	126
Apéndice III	130
Apéndice IV	133
Agradecimientos	158
Referencias	160

## INTRODUCCIÓN

El objetivo de este trabajo es elaborar un estudio acerca de algunos problemas estadísticos que surgen del estudio de distribuciones de puntos en el espacio. Es una tesis de tipo más bien técnico que teórico, y por lo tanto, se centra en problemas de procedimiento tratando los problemas conceptuales muy superficialmente.

Aunque desde principios de este siglo se intentó dar una solución matemática al problema de hacer evidente la aleatoriedad (definida de acuerdo a algún criterio simple) o no aleatoriedad (es decir, existencia de algún comportamiento sistemático) de un patrón espacial de puntos, es hasta la década de los cincuenta que se construyen modelos estadísticos para resolver este problema que no se basen en conteos de ocurrencias de eventos sobre cuadrantes limitados arbitrariamente.

Estos primeros modelos fueron desarrollados especialmente por botánicos que buscaban describir poblaciones de plantas sobre regiones perfectamente limitadas. Como muestran Bartlett (1975) y Ripley (1982) el grado de diversificación, tanto en aplicaciones como en complejidad matemática que han alcanzado estos modelos a partir de la década de los setenta es muy grande. En particular, los modelos de vecino más cercano han sido utilizados en geografía (Dacey (1964), Garner (1967), King (1968), Harvey (1973)), ecología de animales (Hansen y Remmenga (1961)) y arqueología (Hodder (1972), Whallon (1974), Washburn (1974)). Otras clases de modelos espaciales se utilizan en geología, (Davis, (1973)), epidemiología (Bartlett (1975)) y física (Lewis (1976)). Todo esto justifica un estudio de los principios de estos modelos para

obtener, posteriormente, un mejor uso de ellos, tomando en cuenta las restricciones y ventajas que producen sus supuestos y consecuencias.

Esta tesis intenta ser, por lo tanto, más bien una iniciación al estudio de los modelos estadísticos espaciales que un estudio exhaustivo de ellos. Por una parte, ya existen trabajos que cubren de una manera general el desarrollo de una gran cantidad de modelos (por ejemplo Bartlett (1975), Hodder y Orton (1976) y Ripley (1982)), además de que no dispongo de suficiente tiempo para esto; por otro lado, la disponibilidad del material bibliográfico es otra limitante real. Es por esto que la tesis se centra en el análisis de dos clases de modelos: los de vecino más cercano y un índice de asociación espacial para dos distribuciones propuesto por Hodder y Okell en 1978. Para el primer tipo de modelos se presenta el desarrollo del modelo original y de varias generalizaciones (entre ellas una original) con un detalle tal que no permite su exposición completa en los artículos originales, presentando algunos ejemplos. Para la segunda clase se presenta el desarrollo del índice así como aproximaciones a su distribución obtenidas por métodos Monte Carlo.

En resumen, la tesis de este trabajo es que ambas clases de modelos resultan parciales, y es necesario construir modelos que involucren más factores para decidir la aleatoriedad de un patrón espacial; es decir, la aleatoriedad señalada por estos modelos se referirá únicamente a la distribución de alguna clase de distancias, sin considerar elementos de otra naturaleza y que podrían influir en la formación de la distribución. Un intento de solución se presenta en la generalización del modelo de Clark y Evans presentada por primera vez en este trabajo, pues permite incluir tantas variables como se desee, aunque limita la relación entre estas variables a algún tipo de distancias (normalmente distancias en forma de métrica de Minkowski).

La tesis consta de tres capítulos y cuatro apéndices. A continuación se describe a cada uno de ellos.

En el primer capítulo discuto el concepto de proceso espacial, algunas interpretaciones de la aleatoriedad en patrones espaciales y el uso que se ha hecho de ellos en arqueología. Se analizan también los supuestos en los que se basan los modelos y los distintos usos que se puede hacer de ellos en general, mostrando, rápidamente, algunos problemas metodológicos que tiene su aplicación.

El objetivo del segundo capítulo es analizar, lo más detalladamente posible, los modelos de vecino más cercano, primero en la forma original desarrollada por Clark y Evans en 1954, y posteriormente, con algunas de las extensiones y modificaciones más notables, como la presentada por Thompson en 1956, que permite considerar los primeros  $n$  vecinos más cercanos, y la hecha por Clark y Evans en 1979 para generalizar el modelo (de primer vecino más cercano) a espacios de dimensión  $k$ . A continuación expongo una nueva generalización del modelo para  $n$  vecinos más cercanos en espacios de dimensión  $k$ , presentando algunos ejemplos que muestran el comportamiento de las estadísticas propuestas para varias distribuciones. Finalmente se comentan el método de la mediana dado por Washburn en 1974 y una extensión del modelo para espacios de dimensión 1 hecha por Stark y Young en 1981.

Por último, en el tercer capítulo examino un modelo completamente distinto a los modelos de vecino más cercano, en cuanto a sus supuestos, contrastes de hipótesis y construcción de la distribución de las estadísticas de prueba y que aún está en una fase de experimentación; el índice  $A$  de Hodder y Okell. Se presenta una descripción del comportamiento del índice con distintas distribuciones espaciales, de los algoritmos usados para obtener su función de distribución bajo hipótesis de aleatoriedad y algunas tablas de los fractiles inferior y superior de la distribución con algunas

hipótesis nulas para distintos valores del nivel de significancia, así como un análisis de los resultados obtenidos.

En el primer apéndice presento dos formas de construir la función de probabilidad de una variable aleatoria generada por un proceso de Poisson; luego doy la demostración de la fórmula para calcular el volumen de una esfera de dimensión  $k$  empleada en el capítulo II; en el tercer apéndice expongo el método utilizado para generar variables aleatorias con distribución Poisson y en el último apéndice se incluyen las tablas con los cuantiles de la distribución del índice  $A$  de Hodder y Okell.

## I.1) ESTADÍSTICAS DE DISTRIBUCIONES ESPACIALES.

Un problema central en la geografía humana es el estudio de los patrones espaciales de objetos localizados en mapas.

Una vez que se conoce el patrón espacial, resulta natural preguntarse qué fue lo que lo causó.

Es posible deducir hipótesis respecto un patrón particular a partir de teorías físicas, económicas, políticas o sociales. Claramente, los fenómenos explicados por estas teorías tienen manifestaciones espaciales, y a estos aspectos se les llama procesos espaciales. (Getis y Boots, 1978, p.1).

Abler, Adams y Gould amplían este punto de vista cuando afirman que "es nuestra limitada perspectiva del tiempo lo que hace que designemos algunos fenómenos espacialmente diferenciados como distribuciones de objetos, mientras que designamos a otras distribuciones que varían más rápidamente como procesos... *distribución espacial* es un término que aplicamos a la organización espacial interna de esas distribuciones de elementos del proceso. Considerada correctamente, la estructura espacial de una distribución es vista como un índice del estado actual de un proceso en acción (ongoing process)". (Abler, Adams & Gould, 1971, pp.60-61).

De lo anterior se desprende que dado un patrón espacial es posible relacionarlo, mediante algún modelo, con un proceso espacial teórico que explique la naturaleza del patrón observado. (Getis & Boots, 1978, p.1).

En este trabajo utilizo los términos *patrón espacial*, *estructura espacial*, y *distribución espacial* como sinónimos,

siempre y cuando el último no pueda confundirse con la acepción probabilística del término *distribución*.

Como señala Harvey, hay dos formas de aplicar modelos probabilísticos en geografía: i) con los datos geográficos, en un nivel descriptivo o inferencial, y ii) definiendo funciones de densidad de probabilidad (o de probabilidad para casos discretos) respecto a los procesos que generaron esos datos. (Harvey, 1973, p.266).

La primera forma ha sido desarrollada con detalle en libros como King (1969) y Hammond & McCullagh (1978). Los modelos estudiados en este trabajo se referirán a la segunda forma dada por Harvey.

Es posible afirmar que, en general, los procesos espaciales producirán distribuciones que oscilan entre dos extremos: aglomeración y dispersión regular de los elementos. (Hodder & Orton, 1976, p.31), (Davis, 1973, p.301). Los problemas principales que las estadísticas de distribuciones espaciales tratan de resolver son: i) clasificar el patrón observado entre estos dos extremos, y ii) dar elementos que permitan contrastar hipótesis estadísticas acerca del proceso que causó al patrón observado.

En general se construirán modelos que reúnan y articulen las implicaciones y relaciones de algún proceso teórico. Estos modelos servirán como conexión entre la teoría y la distribución observada para poder determinar (con alguna función de probabilidad) qué tan probable es que el patrón haya sido generado por un proceso como el que se utilizó para construir el modelo. De

esta forma se solucionan ambos problemas, pues los modelos producirán, por una parte, algunas estadísticas cuyo valor indique la situación del patrón en el continuo con extremos en situaciones de máxima agrupación (i.e., todos los elementos ocupan un solo punto) y de máxima regularidad (i.e., una distribución hexagonal; más adelante se justificará esta afirmación), y por otra, incluirán pruebas estadísticas de hipótesis acerca de la naturaleza del proceso generador.

Hay buenas razones para suponer que estos modelos serán de naturaleza estocástica: si bien "todos los patrones de asentamiento suponen un orden medible en la conducta espacial" (Garner, 1971, p.311), este orden no es necesariamente determinista, sino, más bien, puede hablarse de un orden estadístico en el que es posible predecir situaciones con un cierto error, debido a la variabilidad intrínseca del fenómeno estudiado. Claramente, es más adecuado considerar este último tipo de orden para tratar con patrones de puntos producidos por actividades humanas, pues éstos varían debido a una gran cantidad de factores, cuya acción conjunta sobre el estado del proceso que se observa es muchas veces imposible de determinar absolutamente. Como se verá más adelante, los modelos estudiados simplifican esta situación.

Todos los modelos de distribuciones espaciales parten de la suposición de que el patrón es aleatorio (Hodder & Orton, 1976, p.31, citando a Ord (1972)). En mi opinión esto se debe a que es muy simple construir modelos que describan estructuras espaciales generadas por algún tipo de proceso aleatorio. Vale la pena citar

a Dacey: "En términos de patrones en mapas, aleatoriedad pura significa que cada locación del mapa tiene una misma probabilidad de recibir un símbolo (elemento del proceso). Puesto que es muy poco probable que distribuciones geográficas, particularmente patrones locacionales que involucran decisiones humanas, sean el resultado de eventos igualmente probables, se esperaría que la mayoría de los patrones en mapas reflejen algún sistema u orden. Es por esta razón que los patrones espaciales son examinados en busca de evidencias de procesos espaciales." (Dacey, 1964, p.559).

Es posible, sin embargo, pensar en algunas situaciones en que sea factible considerar a un proceso aleatorio como generador del patrón observado. Este problema se analizará más adelante.

Hammond y McCullagh (1978, p.48) señalan que hay dos enfoques para analizar eventos locacionales: el primero consiste en identificar intuitivamente características de las distribuciones y describirlas de una manera más bien verbal; el otro utiliza medidas objetivas de algunas características del patrón (por ejemplo densidad, distancias entre puntos, número de eventos por cuadrante, etc.) y posteriormente las presenta resumidas en conjuntos de estadísticas, muchas veces incluyendo sus funciones de probabilidad. Estos autores dan tres atributos ("no necesariamente ventajas" op. cit., p.49) del segundo enfoque: i) añade precisión a descripciones cualitativas; ii) al ofrecer criterios objetivos facilita la comparación entre distintas distribuciones y iii) resaltan características (como por ejemplo la aleatoriedad) que son difíciles de notar con una simple inspección intuitiva.

En resumen, estos modelos simplifican (como todo modelo) la

situación real (i.e., que la varibilidad de una estructura espacial se debe a procesos que involucran una gran cantidad de factores) al medir únicamente qué tan agrupado o qué tan regularmente distribuido o en qué medida se parece a una distribución aleatoria el patrón observado. Esta medición se basa, por lo general, en una o dos características de la distribución (densidad, distancias al vecino más cercano...), lo que origina una segunda (y más fuerte) simplificación. Cualquier uso que de ellos se haga debería considerar estas dos observaciones.

## 1.2) INTERPRETACIONES DE LA ALEATORIEDAD.

Como se dijo, al tratar con patrones espaciales que han sido producidos por decisiones humanas resulta poco probable que tales decisiones (consideradas en conjunto) formen un proceso aleatorio.

Entonces, ¿qué significa el hecho de que la estadística obtenida con algún modelo indique que el patrón no difiere significativamente\* de un patrón generado por un proceso aleatorio? .

Es posible responder a esta pregunta de muchas formas:

Para King (1968, p. 166), "el concepto de aleatoriedad para patrones de asentamiento puede ser desechado, excepto por el hecho de que los valores para los cuales las estadísticas señalan aleatoriedad forman un origen conveniente y útil para medir tendencias hacia la agrupación o el espaciamiento regular de los elementos", una especie de frontera.

Getis y Boots (1978, pp.15-16) van más allá cuando afirman que

se abstienen de utilizar el calificativo de aleatorio para un patrón de asentamiento, pues se implicaría que si una distribución espacial no es ni agrupada ni regularmente espaciada es, automáticamente, considerada como aleatoria, aunque se esté de acuerdo en que el proceso que la generó no es aleatorio.

Según Dacey (1964, p.559), "decir que una distribución es aleatoria en un sentido no técnico es decir que el patrón no tiene un orden discernible y que su causa es indeterminable". Podría agregarse

\* La palabra "significativamente" está empleada aquí en el sentido estadístico; i.e., existe una probabilidad bastante pequeña de rechazar la hipótesis de que el patrón provenga de un proceso aleatorio cuando esto no sea cierto.

que ni el orden es discernible ni la causa es indeterminada cuando únicamente se utilizan una o dos características del patrón como medidas. Esto último lleva al concepto de economía espacial aleatoria de Curry (1964), quien afirma que si un proceso espacial es el resultado de un conjunto de decisiones, todas ellas racionales, la combinación de éstas puede producir un patrón similar a algunos patrones obtenidos de un proceso aleatorio dada la cantidad y la diversidad de los factores que afectan a estas decisiones.

Por lo anterior se puede ver que "sólo la no aleatoriedad da información" (Hodder & Orton, 1976, p.51), y que el hecho de haber identificado a un patrón como no aleatorio no significa que se tenga una explicación del proceso (ibid., citando a Stanislawki (1973)). Esta explicación deberá referirse, como se mencionó al principio de este capítulo, a alguna teoría, pues en este caso, la utilidad de los modelos es la identificación objetiva de tendencias existentes en la estructura espacial observada.

Hay que hacer notar que el problema de discernir cuando un patrón es o no aleatorio está inmerso en un problema muy general: la cuestión de si es posible o no considerar procesos cuyos resultados sean independientes entre sí. Hacking dice que "En la práctica es muy difícil encontrar experimentos independientes. La solución es tener un sistema estandar de probabilidades (standard chance set-up) y sujetarlo con la teoría del muestreo aleatorio. La estructura de esto la conforma la teoría de los números aleatorios, considerados como sucesiones de dígitos obtenidos suponiendo que cada dígito tiene una probabilidad igual de ocurrir, y que lo hace independientemente de los demás". (Hacking, 1965, p.129).

Esta discusión lleva directamente al problema de si la estadística tiene una base suficientemente firme en el enfoque frecuentista de la teoría de la probabilidad, lo que está fuera del objetivo de esta trabajo. Para un tratamiento más completo del problema, véase por ejemplo el primer capítulo del libro de Papoulis (1972).

Por mi parte, estoy de acuerdo con las observaciones de King y Getis y Boots en cuanto al uso del término "aleatorio" en referencia a patrones espaciales, pero esto no implica un rechazo a las posiciones de Dacey y Curry. Más bien, me parece que éstas últimas proporcionan una interpretación de la aleatoriedad coherente con la proposición general de que los patrones espaciales, producto de procesos que involucran decisiones humanas, no pueden ser aleatorios, al menos en el sentido "técnico" de la palabra, y aunque lo sean en un sentido "no técnico", se sabe que ésto se debe a las simplificaciones hechas por los modelos.

### 1.3) ESTADISTICAS ESPACIALES EN ARQUEOLOGIA.

Es notable que, a pesar de la gran importancia que las distribuciones espaciales tienen en arqueología (Clarke, 1977), sean bastante escasas las aplicaciones de estadísticas espaciales en esta disciplina. Esto podría deberse al escepticismo que respecta al uso de la estadística hay en una gran proporción de arqueólogos más que a las dificultades que existen para efectuar los cálculos o para interpretar arqueológicamente las conclusiones sugeridas por los modelos.

En particular me refiero únicamente a las aplicaciones del análisis de vecino más cercano desarrollado por Clark y Evans (1954), sus modificaciones (por ejemplo Thompson (1956), Washburn (1974), Clark y Evans (1979), Young (1980)) y del índice A de Hodder y Okell (1978). Esto último no quiere decir que estos modelos sean los únicos que existen o que se han aplicado a la arqueología; simplemente, éstos son los modelos estudiados en esta tesis. Para tener una idea más completa de la diversidad y complejidad matemática de modelos de procesos espaciales véanse los trabajos de Lewis (1972), Holgate (1972), Bartlett (1974, 1975, 1978), Cox y Lewis (1976), y Ripley (1977, 1979, 1982).

Los artículos con aplicaciones de modelos espaciales en arqueología que fueron consultados pueden dividirse en dos grupos. Por supuesto, estos dos grupos no son, de ninguna manera, exhaustivos. Por una parte están los trabajos que únicamente aplican las estadísticas y pruebas de hipótesis de los modelos para obtener conclusiones relativas a materiales arqueológicos. Ejemplos de esta categoría son los trabajos de Hodder (1972, 1977), Whallon

(1974), Chadwick (1978, 1979) y Sugiyura y Cortina (1983).

El otro grupo incluiría trabajos en los que, con mayor o menor acierto, se han propuesto modificaciones de orden estadístico para los modelos. Elementos de esta categoría son los trabajos de Washburn (1974), Donnelly (1978), Pinder, Shimada y Gregory (1979), Stark y Young (1980) y McNutt (1981).

Los trabajos de este último grupo tienen en común, además de la característica mencionada anteriormente, que han sido desarrollados conjuntamente por arqueólogos y matemáticos, o bien, por matemáticos que buscan adecuar los modelos a situaciones arqueológicas. En adelante, me referiré sólo a los trabajos del segundo grupo.

En mi opinión, lo más importante de estos artículos es el hecho de que evidencian una colaboración entre arqueólogos y matemáticos que, por parte de los primeros, va más allá de la simple aplicación de técnicas matemáticas que, a veces mágicamente, hacen rechazar o aceptar hipótesis, y que, por parte de los matemáticos los lleva a genera nuevos modelos o modificaciones a los ya existentes, más bien que a producir números (a veces en cantidades enormes) sin saber qué significan para los arqueólogos, si es que algo significan.

La calidad de los artículos es bastante alta, excepto por el de Pinder et al., que es innecesariamente complicado, y, peor aún, obtiene conclusiones equivocadas. Esto último tal vez se deba a la poca profundidad con que estos autores manejan la construcción del modelo. En este trabajo se intenta establecer una corrección a la expresión para la media esperada de las distancias al vecino más cercano bajo la hipótesis de aleatoriedad dada por Clark y Evans

(1954) usando simulaciones. El problema es que todo parece indicar que la fórmula de Clark y Evans es correcta analíticamente. McNutt (1981) ha refutado ya a Pinder et al., por lo que no repetiré sus argumentos.

Como se dijo en las secciones anteriores, los modelos espaciales analizados en esta tesis presentan algunas limitaciones bastante serias: a) parten del supuesto de que el proceso generador es aleatorio; b) sólo utilizan una o dos características del patrón en sus construcciones; c) como se verá más adelante, tienen problemas de escala, distancias y frontera.

En vista de esto es necesario preguntarse si estos modelos sirven de algo, en particular en la arqueología.

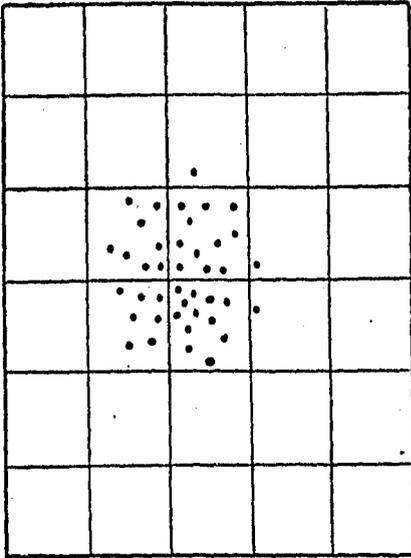
Debido a la poca dificultad matemática que involucra su construcción pueden verse como una introducción al estudio de modelos espaciales más complicados: tal es el caso de esta tesis.

En segundo lugar, las tres características dadas por Hammond y McCullagh (véase pag. 8) para los métodos estadísticos de análisis de mapas se cumplen, en particular, para estos modelos, lo que les confiere, en mi opinión, una clara ventaja sobre las descripciones verbales de distribuciones de materiales o sitios arqueológicos, al menos en el aspecto de facilidad de comparación con otras regiones.

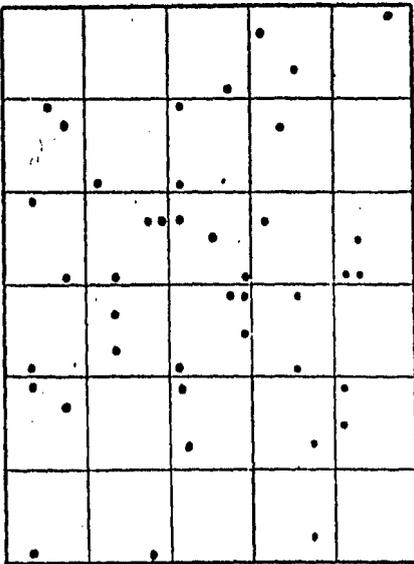
Por otra parte, la simplicidad, tanto de sus supuestos como de sus conclusiones los convierte en técnicas fácilmente utilizables por arqueólogos como un primer nivel de aplicación de modelos matemáticos en su trabajo. Si se acepta

entonces que las conclusiones producidas por los modelos tienen únicamente un valor descriptivo. Sin embargo, con distribuciones arqueológicas un valor de la estadística que indique aleatoriedad en la distribución podría interpretarse como que, o no se tiene registrada una parte significativa de los elementos del proceso, o que ha existido una eliminación de los mismos (por saqueo, expansión de poblaciones actuales, construcción de presas o carreteras, etc.), o bien, ambas cosas han ocurrido. (Modder y Orton, 1976, p. 54). Greig-Smith ha estudiado situaciones en las que, a partir de una distribución no aleatoria, eliminando algunos puntos se obtiene una distribución aleatoria. En resumen, una indicación de aleatoriedad por parte del modelo podría significar pérdida de información.

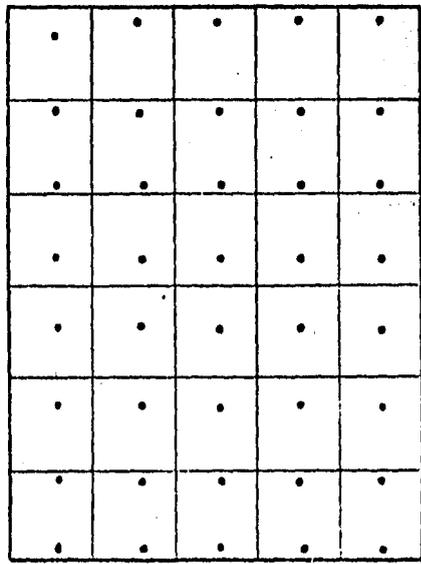
De lo anterior se obtienen dos conclusiones: a) Los problemas que surgen a raíz del análisis de distribuciones espaciales de materiales arqueológicos han generado una colaboración entre arqueólogos y matemáticos con resultados interesantes para las disciplinas de ambos y b) debido a las limitaciones propias de los modelos que parten del supuesto de aleatoriedad para el proceso generador de la distribución, es necesario construir modelos matemáticos a partir de teorías arqueológicas con implicaciones espaciales. Estos modelos deberán ser lo suficientemente simples de modo que puedan construirse y manipularse con las herramientas matemáticas disponibles; y lo suficientemente complejos para que las consecuencias de las teorías estén siendo representadas con suficiente realismo.



DISTRIBUCION CONCENTRADA



DISTRIBUCION ALEATORIA



DISTRIBUCION REGULAR

## II.1) MODELOS DE VECINO MAS CERCANO.

El análisis de vecino más cercano es un método que cuantifica las diferencias de un patrón de puntos en el espacio respecto a un patrón generado por un proceso aleatorio. Estas diferencias pueden tender hacia un patrón de puntos agrupado o bien hacia una distribución regularmente espaciada.

El modelo propuesto por Clark y Evans vino a resolver el problema de la descripción de distribuciones espaciales sin basarse en conteos de puntos en cuadrante. El mayor problema que tienen los métodos de cuadrantes es que, por lo general, el tamaño de cada cuadrante es elegido arbitrariamente, y como muestra Gómez López (1978), al variar el tamaño de los cuadrantes varía completamente el resultado de la prueba. La idea fundamental es comparar el número de puntos de cada cuadrante con el que se esperaría que hubiera si la distribución fuera aleatoria. Estas comparaciones se pueden efectuar usando alguna prueba convencional de bondad de ajuste. La alternativa para los métodos de cuadrantes son los llamados métodos de distancias, basados en alguna medida de las distancias del patrón, ya sea entre todos los puntos o bien referidas a un solo lugar.

Aunque existe una gran cantidad de modelos de distancias, (Pielou(1977)), sólo analizaré el propuesto por Clark y Evans en 1954, debido a su importancia, evidenciada por la gran cantidad de aplicaciones que de él se han hecho en campos tan variados como ecología (Pielou 1969, 1977), poblaciones animales (Hansen y Remmenga 1961, Miller y Stephen 1966), geografía humana (Dacey 1960, 1964; Getis 1964; King 1968, 1969; Getis y Boots 1978), geología (Miller 1971) y arqueología (ya citados).

La estadística propuesta por Clark y Evans (en adelante C&E) puede escribirse como

$$R = \frac{r_A}{r_E},$$

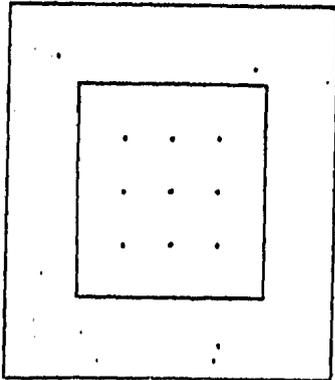
donde  $r_A$  es el promedio de las distancias de cada punto a su vecino más cercano, y  $r_E$  es "la distancia media al vecino más cercano esperada en una distribución infinitamente grande que fuera aleatoria y con densidad  $\rho$ " (C&E (1954), p. 447).

El principio según el cual el modelo mide la desviación del patrón respecto a uno aleatorio es muy simple: si la distribución tiende hacia el agrupamiento entonces las distancias de cada punto a su vecino más cercano serán muy pequeñas, y  $R$  tomará valores cercanos a cero; si la distribución fuera aleatoria, se esperaría que en promedio, las distancias al vecino más cercano fueran iguales a las que se esperaría en una distribución aleatoria, por lo que  $R \approx 1$ ; por último, si la distribución fuera regularmente espaciada, entonces la distancia de cada punto a su vecino más cercano sería constante, y si se conoce el área y el número de puntos regularmente espaciados, es fácil calcular el valor que tomaría  $R$  en este caso.

Una primera observación general al método es que se asume que el área considerada es de tamaño infinito (es decir, sin fronteras). Desde luego, en la práctica siempre se tiene una región limitada. Sin embargo, esto no necesariamente produce problemas, pues, como se verá en el desarrollo del modelo, resulta conceptualmente más fácil trabajar con regiones de tamaño infinito (en particular con sectores de círculo de radio infinito) en las que a partir de algún cierto punto se tiene

probabilidad cero de que haya elementos del proceso, por lo que no importa lo grande del tamaño utilizado en la construcción del modelo.

Un primer problema metodológico sería el llamado problema de escala, y que consiste en que si bien un patrón puede ser regularmente espaciado si se le considera inmerso en una región dada, al colocarlo en una región mayor que la original muy probablemente resultaría conglomerado. Harvey. (1969, p. 385-386) no duda en afirmar que este problema es inevitable siempre que se trabaje con distribuciones espaciales: "las distribuciones de formas espaciales son enteramente dependientes de la escala,... (y por tanto), los procesos espaciales son relevantes sólo en ciertas escalas".



En la figura se ve que considerando únicamente a los puntos dentro del rectángulo interior, se tiene un valor de  $R$  que claramente indica regularidad (por ser mayor que 1). Si se quisiera incluir dentro del estudio a los tres puntos restantes, se tendría una medida que indicaría tendencia hacia el agrupamiento.

Claramente el problema se resolvería si se tuviera algún procedimiento para decidir cuál es la frontera de la región en la que ocurre el proceso, pero esto no es siempre posible. C&E (1954, p.449-450) notaron este problema y sugirieron que en caso de haber puntos del proceso fuera de la región especificada que fueran vecinos más cercanos de algunos puntos de adentro no se tomaran medidas a partir de ellos, pero sí se considerase su distancia a puntos de adentro.

Otra posibilidad es señalada por Hammond y McCullagh (1978, p.276-277), y consiste en extender la frontera un poco y no considerar a puntos que queden fuera de esta frontera ampliada. Desgraciadamente no hay criterios generales para marcar la extensión de la frontera que pudiera resultar adecuada. En esta misma línea Donnelly (1978) construye, usando métodos de simulación) estimadores para la media y la varianza de las distancias esperadas al vecino más cercano bajo la hipótesis de aleatoriedad considerando una extensión irregular de la frontera. McNutt (1981) presenta una solución analítica general para el problema de estimadores en extensiones de frontera si la región es circular, rectangular o triangualar.

Otro problema metodológico es el de la definición de distancia entre dos elementos del proceso. Este problema puede verse en dos niveles:

1) Si la escala lo permite, los elementos pueden considerarse como puntos en un mapa, y el problema de calcular su distancia se reduce notablemente. Sin embargo, comúnmente puede haber dificultades, puesto que la distancia lineal raras veces coincide con la distancia real entre dos puntos, y casi nunca se puede contar con los valores de esta última. Si los puntos sólo difieren en cuanto a curvas de nivel, y no hay accidentes topográficos considerables entre los dos, es posible utilizar, como distancia corregida, a la hipotenusa del triángulo rectángulo formado con catetos iguales a la distancia lineal y a la diferencia en curvas de nivel (Sugiura y Cortina (en prensa)); empero, esto todavía está en discusión.

2) Si la escala es tal que los elementos del proceso no pueden representarse como puntos, sino más bien como superficies, la situación se complica, porque además de los problemas producidos por los accidentes topográficos señalados en el punto anterior, la definición de la distancia entre dos superficies no es tan inmediata ni tan clara como la de distancia entre dos puntos. Hay varias alternativas para resolver este problema, pero ninguna de ellas definitiva. Una sería definir a la distancia entre dos superficies tal y como se define en matemáticas a la distancia entre dos conjuntos; es decir, como la mínima entre sus fronteras (Loomis y Sternberg (1968)), si es que éstas pueden ser definidas claramente. Esta definición es intuitivamente atractiva, pues si, por ejemplo, se desea ir de una ciudad a otra, se es fácil suponer que se elegirá la ruta más corta.

La otra alternativa es definir la distancia entre dos superficies como la distancia entre sus centroides, y aquí hay dos problemas: el primero es calcular el centroide de una superficie

irregular puede ser bastante complicado; el otro problema es que es posible que el centroide sea un punto fuera de la superficie considerada.

Una última alternativa es la propuesta por Alden (1979, p.171,175), quien define una medida de la distancia proporcional al tiempo de traslado entre sitios. La medida sugerida es:

$$t = C \cdot l + .4 \cdot n + .5 \cdot a,$$

donde  $C$  es una constante,  $l$  es la distancia lineal (entre fronteras o entre centroides),  $n$  es el número de curvas de nivel de más de 50 m cruzadas y  $a$  es la distancia recorrida por agua.

Los coeficientes  $C$ ,  $.4$  y  $.5$  fueron asignados por estimaciones personales de Alden, siendo obtenido el último "de comentarios españoles acerca de viajes en canoa" (Alden cita a Hernán Cortés y a Gibson). Esta proposición podría hacerse más general, y sobre todo, se debería justificar con más detalle el cálculo de los coeficientes de la ecuación, pero, en principio, me parece la más adecuada, siempre y cuando se resolviera el problema de si la distancia se mide entre puntos más cercanos sobre las fronteras o entre los centroides de las superficies.

Un último problema metodológico es el del cálculo del área de la región considerada. Este consiste en que no siempre es fácil determinar si algunas partes contenidas totalmente en la región deben incluirse o no para efectos del cálculo del área. Por ejemplo, en un trabajo referente al Valle de Toluca, (Sugiura y Cortina 1983), no se tuvo problema para señalar las fronteras, pues esto se hizo con criterios geográficos más o menos cla-

ros; se excluyeron de dicho cálculo zonas como presas y poblaciones nucleadas, pues resultaba imposible estudiar los restos arqueológicos en los sitios allí presentes; se excluyeron las zonas situadas a más de 2800 m de altura sobre el nivel del mar, pues no fueron recorridas. Evidentemente, esto afectó el valor de R, que finalmente indicó una tendencia hacia la nucleación de los asentamientos. De otra forma, R podría haber significado aleatoriedad, en vista de los argumentos expuestos en los capítulos anteriores. Como se ve, tampoco hay una solución general para este problema.

## II.2) MODELO ORIGINAL DE CLARK Y EVANS.

El modelo de Clark y Evans (1954) es importante por varias razones: i) resuelve el problema de la determinación del tamaño de los cuadrantes; ii) proporciona, además de una estadística descriptiva ( $R$ ), una función de esta última que tiene densidad asintóticamente conocida; iv) a partir de una aplicación del análisis de varianza sugerida en el artículo original se pueden realizar comparaciones entre patrones con más precisión.

Hay que hacer notar que existe una gran cantidad de modelos alternativos al aquí estudiado (por ejemplo, Pielou (1959, 1961, 1962), Diggle, Besag y Gleaves (1975)), pero por razones de espacio, no serán mencionados en esta tesis. Igualmente, es necesario señalar que únicamente se examinarán aspectos como la construcción de los estimadores para la media y la varianza de las distancias al vecino más cercano que se tendrían bajo la hipótesis de aleatoriedad, sin considerar problemas como el de la eficiencia de tales estimadores (Holgate, 1965), la potencia de algunas posibles pruebas de hipótesis, los problemas de frontera (Donnelly, 1978, McNutt, 1981) y la aplicación del análisis de varianza a la comparación entre  $p$  distribuciones (Clark y Evans, 1954, p.452, Washburn, 1974) entre otros problemas relacionados con el modelo. La razón es que preferí estudiar las modificaciones al modelo original (aunque fuera de la misma forma) y el índice  $A$  de Hodder y Okell en vez de realizar un análisis exhaustivo del modelo original de Clark y Evans, que ya ha sido bastante discutido. En su trabajo original, Clark y Evans (en adelante C&E) afirman que para propósitos de exposición era suficiente con expo-

ner la construcción del modelo para espacios de 2 dimensiones, mencionando que una generalización para espacios de dimensión finita mayor había sido desarrollada y sería presentada más tarde. Esto último se hizo hasta 1979, quince años después de la muerte de P.J. Clark. Esta generalización se expone en la siguiente sección.

Como ya se dijo, la fórmula propuesta por C&E es

$$R = \frac{r_A}{r_E},$$

donde  $r_A$  es el promedio de las distancias de cada punto a su vecino más cercano (en cualquier dirección), y  $r_E$  es definida como

$$r_E = \frac{1}{2\sqrt{\rho}},$$

siendo  $\rho$  la densidad observada (i.e., el número de puntos por unidad de área). Puede demostrarse que  $r_E$  es la media de las distancias al vecino más cercano cuando éstas se distribuyen según una función de probabilidad de Poisson. (C&E, 1954, p. 451).

Sería más preciso decir que lo que sigue una distribución de Poisson es el número de puntos (elementos del proceso) que hay en un área determinada. Si sucede esto, se dirá que el patrón es aleatorio. Es inmediata la relación entre el concepto de aleatoriedad expresado en este párrafo y los procesos de Poisson, por la misma construcción de estos procesos, lo cual justifica el uso de esta distribución en la construcción del modelo.

C&E no presentan una deducción completa de las fórmulas, sino que únicamente indican la forma en que se

obtuvieron.

Se sabe que si  $X$  es una variable aleatoria con distribución Poisson con parámetro  $\lambda$ , entonces,

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}.$$

Considérese al círculo de radio  $r$  como unidad de área, y sea  $\rho$  el número promedio de puntos por esta unidad. Entonces, en un círculo de radio  $r$  dividido en  $k$  sectores de igual tamaño habrá, en promedio,  $\rho \pi r^2 / k$  puntos en cada sector; entonces,

la probabilidad de encontrar  $x$  puntos en algún sector es, si se piensa en un patrón aleatorio (en el sentido explicado en la página anterior):

$$P(X = x) = \frac{e^{-\rho \pi r^2 / k} (\rho \pi r^2 / k)^x}{x!}, \quad x = 0, 1, 2, \dots$$

0 e.o.c.

Por lo tanto,  $P(X = 0) = \exp(-\rho \pi r^2 / k)$ , que puede ser interpretada como la probabilidad de que no haya otro punto en un  $(1/k)$ -sector en una distribución menor o igual a  $r$ , si se centra un círculo de radio  $r$  en cualquier punto.

Considerada como una función de  $r$ , la expresión  $P(X = 0)$  puede verse como la proporción de distancias al vecino más cercano (en el sector) que son mayores o iguales que  $r$ , y, por lo tanto, su complemento será la proporción de distancias al vecino más cercano menores que  $r$ . Esto es:

$$F(r) = 1 - \exp(-\rho \pi r^2 / k).$$

La ecuación anterior es la función de distribución de las distancias al vecino más cercano que son menores que  $r$ , por lo que su derivada (respecto a  $r$ ) será la función de densidad de probabilidad de estas distancias:

$$\frac{dF(r)}{dr} = f(r) = (2r\rho\pi/k) \exp(-\rho\pi r^2/k),$$

es decir,  $f(r)$  indica la probabilidad de que el vecino más cercano de cualquier punto esté contenido en un  $(1/k)$ -sector de un círculo de radio  $r$ .

Es inmediato verificar que  $f(r)$  cumple con las condiciones para ser una función de densidad de probabilidad (continua):

i)  $f(r) \geq 0$ , porque  $r, \rho$  y  $k \geq 0$ ;

$$\begin{aligned} \text{ii) } \int_{-\infty}^{\infty} f(r) dr &= \int_0^{\infty} (2r\rho\pi/k) \exp(-\rho\pi r^2/k) dr = \\ &= \int_0^{\infty} -D_r (\exp(-\rho\pi r^2/k)) = \\ &= -\exp(-\rho\pi r^2/k) \Big|_0^{\infty} = 1. \end{aligned}$$

Para obtener la media de la distribución es necesario evaluar:

$$E(r) = \int_0^{\infty} (2r^2\rho\pi/k) \exp(-\rho\pi r^2/k) dr,$$

$$\begin{aligned} \text{sea } \alpha = \rho\pi/k, \text{ entonces, } E(r) &= \int_0^{\infty} 2r^2\alpha \exp(-ar^2) dr = \\ &= \int_0^{\infty} -r D_r (\exp(-ar^2)); \end{aligned}$$

integrando por partes:

$$u(r) = r \qquad dv = D_r (\exp (-ar^2)).$$

$$du = dr \qquad v = \exp (-ar^2),$$

entonces:

$$E(r) = - (r \exp (-ar^2)) \Big|_0^\infty + \int_0^\infty \exp (-ar^2) dr.$$

Como la función exponencial negativa tiende a cero más rápido de lo que la función idéntica tiende a menos infinito cuando la variable independiente se va a infinito,

$$E(r) = \int_0^\infty \exp (-ar^2) dr,$$

sea  $u^2 = ar^2$ , y  $dr = \frac{du}{\sqrt{a}}$ , entonces;

$$E(r) = \frac{1}{\sqrt{a}} \int_0^\infty \exp (-u^2) du, \text{ pero } \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp (-u^2/2) du = 1,$$

y por la simetría de la función  $\exp (-u^2/2)$  respecto a 0,

$$\int_0^\infty \exp (-u^2/2) du = \sqrt{\pi/2}, \text{ y si } t = u^2/2, \text{ y } du = \sqrt{2} dt,$$

$$\text{entonces, } \sqrt{2} \int_0^\infty \exp (-t^2) dt = \sqrt{\pi/2}, \text{ y } \int_0^\infty \exp (-t^2) dt = \frac{\sqrt{\pi}}{2}.$$

Finalmente, se tiene que:

$$E(r) = \frac{1}{\sqrt{a}} \frac{\sqrt{\pi}}{2} = \frac{\sqrt{\pi}}{2\sqrt{\rho\pi/k}} = \frac{\sqrt{k}}{2\sqrt{\rho}}, \text{ y si } k = 1 \text{ (i.e., si se$$

considera al círculo completo), se tiene que:

$$r_E = \frac{1}{2\sqrt{\rho}},$$

que es la expresión obtenida por C&E. Una deducción ligeramente diferente de esta fórmula puede verse en McNutt (1981).

Para calcular la varianza hay que resolver:

$$E(r^2) = \int_0^{\infty} 2ar^3 \exp(-ar^2) dr$$

$$\int_0^{\infty} -r^2 D_r (\exp(-ar^2)) \cdot$$

Integrando por partes:

$$u(r) = r^2 \quad dv = D_r (\exp(-ar^2))$$

$$du = 2rdr \quad v = \exp(-ar^2),$$

entonces:

$$E(r^2) = -(r^2 \exp(-ar^2)) \Big|_0^{\infty} - \int_0^{\infty} 2r \exp(-ar^2) dr,$$

y el primer sumando se anula por el argumento anterior, y se tiene que:

$$E(r^2) = \int_0^{\infty} 2r \exp(-ar^2) dr = \frac{-1}{\alpha} \int_0^{\infty} D_r \exp(-ar^2)$$

$$= \frac{-1}{\alpha} \exp(-ar^2) \Big|_0^{\infty} = \frac{1}{\alpha},$$

y como  $\text{Var}(r) = E(r^2) - (E(r))^2$ , se tiene que:

$\text{Var}(r) = \frac{k}{\pi\rho} - \frac{k}{4\rho} = \frac{k(4 - \pi)}{4\pi\rho}$ , y si  $k = 1$ , la expresión es la misma que la obtenida por C&E (1954), p.452).

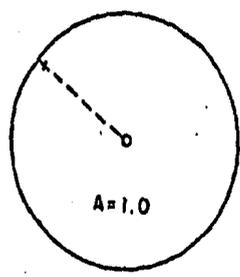
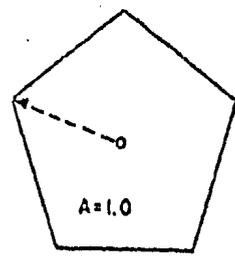
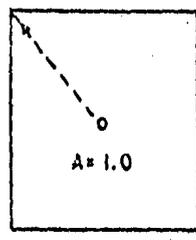
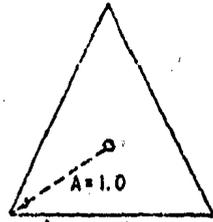
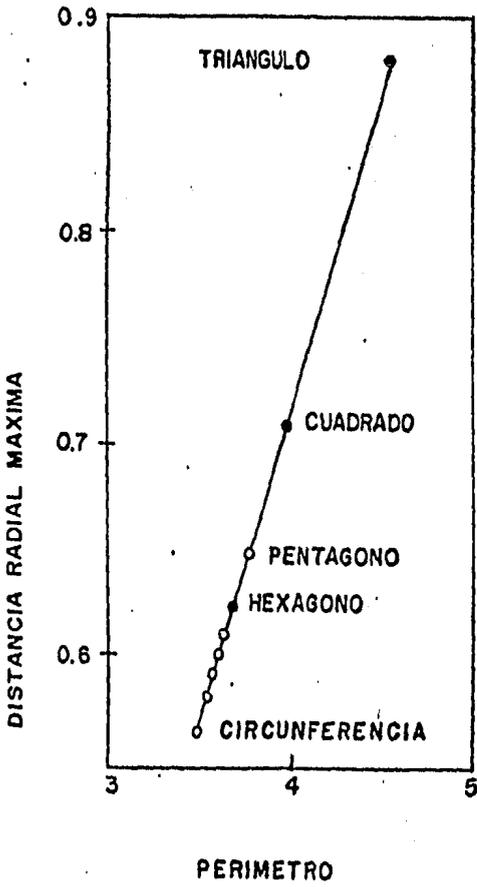
El error estandar de la media esperada es:

$$\sigma_{r_E} = \frac{\sqrt{k(4 - \pi)}}{\sqrt{4n\pi\rho}} = \frac{.2613616 \sqrt{k}}{\sqrt{\rho n}}$$

donde  $n$  es el número de puntos considerados.

Como ya se dijo, si la media de las distancias observadas al vecino más cercano coincidiera con las distancias esperadas suponiendo aleatoriedad, el índice tomaría el valor 1. Si los puntos estuvieran agrupados,  $r_A$  sería más pequeña que  $r_E$ , por lo que el valor de  $R$  estaría cerca de cero. Por último, si el patrón tiende a estar regularmente espaciado,  $R$  tomará valores mayores que 1. C&E afirman que "bajo condiciones de máximo espaciamiento los individuos se distribuirán en un patrón uniforme hexagonal, y cada individuo (excepto aquéllos en la periferia de la población) estará equidistante de otros seis individuos... Cuando esto sucede,  $R$  toma el valor 2.1491" (que es el máximo según estos autores) (C&E, (1954), p.447).

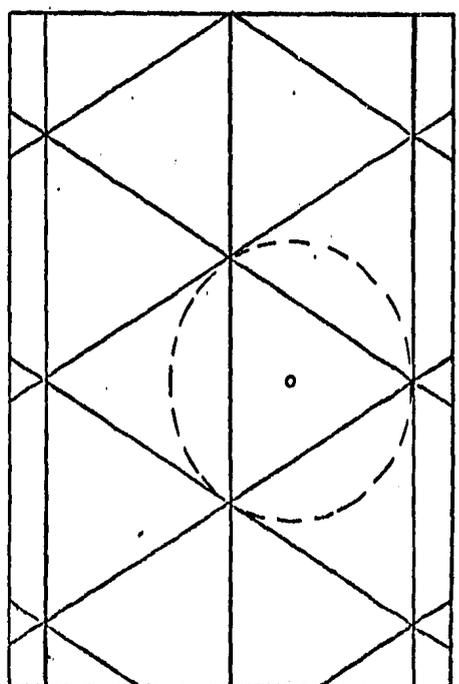
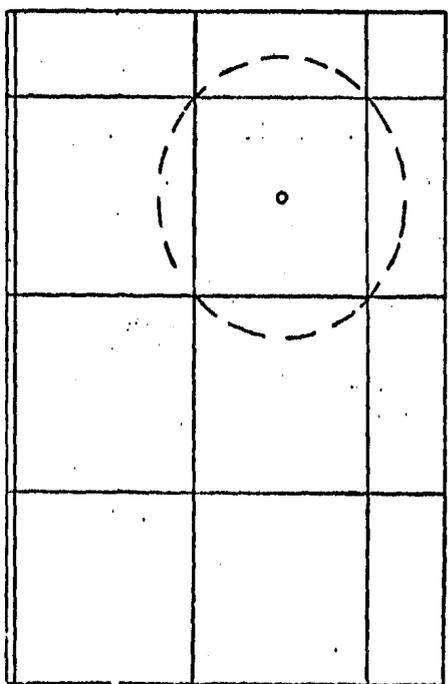
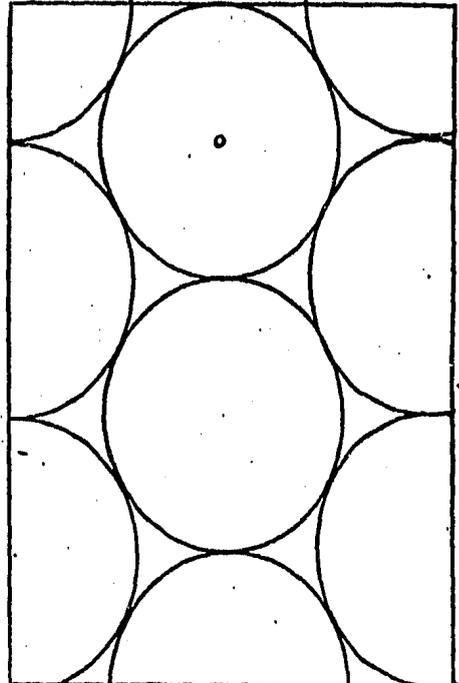
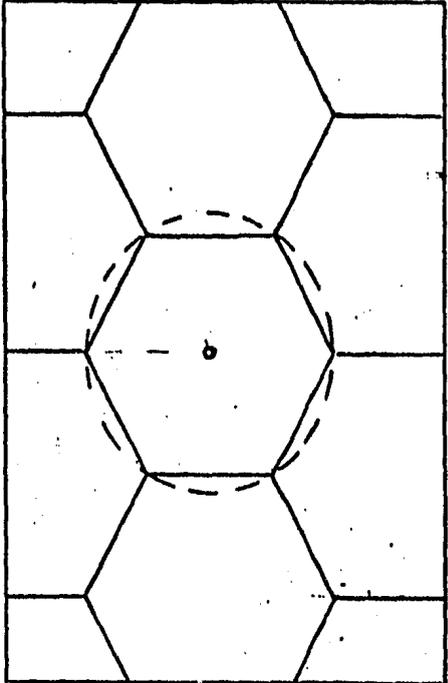
La razón que hay para considerar a un patrón regular hexagonal como de máxima regularidad está dada por ejemplo en Haggett et al. (1975, vol, I, pp.55-57), y está expresada en la siguiente gráfica:



Como se ve en la parte superior, las únicas figuras regulares que cubren completamente un área cualquiera son el triángulo, el cuadrado y el hexágono \*. Lo que se busca tener con un patrón regularmente espaciado es eficiencia, definida de dos maneras: i) *eficiencia de movimiento*, medida por la distancia del centro al perímetro de la región y ii) *eficiencia de fronteras*, medida por la longitud del perímetro de la región. (Haggett et al., op. cit. p.55). Como apuntan Haggett et al. (p.56), "experimentaciones han demostrado que mientras mayor sea el contraste entre los lados de (alguna figura irregular), ésta se hace cada vez menos económica, en términos de accesibilidad del centro (al perímetro) y de longitud de perímetro". Esto último da elementos para apoyar la idea de que un patrón de espaciamiento regular máximo debe estar formado por figuras regulares, que además cubran completamente cualquier región. Los círculos son los polígonos más económicos en ambos términos, pero no cubren totalmente una región. En este sentido, el hexágono es el polígono que además de cumplir con esta característica, se parece más al círculo (véase la gráfica), por lo que se le utiliza como modelo de máximo espaciamiento en patrones de puntos.

Existen otras razones (de tipo económico) para considerar a los patrones hexagonales como los que minimizan costos, y, por lo tanto, como modelos hacia los cuales deberían tender los patrones de asentamiento humanos. Todo esto está relacionado con la teoría de lugar central de Christaller y Lösch (Chisholm, 1979,

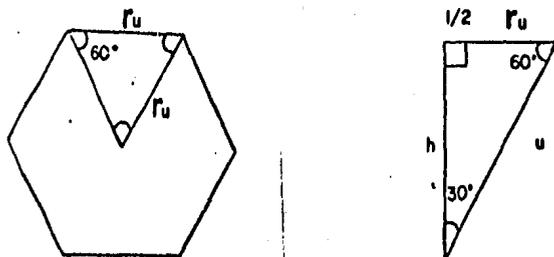
\* Este problema fue resuelto primeramente por Kepler.



p. 143-147). Sin embargo, no se discutirá en este trabajo.

Es posible obtener patrones en los que  $R > 2.1491$ . A continuación se justificará el valor de 2.1491 para un patrón hexagonal regular, se construirá el valor para una distribución cuadrada regular y se verá (con un ejemplo) que, como mencionan C&E (1954, p. 447), para tener un valor de  $R$  mayor que 2.1491 es necesario que los puntos queden "en la periferia de la población".

Sea  $r_u$  la distancia (constante) entre vecinos más cercanos para una distribución hexagonal regular.



Por trigonometría se obtiene que:

$$h = \frac{.5 r_u}{\tan(30)} = \frac{.5 r_u}{1/\sqrt{3}} = \frac{r_u \sqrt{3}}{2}$$

entonces, el área del triángulo mostrado en la figura es:

$$A = \frac{.5 r_u h}{2} = \frac{\sqrt{3} r_u^2}{4}, \text{ y la del triángulo inscrito en}$$

el hexágono es, simplemente  $2A$ .

Puede concluirse que la densidad en tal patrón es de un punto por cada triángulo equilátero como el inscrito en el hexágono.

Entonces:

$$\rho = \frac{2}{\sqrt{3} r_u^2} \quad \Leftrightarrow \quad r_u = \frac{\sqrt{2}}{\sqrt{\rho} \sqrt{3}} .$$

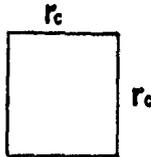
Como  $r_u$  es el valor de  $r_A$  en una distribución hexagonal regular,

$$R = \frac{r_u}{r_E} = \frac{\frac{\sqrt{2}}{2}}{\sqrt{3} \sqrt{k}} = \frac{2.149139864}{\sqrt{k}} .$$

La construcción del valor de  $R$  es análoga a la anterior cuando la distribución espacial es cuadrada regular.

Sea  $r_c$  la distancia (constante) al vecino más cercano de cada punto para una distribución cuadrada regular.

Entonces,  $A = r_c^2$ ; análogamente,  $\rho = 1/r_c^2 \Leftrightarrow r_c = 1/\sqrt{\rho}$ .



Como  $r_c$  es  $r_A$  en tal patrón,

$$R = \frac{r_c}{r_E} = \frac{1/\sqrt{\rho}}{\sqrt{k}/2\sqrt{\rho}} = \frac{2}{\sqrt{k}} .$$

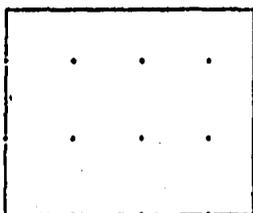
Para encontrar un valor de  $R$  mayor que 2.14914 basta que:

$$R = \frac{\frac{1}{n} \sum_{i=1}^n r_i}{1/2\sqrt{\rho}} = \frac{2 \sum_{i=1}^n r_i}{\sqrt{na}} , \text{ donde } n \text{ es el número de}$$

puntos, y  $a$  es el área de la región ( $\rho = n/a$ ), y  $R > 2.14914$  si y

sólo si  $\sum r_i > \frac{2.149139864}{2} \sqrt{n \cdot a}$ , siendo  $r_i$  la distancia del  $i$ -ésimo punto a su vecino más cercano.

Por ejemplo:



Para este patrón se tienen los siguientes valores:

$$a = 80, n = 12, \sum r_i = 36, R = 2.3238.$$

Como se ve, los puntos están situados (en su mayoría) en la frontera de la región, y podría considerarse a esta distribución como "patológica".

Un problema que sería interesante tratar es clasificar las distribuciones que produzcan un mismo valor de  $R$ . Sin embargo, la falta de una caracterización algebraica para la función mínimo dificulta esto. Por ejemplo, es posible demostrar que el conjunto de muestras de tamaño  $n$  que tienen la misma varianza forma una variedad diferenciable. Para esto es necesario usar un teorema de Milnor (Milnor, 1978, p. ). Este hecho permite, por ejemplo, hablar de "ángulos" entre distintas muestras con igual varianza, lo que podría en un momento dado ser de utilidad para encontrar otro tipo de diferencias entre ellas. Por el momento, no puedo más que plantear el problema, que debo a una sugerencia de los profesores Guillermo Espinosa y Jesús López Estrada.

### II.3) GENERALIZACION PARA $n$ VECINOS MAS CERCANOS.

Supóngase que  $X$  es una variable aleatoria Poisson con parámetro  $\lambda$ , y sea  $P(X = n)$  la probabilidad de que existan al menos  $n$  puntos en un  $(1/k)$ -sector de un círculo de radio  $r$ .

Primeramente, es necesario construir la probabilidad de que haya a lo más  $n-1$  puntos en esta área, esto es:

$$\sum_{x=0}^{n-1} \frac{\exp(-\rho\pi r^2/k)}{x!} (\rho\pi r^2/k)^x$$

La suma anterior es la probabilidad de la unión de los eventos (ajenos) "sector con 0 puntos (excluyendo al centro)", "sector con 1 punto (excluyendo al centro)", ..., "sector con  $n-1$  puntos (excluyendo al centro)".

Sea  $m = ar^2 = \rho\pi r^2/k$ . Entonces,

$$1 - \sum_x \frac{\exp(-m)}{x!} m^x,$$

vista como función de  $r$  es la proporción de distancias al  $n$ -ésimo vecino más cercano que son menores o iguales que  $r$ .

Es posible escribir:

$$\begin{aligned} F_n(r) &= 1 - \sum_x \frac{\exp(-ar^2)}{x!} (ar^2)^x \\ &= 1 - \exp(-ar^2) \cdot \left( 1 + ar^2 + \dots + \frac{(ar^2)^{n-1}}{(n-1)!} \right) \end{aligned}$$

Como  $r$  es una variable  $r$  es continua con rango en los reales

no negativos, es posible obtener su función de densidad de probabilidad derivando (respecto a  $r$ ) a  $F_n(r)$ :

$$\begin{aligned}
 f_n(r) &= \frac{dF_n(r)}{dr} = 2ar \cdot \exp(-ar^2) \cdot \sum_{x=0}^{n-1} \frac{(ar^2)^x}{x!} \\
 &= \exp(-ar^2) \cdot (2ar + \dots + \frac{2(n-1)ar}{(n-1)!} \cdot (ar^2)^{n-2}) \\
 &= 2ar \cdot \exp(-ar^2) \cdot \frac{(ar^2)^{n-1}}{(n-1)!} \\
 &= \frac{2a^n \cdot r^{2n-1} \cdot \exp(-ar^2)}{(n-1)!} \\
 &= \frac{2(\rho\pi/k)^n r^{2n-1} \exp(-\rho\pi r^2/k)}{\Gamma(n)},
 \end{aligned}$$

donde  $\Gamma(n) = \int_0^{\infty} e^{-y} y^{n-1} dy$  es la función gamma. Esta función es una de las más notables en el análisis matemático. Más adelante se utilizarán algunas de sus propiedades.

La función  $f_n(r)$  es la densidad de las distancias en las que hay un máximo de  $n-1$  vecinos más cercanos. Si  $r$  tiende a infinito,  $f_n(r)$  tiende a 0 (porque la función exponencial negativa tiende a cero más rápido que cualquier polinomio cuando la variable independiente tiende a infinito).

Esto significa que la probabilidad de que no haya a lo más  $n-1$  vecinos más cercanos en un  $(1/k)$ -sector de un círculo de radio  $r$  siendo  $r$  muy grande es prácticamente 0.

Por otra parte, si  $r$  tiende a 0 (que es el mínimo valor que puede tomar), se tiene que  $f_n(r)$  tiende al número  $2(\rho\pi/k)^n/\Gamma(n)$ , que decrece cuando  $n$  crece, es decir, para áreas muy pequeñas

la probabilidad de tener un número de vecinos más cercanos que sea grande es inversamente proporcional a éste último.

Con estas dos observaciones se ve que, intuitivamente, la función tiene sentido. Thompson (1956) obtuvo esta función por otra forma, en mi opinión, más complicada.

Claramente  $f_n(r)$  es una función de densidad de probabilidad, pues:

i)  $f_n(r) \geq 0$ , ya que  $\alpha > 0$ ,  $r \geq 0$ , y

ii)  $\int_0^{\infty} \frac{2 \alpha^n r^{2n-1} \exp(-\alpha r^2)}{\Gamma(n)} dr$ , sustituyendo  $u = \alpha r^2$  se

tiene que  $du = 2\alpha r dr$ , y la integral anterior se transforma en:

$$\frac{1}{\Gamma(n)} \int_0^{\infty} e^{-u} u^{n-1} du = 1 \text{ por definición de}$$

la función gamma.

Para calcular la media de la distribución es necesario obtener:

$$E(r) = \int_0^{\infty} \frac{2\alpha^n}{\Gamma(n)} r^{2n} \cdot \exp(-\alpha r^2) dr$$

Con la sustitución anterior se tiene que  $r = \sqrt{u/\alpha}$ , y entonces:

$$E(r) = \frac{1}{\Gamma(n)} \int_0^{\infty} \sqrt{u/\alpha} u^{n-1} e^{-u} du,$$

$$= \frac{1}{\sqrt{\alpha} \Gamma(n)} \int_0^{\infty} u^{n-\frac{1}{2}} e^{-u} du,$$

y, por definición de la función gamma:

$$E(r) = \frac{1}{\sqrt{\alpha}} \frac{\Gamma(n + \frac{1}{2})}{\Gamma(n)}, \text{ pero como } (n + \frac{1}{2})! = \frac{(2n)! \sqrt{\pi}}{4^n n!}$$

(Buck, 1977, p.297), se tiene que:

$$E(r) = \frac{\sqrt{k} (2n)! \sqrt{\pi}}{\sqrt{\rho\pi} \Gamma(n) 4^n n!} = \frac{(2n)! \sqrt{k}}{\sqrt{\rho} n! (n-1)! 4^n} ..$$

donde  $\Gamma(n) = (n-1)!$ , por ser  $n$  un entero positivo.

Esta última fórmula es más conveniente para fines de cálculo, a menos que se cuente con una computadora que efectúe directamente la evaluación de la función gamma.

Para calcular la varianza es necesario efectuar:

$$\begin{aligned} E(r^2) &= \int_0^{\infty} \frac{2\alpha^n}{\Gamma(n)} r^{2n+1} \exp(-\alpha r^2) dr \\ &= \frac{1}{\alpha \Gamma(n)} \int_0^{\infty} 2\alpha r (\alpha r^2) \exp(-\alpha r^2) dr. \end{aligned}$$

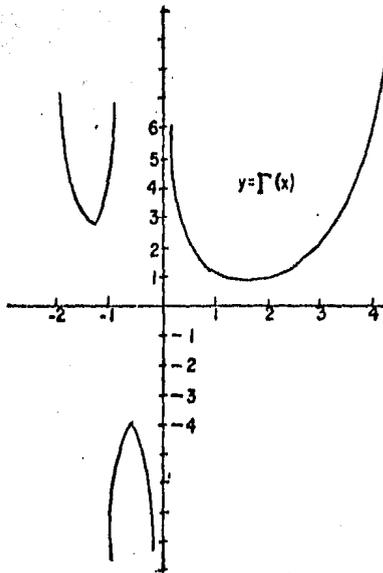
Sea  $u(r) = \alpha r^2$ , y  $du = 2\alpha r dr$ ; Entonces:

$$E(r^2) = \frac{1}{\alpha \Gamma(n)} \int_0^{\infty} u^n e^{-u} du = \frac{1}{\alpha \Gamma(n)} \Gamma(n+1) = \frac{k n}{\rho \pi} .$$

Por lo tanto,

$$\text{Var}(r) = E(r^2) - (E(r))^2 = \frac{k n}{\rho \pi} - \left( \frac{(2n)! \sqrt{k}}{\sqrt{\rho} n! (n-1)! 4^n} \right)^2$$

$$= \frac{n k}{\rho \pi} - \left( \frac{\Gamma(n + \frac{1}{2})}{\sqrt{\alpha} \Gamma(n)} \right)^2 .$$



Función gamma.

C&E sugieren que la estadística

$$C = \frac{r_A - r_E}{\sigma_{r_E}}$$

tiene una distribución normal estandar bajo la hipótesis de aleatoriedad si el número de puntos es grande. Aunque estos autores no lo mencionan, es claro que esto se debe al teorema del límite central. Si el número de puntos es pequeño (menor que 100 según C&E p448) es mejor utilizar una distribución Pearson tipo III (gamma) para contrastar la hipótesis de aleatoriedad.

Sin embargo, como Ord (citado por Haggett et al., vol.II, p.440) señala, no es inmediato que la distribución de  $r_A$  sea normal con media  $r_E$  y varianza  $\sigma_{r_E}$  cuando el número de puntos tiende a infinito. Puesto que hay una alternativa menos complicada (sugerida muy levemente por C&E p.448) no analizaré el comportamiento de la estadística  $C$ .

Resulta menos complicado construir una estadística propuesta por Thompson (1956) que tiene distribución  $\chi^2$ . A continuación se presenta su desarrollo.

Se sabe que

$$f_n(r) = \frac{2\alpha^n \exp(-\alpha r^2) r^{2n-1}}{\Gamma(n)}$$

es la función de densidad de probabilidad para la distancia al  $n$ -ésimo vecino más cercano de alguna distribución puntual. Si se hace la transformación

$$u = 2\alpha r^2$$

se tiene que :

$$f_n(u) = \frac{\exp(-\frac{u}{2}) (\frac{u}{2})^{n-1}}{\Gamma(n)}$$

que es la función de densidad de probabilidad de una variable aleatoria con distribución  $\chi^2$  con  $2n$  grados de libertad (Hogg & Craig, 1977, p.107), lo que muestra que

$$2\alpha r^2 \sim \chi^2_{2n}.$$

Esta estadística me parece preferible a la estadística  $C$  propuesta por C&E, principalmente porque su distribución es exacta, no asintótica como en el caso de  $C$ . Tal vez su poco uso se deba a la escasísima difusión que ha tenido el resultado obtenido por Thompson (1956).

Si el tamaño de la población es  $n$  individuos, entonces la estadística  $n\bar{x}_n$  tiene una distribución xi-cuadrada con  $2n$  grados de libertad, siendo  $\bar{x}_n$  la media de las distancias de cada punto a su  $n$ -ésimo vecino más cercano (Thompson, 1956, p.392).

Cuando el tamaño de la población es suficientemente grande es posible construir un intervalo de confianza para la media utilizando la aproximación normal a la xi-cuadrada (Larson, 1978, p. 124); el intervalo está dado por la expresión

$$(\sqrt{4nn-1} \pm t_{1-\alpha/2})^2/2n,$$

donde  $t_{1-\alpha/2}$  es el punto bajo una normal  $(0,1)$  hasta el cual se acumula el  $1-\alpha/2$  de probabilidad, siendo  $\alpha$  el nivel de significancia del intervalo.

Nótese que si  $n = 1$  es el caso del primer vecino más cercano, por lo que debería utilizarse esta prueba en vez de la propuesta por Clark y Evans. Por otra parte, es más fácil calcular la estadística de prueba  $n\bar{x}_n$  que la estadística  $C$ , y la distribución de la primera es exacta, mientras que la de la segunda es asintótica.

Es claro que hay una relación entre el tamaño promedio de grupos nucleados dentro de un patrón espacial y la estadística  $R$  para  $n$  vecinos más cercanos.

Si se piensa en un patrón formado por conglomerados aislados distribuidos aleatoriamente (como conglomerados), la estadística para los primeros vecinos más cercanos indicaría nucleación, pues las distancias observadas serían pequeñas (i.e., no se tiene que salir de los conglomerados para encontrar vecinos más cercanos de órdenes pequeños). De hecho, como señala Thompson (1956, p.393),  $R$  empezaría a indicar aleatoriedad cuando el orden de los vecinos más cercanos se aproximara al tamaño promedio de los conglomerados.

Una posibilidad de analizar esto es la sugerida por Pielou (1977; p.110), y consiste en graficar el orden de los vecinos más cercanos contra los valores de  $R$  para cada orden. Evidentemente habría una diferencia bastante notable cuando se esté cerca del número promedio de elementos en los conglomerados, ya que los vecinos más cercanos empezarían a estar fuera de los núcleos aislados.

Sin embargo, no se ha podido encontrar una solución analítica al problema de detectar el orden de vecino más cercano en el que se produce ese cambio, en parte por la dificultad para tratar con funciones de orden.

Como se mencionó anteriormente, todo esto se aplicaría para contrastar hipótesis acerca de que la distribución está formada por núcleos aislados distribuidos aleatoriamente (en el sentido descrito en la sección II.2) y que estos núcleos tienen en promedio algún número de puntos que se desea estimar. "El problema de determinar la distribución exacta de las distancias a vecinos más cercanos en modelos no-aleatorios requiere aún de atención, pero

parece muy difícil de solucionar" (Thompson, 1956, p.393).

Como se dijo en la sección I.3, este problema continúa sin solución.

La forma en que la estadística  $n\bar{x}_n$  contrasta la hipótesis de aleatoriedad es la siguiente: si la estadística toma valores mayores que el punto bajo una distribución xi-cuadrada con  $2n$  grados de libertad hasta el cual se acumula el  $1 - \alpha/2$  de probabilidad se rechaza la hipótesis nula con un nivel de significancia  $\alpha$  afirmando que existe una tendencia hacia la regularidad; si la estadística resulta ser menor que el punto hasta el cual se acumula el  $\alpha/2$  de probabilidad se rechaza la hipótesis nula concluyendo que hay una tendencia hacia el agrupamiento en conglomerados.

#### II.4) GENERALIZACION PARA K DIMENSIONES DEL METODO PARA EL PRIMER VECINO MAS CERCANO.

En un artículo aparecido en 1979 C&E presentan una generalización de su método para usarla en espacios de dimensión finita  $k$ . Hay que hacer notar que la medida original se construyó para espacios bidimensionales, sin embargo, resulta útil pensar en medir la desviación de la aleatoriedad para patrones puntuales en espacios de, por ejemplo, 3 dimensiones (como la localización de nidos en un árbol (C&E, 1979, p.316), o la distribución de asentamientos considerando su altura sobre el mar), o en espacios de dimensión 1 (como el número de nacimientos en 24 horas, el número de accidentes en una carretera, o una distribución de asentamientos a lo largo de un río (Stark & Young, 1980)).

El principio para medir la desviación de aleatoriedad es el mismo utilizado en las secciones anteriores, es decir, el cociente

$$R = \frac{r_A}{r_E} .$$

A continuación se presenta la derivación de las fórmulas para la media y los momentos de orden superior de los valores esperados bajo la hipótesis de aleatoriedad.

El volúmen de una esfera  $k$ -dimensional de radio  $r$  está dado por la expresión:

$$V_k(r) = \frac{\pi^{k/2} r^k}{\Gamma(k/2 + 1)} .$$

(para la demostración de esta fórmula, véase el apéndice II).

Análogamente, la probabilidad de que tal esfera centrada en

cualquier individuo no contenga a otro individuo es:

$$P(X = 0) = \exp\left(-\frac{\rho \pi^{k/2} r^k}{\Gamma(k/2 + 1)}\right),$$

donde  $\rho$  es la densidad de la población. Esta expresión puede ser vista como función de  $r$ , siendo entonces la proporción de distancias al vecino más cercano mayores que  $r$ . Si se considera, como en las secciones anteriores no a la esfera completa, sino a un  $1/\ell$ -sector se tiene que:

$$F(r) = 1 - \exp(-\alpha r^k)$$

representa la proporción de distancias al vecino más cercano menores o iguales que  $r$ , siendo

$$\alpha = \frac{\rho \pi^{k/2}}{\Gamma(k/2 + 1) \ell}$$

Entonces:

$$\frac{dF(r)}{dr} = \alpha k r^{k-1} \exp(-\alpha r^k) = f_k(r)$$

es la función de densidad de tales distancias.

Claramente  $f_k(r)$  es no negativa, pues  $k$  es un número natural y  $r, \rho \geq 0$ . Además,

$$\int_0^{\infty} \alpha k r^{k-1} \exp(-\alpha r^k) dr =$$

$$\int_0^{\infty} -D_r (\exp(-ar^k)) = \exp(-ar^k) \Big|_0^{\infty} = 1.$$

Para calcular los momentos de cualquier orden se tiene que:

$$E(r^s) = \int_0^{\infty} kar^{k-1+s} \exp(-ar^k) dr; \quad \text{sea } u = ar^k, \text{ y } du = kar^{k-1}$$

y despejando  $r$  se tiene que  $r = (u/a)^{1/k}$ . Entonces:

$$E(r^s) = \int_0^{\infty} e^{-u} r^s du = \int_0^{\infty} e^{-u} (u/a)^{s/k} du = \frac{\Gamma(s/k + 1)}{a^{s/k}}$$

Por lo tanto, la media de la distribución es:

$$E(r) = \frac{\lambda^{1/k} (\Gamma(k/2 + 1))^{1/k} \Gamma(1/k + 1)}{\rho^{1/k} \sqrt{\pi}}, \text{ y la varianza es:}$$

$$E(r^2) - E^2(r) = \frac{\lambda^{2/k} (\Gamma(k/2 + 1))^{2/k} (\Gamma(2/k + 1) - (\Gamma(1/k + 1))^2)}{\rho^{2/k} \Pi}$$

que son las fórmulas obtenidas por Clark y Evans.

## II.5) GENERALIZACION PARA K DIMENSIONES DEL METODO PARA N VECINOS MAS CERCANOS.

La derivación de las fórmulas para esta generalización es análoga a las presentadas en las secciones anteriores. La notación es la misma.

$$\text{Sean } m = \frac{\rho}{\ell} \frac{\prod^{k/2} r^k}{\Gamma(k/2 + 1)}, \text{ y } \alpha = \frac{\rho}{\ell} \frac{\prod^{k/2}}{\Gamma(k/2 + 1)}.$$

Entonces:  $\frac{dm}{dr} = \alpha k r^{k-1}$ , y  $m = \alpha r^k$ .

$\sum_{x=0}^{n-1} \frac{e^{-m} m^x}{x!}$  es la probabilidad de la unión de los eventos

ajenos "1/ℓ-sector con 0 puntos (excluyendo al centro)",  
 ..., "1/ℓ-sector con n-1 puntos (excluyendo al centro)".

Entonces el complemento de la suma anterior (vista como función de r) es la proporción de distancias al n-ésimo vecino más cercano que son menores que r en un 1/ℓ-sector de una esfera n-dimensional de radio r, por lo que puede considerarse como la función de distribución de probabilidad de estas distancias:

$$F_{n,k}(r) = 1 - \sum_{x=0}^{n-1} \frac{e^{-m} m^x}{x!} = 1 - e^{-m} \left( 1 + m + \frac{m^2}{2!} + \dots + \frac{m^{n-1}}{(n-1)!} \right)$$

La función de densidad de probabilidad de estas distancias es la derivada (respecto a r) de  $F_{n,k}(r)$ :

$$\begin{aligned} f_{n,k}(r) &= \frac{dF_{n,k}(r)}{dr} \\ &= \alpha k r^{k-1} \exp(-\alpha r^k) \cdot \sum_{x=0}^{n-1} \frac{(\alpha r^k)^x}{x!} \exp(-r^k) (k r + \dots \\ &\quad + (r^k)^{n-2} \frac{k(n-1)r}{(n-1)!}) \end{aligned}$$

Reduciendo:

$$f_{n,k}(r) = k \alpha r^{k-1} \exp(-\alpha r^k) \frac{(\alpha r^k)^{n-1}}{(n-1)!}$$

$$= \frac{k \alpha^n r^{kn-1} \exp(-\alpha r^k)}{(n-1)!}.$$

Es fácil verificar que la expresión anterior es, en efecto, una función de densidad de probabilidad:

i)  $f_{n,k}(r) \geq 0$ , porque  $r \geq 0, \alpha \geq 0$  y  $k, n \in \mathbb{N}$ .

ii)  $\int_0^{\infty} \frac{k \alpha^n r^{kn-1} \exp(-\alpha r^k)}{\Gamma(n)} dr = 1$ , porque:

$$= \frac{1}{\Gamma(n)} \int_0^{\infty} \alpha^{n-1} k^{(n-1)} (-D_r(\exp(-\alpha r^k))) dr;$$

integrando por partes:

$$u(r) = \alpha^{n-1} r^{k(n-1)}, \quad du = \alpha^{n-1} (k(n-1)) r^{k(n-1)-1} dr$$

$$dv = -D_r(\exp(-\alpha r^k)), \quad v = -\exp(-\alpha r^k),$$

entonces:

$$= \frac{1}{\Gamma(n)} \left\{ (-\alpha^{n-1} r^{k(n-1)} \exp(-\alpha r^k)) \Big|_0^{\infty} + \int_0^{\infty} \alpha^{n-1} r^{k(n-1)-1} (k(n-1)) \exp(-\alpha r^k) dr \right\}.$$

El primer sumando se anula, y al hacer la transformación

$$u = \alpha r^k, \quad du = k \alpha r^{k-1} dr. \quad \text{y } r = \left(\frac{u}{\alpha}\right)^{1/k} \text{ se tiene:}$$

$$= \frac{1}{\Gamma(n)} \int_0^{\infty} e^{-u} \alpha^{n-2} (n-1) r^{k(n-2)} du, \quad \text{y sustituyendo } r:$$

$$= \frac{1}{\Gamma(n)} \int_0^{\infty} e^{-u} \alpha^{n-2} (n-1) \frac{u^{n-2}}{\alpha^{n-2}} du$$

$$= \frac{1}{\Gamma(n)} (n-1) \Gamma(n-1) = 1, \text{ por propiedad de la función}$$

gamma.

Para obtener los momentos de cualquier orden basta efectuar:

$$E(r^s) = \frac{1}{\Gamma(n)} \int_0^{\infty} k \alpha^n r^{kn-1+s} \exp(-\alpha r^k) dr.$$

Sea  $u = \alpha r^k$ ;  $du = k\alpha r^{k-1} dr$  y  $r = \frac{u^{1/k}}{\alpha^{1/k}}$ , entonces:

$$E(r^s) = \frac{1}{\Gamma(n)} \int_0^{\infty} e^{-u} \alpha^{n-1} r^{k(n-1)+s} du, \text{ y sustituyendo } r:$$

$$= \frac{1}{\Gamma(n)} \int_0^{\infty} e^{-u} \alpha^{n-1} \frac{u^{n-1+(s/k)}}{\alpha^{n-1+(s/k)}} du$$

$$= \frac{1}{\Gamma(n) \alpha^{s/k}} \int_0^{\infty} e^{-u} u^{n-1+(s/k)} du$$

$$= \frac{\Gamma(n + (s/k))}{\Gamma(n) \alpha^{s/k}}.$$

A partir de esta última fórmula se construirán los valores para la media, la varianza y los coeficientes de sesgo y kurtosis de la distribución para cualquier  $n$  y  $k$ .

Siguiendo la notación de Rahman (1978, p. 104-106), sean

$$\mu'_s = E(x^s), \text{ y}$$

$$\mu'_s = E((r - E(r))^s)$$

Entonces, la media de la distribución,  $E(r)$  es:

$$\mu'_1 = \frac{\Gamma(n + 1/k)}{\Gamma(n) \alpha^{1/k}} = \frac{\Gamma(n + 1/k) \rho^{1/k} (\Gamma(k/2 + 1))^{1/k}}{\Gamma(n) \alpha^{1/k} \sqrt{\pi}}$$

Nótese que si  $n = 1$  y  $k = 2$  (i.e. primer vecino más cercano en una esfera completa en dos dimensiones) se tiene que

$$\mu'_1 = \frac{1}{2\sqrt{\rho}}, \text{ que es la fórmula original de Clark y Evans (1954).}$$

La varianza está dada por la expresión

$$\begin{aligned} \text{Var}(r) &= \mu'_2 - (\mu'_1)^2 = \frac{\Gamma(n + 2/k)}{\Gamma(n) \alpha^{2/k}} - \left( \frac{\Gamma(n + 1/k)}{\Gamma(n) \alpha^{1/k}} \right)^2 \\ &= \frac{(\Gamma(n) \Gamma(n + 2/k) - (\Gamma(n + 1/k))^2) \rho^{2/k} (\Gamma(k/2 + 1))^{2/k}}{(\Gamma(n))^2 \rho^{2/k} \pi} \end{aligned}$$

El coeficiente de sesgo se define como  $Sk = \frac{\mu'_3}{(\mu'_2)^{1.5}}$  (Rahman, 1978, p. 104), Es posible demostrar que  $\mu'_3$  puede expresarse como

$$\mu_3 = E(r^3) - 3E(r)E(r^2) + 2(E(r))^3 \text{ (ibid., p. 128).}$$

Entonces,  $Sk$  queda definido por la siguiente expresión:

$$Sk = \frac{(\Gamma(n))^2 \Gamma(n + 3/k) - 3\Gamma(n) \Gamma(n + 1/k)\Gamma(n + 2/k) + 2((n+1/k))^3}{\left[ \Gamma(n) \Gamma(n + 2/k) - (\Gamma(n + 1/k))^2 \right]^{1.5}}$$

Nótese que, en particular, si  $n = k = 1$  se tiene que

$$Sk = \frac{\Gamma(4) - 3\Gamma(2) \Gamma(3) + 2(\Gamma(2))^3}{(\Gamma(3) - \Gamma(2))^2} = 2.$$

El coeficiente de Kurtosis se define como  $Ku = \frac{\mu_4}{(\mu_2)^2} - 3;$

(ibid., p. 106). Rahman demuestra (p. 128) que

$$\mu_4 = E(r^4) - 4E(r)E(r^3) + 6E(r^2)(E(r))^2 - 3(E(r))^4.$$

Evitando los cálculos,

$$3(\Gamma(n+1/k))^4$$

$$\frac{(\Gamma(n))^3 \Gamma(n+4/k) - 4(\Gamma(n))^2 \Gamma(n+1/k) \Gamma(n+3/k) + 6\Gamma(n)\Gamma(n+2/k)\Gamma(n+1/k)^2 - \left[ \Gamma(n) \Gamma(n + 2/k) - (\Gamma(n + 1/k))^2 \right]^2}{}$$

-3.

A continuación se presenta una tabla con los valores del sesgo y la kurtosis para los primeros diez vecinos más cercanos en dimensiones 1, 2, 3, 4, y 5. La nota "sesgo no calculable" se debe a que

$\Gamma(n) \Gamma(n + 2/k) \geq (\Gamma(n + 1/k))^2$  y los números  $n$  y, puesto que los números negativos no tienen raíces cuadradas reales definidas, no es posible evaluar a  $Sk$ .

## TABLA DE SESGOS Y KURTOSIS.

orden	dimension	sesgo	kurtosis
1	1	2.0000	6.0000
1	2	0.6311	0.2451
1	3	0.1681	-0.2705
1	4	no calculable	-0.2521
1	5	no calculable	-0.1197
2	1	1.4142	3.0000
2	2	0.4056	0.0593
2	3	0.0598	-0.1427
2	4	-0.1249	-0.0805
2	5	-0.2415	0.0221
3	1	1.1547	2.0000
3	2	0.3179	0.0251
3	3	0.0317	-0.0913
3	4	-0.1182	-0.0371
3	5	-0.2115	0.0378
4	1	1.0000	1.5000
4	2	0.2692	0.0136
4	3	0.0202	-0.0661
4	4	-0.1088	-0.0209
4	5	-0.1884	0.0370
5	1	0.8944	1.2000
5	2	0.2374	0.0085
5	3	0.0143	-0.0515
5	4	-0.1006	-0.0135
5	5	-0.1711	0.0337
6	1	0.8165	1.0000
6	2	0.2147	0.0058
6	3	0.0107	-0.0421
6	4	-0.0938	-0.0092
6	5	-0.1577	0.0302

orden	dimension	sesgo	kurtosis
7	1	0.7559	0.8571
7	2	0.1974	0.0042
7	3	0.0084	-0.0355
7	4	-0.0880	-0.0068
7	5	-0.1468	0.0272
8	1	0.7071	0.7500
8	2	0.1836	0.0032
8	3	0.0069	-0.0307
8	4	-0.0832	-0.0052
8	5	-0.1379	0.0247
9	1	0.6667	0.6667
9	2	0.1724	0.0025
9	3	0.0057	-0.0270
9	4	-0.0790	-0.0041
9	5	-0.1305	0.0225
10	1	0.6325	0.6000
10	2	0.1630	0.0020
10	3	0.0049	-0.0241
10	4	-0.0754	-0.0033
10	5	-0.1241	0.0207

La utilidad de esta tabla está en que amplía la información disponible acerca de la función de distribución de las distancias al  $n$ -ésimo vecino más cercano. Clark y Evans (1979) dan los resultados de los coeficientes de sesgo para el primer vecino más cercano en espacios de dimensión 1, los cuales sirvieron para validar los cálculos presentados en la tabla.

Clark y Evans (1979) afirman que "los ecólogos no parecen estar muy interesados en valores de  $k$  mayores que 3" (p. 317). Sin embargo, es posible pensar en una aplicación del método en la que se utilicen espacios de dimensión posiblemente mayor que 3. Por ejemplo, si se consideran puntos descritos por dos coordenadas espaciales (como localización en planta y altura sobre el nivel del mar) y además se tienen medidas de algún conjunto de variables en cada punto sería posible contrastar hipótesis acerca de la aleatoriedad de la distribución definida en un espacio de dimensión  $3 + v$ , donde  $v$  es el número de variables medidas en cada punto además de las variables de localización.

Evidentemente, dichas variables deben ser medibles con escalas al menos de intervalo, pues de lo contrario no tendría sentido hablar de distancias entre los puntos.

Podría decirse que, en general, la métrica de Minkowski debería ser utilizada para medir las distancias entre puntos, Sin embargo, sería necesario investigar antes si existe alguna relación entre la aleatoriedad (medida en términos de una cierta función  $R$  análoga a la propuesta originalmente por Clark y Evans) y el orden de la métrica de Minkowski que se utilice, definido por

el valor  $p$  en la fórmula

$$D_p(x_j, x_h) = \left( \sum_{i=1}^k |x_{ij} - x_{ih}|^p \right)^{1/p} \quad (p \geq 1)$$

para un espacio de dimensión  $k$ .

Esta sugerencia de utilizar métricas de este tipo para calcular las distancias entre puntos se basa en que estas medidas satisfacen condiciones que a menudo resultan necesarias para describir ciertas distribuciones espaciales. Ejemplos de estas condiciones son la simetría, la desigualdad del triángulo y la ultrametría (Anderberg, 1973, p. 99-103).

A continuación se presentan algunos ejemplos.

## EJEMPLO 1

Se generaron 256 puntos distribuidos de la siguiente manera: se calculó la localización de un primer punto utilizando la función RANDOM; a partir de estas coordenadas se generó (también con RANDOM) una dirección (i.e. un ángulo entre 0 y  $2\pi$ ). En esta dirección se "tiró" una línea de longitud variando según una distribución de Poisson con parámetro 10. De esta forma se produjeron las coordenadas de los 255 puntos restantes.

El hecho de que la mediana observada sea cero se debe a que hubo puntos con las mismas coordenadas, lo que trae problemas para el cálculo del índice R de Washburn.

Es interesante notar que aunque con este método se tienen puntos cuya distancia a su vecino más cercano en cualquier dirección es una variable aleatoria con distribución Poisson el índice C de Clark y Evans señaló concentración del patrón para todos los órdenes de vecino más cercano. Una posible explicación de esto es que como el número de puntos es bastante grande, la prueba se vuelve más potente (en el sentido de ser menos tolerante con el error tipo I). Si se observa únicamente el valor de R para cada orden se tiene que casi todos ellos distan de 1 en menos de .1, por lo que, en estos casos no es suficiente esta información para discriminar patrones aleatorios y no aleatorios.

Utilizando las estadísticas de Thompson se ve que para los primeros cinco vecinos más cercanos existe aleatoriedad, indicando agrupación para los órdenes superiores. Esto podría significar que hay grupos dentro de la distribución con un tamaño promedio de seis elementos por grupo. El nivel de significancia empleado fue de .05.



POISSON/R

256 PUNTOS

AREA=

961

DENSIDAD= 1.2663892

MEDIANA OBSERVADA= 1.9161

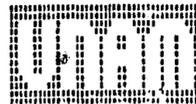
MEDIANA ESPERADA= 0.9161

R= 0.0000

	Obs.	ESP.	REI1	D.S.	C
1	0.3021	1.9687	0.3290	1.5164	-3.7471
2	1.4042	1.4531	0.9663	1.5275	-1.0772
3	1.6729	1.8164	0.9210	1.5342	-3.1411
4	1.7431	2.1191	0.9169	1.5375	-3.8415
5	2.1875	2.3940	0.9176	1.5394	-4.2817
6	2.4015	2.6224	0.9194	1.5406	-4.8287
7	2.5386	2.8410	0.9112	1.5415	-5.4881
8	2.9132	3.0439	0.9242	1.5421	-5.0129
9	2.9957	3.2341	0.9265	1.5426	-5.1800
10	3.1923	3.4130	0.9251	1.5431	-4.8092
11	3.3619	3.5845	0.9379	1.5434	-4.8312
12	3.4963	3.7474	0.9330	1.5436	-5.4502
13	3.6577	3.9036	0.9370	1.5439	-5.3550
14	3.8226	4.0537	0.9430	1.5441	-5.0137
15	3.9697	4.1990	1.9455	1.5442	-4.9618
16	4.1149	4.3394	0.9435	1.5444	-4.8478
17	4.2542	4.4740	0.9519	1.5445	-4.7663
18	4.3661	4.6156	0.9440	1.5446	-5.1936
19	4.4926	4.7335	0.9670	1.5447	-5.4464
20	4.6249	4.8531	0.9520	1.5448	-5.0557
21	4.7470	4.9796	0.9534	1.5449	-5.0299
22	4.8546	5.0991	0.9522	1.5450	-5.2784

23	4.9586	5.2143	0.9511	0.5451	-5.5347
24	5.11593	5.3273	0.9497	0.5451	-5.8076
25	5.1944	5.4383	0.9551	0.5452	-5.2068
26	5.3167	5.5471	0.9535	0.5452	-4.9930
27	5.4162	5.6538	0.9580	0.5453	-5.1484
28	5.5221	5.7545	0.9590	0.5453	-5.1213
29	5.5958	5.8613	0.9552	0.5454	-5.6879





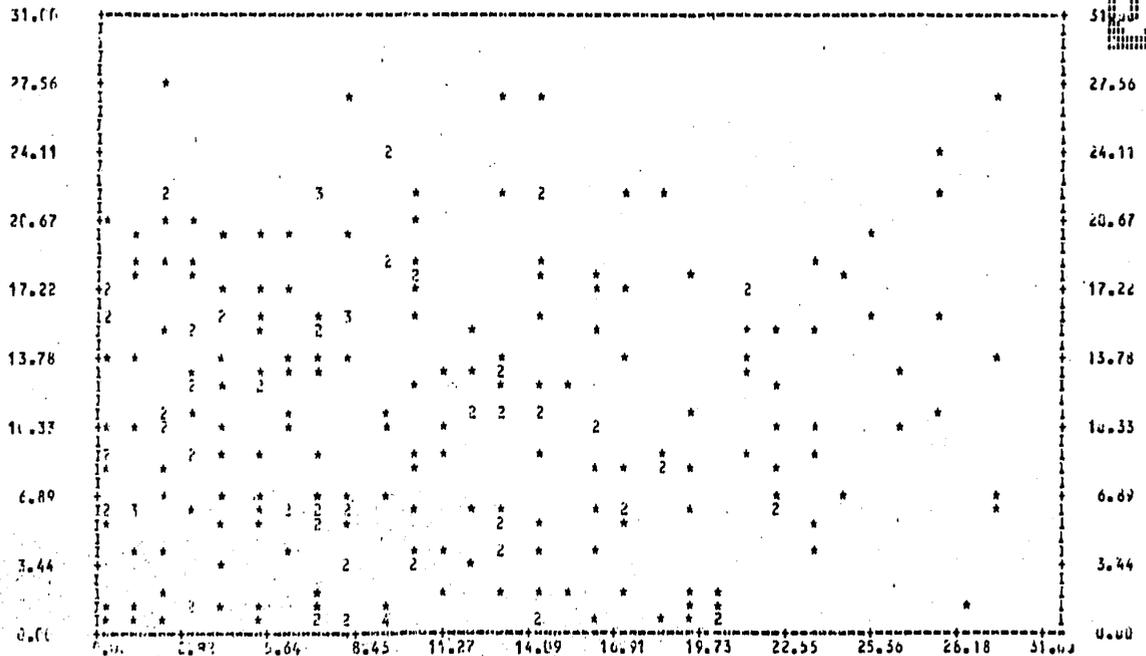
## POISSON/2

	X44	LIMINF	LIMSUP	THMP.	GL	QALFA	Q1-ALFA
1	2.112	1.794	2.277	547.629	512	460.239	565.467
2	4.629	3.713	4.274	1185.631	1024	950.427	1099.279
3	6.375	5.647	6.359	1631.928	1536	1445.093	1628.113
4	8.375	7.592	8.415	2130.712	2048	1943.536	2154.12.
5	11.14	9.544	11.463	2616.064	2560	2443.153	2678.548
6	12.125	11.511	12.507	3104.348	3072	2943.922	3201.784
7	13.760	13.459	14.547	3524.965	3584	3445.590	3724.116
8	16.174	15.422	16.585	4117.481	4096	3947.974	4245.732
9	18.132	17.386	18.621	4416.265	4608	4456.942	4766.764
10	21.157	19.353	20.654	5166.241	5120	4954.399	5287.397
11	22.217	21.321	22.689	5637.479	5632	5458.274	5817.432
12	23.277	23.291	24.716	6112.517	6144	5962.510	6327.196
13	26.275	25.267	26.745	6667.962	6656	6467.104	6846.642
14	27.151	27.234	28.773	7128.853	7168	6971.899	7365.807
15	28.115	29.217	30.800	7709.396	7680	7476.986	7884.720
16	32.148	31.181	32.926	8229.940	8192	7982.299	8403.407
17	34.221	33.156	34.851	8760.526	8704	8487.819	8921.887
18	36.291	35.131	36.876	9239.225	9216	8993.526	9442.180
19	38.131	37.107	38.900	9761.442	9728	9499.400	9956.500
20	41.472	39.184	41.923	10357.305	10240	10005.446	10476.261
21	42.475	41.161	42.946	10381.195	10752	10511.032	10994.074
22	44.315	42.039	44.969	11343.157	11264	11017.955	11511.751
23	46.264	43.917	46.939	11843.615	11776	11524.416	12029.360
24	48.195	45.896	48.911	12312.271	12288	12030.976	12546.730
25	51.032	47.875	51.831	12961.695	12800	12537.658	13064.648
26	52.789	49.855	53.852	13514.040	13312	13044.446	13581.260

27	54.751	52.935	55.072	14016.171	13924	13551.532	14098.374
28	56.974	54.915	57.191	14585.254	14336	14058.313	14615.393
29	59.523	56.896	59.111	14981.938	14848	14565.583	15132.323



POISSON/2



64

En mi opinión se debería considerar más creíble el resultado obtenido con las estadísticas de Thompson, debido a que la distribución de la estadística es conocida exactamente. Este ejemplo muestra que estas estadísticas discriminan mejor que las de Clark y Evans.

#### EJEMPLO 2

En este ejemplo se generaron 300 puntos distribuidos regularmente sobre un rectángulo de 25x27 unidades.

Como se ve, para el primer vecino más cercano el valor de C indica regularidad, lo mismo que el valor de las estadísticas de Thompson. Para el segundo orden de vecino más cercano el valor de C indica aleatoriedad, mientras que el valor de Thmp marca concentración. Una vez más se muestra que estas estadísticas detectan mejor la verdadera situación en la distribución espacial, pues, claramente, es inadmisibile cualquier aleatoriedad en este patrón.

A partir de este orden, Thmp indica regularidad en todo caso, que es lo que se esperaría, mientras que C tiene algunas fluctuaciones: para los órdenes 6,8,14 y 16 indica, una vez más, aleatoriedad. El índice de Washburn también señaló regularidad (para el primer vecino más cercano).

#### EJEMPLO 3

Se generaron 50 puntos en un rectángulo de 100x44 unidades de la siguiente forma: se generó aleatoriamente un número entre 4 y 20 (utilizando la función RANDOM); este número fue el número de conglomerados en la distribución. Posteriormente se calcularon



REGULARES

3FD PUNTOS

AREA=

672

DENSIDAD= (.4464286

MEDIANA OBSERVADA= 1.0000

MEDIANA ESPERADA= 0.7030

R= 1.4225

	ONS.	ESP.	RCI1	D.S.	C
1	1.0000	0.7483	1.3363	0.3912	6.9696
2	1.0909	1.1225	0.9621	0.4075	-1.1531
3	2.0000	1.4031	1.4254	0.4127	16.0934
4	2.0031	1.6270	1.2237	0.4152	9.8431
5	2.0259	1.8416	1.1033	0.4166	5.0537
6	2.0310	2.1259	1.0273	0.4176	1.4808
7	2.3017	2.1946	1.0438	0.4183	2.8695
8	2.3922	2.2513	1.0174	0.4188	1.0938
9	3.4117	2.4983	0.9659	0.4192	-2.3161
10	2.5025	2.6371	0.9565	0.4195	-3.0644
11	2.9459	2.7639	1.0653	0.4197	4.8372
12	3.0712	2.8948	1.0009	0.4200	4.7152
13	3.1186	3.0154	1.0342	0.4201	2.7579
14	3.1725	3.1314	1.0144	0.4203	1.2058
15	3.3343	3.2432	1.0296	0.4204	2.5657
16	3.3904	3.3513	1.0135	0.4205	1.2090
17	3.3199	3.4561	1.1050	0.4206	9.6900
18	3.8686	3.5577	1.0074	0.4207	8.3015
19	3.9246	3.6565	1.0733	0.4208	7.1566
20	3.9301	3.7528	1.0497	0.4209	4.9761
21	4.2358	3.8466	1.1012	0.4209	10.3922
22	4.2755	3.9312	1.0857	0.4210	9.0065

23	4.3512	4.277	1.6801	6.4211	8.6113
24	4.4487	4.1152	1.7713	6.4211	7.8328
25	4.6064	4.2010	1.6703	6.4211	7.8654
26	4.5137	4.2851	1.1743	6.4212	8.4942
27	4.6507	4.3674	1.1666	6.4212	7.7665
28	4.7123	4.4433	1.1594	6.4212	7.0459
29	4.9773	4.5277	1.1915	6.4215	10.9314





## REGULARES

	XHQ	LINIF	LINSUP	THMP.	GL	QALFA	Q1-ALFA
1	2.375	1.779	2.231	841.498	600	533.553	669.289
2	3.478	3.685	4.325	1043.458	1200	1165.421	1297.421
3	11.221	5.613	6.327	3365.992	1800	1683.837	1919.064
4	11.257	7.552	8.457	3377.212	2400	2265.642	2537.199
5	11.594	9.499	10.511	3478.192	3000	2849.615	3153.229
6	12.211	11.455	12.559	3663.321	3600	3435.121	3767.721
7	14.997	13.416	14.603	4499.209	4200	4021.795	4381.047
8	16.325	15.365	16.645	4897.519	4800	4609.591	4993.451
9	16.462	17.326	18.684	4998.498	5400	5197.741	5605.101
10	18.513	19.299	20.720	5553.887	6000	5786.723	6216.119
11	24.665	21.254	22.755	7399.573	6600	6376.242	6826.599
12	26.944	23.221	24.749	8095.211	7200	6966.229	7436.613
13	27.919	25.189	26.341	8375.710	7800	7556.625	8046.217
14	29.172	27.156	28.852	9751.590	8400	8147.383	8655.458
15	32.172	29.124	30.881	9632.347	9000	8738.467	9264.375
16	33.915	31.099	32.915	9991.307	9600	9329.342	9872.999
17	41.364	33.072	34.939	12409.291	10200	9921.484	10481.358
18	42.524	35.045	36.965	12757.110	10800	10513.567	11089.474
19	43.926	37.013	38.991	13177.859	11400	11105.474	11697.368
20	44.174	38.983	41.017	13312.499	12000	11697.785	12305.056
21	51.939	40.963	43.043	15281.604	12600	12290.287	12912.555
22	51.986	42.943	45.066	15595.764	13200	12882.965	13519.877
23	54.002	44.919	47.093	16218.472	13800	13475.807	14127.035
24	55.688	46.896	49.113	16706.541	14400	14069.804	14734.038
25	57.320	48.873	51.136	17346.079	15000	14661.944	15341.997
26	60.719	51.851	53.159	18215.627	15600	15255.221	15947.629

67	62.482	52.824	59.131	19720.526	16290	15840.626	16554.215
78	64.135	54.817	57.322	19225.425	16336	16442.152	17160.687
29	69.075	56.786	59.223	20992.571	17490	17035.793	17767.049





(también con RANDOM) las coordenadas de un punto para cada conglomerado. A partir de estas coordenadas se generó un conglomerado cuyo número de elementos se obtuvo con una variable aleatoria Poisson con media 9. Los conglomerados se generaron de la misma forma en que se construyeron las coordenadas de los puntos del ejemplo 1, utilizando una Poisson con parámetro 4.

Para los primeros 14 vecinos más cercanos C indica concentración. Es notable, sin embargo, la diferencia entre el valor de la estadística entre los órdenes 3 y 4. Esto señala el paso de una distribución muy concentrada a una que también lo es, pero en mucho menor grado. Con Thmp sucede lo mismo, pero mucho más drásticamente: de una distribución concentrada, el índice, para el valor de orden 4 pasa a indicar una distribución regular.

Este hecho podría indicar que el tamaño promedio de los grupos tal vez sea 4, aunque este número haya sido generado utilizando una variable aleatoria Poisson con media 9.

Para los órdenes entre 1 y 14 C indica concentración. En estos mismos órdenes Thmp tiene fluctuaciones entre Concentración, regularidad (para los órdenes 4, 5, 6 y 7) aleatoriedad (del 8 al 11) concentración (12 y 13) y aleatoriedad (para el 14).

Entre los órdenes 15 y 18 (inclusive), C señala aleatoriedad, mientras que Thmp indica regularidad entre 15 y 17 y aleatoriedad sólo en 18. A partir de este punto, para todos los órdenes ambos índices marcan regularidad.

No es fácil interpretar estos resultados. Tal vez lo más claro sea lo referente al salto entre los órdenes 3 y 4.

Thompson afirma que la utilidad de analizar  $n$  vecinos más cercanos está en que "se incrementa la exactitud para la determinación de la densidad del patrón y ... sería posible detectar heterogeneidad



CONGLOMERADOS/1

SC PUNTOS

AREA=

4400

DENSIDAD= 0.0113636

MEDIANA OBSERVADA= 1.4142

MEDIANA ESPERADA= 4.4064

R= 0.3209

	ODS.	ESP.	R[1]	D.S.	C
1	1.3231	4.2904	0.2321	2.4518	-15.2063
2	2.6045	7.0356	0.3705	2.5539	-19.5974
3	4.2962	8.7945	0.4885	2.5865	-19.7778
4	6.5807	10.2603	0.9333	2.6022	-2.9787
5	9.5177	11.5428	0.9112	2.6114	-4.4854
6	11.1683	12.6971	0.9796	2.6175	-6.6818
7	11.7448	13.7552	0.8538	2.6217	-8.7795
8	12.2809	14.7377	0.3333	2.6249	-11.7227
9	12.8043	15.6593	0.2177	2.6273	-12.4518
10	12.5188	16.5287	0.8179	2.6293	-13.1259
11	14.1446	17.3552	0.9157	2.6309	-13.9964
12	14.6553	18.1441	0.8077	2.6322	-15.2055
13	15.6359	18.9171	0.3273	2.6333	-14.2234
14	18.3714	19.6277	0.9360	2.6342	-5.4703
15	20.5110	20.3279	1.0791	2.6351	0.8014
16	21.0486	21.0155	1.0720	2.6358	0.1875
17	21.5150	21.6620	0.9932	2.6364	-0.6399
18	22.1817	22.2391	0.9947	2.6370	-0.5113
19	22.6163	22.8125	1.0374	2.6375	3.0380
20	24.2653	23.5216	1.0313	2.6379	3.2552
21	25.1472	24.1197	1.0339	2.6383	4.0816
22	27.7302	24.6157	1.1234	2.6387	13.2615
23	28.3548	25.2447	1.1232	2.6390	13.5274

24	21.7026	25.7935	1.2291	2.6393	25.7194
25	33.0054	26.3309	1.2535	2.6396	29.0496
26	34.6550	26.8575	1.2920	2.6399	34.1269
27	35.4443	27.3740	1.2943	2.6401	35.1212
28	39.0355	27.8009	1.4301	2.6403	48.5414
29	39.5317	28.3768	1.3930	2.6405	48.5321





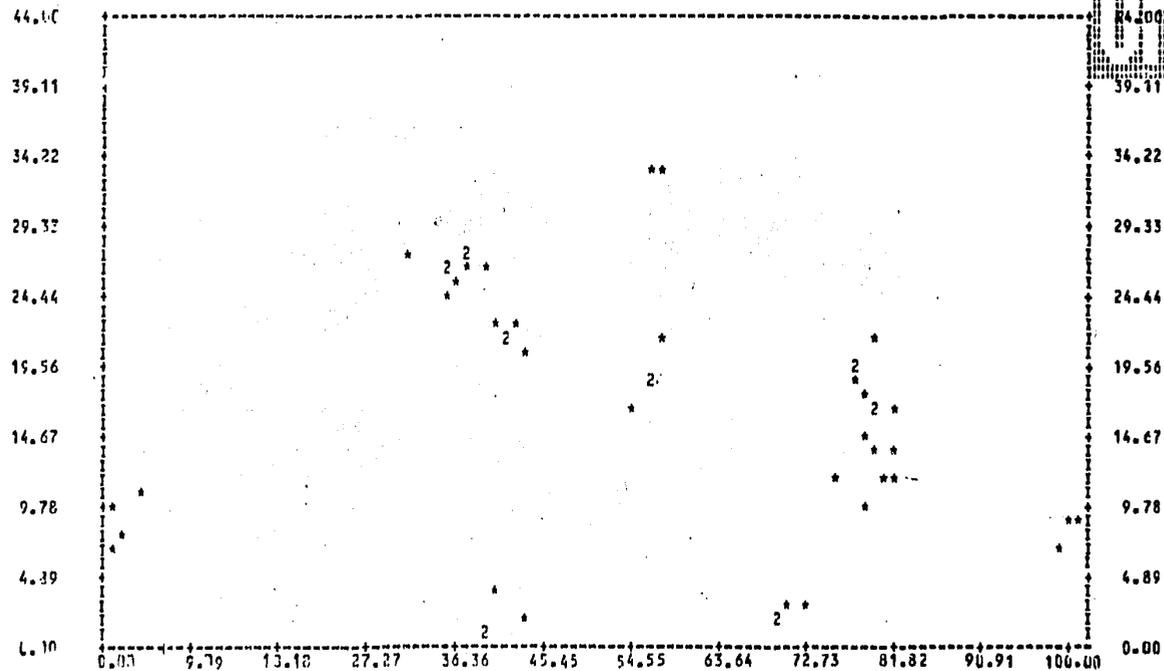
## CONGLOMERADOS/1

	XNB	LIFINF	LINSUP	THMP.	CL	QALFA	01-ALFA
1	0.222	1.475	2.591	11.138	100	73.772	129.070
2	0.407	3.247	4.811	40.341	200	162.270	240.572
3	2.720	5.069	6.908	136.017	300	253.451	349.391
4	13.110	6.927	9.136	655.522	400	346.018	456.823
5	14.841	8.789	11.267	742.058	500	439.471	563.370
6	10.024	10.671	13.386	801.178	600	533.553	669.289
7	16.746	12.562	15.495	837.306	700	628.111	774.731
8	17.503	14.461	17.596	875.148	800	723.045	879.796
9	18.393	16.366	19.691	919.630	900	818.288	984.553
10	19.575	18.276	21.791	978.749	1000	913.789	1089.053
11	20.774	20.190	23.867	1038.725	1100	1009.510	1193.332
12	22.074	22.103	25.948	1103.699	1200	1105.421	1297.421
13	22.725	24.030	28.027	1186.237	1300	1201.499	1401.342
14	25.427	25.955	30.132	1471.334	1400	1297.726	1505.116
15	33.712	27.332	32.175	1685.607	1500	1394.085	1608.757
16	35.159	29.811	34.246	1757.935	1600	1490.564	1712.278
17	36.354	31.743	36.314	1917.697	1700	1587.151	1815.691
18	38.334	33.677	38.380	1916.300	1800	1683.837	1919.004
19	42.474	35.612	40.445	2123.717	1900	1780.614	2022.227
20	44.735	37.551	42.507	2236.957	2000	1877.475	2125.367
21	47.407	39.483	44.569	2370.332	2100	1974.413	2228.428
22	60.710	41.428	46.628	3035.493	2200	2071.424	2331.418
23	63.692	43.370	48.687	3194.575	2300	2168.502	2434.340
24	77.713	45.313	50.744	3885.650	2400	2265.642	2537.199
25	84.200	47.257	52.800	4210.020	2500	2362.842	2640.000
26	92.781	49.202	54.855	4639.061	2600	2460.097	2742.745
27	06.675	51.143	56.910	4877.769	2700	2557.404	2845.437

28	117.344	53.095	58.962	5867.210	2800	2654,761	2948.081
29	127.029	55.043	61.014	6001.442	2900	2752,165	3050.677



CONGLOMER ADCS/1



44.00  
39.11  
34.22  
29.33  
24.44  
19.56  
14.67  
9.78  
4.89  
0.00

en una escala mayor (de la que se detectaría) utilizando sólo al primer vecino más cercano" (1956, p. 391-392). Este mismo autor no muestra más análisis de sus ejemplos más que el siguiente: "... si están concentrados en conglomerados distribuidos aleatoriamente, entonces, dependiendo del tamaño de estos conglomerados esperaríamos tener desviaciones de la aleatoriedad para luego aproximarse a esta" (ibid. p. 393).

#### EJEMPLO 4

Se tienen 88 puntos de la misma forma que en el ejemplo anterior y en la misma superficie. Se observa un patrón semejante al anterior: para los órdenes 1-18 (para C) y 1-9 (para Thmp) se tienen valores de ambos índices que indican concentración.

Como en el ejemplo anterior, hay un brinco entre los órdenes 8 y 9 que es señalado por C con una disminución de -27 a -14, y por Thmp de 624 a 1429. Esto podría decir algo sobre el tamaño promedio de los conglomerados, que en este caso sí coincide con el parámetro de la Poisson utilizado. Para el orden 10, Thmp muestra aleatoriedad, manteniéndola hasta el orden 12. Para todos los siguientes órdenes, Thmp indica regularidad. C tiene indicación de aleatoriedad en los órdenes 19 y 20, manteniendo la regularidad para los siguientes.

En general, puede afirmarse que sería necesario construir un contraste de hipótesis para probar ideas acerca del tamaño promedio de los conglomerados. Esto se dificulta debido a la falta de caracterización de la función mínimo, por lo que, por el momento, sólo se dispone de estos modelos para ser utilizados en forma

EE PUNTOS

AREA=

4499

DENSIDAD= 0.0200000

MEDIANA OBSERVADA= 1.7971

MEDIANA ESPERADA= 3.3214

R= 0.5140

	OBS.	ESP.	REII	D.S.	C
1	1.6587	3.5355	0.4691	1.8481	-12.9513
2	2.7353	5.3033	0.5153	1.9231	-17.3623
3	3.3023	6.6291	0.4982	1.9497	-22.3505
4	4.1949	7.7349	0.5424	1.9615	-23.7049
5	4.9078	8.7007	0.5641	1.9685	-25.3599
6	5.4029	9.5708	0.5645	1.9730	-27.8349
7	6.2338	10.3684	0.6012	1.9762	-27.5900
8	6.9232	11.1090	0.6232	1.9786	-27.9148
9	9.6689	11.8033	0.8192	1.9804	-14.2276
10	10.8648	12.4590	0.8720	1.9819	-10.6230
11	11.5577	13.0829	0.8035	1.9831	-10.1539
12	12.3714	13.6765	0.9046	1.9841	-8.6926
13	13.1001	14.2465	0.9195	1.9849	-7.6329
14	13.8001	14.7944	0.9323	1.9856	-6.6190
15	14.5523	15.3223	0.9497	1.9862	-5.1283
16	15.2834	15.8335	0.9653	1.9868	-3.6615
17	15.8124	16.3233	0.9604	1.9873	-3.4332
18	16.3327	16.8036	0.9717	1.9877	-3.1663
19	17.4004	17.2755	1.0072	1.9881	0.8309
20	17.8786	17.7301	1.0084	1.9884	0.9880
21	18.5609	18.1733	1.0213	1.9887	2.5778
22	18.9405	18.6060	1.0180	1.9890	2.2249
23	19.4815	19.0239	1.0233	1.9892	3.0104

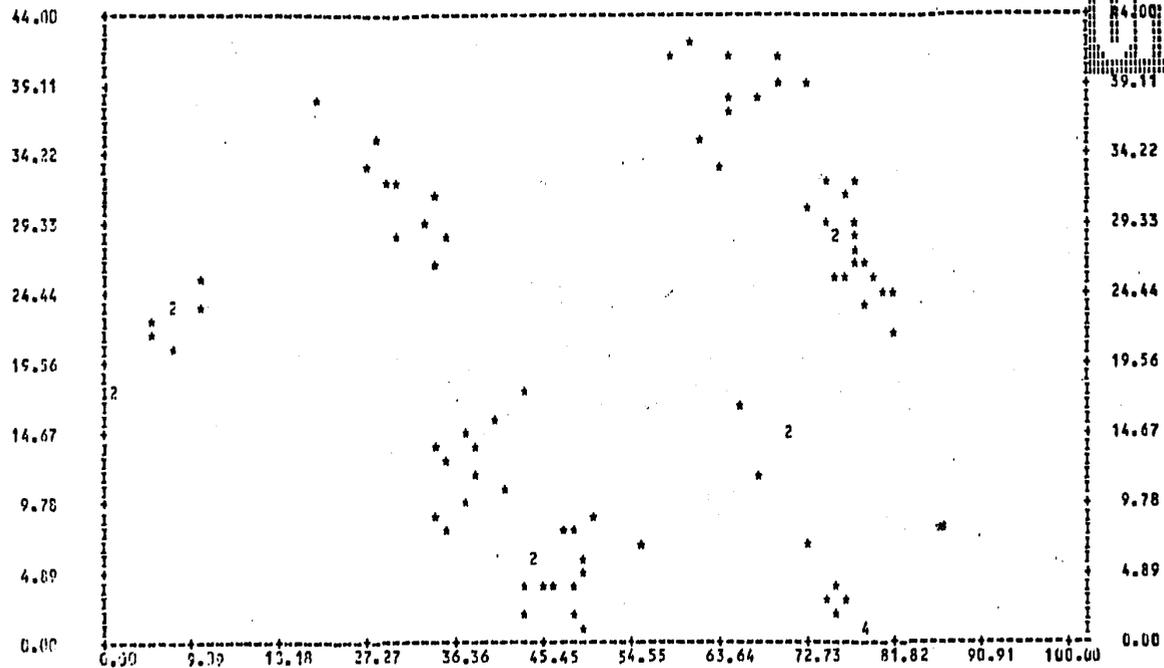
25	20.6254	19.8476	1.0392	1.9897	5.1726
26	21.5907	20.2046	1.0363	1.9899	11.6119
27	22.4668	20.6339	1.0383	1.9900	12.1887
28	23.4101	21.0169	1.1234	1.9902	17.2493
29	24.1617	21.3913	1.1295	1.9904	18.4214

	XHB	LIMINF	LIMSUP	THMP.	GL	QALFA	01-ALFA
1	0.544	1.599	2.433	47.878	176	140.700	214.141
2	1.184	3.426	4.637	104.175	352	301.453	405.389
3	1.681	5.293	6.740	147.906	528	465.759	593.083
4	2.665	7.181	8.852	234.488	704	631.901	778.940
5	3.787	9.082	10.950	333.260	880	799.218	963.624
6	4.455	10.993	13.039	392.071	1056	967.367	1147.474
7	5.868	12.911	15.122	516.352	1232	1136.149	1330.693
8	7.055	14.834	17.198	624.674	1408	1305.430	1513.412
9	10.239	16.763	19.270	1429.048	1584	1475.120	1695.722
10	20.018	18.695	21.337	1761.554	1760	1645.151	1877.690
11	22.201	20.630	23.402	1953.694	1936	1815.475	2059.367
12	25.198	22.569	25.464	2216.582	2112	1986.051	2240.791
13	27.642	24.510	27.523	2432.472	2288	2156.840	2421.993
14	29.984	26.453	29.580	2638.561	2464	2327.843	2602.998
15	32.758	28.398	31.634	2882.725	2640	2499.014	2783.828
16	35.474	30.345	33.687	3121.738	2816	2670.342	2964.499
17	37.582	32.293	35.739	3307.217	2992	2841.815	3145.027
18	39.506	34.243	37.789	3476.486	3168	3013.419	3325.423
19	42.951	36.195	39.837	3783.232	3344	3185.144	3505.698
20	45.302	38.147	41.885	3936.555	3520	3356.979	3685.862
21	48.475	40.101	43.931	4265.760	3696	3528.918	3865.924
22	50.213	42.056	45.976	4418.713	3872	3700.952	4045.890
23	52.912	44.012	48.020	4656.217	4048	3873.075	4225.766
24	55.495	45.969	50.063	4883.543	4224	4045.282	4405.560
25	58.512	47.927	52.105	5149.070	4400	4217.567	4585.275
26	65.068	49.886	54.147	5725.992	4576	4389.926	4764.916
27	47.044	51.845	56.189	6000.000			

09

28	14.710	55.000	20.227	6574.821	4928	4734.847	5123.995
29	78.254	55.766	63.266	6336.371	5104	4907.403	5303.439

CONGLOMERATES/2



44.00  
39.11  
34.22  
29.33  
24.44  
19.56  
14.67  
9.78  
4.89  
0.00

heurística.

Una posibilidad para afinar esto sería experimentar con los modelos, intentando descubrir patrones para algunas clases de distribuciones espaciales "típicas".

#### EJEMPLO 5

Para mostrar el funcionamiento de la generalización propuesta en este trabajo se presenta este ejemplo. Considérese un cubo, de manera que en los puntos medios de cada arista, en los puntos medios de cada plano y en cada vértice se tiene una medida en una cuarta coordenada. Los valores para esta coordenada se generaron con una distribución Poisson con parámetro 15.

El "área" se consideró de la siguiente forma: el cubo tiene un volumen de 216 (i.e., 6 unidades por arista) unidades cúbicas, y los puntos de la cuarta coordenada se conceptualizaron como dentro de un plano de "área" 40. Esto da un total de 8640 unidades a la cuarta. En este caso, C midió mejor lo que estaba sucediendo con la distribución. Para los primeros tres órdenes indicó aleatoriedad, pasando después a la regularidad para todos los siguientes.

En cambio, Thmp marcó, para todos los órdenes, concentración. Este ejemplo muestra que la afirmación de que es mejor utilizar a Thmp no es cierta en general.



(LPC(4-14)

17 PUNTO

18 RA

19 TOS

DESIORAR 1115677

REPLANT DEB 2.1747

ALICHA 2687000 11.5635

PR 1.2137

	CRG.	CRP.	CRIT	CRP.	C
1	5.7744	5.7744	1.534	1.962	-2.4912
2	7.1771	7.1771	1.037	1.7619	-2.5773
3	8.3721	8.3721	1.067	1.636	-1.5714
4	9.667	9.667	1.277	1.5368	1.8874
5	11.0656	11.0656	1.1713	1.411	4.2713
6	12.452	12.452	1.2433	1.2757	9.1646
7	13.8381	13.8381	1.311	1.1515	11.2566
8	15.2244	15.2244	1.3713	1.0231	14.1814
9	16.6107	16.6107	1.4274	0.897	17.7241
10	18.0074	18.0074	1.4773	0.7717	21.9762
11	19.4044	19.4044	1.5243	0.6415	26.9751
12	20.8012	20.8012	1.5671	0.5118	31.8543
13	22.1976	22.1976	1.6063	0.4714	36.6711
14	23.5938	23.5938	1.6419	0.4604	41.4962
15	24.9897	24.9897	1.6741	0.4544	46.3267
16	26.3851	26.3851	1.7027	0.4477	51.1729
17	27.7801	27.7801	1.728	0.4412	56.0259
18	29.1746	29.1746	1.7514	0.435	60.8887
19	30.5684	30.5684	1.7719	0.4281	65.7619
20	31.9615	31.9615	1.7897	0.4226	70.6457
21	33.3538	33.3538	1.8053	0.4174	75.5396
22	34.7453	34.7453	1.8187	0.4128	80.4433
23	36.136	36.136	1.8297	0.4079	85.3477
24	37.5258	37.5258	1.8382	0.4036	90.2527

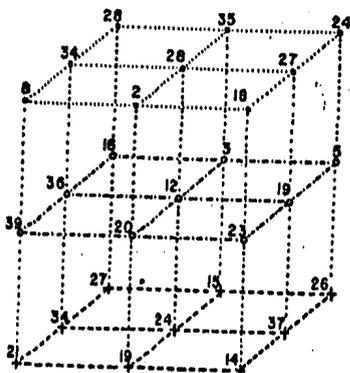
25	26.8103	7.9444	3.7600	1.3994	166.0847
26	27.2847	8.1371	3.4710	1.3935	159.1371





(U.S.G.-111)

	1961	1962	1963	1964	GL	QALFA	01-ALFA
1	1,000	1,000	1,000	1,000	54	35,146	75,695
2	1,000	1,000	1,000	1,000	103	30,682	133,164
3	1,000	1,000	1,000	1,000	162	120,193	198,646
4	1,000	1,000	1,000	1,000	211	176,730	259,111
5	1,000	1,000	1,000	1,000	270	225,917	316,625
6	1,000	1,000	1,000	1,000	329	275,566	375,276
7	1,000	1,000	1,000	1,000	378	325,535	433,276
8	1,000	1,000	1,000	1,000	422	375,942	490,999
9	1,000	1,000	1,000	1,000	466	426,745	548,492
10	1,000	1,000	1,000	1,000	509	477,129	605,603
11	1,000	1,000	1,000	1,000	554	527,853	662,948
12	1,000	1,000	1,000	1,000	598	578,848	719,534
13	1,000	1,000	1,000	1,000	642	629,816	776,836
14	1,000	1,000	1,000	1,000	686	681,232	833,609
15	1,000	1,000	1,000	1,000	729	732,557	890,285
16	1,000	1,000	1,000	1,000	773	783,969	946,873
17	1,000	1,000	1,000	1,000	816	835,405	1,003,381
18	1,000	1,000	1,000	1,000	859	887,225	1,059,917
19	1,000	1,000	1,000	1,000	902	938,656	1,116,185
20	1,000	1,000	1,000	1,000	945	990,549	1,172,492
21	1,000	1,000	1,000	1,000	988	1,042,199	1,228,742
22	1,000	1,000	1,000	1,000	1,031	1,093,902	1,284,939
23	1,000	1,000	1,000	1,000	1,074	1,145,755	1,341,187
24	1,000	1,000	1,000	1,000	1,117	1,197,657	1,397,185
25	1,000	1,000	1,000	1,000	1,160	1,249,595	1,453,247
26	1,000	1,000	1,000	1,000	1,203	1,301,575	1,509,266



## EJEMPLO 6

Con el objeto de examinar el comportamiento de las estadísticas en situaciones en que el proceso aleatorio que generó la distribución espacial no fue un proceso de Poisson se utilizó la distribución de probabilidad uniforme en el intervalo  $(0,1)$  cerrado.

Se generaron 120 puntos, obteniéndose sus coordenadas a partir de la función RANDOM de la Burroughs B7800, multiplicando estos números (distribuidos uniformemente en el intervalo  $(0,1)$  cerrado) por las longitudes de los lados de un rectángulo de  $100 \times 44$ .

La estadística de Washburn indica un valor adecuado para una distribución aleatoria, lo mismo que las estadísticas C y Thmp para los primeros dos vecinos más cercanos. El hecho de que la mayor diferencia entre cambios de orden de vecino más cercano se de entre el tercero y el segundo en ambas estadísticas puede interpretarse como la existencia de conglomerados de tamaño promedio 3 dentro de la distribución. Los demás valores de C y Thmp indican regularidad.



UNIFORME/1

120 PUNTOS

AREA=

4400

DENSIDAD= 0.0272727

MEDIANA OBSERVADA= 3.4853

MEDIANA ESPERADA= 2.8443

R= 1.2256

	CNS.	ESP.	RCI3	D.S.	C
1	3.1644	3.0277	1.0452	1.5826	1.1908
2	4.6657	4.5415	1.0274	1.6486	1.0602
3	6.1979	5.6768	1.0913	1.6696	4.4177
4	7.4451	6.6230	1.1241	1.6797	6.9450
5	8.2753	7.4509	1.1113	1.6857	6.9857
6	9.1865	8.1959	1.1239	1.6896	8.3476
7	9.9023	8.8739	1.1153	1.6923	8.6178
8	10.5202	9.5131	1.1059	1.6944	8.4752
9	11.2555	10.1077	1.1136	1.6959	9.6547
10	11.9133	10.6693	1.1166	1.6972	10.4604
11	12.4204	11.2027	1.1087	1.6982	10.2360
12	12.6615	11.7119	1.1101	1.6991	10.8372
13	13.5324	12.1999	1.1092	1.6998	11.1935
14	14.0560	12.6692	1.1095	1.7004	11.6505
15	14.6633	13.1216	1.1175	1.7019	12.9454
16	15.1934	13.5500	1.1205	1.7014	13.7262
17	15.7049	13.9827	1.1232	1.7016	14.4607
18	16.1926	14.3940	1.1250	1.7021	15.1019
19	16.8085	14.7939	1.1362	1.7025	16.9146
20	17.3572	15.1831	1.1432	1.7028	18.2511
21	18.0081	15.5627	1.1571	1.7030	20.5274
22	18.5857	15.9333	1.1667	1.7033	22.2973
23	19.1246	16.2954	1.1736	1.7035	23.7463

25	20.1720	16.9265	1.1960	1.7039	26.6491
26	21.6683	17.3364	1.1922	1.7040	27.9605
27	21.2575	17.6693	1.2337	1.7042	30.1059
28	21.7962	17.9970	1.2111	1.7043	31.8797
29	22.3565	18.3184	1.2204	1.7045	33.8827

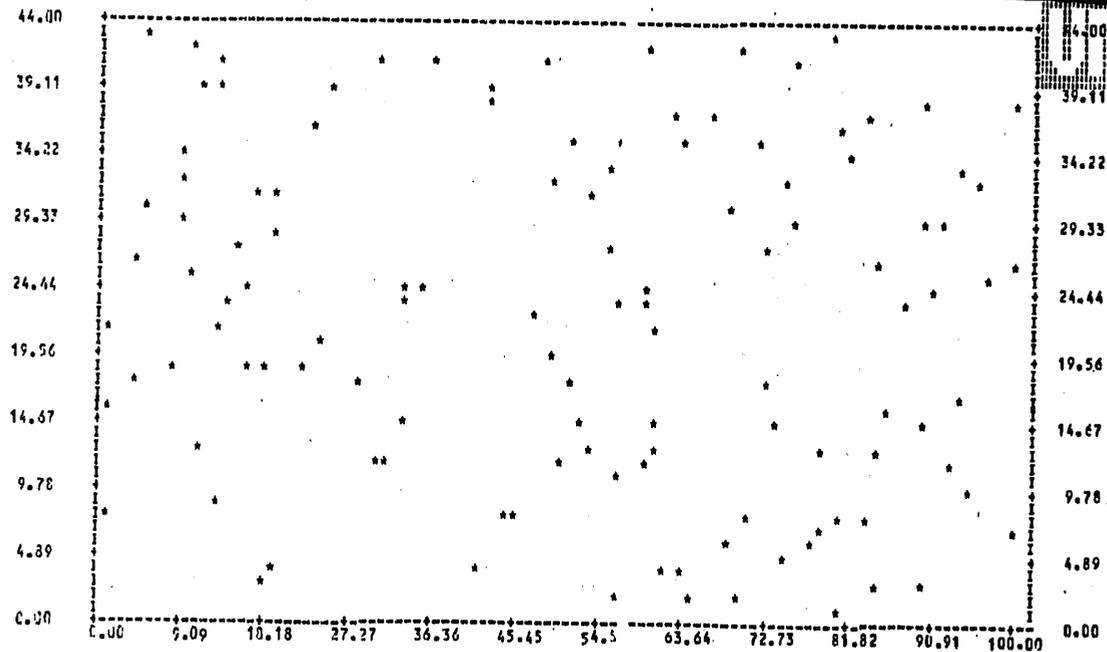


## UNIFORME/1

	XNB	LIMINF	LIMSUP	TIMP.	GL	ALFA	Q1-ALFA
1	2.089	1.654	2.369	250.675	240	198.524	284.317
2	4.212	3.536	4.518	505.383	480	420.724	542.118
3	7.224	5.392	6.631	866.871	720	647.070	795.772
4	10.290	7.296	8.727	1234.805	960	875.560	1047.281
5	12.443	9.212	10.812	1493.444	1200	1105.421	1297.421
6	15.301	11.135	12.808	1836.137	1440	1336.254	1546.587
7	17.736	13.065	14.958	2129.369	1680	1567.825	1795.016
8	19.881	15.000	17.024	2385.722	1920	1799.980	2042.862
9	22.677	16.939	19.085	2721.239	2160	2032.611	2290.230
10	25.328	18.890	21.143	3039.315	2400	2265.642	2537.199
11	27.467	20.825	23.199	3298.430	2640	2499.014	2783.828
12	30.021	22.772	25.251	3602.465	2880	2732.680	3030.161
13	32.411	24.722	27.302	3889.303	3120	2966.606	3276.236
14	34.852	26.673	29.351	4187.100	3360	3200.761	3522.081
15	38.063	28.626	31.298	4567.552	3600	3435.121	3767.721
16	40.839	30.581	33.443	4900.700	3840	3669.666	4013.175
17	43.567	32.526	35.407	5228.071	4080	3904.380	4258.462
18	46.268	34.494	37.530	5552.212	4320	4139.246	4503.595
19	49.839	36.452	39.572	5980.677	4560	4374.254	4748.588
20	53.139	38.412	41.612	6376.635	4800	4609.391	4993.451
21	57.173	40.372	43.652	6860.706	5040	4844.648	5238.193
22	60.843	42.333	45.697	7301.381	5280	5080.017	5482.825
23	64.316	44.296	47.723	7717.165	5520	5315.490	5727.351
24	68.021	46.259	49.765	8162.374	5760	5551.061	5971.781
25	71.458	48.225	51.801	8575.013	6000	5786.723	6216.119
26	75.052	50.187	53.836	9006.251	6240	6022.470	6460.371
27	78.275	52.152	55.871	9512.992	6480	6258.299	6704.543
28	82.262	54.110	57.905	9991.453	6720	6494.204	6948.637
29	87.616	56.085	59.939	10513.923	6960	6730.182	7192.659



UNIFORME/1



## II.6) METODO DE LA MEDIANA

Otra modificación al método de Clark y Evans es la propuesta por Washburn (1974). La diferencia con el método original consiste en que en este la medida usada para comparar la distribución observada con un patrón aleatorio es la media, mientras que Washburn utiliza la mediana.

La ventaja de este método es evidente: cuando se tiene una distribución sesgada o bimodal puede decirse que, en general, la mediana es la medida de tendencia central que mejor representa los datos.

Tal es el caso de las distribuciones de frecuencia obtenidas por Washburn para las distancias al vecino más cercano entre sitios arqueológicos en Nuevo México.

En esta sección expongo la derivación de la fórmula para la mediana esperada en patrones aleatorios y el cálculo del valor máximo del índice

$$R = \frac{r_m}{r_a} ,$$

siendo  $r_m$  la mediana esperada y  $r_a$  la mediana observada. Cabe mencionar que Washburn presenta, además, un intervalo de confianza para la mediana poblacional (basado en la distribución normal) y la modificación de las pruebas de análisis de varianza para varios patrones sugerida por Clark y Evans (1954). Sin embargo, éstos no serán analizados en este trabajo.

Otra posibilidad es modificar los resultados obtenidos en las secciones anteriores trabajando con la mediana en lugar de la media.

Se sabe que  $\exp(-\rho\pi r^2/k)$  es la probabilidad de que en un área de tamaño  $\pi r^2/k$  no se encuentre más que un solo punto, siendo la densidad de la población. Puesto que esta expresión puede considerarse como la función de distribución de las distancias al vecino más cercano que son mayores o iguales que  $r$ , se puede escribir:

$$F(r) = \exp(-\rho\pi r^2/k).$$

Por definición, la mediana de una distribución es el punto  $r_m$  tal que  $F(r_m) = .5$ . Entonces, se tiene que:

$$F(r_m) = \exp(-\rho\pi r_m^2/k) = .5, \text{ y despejando a } r_m:$$

$r_m = (k \ln(2)/\rho\pi)^{1/2}$ , considerando sólo a la parte positiva de la raíz por no tener sentido las distancias negativas.

Como en el modelo original,  $R$  alcanzará su máximo valor cuando el patrón observado sea de tipo hexagonal. Por un argumento análogo al presentado en la sección II.2 se tiene que:

$$\rho = \frac{2}{r_u^2} \cdot 3, \text{ y entonces, } r_u = (2/\rho \sqrt{3})^{1/2},$$

donde  $r_u$  es el valor que se esperaría para la mediana de las distancias al vecino más cercano si el patrón fuera aleatorio.

Entonces:

$$R = \frac{r_u}{r_a} = \frac{\rho^{-1/2} \sqrt{2} \sqrt[4]{3}}{\pi^{-1/2} \rho^{-1/2} (\ln(2))^{1/2}} = 2.2876885.$$

Finalmente, habría que señalar que Stark y Young (1981) y Young (1982) discuten una variante de la estadística de Clark y Evans para espacios de dimensión 1. La estadística propuesta está basada en estadísticas de orden y se presenta como análoga al modelo original de Clark y Evans.

Young (1982) obtiene los primeros dos momentos de la estadística así como ajustes con densidades beta y normal para datos simulados, afirmando que aparentemente la distribución es unimodal y simétrica.

Un trabajo que queda pendiente es comparar la estadística propuesta por Young con la obtenida para dimensión 1 en la sección II.5 con el objeto de indagar si efectivamente son equivalentes. De ser cierto esto último, habría que explicar porque la forma en que se supone al patrón aleatorio en el modelo de Clark y Evans (utilizando un modelo de Poisson) genera resultados iguales que la utilizada por Young (uniforme), lo que lleva a problemas de interpretación del término "aleatoriedad".

Otra posibilidad es, empleando el método con el que Young construye la estadística del vecino más cercano lineal, intentar construir generalizaciones para  $n$  vecinos más cercanos en espacios de dimensión  $k$  mayor que uno, comparando los resultados con los obtenidos en la sección II.5.

## III.1) INDICE A.

Un problema relacionado con las medidas para desviaciones de la aleatoriedad en patrones puntuales es el establecer si dos distribuciones espaciales situadas en una misma área tienen alguna asociación. Este problema ha sido tratado por Pielou (1961,1977) y por Greig-Smith (1964).

Según Hodder y Okell (1978, p.97) (en adelante, esta referencia se denotará como H&O) (citando a Hodder y Orton, 1976), los métodos propuestos por Pielou y Greig-Smith tienen tres desventajas:

i) Dependencia de escala: como se vió en la sección II.3, los modelos de vecino más cercano dependen del orden de vecino considerado (Pielou se basó en estas distancias para proponer su índice S de asociación espacial en 1961), y los métodos de cuadrantes dependen de la elección del tamaño de cuadrante (Greig-Smith, 1964).

ii) Uso de una cantidad restringida de información: normalmente, estos modelos sólo utilizan uno o dos aspectos de la distribución (véase la sección I.3).

iii) Los índices propuestos proporcionan poca información acerca del tipo de asociación o segregación existente.

En vista de esto, Hodder y Okell propusieron un modelo para establecer la asociación entre distribuciones que "no depende de la escala, utiliza los datos disponibles en una forma más completa y proporciona mayor información acerca de las distribuciones" (H&O, p.97). Hay que hacer notar que la no dependencia de escala se refiere únicamente a que no se consideraron órdenes de vecinos

más cercanos ni cuadrantes, más bien que a la dependencia de escala que menciona Harvey (1969).

El modelo funciona de la siguiente forma: supóngase que en una región cualquiera se tienen dos tipos de puntos: A y B. Sean  $n$  y  $m$  el número de puntos del tipo A y del tipo B respectivamente. Se definen entonces las siguientes estadísticas:

$$\bar{r}_{aa} = \frac{\sum_{i=1}^n \sum_{j>i}^n r_{ij}}{(.5)n(n-1)},$$

$$\bar{r}_{bb} = \frac{\sum_{i=1}^m \sum_{j>i}^m r_{ij}}{(.5)m(m-1)},$$

$$\bar{r}_{ab} = \frac{\sum_{i=1}^n \sum_{j=1}^m r_{ij}}{mn},$$

siendo  $r_{ij}$  la distancia entre los puntos  $i$  y  $j$ .

De esta forma se consideran las tres clases de distancias que puede haber entre puntos de dos distribuciones: distancias entre puntos tipo A, entre puntos tipo B y entre puntos de tipo A con puntos de tipo B. Las fórmulas que presento aquí son ligeramente distintas que las dadas por H&O. Sin embargo, si se calcula  $\bar{r}_{aa}$  y  $\bar{r}_{bb}$  con las expresiones aquí obtenidas se eliminan las distancias de regreso, disminuyendo así el tiempo de máquina necesario para los cálculos. Los resultados son equivalentes a los obtenidos con la forma dada por H&O; en lo único que varían

las fórmulas es en los (.5) en los denominadores de las dos primeras expresiones y en que los índices para las segundas sumas en  $\bar{r}_{aa}$  y  $\bar{r}_{bb}$  van desde 1 hasta n y m respectivamente, y siendo  $i \neq j$ .

Finalmente se define al índice como:

$$A = \frac{\bar{r}_{aa} \cdot \bar{r}_{bb}}{(\bar{r}_{ab})^2}.$$

El índice  $A$  mide la asociación entre dos distribuciones espaciales. Según H&O,  $A$  tomará valores cercanos a 1 si los puntos tipo A y tipo B están "bien mezclados"; si las distribuciones se encuentran muy separadas y agrupadas en dos grandes grupos,  $A$  valdrá casi 0, y, finalmente, cuando las distribuciones consistan en pares aislados formados por un punto tipo A y un punto tipo B el índice será mayor que 1. Las razones por las que se esperaría ese comportamiento del índice son las siguientes:

En primer lugar, H&O definen lo que se entiende por "distribuciones bien mezcladas": supóngase que se divide el área considerada en dos partes iguales 1 y 2. Sea  $p$  la probabilidad de que un punto tipo A caiga en la parte 1, y sea  $q$  la probabilidad de que un punto tipo B esté en la parte 2. Si se cumple que  $p = q = .5$ , se tiene que la mezcla de las dos distribuciones es perfecta, en el sentido de que: i) se esperaría que no hubiera dos grupos separados para cada tipo de punto, y ii) no deberían tenerse pares aislados formados con puntos de los dos tipos. Es decir, los puntos de ambas clases estarían colocados aleatoriamente con igual probabilidad de recibir un punto de cualquier tipo



para todo lugar del área. En estas condiciones, las distancias dentro de los puntos tipo A serían iguales a las distancias entre los puntos tipo A y los puntos de tipo B, e igualmente, las distancias dentro de los puntos de tipo B serían iguales a las distancias calculadas entre los puntos de diferente clase, lo que causaría que  $\bar{r}_{aa} = \bar{r}_{bb} = \bar{r}_{ab}$ , esto es, que  $A = 1$ . Hodder y Okell definen otra medida, llamada  $D$ , y definida como

$$D = \frac{\bar{r}_{aa}}{\bar{r}_{bb}}.$$

Este último índice, llamado índice de dispersión será utilizado más adelante en contrastes de hipótesis para aleatoriedad. Claramente, si ésta existe en la distribución, los valores de  $D$  estarán cerca de 1.

Por otra parte, si los puntos estuvieran divididos en dos grupos excluyentes, de tal forma que en el primero sólo hubiera puntos tipo A y en el segundo sólo puntos tipo B se tendría que  $A$  tendería a 0 conforme los grupos se hicieran más compactos, pues las distancias entre los grupos serían grandes y las distancias dentro de los grupos serían pequeñas. Por último, si las dos clases de puntos estuvieran colocadas en parejas de puntos de diferente clase, la situación sería la inversa de la anterior, lo que implicaría que el índice tomaría valores cada vez más grandes si las distancias entre puntos de distintas clases son menores.

Lo anterior indica un uso descriptivo de ambos índices. A continuación expondré algunos contrastes de hipótesis que es posible plantear.

Una primera hipótesis sería probar que la distribución de

de ambos tipos de puntos es aleatoria. Esto se puede hacer utilizando como estadística de prueba al índice  $D$  y se rechazaría la hipótesis si éste es mayor o menor que los valores de tabla para  $\alpha$  y  $1-\alpha$  respectivamente. Posteriormente, si esta primera hipótesis no fue rechazada, se puede contrastar hipótesis de la forma  $p = p'$ ,  $q = q'$  para algunos valores de  $p'$  y  $q'$  dados en tablas.

Por ejemplo, se podría contrastar que  $p = q = .5$  contra la alternativa de que existe una tendencia hacia el agrupamiento en un solo grupo para cada tipo de puntos o bien hacia pares aislados con puntos de diferente tipo. Si la hipótesis nula es rechazada se puede tratar con otras hipótesis que precisen más la idea sobre la asociación de los puntos. Por ejemplo, podría pensarse en probar que  $p = q = .75$  (que de no rechazarse indicaría una tendencia hacia la separación de los dos tipos de puntos en dos grupos espaciales más o menos diferenciados), o en otras combinaciones de valores para  $p$  y  $q$ .

Para poder efectuar estos contrastes se necesita conocer los comportamientos de los índices suponiendo que las hipótesis nulas (referentes a la aleatoriedad) son verdaderas, es decir, su función de distribución bajo la hipótesis nula. Si ésta se quisiera encontrar por métodos de la estadística matemática habría tres problemas: i) si se asume que hay algún tipo de correlación entre los puntos de ambos tipos, se tiene que "poco se sabe acerca del complejo patrón de correlaciones que hay entre varias distancias" (H&O, p. 98); ii) por otra parte, si se considera que estas distancias son independientes, no siempre es fácil determinar la función de distribución para cocientes y productos de variables aleatorias, y menos si no se asume otra distribución para ellas y iii), la distribución de

los índices debería ser diferente para cada combinación de  $p$  y  $q$ .

Por estas razones se utilizó la simulación (o más bien algunos métodos Monte Carlo) para aproximar esta distribución.

El programa fue el siguiente:

Para  $n:=2,4,6,8,10,15,20,30,40$  hacer

Para  $m:=2,4,6,8,10,15,20,30,40$  hacer

si  $m \geq n$  entonces hacer

$n$  veces hacer

genera patrón;

calcula  $A; D$

ordena;

estadísticas;

En el procedimiento genera patrón se asignaron  $(n \cdot p)$  puntos de tipo A en el lado izquierdo de un rectángulo y  $(n \cdot (1-p))$  puntos de este tipo en el lado derecho de la misma área. Para puntos tipo B los números de puntos asignados en cada mitad del rectángulo fueron  $(m \cdot (1-q))$  y  $(m \cdot q)$  respectivamente.

En el párrafo anterior la notación  $(a)$  indica el máximo entero menor o igual que  $a$ . También es necesario aclarar que para todas las simulaciones se consideró como área en la que se distribuyeron los  $a$  un rectángulo de 20 unidades de largo por 10 de ancho, debido a que esto facilitaba los cálculos y no afectaba a los valores del índice (al menos directamente).

En el procedimiento calcula se utilizó la distancia euclidiana (métrica de Minkowski de orden 2) para construir la matriz de distancias entre puntos. La forma en que se calculó el índice fue la siguiente:

Si  $a_i$  y  $b_i$  denotan puntos de tipo A y B respectivamente, definidos como  $a_i = (a_i^1, a_i^2)$  y  $b_i = (b_i^1, b_i^2)$  y  $d(a_i, b_i)$  es la distancia entre los dos puntos definida como

$$d(a_i, b_i) = \left[ (a_i^1 - b_i^1)^2 + (a_i^2 - b_i^2)^2 \right]^{.5}$$

Entonces:

$$A = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n d(a_i, a_j) \cdot \sum_{i=1}^{m-1} \sum_{j=i+1}^m d(b_i, b_j)}{.5 \cdot n \cdot (n-1) \cdot .5 \cdot m \cdot (m-1)} \cdot \left( \frac{\sum_{i=1}^n \sum_{j=1}^m d(a_i, b_j)}{m \cdot n} \right)^2$$

Análogamente, el índice de Dispersión se define como

$$D = \frac{\bar{r}_{aa}}{\bar{r}_{bb}}$$

Esta forma de efectuar los cálculos reduce notablemente el tiempo de máquina en comparación con la propuesta originalmente por Hodder y Okell.

Utilizando estos dos procedimientos se generaron 200 patrones aleatorios para distintos valores de  $p$  y de  $q$  y se calcularon los valores de  $A$  y de  $D$  para cada patrón generado. Posteriormente estos 200 valores de cada índice fueron ordenados con un algoritmo de tipo burbuja en el procedimiento ordena. Esto permitió encontrar los fractiles .005, .025, .05, .95, .975 y .995 que servirán para contrastar hipótesis con niveles de significancia de .01, .05 y .1, puesto que la prueba es de dos colas.

Finalmente, en el procedimiento estadísticas se calcularon (para cada combinación de  $m$  y  $n$ ) la media, varianza, desviación estándar y coeficiente de variación para  $A$  y  $D$ . Los valores de  $p$  y  $q$  considerados fueron (.5, .5), (1, 1), (.25, .25) y (.5, .75). Para cada combinación se observó que al aumentar el número de puntos (y sobre todo, mientras más cercanos eran  $m$  y  $n$ ) el coeficiente de variación disminuyó. Por ejemplo, para  $p = .5$  y  $q = .5$ , con  $m = n = 2$ , se tiene que  $CV = .31123$ , mientras que si  $m = n = 4$ ,  $CV = .15822$ ; si  $m = 40$ ,  $n = 2$ ,  $CV = .17277$ , y finalmente, si  $n = m = 15$ ,  $CV = .04198$ , lo que

ilustra este hecho. Este mismo patrón se presentó para el índice de dispersión.

Por otra parte, los valores de la media de  $\bar{A}$  (en 200 simulaciones) para diferentes combinaciones de  $p$  y  $q$  cuando  $m$  y  $n$  son grandes vienen a confirmar las proposiciones hechas acerca de los valores extremos del índice. Por ejemplo, para valores de  $\bar{A}$ :

$p$	$q$	$\bar{x}$
.5	.5	1.01693
.5	.75	0.92435
.25	.5	0.71032
1.0	1.0	0.23230

Hay que hacer notar que el hecho de que los valores de  $\bar{A}$  cuando  $p=q = 1$  fueran relativamente mayores que 0 se debe a que aunque todos los puntos tipo A están en una sola mitad del rectángulo y todos los de tipo B están en la otra mitad, no hay un patrón conglomerado para cada mitad, sino un patrón aleatorio. Este patrón fue generado utilizando el generador RANDOM de la Burroughs B-7800 empleando como semilla la hora del día dada en microsegundos. Posteriormente me enteré que era posible mejorar la calidad de las sucesiones de números aleatorios generados por RANDOM si se usaba como semillas números impares cercanos a 524288, 1048576 o 2097152. "Se dice que una semilla es mejor que otra si hace que el período sea suficientemente largo para finalidades prácticas, además de que maximiza la velocidad de computación y la sucesión satisface pruebas de independencia en términos estadísticos" (Aranda y Ruiz-Velasco, 1984, p. 29). La función RANDOM produce números distribuidos uniformemente en el intervalo cerrado  $(0,1)$ . Sin embargo, no incorporé estas sugerencias al programa, pues de haberlo hecho hubiera sido necesario repetir las simulaciones con el consiguiente gasto de tiempo de máquina.

En general la hipótesis nula puede plantearse como:

$H_0$ : ambos tipos de puntos están localizados aleatoriamente con probabilidades  $p$  y  $q$  de estar localizados en una u otra mitad de la región.

Esta hipótesis se escribirá como  $H_0: p = p ; q = q$

Para Hodder y Okell la forma de contrastar esta hipótesis sería la siguiente: se calcula el valor de  $A$  y se compara contra los límites (para algún nivel de significancia dado) con los valores particulares de  $p$  y  $q$ . "Entonces, por construcción de estos límites, se esperaría rechazar  $H_0$  cuando esta es cierta una de cada  $(1 - \alpha) \cdot 100$  veces. Sin embargo, si la hipótesis alterna es verdadera nos gustaría que la probabilidad de rechazarla también fuera pequeña" (Hodder y Okell, 1978, p.98-99). Sin embargo, aunque presentan dos ejemplos de esto, no mencionan la forma en que se estimarían (de manera general) estas probabilidades. En este sentido, creo que sería necesario aproximar la función potencia de la prueba, pero como únicamente he obtenido aproximaciones muy empíricas para ciertos casos, espero tratar este problema posteriormente.

En mi opinión, y dado que las tablas fueron generadas asumiendo que la distribución espacial de ambos tipos de puntos era aleatoria, sería conveniente contrastar la hipótesis de aleatoriedad. Para esto se utilizaría el índice  $D$  con las tablas construidas al efecto. Posteriormente se podría verificar si el patrón no se desvía significativamente de un patrón generado con valores de  $p$  y de  $q$  particulares.

Hodder y Orton consideran otro tipo de hipótesis que consiste en construir una región acotada dentro de la región original y calcular los índices basándose únicamente en las distancias entre cada punto dentro del área acotada y los puntos dentro y fuera de la misma. "El objetivo de este procedimiento es evitar los 'efectos de frontera' (Donnelly, 1978) de las pruebas de vecino más cercano" (véase también el capítulo II sección 1 de esta tesis). (ibid., p. 99). Estos autores afirman que se efectuaron simulaciones en esta forma, y que, en general, los rangos de la región de aceptación fueron menores que los obtenidos sin considerar una región dentro de la región total.

La conclusión de Hodder y Okell es que el uso de esta técnica (descrita como "buffer zone" por Hammond y McCullagh (1978)) reducirá el número de puntos, por lo que para muestras pequeñas será necesario calcular los valores de los índices como en la primera hipótesis.

En este trabajo no se consideraron esta clase de procedimientos ya que no encuentro ninguna utilidad en hacerlo por las siguientes razones: i) este método no resuelve ningún problema muestral, pues de cualquier forma se debe considerar (de una u otra manera) todos los puntos disponibles, aunque sólo se tomen distancias a partir de los situados en el área acotada; y ii) la técnica de crear una zona intermedia entre el área de interés y el área en la que están localizados todos los elementos tiene una utilidad en modelos como los de vecino más cercano que dependen directamente del área considerada. Sin embargo, para el caso de A esta dependencia no es directa, pues su cálculo no involucra explícitamente ningún supuesto de área, excepto que, evidentemente, mientras mayor sea el área habrá posibilidad de tener mayores distancias en ambos tipos de puntos.

En el artículo de Hodder y Okell tampoco se le da mucha importancia a esta discusión, al parecer debido a que "en muchas situaciones arqueológicas el colocar un área acotada en el área total limitará severamente el número de puntos que puedan ser considerados" (ibid. p. 99), más bien que a las dos consideraciones anteriores.

Las tablas presentadas por Hodder y Okell son combinaciones de los siguientes valores:

$\alpha$	$p$	$q$	$n$	$m$	$nsim$
.1	.5	.5	2,4,6,8,10,15,20,30,40	=	200
.1	.75	.5	"	"	200
.05	.5	.5	4,5,6,7,8,,...,14,15	=	500
.05	.75	.5	"	"	500
.1	.5	.75	15 y 8	10 y 20	200
.1	.5	1.0	15 y 8	10 y 20	200
.1	.1	.75	15 y 8	10 y 20	200
.1	.1	1	15 y 8	10 y 20	200
.1	1.0	1.0	15 y 8	10 y 20	200
.1	1.0	.5	15 y 8	10 y 20	200
.1	1.0	.75	15 y 8	10 y 20	200

En este trabajo presento tablas para los fractiles .005, .05, .1, .95, .975 y .995 de  $A$  y  $D$  con  $m$  y  $n = 2,4,6,8,10,15,20,30,40$ ,  $nsim = 200$  y las siguientes combinaciones de  $p$  y  $q$ : (.5,.5), (1,1), (.5,.75), (.25,.25).

Como se dijo anteriormente, las tablas del índice  $D$  proporcionan límites de confianza para seguir sosteniendo la hipótesis de aleatoriedad para ambos tipos de puntos sin considerar su interacción. En caso de aceptarse esta hipótesis convendría calcular  $A$  y efectuar el contraste correspondiente para alguna combinación particular de  $p$  y  $q$ .

Hodder y Okell afirman que "la principal ventaja de este índice para el arqueólogo es que ha sido desarrollado específicamente para situaciones arqueológicas y para problemas particulares de la arqueología" (H&O, p. 106). Lo menos que produce este comentario es sorpresa, pues el desarrollo del índice es (debe ser) independiente de si los puntos representan distribuciones producidas por algún proceso espacial particular. Aún considerando que a lo que se referían los autores sea al hecho de que el desarrollo del índice fue motivado por alguna problemática arqueológica particular

(lo cual aparentemente tampoco es cierto), no se sigue que esto sea la principal ventaja del índice. Afirmarlo sería equivalente a decir que la principal ventaja de la distribución  $t$  de Student para los cerveceros es que ésta surgió a partir de problemas relacionados con la producción de cerveza.

En lo que sí es posible estar de acuerdo con Hodder y Okell es en que las principales desventajas de estos índices se refieren a que su cálculo es complicado (sin la ayuda de una computadora) y de que las tablas para los contrastes fueron generadas empíricamente (ibid., p. 107). Ellos mismos mencionan que es "preferible efectuar otras simulaciones para establecer la verosimilitud (bajo una hipótesis nula de aleatoriedad) de obtener los resultados esperados del índice" (Ibid., p. 107) (el paréntesis es mío).

Algo que es necesario señalar es que los resultados de mis simulaciones no fueron idénticos a los obtenidos por Hodder y Okell: en general, los fractiles superiores (.90 y .95) coinciden casi perfectamente con los dados por H&O; por otra parte, los fractiles superiores resultaron ser casi siempre mayores que los de los autores. Consecuencia de esto es que las regiones de aceptación sean más chicas en mis tablas. Una ventaja de esto es que casi seguramente el coeficiente de variación de mis tablas será menor que el de H&O (aunque no puedo comprobarlo por no tener datos), y este es un criterio para elegir entre dos simulaciones (Kleijnen, 1974). Además, el hecho de que los valores estén más cerca de 1 (para  $p = q = .5$ , i.e. un patrón con dos distribuciones espaciales "bien mezcladas") y más cerca de 0 (para  $p = q = 1$ ) también aporta elementos a favor de estas tablas. Hay que reconocer que la elección del número de simulaciones fue arbitraria y que no se hicieron pruebas con diferentes valores de este parámetro. En resumen, aún queda mucho por hacer con estas distribuciones.

### III.2) APROXIMACIONES A LA DISTRIBUCION DE A BAJO ALGUNAS HIPOTESIS.

Como se vió en la sección anterior hay serios problemas para determinar analíticamente la función de distribución de los valores del índice A bajo hipótesis que impliquen aleatoriedad en algún sentido. La alternativa dada por Hodder y Okell para conocer (aunque fuera en una forma aproximada) esta función y así utilizarla en contrastes de hipótesis es generar aleatoriamente pares de distribuciones espaciales, calcular el índice A para cada patrón resultante y examinar la distribución de frecuencias de los valores del índice en cada serie de simulaciones.

El término "simulación" tiene muchas acepciones. No pretendo por lo tanto presentar todas las interpretaciones que se han hecho de él. En esta sección únicamente menciono algunos usos del término para posteriormente detallar el método utilizado en las aproximaciones a la distribución de A.

Una primera aproximación al término "simulación" podría ser la dada por Sullivan y Schuster (1975), para quienes la simulación es "el modelaje de un proceso por otro proceso". En esta misma línea, Zubrow y Harbaugh (1978, p.115) afirman que "simular es el proceso de realizar experimentos sobre un modelo de la realidad en lugar de i) experimentar directamente con la realidad misma o ii) obtener soluciones analíticas directas.

Por otro lado, Tocher (1969, p.3) define a la simulación como "un conjunto de técnicas para generar realizaciones de una variable aleatoria", y Doran y Hodson (1975, p. 222) restringen la simulación a la operación de procesos estocásticos.

Nótese que en ninguna de estas descripciones se ha mencionado a la computadora como elemento sine qua non de la simulación. Sin embargo, House y MacLeod (1972, citados por Aldenderfer (1981, p.13) afirman que "la simulación se refiere al proceso de construir y usar modelos en una computadora", mientras que Renfrew (1979, p.202) resume: "simulación es, simplemente, una técnica para explorar las implicaciones de un modelo en una forma que, sin una computadora sería muy tediosa".

Otros autores (por ejemplo, Bell, 1981 p. 57) señalan que existen dos clases de modelos de simulación: "por una parte están modelos en los que cada iteración corresponde a la realidad,..., y por otra están los modelos de simulación que intentan simular sólo una salida (outcome) y no intentan imitar a un proceso en diferentes etapas.

Considerando, en principio, estas opiniones acerca de la simulación se puede observar dos cosas: i) la simulación tiene que ver con modelos. No pretendo estudiar a fondo las implicaciones de esta afirmación, aunque esto no quiere decir que niegue la importancia de contar con una definición precisa del término "modelo". Simplemente, se entiende por modelo cualquier representación de la realidad, independientemente de la naturaleza de esta representación. ii) las simulaciones efectuadas para aproximarse a la distribución de A: a) no tienen una solución analítica directa, b) se basan en técnicas para generar realizaciones de variables aleatorias (en particular uniformes), c) sería sumamente tedioso y largo efectuarlas sin computadora y d) únicamente pretenden obtener un resultado, que es independiente de iteraciones "reales" de tiempo.

En este trabajo utilizo el término "simulación" como sinónimo

equivalente de lo que Kleijnen (1974, p. 6) llama "métodos Monte Carlo en un sentido amplio" (los subrayados son de este autor), definidos como "cualquier técnica para la solución de un modelo utilizando números aleatorios".

La distinción entre métodos Monte Carlo y simulación hecha por Kleijnen es importante: "la simulación (en un sentido amplio) es experimentar con modelos respecto al tiempo". Es decir, "la simulación implica experimentación. Sin embargo, en lugar de experimentar con el mundo real experimentamos por medio del modelo del objeto. La conducta del objeto es modelada sobre el tiempo." (Kleijnen, 1974, p.12).

Las simulaciones aquí presentadas se refieren al espacio muestral del experimento aleatorio "modelar patrones espaciales consistentes en dos distribuciones de puntos y calcular el valor del índice A para cada uno de ellos", con el objeto de explorar el comportamiento de la función de distribución del índice bajo ciertas hipótesis nulas.

La forma de contrastar hipótesis en estadística es como sigue: Se proponen dos hipótesis (llamadas nula y alternativa) excluyente y exhaustivas (respecto a un cierto espacio de parámetros) acerca de la función de distribución de la variable aleatoria de interés. Utilizando la información muestral se calcula el valor de una cierta función llamada estadística de prueba. Claramente esta estadística es una variable aleatoria, pues es una función de los valores muestrales, que, por definición, son realizaciones de una variable aleatoria. Esta función cumple con las siguientes características: i) mientras mayor es el valor de la estadística de prueba, la evidencia de discrepancia con la hipótesis nula es más fuerte, y b) si la hipótesis nula es cierta, entonces la

distribución de la estadística de prueba es conocida, al menos aproximadamente. (Cox & Hinkley, 1978, p. 19).

Como lo que se pretende es inferir con base en la información muestral acerca del comportamiento de la variable en la población y esta información es, por definición, incompleta, existe el riesgo de cometer dos clases de errores:

El error tipo I consiste en rechazar la hipótesis nula ( $H_0$ ) cuando esta es, en la realidad, correcta; el error tipo II se comete al aceptar una hipótesis nula que en realidad es falsa. A las probabilidades de incurrir en los errores tipo I y II se les llama  $\alpha$  y  $\beta$  respectivamente. Lo óptimo sería que al hacer un contraste de hipótesis se pudiera minimizar ambas probabilidades. Sin embargo, "cuando el número de observaciones es dado, ambas probabilidades no pueden controlarse simultáneamente" (Lehmann, 1959, p. 61). De hecho, cuando  $\alpha$  decrece  $\beta$  crece y viceversa.

Un ejemplo simple servirá para ilustrar este hecho: supóngase que se va a comprar un lote de  $N$  artículos. Se extrae una muestra aleatoria de tamaño  $n$  con el objeto de examinar qué proporción de artículos en la muestra están defectuosos. Si esta proporción es "suficientemente grande" no se comprará el lote. La pregunta sería ¿qué tan grande debe ser esta proporción para considerar que el lote en general es inadecuado?. Si el lote fuera bueno se esperaría que la mayor parte de los artículos seleccionados en la muestra no estuvieran defectuosos. Sin embargo, es posible que, por puro efecto del azar, se tenga una muestra aleatoria con una proporción de defectuosos mayor que la esperada asumiendo que el lote es bueno. Así, una probabilidad pequeña de cometer el error tipo I implicaría que se está dispuesto a aceptar que ciertamente

el lote es adecuado aunque salga una proporción tan grande que haga poco plausible esta afirmación; i.e., hay poco riesgo de equivocarse en la decisión si en efecto el lote es bueno, pues en muy pocas situaciones (verdaderos extremos) esta hipótesis sería rechazada. Esto implica que la probabilidad de aceptar un lote defectuoso se incrementa. Por lo tanto, es necesario modificar la pretensión original de minimizar  $\alpha$  y  $\beta$  simultáneamente. La modificación (sugerida por Neyman y Pearson a principios de los años treinta) "está basada en el hecho de que en la mayoría de las circunstancias, la actitud hacia la hipótesis nula es diferente de la actitud hacia la hipótesis alterna: a menudo nos preguntamos si hay suficiente evidencia para rechazar la hipótesis nula. En términos de dos errores posibles esto puede traducirse en el enunciado de que el error tipo I es más serio que el error tipo II" (Silvey, 1975, p. 97). Por lo tanto, "se acostumbra asignar una cota a la probabilidad de rechazar incorrectamente la hipótesis nula cuando es cierta y se intenta minimizar a  $\beta$  sujeta a esta condición" (Lehmann, 1959, p. 61). Esto se hace calculando la función potencia de la prueba, que mide el comportamiento de esta cuando la hipótesis nula es falsa en la población. Hay pruebas que detectan esta situación mejor que otras dada una cierta probabilidad de cometer el error tipo I llamada nivel de significancia de la prueba. La elección de esta probabilidad se hace de manera más o menos arbitraria, pues en la mayoría de los casos se desconoce qué valor de  $\alpha$  se está dispuesto a aceptar. Lo usual es considerarlo cercano a .05. Sin embargo, no hay justificación a este hecho, excepto que la mayoría de las tablas de las funciones de distribución de estadísticas de prueba tienen los valores de estas funciones para un número restringido de posibles valores de  $\alpha$ , en general .01,

.05 y .1. Es necesario hacer notar que esto implica conocer no sólo la distribución de la estadística de prueba bajo la hipótesis nula (que mide la probabilidad de que determinado valor de la estadística haya sido obtenido y la hipótesis nula continde considerandose consistente con los datos muestrales), sino también la potencia de la prueba. Regresando al ejemplo del lote de artículos, la estadística de prueba sería el porcentaje de defectuosos que hay en la muestra. Si la hipótesis nula (lote no defectuoso) fuera cierta y se obtuviera una proporción de .9 defectuosos, se diría que hay poca probabilidad de obtener este resultado (asumiendo que la muestra es suficientemente representativa de la población). Sin embargo, la hipótesis nula se aceptaría si el nivel de significancia fijado fuera muy pequeño (asumiendo que la potencia de la prueba es también suficientemente grande).

Para el caso del índice  $A$ , la distribución de la estadística de prueba es conocida (al menos aproximadamente), sin embargo, no se conoce la potencia de la prueba. Una posible solución a este problema sería considerar hipótesis alternas (simuladas) y calcular probabilidades de aceptarlas utilizando a  $A$  como estadística de prueba. Lo mismo sucede para el índice  $D$ .

En la práctica, se ha visto que, si la prueba es suficientemente potente, un valor de  $\alpha$  cercano a .05 es aceptable; i.e., se está dispuesto a tener una decisión errónea en una de cada veinte veces que se efectuara la prueba de hipótesis. Puesto que no se conoce la potencia de la prueba, las tablas generadas para los índices deben considerarse preliminares, aunque se debe notar que los valores de  $\alpha$  fueron .01, .05 y .1.

Una vez calculado el valor de la estadística de prueba se ve, utilizando alguna regla de decisión deducida bajo el supuesto de que la hipótesis nula es verdadera, si se acepta o se rechaza  $H_0$ . Los términos "acepta" y "rechaza" no significan que se desecha o se admite de una manera absoluta que la hipótesis nula es verdadera o falsa en términos de la población. La única forma de saber esto es analizar a todos los elementos de la población y así concluir; desde luego, en este caso la estadística no tiene un sentido inferencial. Más bien, si se aceptara la hipótesis nula se debería decir que no se encontró (en la muestra) suficiente información consistente con la hipótesis alterna; es decir, no se puede afirmar que la hipótesis nula esté correcta, sino más bien, se considera que la hipótesis alterna no fue consistente (en cierto grado) con los valores muestrales, todo esto con una probabilidad de error asociada.

Esta forma de inferencia es el modus tollens de la lógica clásica, y puede expresarse de la siguiente forma:

$((t \rightarrow p) \wedge \bar{p}) \rightarrow \bar{t}$  donde  $\bar{t}$  es "no t", t, p proposiciones cualquiera. Popper (1975, p. 76-77) describe a tollens en los siguientes términos: "Sea p una conclusión del sistema de enunciados t, que puede consistir de teorías y condiciones iniciales... Es posible simbolizar la relación de derivabilidad (implicación analítica) como  $t \rightarrow p$ , que puede leerse como "p se sigue de t". Supóngase que p es falsa, es decir,  $\bar{p}$ . Dada la relación de deducibilidad  $t \rightarrow p$  y el supuesto  $\bar{p}$ , se puede inferir  $\bar{t}$ ; esto es, se considera que t como falsa (we regard t as falsified)... o, en palabras, 'si p es derivable de t y si p es falsa, entonces t es también falsa.'".

De otra forma, es posible transmitir falsedad a enunciados universales (teorías, o bien, enunciados acerca de una población) a partir de enunciados particulares (resultados empíricos o valores muestrales); lo que no se puede hacer (sin caer en la falacia de afirmar el consecuente:  $((t \rightarrow p) \wedge p) \rightarrow t$ ) es transmitir verdad de enunciados particulares a enunciados universales.

De todo esto se desprende que lo que da información es el rechazar la hipótesis nula, pues el aceptarla no garantiza nada, mientras que el encontrar una muestra que no es consistente con la hipótesis implica (con probabilidad de error asociada debido a la aleatoriedad de la muestra) que ésta no es verdadera en la población.

Si  $T$  es una estadística de prueba,  $R$  una región de rechazo (definida por el nivel de significancia) y el símbolo  $\sim$  denota negación, se puede sintetizar lo anterior en dos formas:

$$\begin{array}{l} H_0 \rightarrow T \notin R \\ T \in R \end{array}$$

$$\begin{array}{l} \sim H_0 \rightarrow T \in R \\ T \notin R \end{array}$$

---

 $\sim H_0$ 


---

 $H_0$ 

Nótese que el segundo renglón de la inferencia corresponde a los datos muestrales.

Como ya se ha mencionado, esta es una tesis técnica, y, por lo tanto, no discutiré con más detalle estos aspectos acerca de los fundamentos del método estadístico para contrastar hipótesis.

Finalmente, es necesario mencionar que los resultados obtenidos fueron ligeramente distintos a los presentados por Hodder y Okell. Para el índice  $A$  se tienen valores mayores que los de estos autores,

para los cuantiles inferiores; para los cuantiles superiores no hay diferencias. Probablemente esto se deba a la función generadora de números aleatorios utilizada por Hodder y Okell; sin embargo, puedo afirmar que mis simulaciones están bien programadas. Para el índice de dispersión se encontraron valores muy ligeramente distintos en ambos extremos de la distribución sin que se pueda encontrar un patrón sistemático de comp

Estas discrepancias podrían servir para concluir que (al menos para el índice A con toda seguridad y muy probablemente para D) los valores presentados en el apéndice IV de este trabajo representan más fielmente la situación de aleatoriedad en los patrones generados, pues sus valores tienen una variación menor respecto a 1 que la de las tablas de Hodder y Okell.

## I.1) PROCESOS DE POISSON

Algunos experimentos se reducen al conteo del número de veces que algún evento ocurre en un tiempo o espacio dado.

Por ejemplo, es posible contar el número de personas que llegan a un banco entre las 9 y la 1:30, el número de puestos de periódicos por  $\text{km}^2$  en la Ciudad de México, etc.

Se sabe por experiencia que un modelo matemático elemental adecuado para tratar con este tipo de fenómenos es la distribución de Poisson.

Conceptualmente es conveniente considerar al resultado del conteo en un intervalo continuo (o tal que se pueda considerar continuo) cualquiera como una variable aleatoria discreta.

El desarrollo del modelo puede hacerse de muchas formas. A continuación expongo dos de ellas. La primera está basada en libros como Larson (1978) y Hogg & Tanis (1977). Para la segunda parte utilicé los textos de Hogg & Craig (1978), Ross (1970) y Lindgren (1970).

En un proceso de Poisson con parámetro  $\lambda$  se generan eventos discretos en un intervalo continuo (tiempo, área, longitud, etc.) de la siguiente forma:

i) Existe un intervalo de longitud  $h$  suficientemente pequeña tal que: a) la probabilidad de que haya solamente una ocurrencia en el intervalo es aproximadamente igual a  $\lambda h$ , y b) la probabilidad de tener dos o más ocurrencias en el intervalo es aproximadamente cero.

ii) La ocurrencia de un evento en un intervalo de longitud  $h$  suficientemente pequeña no tiene ningún efecto en la ocurrencia o no ocurrencia de otro evento en cualquier otro intervalo ajeno de igual longitud; i.e., las ocurrencias son consideradas como estocásticamente independientes.

Si es posible aceptar que el experimento satisface los tres supuestos anteriores, entonces los eventos pueden considerarse como generados por un proceso de Poisson.

Cuando se observa un proceso de Poisson en algún intervalo del continuo con longitud unitaria, el número de eventos que ocurren en él es una variable aleatoria, y si además se asume que ocurren en una cantidad constante  $\lambda$  (lo que no implica, por la naturaleza del fenómeno, que ocurran siempre  $\lambda$  eventos por cada intervalo unitario, sino, más bien, que en promedio esto sucede), la distribución queda completamente determinada de la siguiente manera:

Si se observa un proceso de Poisson con parámetro  $\lambda$  durante  $s$  unidades y  $X$  denota el número de eventos que ocurren, entonces  $X$  es una variable aleatoria

Lo anterior significa que si se esperan (en el sentido de esperanza matemática)  $\lambda$  eventos por unidad, es fácil pensar que se esperan  $\lambda s$  eventos en  $s$  unidades.

Entonces, es posible establecer el siguiente

**TEOREMA:** Si  $X$  es una variable aleatoria de Poisson con parámetro igual a  $\lambda s$ , entonces, su función de probabilidad (no de densidad por ser  $X$  una variable aleatoria discreta) está dada por:

$$f_X(x) = P\{X = x\} = \frac{(\lambda s)^x e^{-\lambda s}}{x!} \quad \text{si } x = 0, 1, 2, 3, \dots$$

$$= 0 \quad \text{e. o. c.}$$

**DEMOSTRACION:** Se tiene un proceso de Poisson con parámetro  $\lambda$  sobre  $s$  unidades, y  $X$  es una variable aleatoria que cuenta el número de eventos que ocurren.

Si el intervalo se divide en  $n$  subintervalos ajenos, cada uno de longitud  $h$  ( $h = \delta/n$ ), se tiene una partición en la que el valor de  $h$  se determina dividiendo al intervalo original (unitario) en subintervalos tales que no haya más que 0 o 1 eventos ocurriendo en cada uno de ellos.

Por la definición de procesos de Poisson se sabe que la probabilidad de que se tenga una ocurrencia es aproximadamente igual a  $\lambda h$  para cada subintervalo. Además, puesto que se trata de un proceso de Poisson, la ocurrencia o no ocurrencia de estos eventos es independiente para cada uno de ellos.

Por todo esto, la ocurrencia o no ocurrencia de un evento en cada subintervalo puede considerarse como una variable aleatoria Bernoulli con parámetro igual a  $\lambda h$ .

Como hay  $n$  eventos Bernoulli con probabilidad de éxito igual a  $\lambda h$ , entonces, si  $X$  denota al número total de ocurrencias en el intervalo original de longitud  $\delta$ ,  $X$  es una variable aleatoria Binomial con parámetros  $n$  y  $\lambda h$ .

Es claro que para llegar a esta última conclusión se utilizó la suposición de independencia de las ocurrencias de los eventos.

Pero:

$$n = \frac{\delta}{h}, \text{ y } h = \frac{\lambda \delta}{n};$$

entonces:

$$f_X(x) = \binom{n}{x} \left( \frac{\lambda \delta}{n} \right)^x \left( 1 - \frac{\lambda \delta}{n} \right)^{n-x} \quad \text{si } x = 0, 1, 2, \dots$$

0

e. o. c.

La construcción del modelo binomial es fácil: si se tienen  $n$  eventos cada uno de los cuales admite únicamente dos posibles resultados (por ejemplo, éxito o fracaso) y se cuenta cuántos resultados fueron "éxitos", la función que da las probabilidades:

la función que da las probabilidades para los valores de ese número es:

$$\binom{n}{x} p \cdot p \cdot p \cdots p \cdot q \cdot q \cdots q$$

donde  $p$  es la probabilidad de éxito para cada uno de los  $n$  experimentos Bernoulli y ocurre  $x$  de las  $n$  veces,  $q$  es la probabilidad de fracaso, y aparece  $n-x$  veces en el producto anterior. Claramente  $q = 1-p$ .

El factor  $\binom{n}{x}$  es el número de formas en que se pueden arreglar los  $x$  éxitos entre los  $n$  experimentos sin considerar el orden con el que van saliendo, y se calcula como:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Como la única forma de garantizar que en todos los casos se cumplan los supuestos del proceso de Poisson es considerando que  $n$ , i.e. el número de intervalos, sea tan grande como se requiera para cualquier situación es considerando  $n$  como infinito, lo cual es posible hacer ya que los eventos ocurren en un continuo, basta con tomar dicho límite para la función de probabilidad anterior.

Haciendo esto último se obtendrá la función de probabilidad exacta para una variable aleatoria de Poisson.

Por una parte:

$$\begin{aligned} \binom{n}{x} \left(\frac{\lambda \Delta}{n}\right)^x \left(1 - \frac{\lambda \Delta}{n}\right)^{n-x} &= \frac{(\lambda \Delta)^x}{x!} \left(1 - \frac{\lambda \Delta}{n}\right)^{-x} \left(1 - \frac{\lambda \Delta}{n}\right)^n \frac{n(n-1) \cdots (n-x+1)}{n^x} \\ &= \frac{(\lambda \Delta)^x}{x!} \left(1 - \frac{\lambda \Delta}{n}\right)^{-x} \left(1 - \frac{\lambda \Delta}{n}\right)^n \cdot 1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \end{aligned}$$

Además se sabe que:  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda \Delta}{n}\right)^n = e^{-\lambda \Delta}$ , y como

$$\lim_{n \rightarrow \infty} \left( \frac{(\lambda \delta)^x}{x} \right) = \frac{(\lambda \delta)^x}{x!},$$

$$\lim_{n \rightarrow \infty} \left( 1 - \frac{\lambda \delta}{n} \right)^{-x} = 1,$$

$$\lim_{n \rightarrow \infty} \left( 1 - \frac{\lambda \delta}{n} \right)^{-n} = e^{-\lambda \delta},$$

$$\lim_{n \rightarrow \infty} \left( 1 - \frac{1}{n} \right) \cdots \left( 1 - \frac{x-1}{n} \right) = 1,$$

entonces: 
$$\lim_{n \rightarrow \infty} \binom{n}{x} \left( \frac{\lambda \delta}{n} \right)^x \left( 1 - \frac{\lambda \delta}{n} \right)^{n-x} = \frac{(\lambda \delta)^x}{x!} e^{-\lambda \delta} \quad x = 0, 1, 2, \dots$$

Para demostrar que la parte derecha de la ecuación anterior es una función de probabilidad basta ver que sea positiva y que su suma sobre todos los valores sea uno:

i) La función es positiva por ser un producto de funciones positivas.

$$ii) \sum_{x=0}^{\infty} \frac{(\lambda \delta)^x}{x!} e^{-\lambda \delta} = e^{-\lambda \delta} \sum_{x=0}^{\infty} \frac{(\lambda \delta)^x}{x!} = 1,$$

pues por definición de exponencial,  $e^u = \sum_{x=0}^{\infty} \frac{u^x}{x!}$ ,

y por lo tanto, la función es una función de probabilidad, lo que termina la demostración.

A continuación expongo otra forma de enunciar los supuestos del proceso de Poisson y de construir la función de probabilidad de Poisson:

Sea  $E$  un experimento tal que sus resultados  $\lambda$  forman un espacio  $S$ . Entonces,  $E$  está especificado por algunos subconjuntos de  $S$  (llamados *eventos*) y por las probabilidades asociadas a estos subconjuntos.

Si  $E$  es un experimento aleatorio (es decir, es tal que es imposible predecir sus resultados con certeza si se le observa  $n$  veces en condiciones "idénticas", aunque se pueden conocer las probabilidades con que ocurren algunos resultados) y se observan los resultados (contenidos en  $S$ ) de  $E$  que ocurren en algún subconjunto de un espacio  $T$ , se tiene un *proceso estocástico*. Si  $T$  es numerable, el proceso se llamará *discreto*, mientras que si  $T$  es continuo (área, longitud, tiempo, etc.), el proceso será continuo.

Si se observa la variación respecto a puntos de  $T$  se tiene una familia de variables aleatorias para cada uno de ellos, siendo el punto de  $T$  considerado un parámetro para cada familia. Esto implica que una observación del proceso completo es una función del espacio  $T$ .

En particular, un proceso estocástico que es continuo y en el que la función  $N(t)$  ( $t \in T$ ) representa el total de eventos ocurridos hasta el punto  $t$  es llamado *proceso estocástico de conteo*.

Si un proceso estocástico de conteo cumple los siguientes supuestos, se tiene un *proceso de Poisson*:

Sea  $g(x, s)$  la probabilidad de tener  $x$  ocurrencias en un intervalo de  $T$  con  $s$  unidades; entonces:

i)  $g(1, h) = \lambda h + o(h)$ , donde  $\lambda$  es una constante positiva y  $h > 0$ ;

ii)  $\sum_{x=2}^{\infty} g(x, h) = o(h)$ ;

iii) el número de eventos ocurridos en intervalos ajenos es estocásticamente independiente.

La utilidad de esta definición de Proceso de Poisson está en que formaliza las nociones de "aproximadamente  $\lambda h$ " y "aproximadamente 0"

que aparecen en el desarrollo anterior; en efecto, una función  $f$  es de clase  $o(x)$  si

$$\lim_{x \rightarrow 0} \frac{f(x)}{x} = 0,$$

es decir,  $f$  pertenece a la clase  $o$ -chica si en una vecindad de 0 es más pequeña que la función idéntica. Por ejemplo,  $f(x) = x^2 \in o(x)$ .

Puede demostrarse que  $o(x) + o(x) = o(x)$  lo cual será muy útil mas adelante. Para mas detalles sobre funciones de esta clase, véase Loomis & Sternberg (1968), §5.1.

Específicamente, lo que se consigue con la inclusión de funciones  $o(x)$  en el desarrollo de la teoría de los procesos de Poisson es establecer las condiciones que garanticen que para intervalos de longitud  $h$  tan pequeña como se quiera, el error que se tiene al afirmar que "la probabilidad de tener una ocurrencia en ese intervalo es aproximadamente  $\lambda h$ " y que "la probabilidad de observar dos o más ocurrencias en un intervalo suficientemente pequeño es aproximadamente 0" es despreciable.

TEOREMA: Si se tiene un proceso de Poisson (definido por los supuestos de la página anterior), la probabilidad de tener  $x$  ocurrencias de algún evento contenido en  $S$  en algún intervalo de  $T$  con medida  $\lambda$  es:

$$f_X(x) = P\{X = x\} = \frac{(\lambda \lambda)^x e^{-\lambda \lambda}}{x!} \quad \text{si } x = 0, 1, 2, 3, \dots$$

$$= 0 \quad \text{e.o.c.}$$

DEMOSTRACION: Claramente,  $g(0,0) = 1$ . La probabilidad de tener al menos una ocurrencia del evento en un intervalo de medida  $h$  es:

$$g(1,h) + \sum_{x=2}^{\infty} g(x,h) = \lambda h + o(h) + o(h) = \lambda h + o(h),$$

y por lo tanto, la probabilidad de tener 0 eventos en ese intervalo es:

$$1 - \lambda h - o(h).$$

Por otra parte, aplicando el supuesto de independencia estocástica:

$$g(0, \delta + h) = g(0, \delta) (1 - \lambda h + o(h)),$$

i.e., la probabilidad de 0 cambios en  $\delta$  unidades y de 0 cambios en  $h$  unidades es el producto de ambas probabilidades; entonces:

$$\frac{g(0, \delta + h) - g(0, \delta)}{h} = -g(0, \delta) - \frac{g(0, \delta) o(h)}{h}.$$

Por el argumento usado en el primer desarrollo, se puede tomar el límite cuando  $h \rightarrow 0$  en ambos lados de la ecuación anterior, asegurando que, puesto que  $\lambda$  es continuo, siempre es posible encontrar un intervalo de longitud tan pequeña que se cumplan los supuestos del proceso de Poisson.

Usando la definición de derivada se tiene que:

$$\frac{d g(0, \delta)}{d \delta} = -\lambda g(0, \delta),$$

o, equivalentemente,

$$\ln g(0, \delta) = -\lambda \delta + c,$$

entonces;

$$g(0, \delta) = c^{\delta} e^{-\lambda \delta}$$

es la solución de la ecuación diferencial.

Como se tiene que  $g(0, 0) = 1$ , entonces,  $c = 1$ , y

$$g(0, \delta) = e^{-\lambda \delta}.$$

Sea  $x \in \mathbb{N}$ ; entonces,  $g(x, 0) = 0$ , y

$$g(x, \delta + h) = g(x, \delta) (1 - \lambda h - o(h)) + g(x-1, \delta) (\lambda h + o(h)) + o(h),$$

que implica

$$\frac{g(x, \delta + h) - g(x, \delta)}{h} = -\lambda g(x, \delta) + g(x-1, \delta) + \frac{o(h)}{h}.$$

Tomando el límite cuando  $h$  tiende a 0 en ambos lados de la ecuación:

$$\frac{d g(x, \delta)}{d \delta} = -\lambda g(x, \delta) + g(x-1, \delta),$$

o, equivalentemente,

$$e^{\lambda \delta} (g'(x, \delta) + \lambda g(x, \delta)) = \lambda e^{\lambda \delta} g(x-1, \delta),$$

lo que implica que:

$$\frac{d}{d \delta} (e^{\lambda \delta} g(1, \delta)) = \lambda, \text{ porque } g(0, \delta) = e^{-\lambda \delta},$$

entonces:

$$g(1, \delta) = e^{-\lambda \delta} (\lambda \delta + c), \text{ y como } g(1, 0) = 0,$$

se tiene que:

$$g(1, 0) = \lambda \delta e^{-\lambda \delta}$$

Para demostrar que  $g(x, \delta) = \frac{e^{-\lambda \delta} (\lambda \delta)^x}{x!}$  se supone que la igualdad se da para  $x-1$ , i.e. :

$$\frac{d}{d \delta} (e^{\lambda \delta} g(x, \delta)) = \frac{\lambda \delta^x e^{-\lambda \delta}}{(x-1)!},$$

lo que implica que:

$$e^{\lambda \delta} g(x, \delta) = \frac{(\lambda \delta)^x}{x!} + c.$$

Finalmente, como  $g(x, 0) = 0$ ,  $c$  debe ser 0, y entonces,

$$g(x, \delta) = \frac{(\lambda \delta)^x}{x!} e^{-\lambda \delta}, \text{ lo que termina la demostración.}$$

## APENDICE II

La siguiente demostración está basada en la propuesta por Apostol (1969, vol. II, sección 11.3).

**TEOREMA:** Sea  $S_k(r)$  la esfera sólida de radio  $r$  con  $k$  dimensiones, definida como:

$$S_k(r) = \{(x_1, \dots, x_k \mid x_1^2 + \dots + x_k^2 \leq r^2)\},$$

y sea  $V_k(r) = \int_{S_k(r)} dx_1 \cdots dx_k$  el volumen de  $S_k(r)$ .

Entonces,

$$V_k(r) = \frac{\pi^{k/2}}{\Gamma(k/2 + 1)} r^k$$

**DEMOSTRACION:**

Si  $k = 1$ , entonces  $V_1(r) = 2r$ , que es la longitud del intervalo cerrado  $[-r, r]$ .

Si  $k = 2$ , entonces  $V_2(r) = \pi r^2$ , que es el área de un círculo de radio  $r$ . Por lo tanto, basta demostrar la fórmula para  $k \geq 3$ .

Antes, es necesario mostrar que para toda  $r > 0$ ,  $V_k(r) = r^k V_k(1)$ .

Para hacerlo, se utiliza el cambio de variable  $\underline{x} = r\underline{u}$ , que mapea  $S_k(1)$  en  $S_k(r)$ . El jacobiano de esta transformación es:

$$\underline{x} = (x_1, \dots, x_k) = r(u_1, \dots, u_k),$$

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \dots & \frac{\partial x_k}{\partial u_1} \\ \dots & \dots & \dots \\ \frac{\partial x_1}{\partial u_k} & \dots & \frac{\partial x_k}{\partial u_k} \end{vmatrix}$$

$$|J| = \begin{vmatrix} r & & & \\ & r & & 0 \\ & & \ddots & \\ 0 & & & r \end{vmatrix} = r^k.$$

Entonces:

$$\begin{aligned} V_k(r) &= \int \cdots \int_{S_k(r)} dx_1 \cdots dx_k \\ &= \int \cdots \int_{S_k(1)} r^k du_1 \cdots du_k = r^k V_k(1), \end{aligned}$$

lo que prueba el primer resultado.

Para demostrar el resultado general basta ver que  $V_k(1) = \frac{\pi^{k/2}}{\Gamma(\frac{1}{2}k + 1)}$ .

$$x_1^2 + \cdots + x_k^2 \leq 1 \iff x_1^2 + \cdots + x_{k-2}^2 \leq 1 - x_{k-1}^2 - x_k^2,$$

$$\text{y } x_{k-1}^2 + x_k^2 \leq 1.$$

Esto implica que:

$$V_k(1) = \int \int_{x_{k-1}^2 + x_k^2 \leq 1} \cdot \left( \int \cdots \int_{x_1^2 \leq 1 - x_{k-1}^2 - x_k^2} dx_1 \cdots dx_{k-2} \right) dx_{k-1} dx_k$$

La integral dentro del paréntesis se evalúa sobre la esfera  $(k-2)$ -dimensional  $S_{k-2}(p)$ , donde  $p = \sqrt{1 - x_{k-1}^2 - x_k^2}$ , es decir:

$$V_{k-2}(p) = p^{k-2} V_{k-2}(1) = (1 - x_{k-1}^2 - x_k^2)^{\frac{n}{2} - 1} \cdot V_{k-2}(1)$$

Sean  $x = x_{k-1}$ ,  $y = x_k$ . Entonces, si  $k \geq 3$  se tiene que:

$$V_k(1) = V_{k-2}(1) \cdot \iint_{x^2 + y^2 \leq 1} \sqrt{1 - x^2 - y^2}^{n/2 - 1} dx dy.$$

Para evaluar la integral anterior, se transforma a coordenadas polares:

$$V_k(1) = V_{k-2}(1) \int_0^{2\pi} \int_0^1 (1 - t^2)^{k/2 - 1} t dt d\theta,$$

donde  $x = t \cos(\theta)$ ,  $y = t \sin(\theta)$  y  $t = \sqrt{x^2 + y^2}$ .

El jacobiano de la transformación es:

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos(\theta) & \sin(\theta) \\ -t \sin(\theta) & t \cos(\theta) \end{vmatrix} = t (\cos^2(\theta) + \sin^2(\theta))$$

$$= t, \text{ y } 0 \leq t \leq 1, 0 \leq \theta \leq 2\pi.$$

Entonces:

$$\int_0^1 (1 - t^2)^{k/2 - 1} t dt = \frac{1}{2} \int_0^1 u^{k/2 - 1} du = \frac{1}{2} \cdot \frac{u^{k/2}}{\frac{k}{2}} \Big|_0^1$$

$$= \frac{1}{k}, \text{ y se tiene que:}$$

$$V_k(1) = V_{k-2}(1) \frac{1}{k} \int_0^{2\pi} d\theta = V_{k-2}(1) \frac{2\pi}{k} \quad (k \geq 3).$$

Ahora sea la sucesión  $\{f(k)\}_k$  definida por  $f(k) = \frac{\pi^{k/2}}{\Gamma(\frac{1}{2}k + 1)}$ .

$\{f(k)\}_k$  satisface la fórmula recursiva  $V_k(1) = \frac{2\pi}{k} V_{k-2}(1)$ ,

es decir:  $f(k) = \frac{2\pi}{k} f(k-2)$ .

Desarrollando la ecuación anterior:

$$\frac{\pi^{k/2}}{\Gamma(\frac{1}{2}k + 1)} = \frac{2\pi}{k} \frac{\pi^{(k-2)/2}}{\Gamma(\frac{1}{2}(k-2) + 1)}$$

Entonces:

$$f(k) = \frac{2\pi^{k/2}}{\Gamma(\frac{1}{2}k)k} = \frac{2\pi^{n/2}}{\Gamma(\frac{1}{2}k) \cdot \frac{1}{2}k + \Gamma(\frac{1}{2}k) \cdot \frac{1}{2}k}, \text{ y como}$$

$$\Gamma(x+1) = x \cdot \Gamma(x), \text{ se tiene la igualdad } f(k) = \frac{2\pi}{k} f(k-2),$$

lo que prueba la igualdad para  $k \geq 3$ .

$$\text{Como } \Gamma(\frac{1}{2}) = \sqrt{\pi}, \Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}, \text{ entonces } f(1) = V_1(1).$$

$$\text{Como } f(2) = V_2(1) = \pi, \text{ se tiene que } f(k) = V_{k-1}(1), \text{ si } k \geq 1,$$

como se quería demostrar.

### APENDICE III

A continuación se presenta el método utilizado para la generación de valores de una variable aleatoria con distribución de Poisson con parámetro  $\lambda$ .

Sea un proceso de Poisson sobre el intervalo cerrado por la izquierda  $(0, \infty)$ . Si  $N(t)$  denota el número de eventos que ocurren en el intervalo cerrado  $(0, t)$ , entonces  $N(t)$  es una variable aleatoria con distribución Poisson con parámetro  $\lambda t$ , donde  $\lambda$  es el parámetro del proceso. Si  $0 \leq s \leq t$ , entonces  $N(t) - N(s)$  es el número de eventos ocurridos en el intervalo cerrado por la derecha  $(s, t)$ , y es una variable aleatoria con distribución Poisson y parámetro  $\lambda(t - s)$  (Hoel, Port & Stone, 1971, p. 230).

Sea  $D_m$  el "tiempo" en el que ocurre el  $m$ -ésimo evento. Es posible demostrar (ibid., p.228) que  $D_m \sim \Gamma(m, \lambda)$ . Sean las variables aleatorias  $W_1, W_2, \dots, W_n, \dots$  definidas como sigue:  $W_1 = D_1, W_n = D_n - D_{n-1}$ , ( $n \geq 2$ ). Entonces,  $D_m = W_1 + \dots + W_m$ . Si se supone que las  $W_1$ 's son independientes, se ve que todas ellas tienen distribuciones exponenciales con parámetro  $\lambda$ , porque la suma de variables aleatorias independientes exponenciales da una variable aleatoria con distribución gamma (ibid., p.229).

Como la variable aleatoria  $W_m$  es el "tiempo" entre los eventos  $(m-1)$  y  $m$ , a las variables  $W_1, W_2, \dots$  se les llama "tiempos de espera" entre los eventos sucesivos de un proceso de Poisson.

Entonces, los tiempos de espera de un proceso de Poisson se distribuyen exponencialmente con parámetro igual al recíproco del parámetro del proceso. (ibid. p.231). (Este resultado no se demostrará).

El siguiente desarrollo está basado en esta idea y ha sido expuesto casi sin variaciones en libros como Tocher (1969, p. 36-39), Hammersley & Handscomb (1979, p. 36-38) y Fishman (1978, p. 250-251).

La idea central consiste en generar  $e_1, e_2, \dots, e_n$ , una sucesión de variables aleatorias con distribución exponencial y parámetro  $1/\lambda$  e ir calculando la suma de valores de estas variables. Cuando se cumpla la condición:

$$\sum_{i=1}^n e_i \leq \lambda < \sum_{i=1}^{n+1} e_i, \quad (1)$$

$n$  será un valor de una variable aleatoria con distribución Poisson y parámetro  $\lambda$ .

Para generar valores de una variable exponencial se utiliza el hecho de que si  $E$  tiene esta distribución, entonces,

$$E = -\ln(U), \quad (2)$$

donde  $U$  tiene una distribución uniforme (Hammersley & Handscomb, 1979, p. 35). Entonces, la condición (1) puede expresarse como

$$\sum_{i=1}^n (-\ln(u_i)) \leq \mu < \sum_{i=1}^{n+1} (-\ln(u_i))$$

o bien,

$$\prod_{i=1}^n u_i > e^{-\mu} \geq \prod_{i=1}^{n+1} u_i$$

donde  $\mu = 1/\lambda$ .

El algoritmo utilizado fue el siguiente:

```
aux:=0;
e:=exp(-λ);
pr:=1;
r:=random(S);
pr:=pr*r;
mientras (pr-e) 0 hacer
    aux:=aux+1;
    pr:=pr*random(S);
Poisson:=aux;
```

Atkinson (1979) demuestra que este algoritmo es el más eficiente para generar valores de variables aleatorias Poisson si el valor de  $\lambda$  es menor que 15. Por esta razón fue utilizado. El lenguaje utilizado fue ALGOL.

ALFA=0.005

P=0.500

Q=0.500

SIMULACIONES= 200

VALORES DEL INDICE A

	2	4	6	8	10	15	20	30	40
2	0.11092 2.99347	0.25763 2.22158	0.39203 1.95316	0.35502 1.85932	0.39786 1.95798	0.29808 1.80428	0.51791 1.71481	0.44126 1.70312	0.16211 1.67300
4	0.00000 0.00000	0.64947 1.55781	0.61936 1.42691	0.52204 1.37435	0.52743 1.34478	0.55820 1.30198	0.67526 1.27341	0.59765 1.26722	0.76008 1.25287
6	0.00000 0.00000	0.00000 0.00000	0.49731 1.31001	0.76143 1.26973	0.75040 1.24354	0.79937 1.21352	0.84066 1.18841	0.64948 1.16549	0.55162 1.16470
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.79226 1.22697	0.83118 1.19478	0.85509 1.16447	0.74779 1.15098	0.79665 1.13213	0.88716 1.12441
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.75880 1.17220	0.82390 1.15349	0.83571 1.12545	0.83604 1.11606	0.91253 1.10262
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.89615 1.10797	0.79263 1.09398	0.87066 1.08226	0.91457 1.07137
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.80580 1.08041	0.87621 1.06413	0.96492 1.05871
30	0.00000 0.00000	0.33333 0.33333	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 1.05311	0.93608 1.05311	0.96383 1.04738
40	0.00000 0.00000	0.92763 1.03243							

	ALFA=0.005		P=0.500		Q=0.500		SIMULACIONES= 200		INDICE DE DISPERSION	
	2	4	6	8	10	15	20	30	40	
2	0.16326 0.88299	0.08757 0.11458	0.21793 0.88054	0.21122 0.76354	0.21552 0.73878	0.25876 0.58987	0.32876 0.37873	0.27357 0.49741	0.13919 0.36893	
4	0.00000 0.00000	0.34966 0.87474	0.48384 1.86706	0.48442 0.00031	0.38315 0.04691	0.31669 0.64238	0.53729 1.79660	0.36668 1.65781	0.48196 1.83527	
6	0.00000 0.00000	0.00000 0.00000	0.41299 0.70777	0.46322 0.67293	0.58166 0.81509	0.60517 0.60191	0.58669 1.53641	0.47263 0.54028	0.56153 0.48341	
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.58153 0.57267	0.57332 0.64885	0.58580 0.55617	0.53523 0.45248	0.61598 0.49082	0.63824 0.63359	
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.63369 0.77057	0.69632 0.44386	0.62256 0.38819	0.63176 0.96005	0.66879 0.38487	
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.70115 0.44591	0.63552 0.38396	0.65010 0.29740	0.73124 0.33113	
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.76441 0.30297	0.70124 0.21676	0.77525 0.27477	
30	0.00000 0.00000	0.74642 0.26314	0.79564 0.18514							
40	0.00000 0.00000	0.80908 0.19531								

ALFA=0.050	P=0.500	Q=0.500	SIMULACIONES= 200		VALORES DEL INDICE A				
	2	4	6	8	10	15	20	30	40
2	0.83441 1.84999	0.73663 1.99185	0.74951 1.83118	0.85848 1.77108	0.90243 1.71412	0.78001 1.69882	0.91017 1.63224	0.81287 1.80039	0.82251 1.88191
4	0.00000 0.00000	0.81661 1.48199	0.82145 1.38435	0.87373 1.34394	0.84612 1.29288	0.89376 1.27540	0.90078 1.29908	0.84688 1.24088	0.82848 1.22448
6	0.00000 0.00000	0.90000 1.30000	0.89545 1.28097	0.82444 1.22315	0.97442 1.22084	0.92848 1.18738	0.95658 1.17748	0.96055 1.18955	0.92522 1.14522
8	0.00000	0.00000	0.00000	0.24086 1.28186	0.93258 1.18182	0.92814 1.15334	0.95484 1.15944	0.96223 1.12027	0.96534 1.10861
10	0.00000	0.00000	0.00000	0.00000	0.91461 1.15986	0.92341 1.12201	0.95089 1.11314	0.94002 1.09951	0.95738 1.09389
15	0.00000	0.00000	0.00000	0.00000	0.00000	0.95513 1.05878	0.95263 1.08263	0.97041 1.07360	0.97305 1.06565
20	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.96683 1.07583	0.96883 1.05883	0.98377 1.05377
30	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.97150 1.04365	0.98123 1.04123
40	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.98465 1.03465

	ALFA=0.050	P=0.500	Q=0.500		SIMULACIONES= 200		INDICE DE DISPERSION		
		2	4	6	8	10	15	20	30
2	0.36865 2.51694	0.41758 2.19760	0.46639 2.11778	0.62287 2.22109	0.57668 2.19029	0.51933 2.08014	0.66263 2.11268	0.69946 2.07946	0.54188 2.14188
4	0.00000 0.00000	0.57936 1.74836	0.59188 1.55388	0.62310 1.62368	0.62121 1.60941	0.63442 1.51402	0.68605 1.59080	0.62924 1.62924	0.71481 1.71481
6	0.00000 0.00000	0.00000 0.00000	0.62782 1.45682	0.63718 1.42464	0.72010 1.38610	0.74332 1.42034	0.75711 1.39751	0.72863 1.34996	0.78024 1.38024
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.71751 1.40246	0.72113 1.33537	0.73247 1.32498	0.74330 1.34440	0.80347 1.29753	0.77365 1.31365
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.74078 1.37724	0.81581 1.27241	0.81645 1.27570	0.74966 1.24966	0.78812 1.27148
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.78398 1.28398	0.72733 1.22733	0.80625 1.23883	0.85006 1.19006
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.83340 1.21674	0.85092 1.17312	0.85892 1.19892
30	0.00000 0.00000	0.86012 1.20912	0.84468 1.14468						
40	0.00000 0.00000	0.00000 0.00000	0.87597 1.13597						

	ALFA=0.025		P=0.500		Q=0.500		SIMULACIONES# 200		INDICE DE DISPERSION		
	2	4	6	8	10	15	20	30	40		
2	0.70544 2.95414	0.27412 2.63139	0.39102 2.34459	0.47476 2.28746	0.45034 2.33828	0.38528 2.34770	0.51795 2.21373	0.31924 2.23894	0.46949 2.28158		
4	0.00000 0.00000	0.47578 1.80996	0.54853 1.64362	0.55844 1.81595	0.48385 1.71084	0.54157 1.54967	0.66373 1.59473	0.51507 1.61013	0.63947 1.52738		
6	0.00000 0.00000	0.00000 0.00000	0.53998 1.55167	0.58580 1.52970	0.60596 1.43538	0.72422 1.45688	0.68176 1.46759	0.68708 1.38741	0.69340 1.39231		
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.43793 1.47839	0.67446 1.42706	0.69975 1.43126	0.70935 1.39872	0.76949 1.35981	0.74501 1.35251		
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.69845 1.41176	0.79466 1.31690	0.79098 1.32357	0.71912 1.29565	0.76921 1.32329		
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.75126 1.33678	0.73505 1.30511	0.77146 1.27116	0.81686 1.24325		
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.78768 1.27643	0.81983 1.19746	0.83593 1.21744		
30	0.00000 0.00000	0.82437 1.22775	0.82412 1.14128								
40	0.00000 0.00000	0.85793 1.14480									

ALFA=0.025

P=0.500

q=0.500

SIMULACIONES= 200

VALORES DEL INDICE A

	2	4	6	8	10	15	20	30	40
2	0.73405 2.77379	0.59573 2.07443	0.67979 1.90062	0.71310 1.83324	0.78527 1.78414	0.62404 1.75015	0.79219 1.68795	0.52353 1.63735	0.75099 1.60898
4	0.60000 0.60000	0.72121 1.51677	0.81323 1.39616	0.81795 1.36150	0.75277 1.31291	0.81819 1.28918	0.85626 1.26006	0.79088 1.25441	0.84692 1.24228
6	0.60000 0.60000	0.80000 0.80000	0.83610 1.29923	0.82462 1.25988	0.82976 1.23840	0.88761 1.19056	0.93414 1.17797	0.89529 1.15803	0.86826 1.14898
8	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.89138 1.21180	0.89920 1.18619	0.88910 1.16107	0.87247 1.14359	0.92983 1.12788	0.95553 1.11951
10	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.86448 1.16537	0.94275 1.13864	0.88094 1.11724	0.91099 1.10865	0.94650 1.09635
15	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.93653 1.10097	0.83056 1.08935	0.94747 1.07788	0.95944 1.07170
20	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.80000 0.80000	0.85299 1.07299	0.86275 1.06065	0.87314 1.05368
30	0.80000 0.80000	0.85624 1.05024	0.97899 1.04403						
40	0.80000 0.80000	0.97819 1.03663							

ALFA=0.050

P=1.000

Q=1.000

SIMULACIONES= 200

VALORES DEL INDICE A

	2	4	6	8	10	15	20	30	40
2	0.02624 0.74479	0.05398 0.65573	0.04308 0.57543	0.04306 0.46796	0.03858 0.49307	0.05478 0.45484	0.07284 0.43227	0.04981 0.49713	0.04370 0.50969
4	0.00000 0.00000	0.12230 0.27272	0.09423 0.39562	0.11313 0.41311	0.11185 0.40990	0.12223 0.37652	0.12445 0.39331	0.11576 0.38916	0.09901 0.35801
6	0.00000 0.00000	0.00000 0.00000	0.12899 0.33979	0.13622 0.42296	0.12775 0.36546	0.11735 0.36158	0.13908 0.35129	0.13757 0.33332	0.13968 0.32381
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.12540 0.36930	0.15146 0.36075	0.14390 0.34530	0.14557 0.32997	0.15084 0.32285	0.14624 0.33005
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.12491 0.32108	0.14125 0.32909	0.15190 0.33066	0.16338 0.30566	0.14523 0.31616
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.16190 0.30638	0.16990 0.30429	0.15950 0.28081	0.16732 0.29464
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.16529 0.30609	0.17136 0.30188	0.16978 0.28928
30	0.00000 0.00000	0.17958 0.27662	0.18118 0.30141						
40	0.00000 0.00000	0.18630 0.28589							

	ALFA=0.050	P=1.000	Q=1.000	SIMULACIONES= 200			INDICE DE DISPERSION		
	2	4	6	8	10	15	20	30	40
2	0.21646 4.49250	0.34192 2.02077	0.21986 1.83788	0.28906 1.89562	0.26362 1.88134	0.28831 1.79940	0.28556 1.79204	0.25502 1.89628	0.27474 1.87412
4	0.60000 0.80000	0.55362 1.90589	0.57079 1.55386	0.55832 1.65583	0.53295 1.45347	0.57021 1.35425	0.64084 1.39135	0.51127 1.36558	0.48522 1.39236
6	0.00000 0.00000	0.00000 0.00000	0.65036 1.55679	0.68402 1.35843	0.65690 1.36896	0.67125 1.40292	0.67961 1.29248	0.68759 1.30228	0.72951 1.33314
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.73684 1.33379	0.69669 1.31599	0.74896 1.31660	0.71256 1.25122	0.76811 1.21118	0.77362 1.21756
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.74985 1.30201	0.73973 1.29799	0.77485 1.25207	0.75461 1.20765	0.76117 1.23954
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.76244 1.22811	0.81110 1.21126	0.81562 1.19924	0.82451 1.15942
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.81437 1.16777	0.84573 1.17165	0.83114 1.13920
30	0.00000 0.00000	0.85126 1.14759	0.88016 1.15326						
40	0.00000 0.00000	0.00000 0.00000	0.88434 1.13132						

ALFA=0.025

P=1.000

Q=1.000

SIMULACIONES= 200

VALORES DEL INDICE A

	2	4	6	8	10	15	20	30	40
2	0.01186 0.81526	0.03357 0.74172	0.02488 0.62942	0.03197 0.52191	0.02217 0.53531	0.04048 0.49459	0.06073 0.48803	0.02762 0.54400	0.03159 0.56482
4	0.00000 0.00000	0.08894 0.51157	0.06204 0.43825	0.09999 0.42060	0.09190 0.49301	0.10173 0.39268	0.10940 0.43429	0.09211 0.38636	0.08677 0.37502
6	0.00000 0.00000	0.06000 0.06000	0.02779 0.43472	0.12773 0.43078	0.12051 0.40945	0.09577 0.40612	0.12563 0.37735	0.12772 0.35292	0.12861 0.33905
8	0.00000 0.30000	0.00000 0.00000	0.00000 0.00000	0.10944 0.39994	0.13411 0.39886	0.13494 0.38920	0.13240 0.36496	0.13721 0.33691	0.13611 0.35539
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.11541 0.33857	0.12724 0.34618	0.13845 0.36233	0.15418 0.32034	0.13394 0.32558
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.09000 0.00000	0.15431 0.32814	0.15213 0.32653	0.14640 0.31864	0.16058 0.32166
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.15670 0.32351	0.16686 0.32707	0.16113 0.31939
30	0.00000 0.00000	0.16460 0.29160	0.17826 0.31366						
40	0.00000 0.00000	0.16460 0.30000	0.18076 0.29651						

ALFA=0.025

P=1.000

Q=1.000

SIMULACIONES= 200

INDICE DE DISPERSION

	2	4	6	8	10	15	20	30	40
2	0.123321	0.189736	0.158999	0.160223	0.147063	0.194966	0.204008	0.176766	0.119321
4	0.000000	0.484779	0.524334	0.477001	0.505061	0.489688	0.522880	0.439662	0.445991
6	0.000000	0.000000	0.572387	0.607119	0.626114	0.627112	0.612994	0.654007	0.663370
8	0.000000	0.000000	0.000000	0.690665	0.681922	0.703885	0.670665	0.704112	0.723881
10	0.000000	0.000000	0.000000	0.000000	0.695882	0.712333	0.733555	0.730229	0.729744
15	0.000000	0.000000	0.000000	0.000000	0.348870	0.361500	0.287666	0.237114	0.278221
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.731266	0.777001	0.781544
30	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.263774	1.243388	1.214224
40	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.783665	0.820445	0.798700
							1.187400	1.206449	1.189966
							0.000000	0.000000	0.000000
							0.000000	0.833228	0.845988
							0.000000	0.177388	0.168113
							0.000000	0.000000	0.000000
							0.000000	0.000000	0.870228
							0.000000	0.000000	0.168878

ALFA=0.005

P=1.000

Q=1.000

SIMULACIONES= 200

INDICE DE DISPERSION

	2	4	6	8	10	15	20	30	40
2	0.01177 *****	0.04577 3.23140	0.00233 2.16510	0.08038 2.26618	0.08008 2.47311	0.12825 2.24977	0.13770 2.20296	0.05494 2.34308	0.03967 2.10301
4	0.00000 0.00000	0.37375 2.56156	0.30968 1.81916	0.46776 1.88552	0.33067 1.97726	0.36263 1.47986	0.35364 1.53973	0.33969 1.46531	0.26046 1.51830
6	0.00000 0.00000	0.00000 0.00000	0.53154 2.15528	0.52878 1.82744	0.55193 1.48664	0.59211 1.94223	0.52503 1.37239	0.52328 1.42943	0.59849 1.50115
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.60124 1.87978	0.51978 1.51123	0.53357 1.38304	0.55887 1.38808	0.59342 1.35718	0.64680 1.33207
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.60011 1.38872	0.64429 1.54456	0.65961 1.38143	0.53180 1.30615	0.65333 1.31778
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.67068 1.45423	0.66898 1.39034	0.74873 1.27398	0.72242 1.26931
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.74896 1.23225	0.78755 1.23921	0.72847 1.26277
30	0.00000 0.00000	0.76058 1.21546	0.80823 1.19274						
40	0.00000 0.00000	0.82418 1.19118							

ALFA=0.005

P=1.000

Q=1.000

SIMULACIONES= 200

VALORES DEL INDICE A

	2	4	6	8	10	15	20	30	40
2	0.00260 0.95070	0.00775 0.81793	0.00645 0.67449	0.02060 0.57284	0.01113 0.67397	0.02220 0.54253	0.02329 0.61541	0.01336 0.64462	0.01671 0.75338
4	0.00000 0.00000	0.06351 0.87750	0.01978 0.54660	0.08203 0.50725	0.05518 0.53470	0.05773 0.45530	0.07586 0.44767	0.06364 0.42704	0.04021 0.38791
6	0.00000 0.00000	0.00000 0.00000	0.06382 0.46027	0.08718 0.48548	0.07205 0.44774	0.06727 0.48916	0.09656 0.44915	0.08163 0.36811	0.09360 0.36457
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.08324 0.43805	0.08946 0.41098	0.06506 0.40937	0.09672 0.38560	0.10127 0.35006	0.11707 0.36663
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.09337 0.41190	0.10745 0.37618	0.16913 0.41707	0.10588 0.34092	0.09652 0.36979
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.13180 0.34885	0.13784 0.35504	0.12265 0.34172	0.13648 0.34509
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.14264 0.36700	0.16283 0.34628	0.13594 0.34162
30	0.00000 0.00000	0.14554 0.30614	0.16797 0.33162						
40	0.00000 0.00000	0.14891 0.30169							

ALFA=0.005

P=0.250

q=0.250

SIMULACIONES= 200

INDICE DE DISPERSION

	2	4	6	8	10	15	20	30	40
2	0.02409 2.37713	0.06335 1.71919	0.02641 1.44997	0.03759 1.58589	0.06402 1.46597	0.04354 1.49335	0.04537 1.63891	0.11030 1.50617	0.06510 1.48643
4	0.00000 0.00000	0.25444 2.59587	0.37899 1.94044	0.55904 1.89253	0.54044 1.75276	0.50986 1.67735	0.55864 1.74127	0.36391 1.67585	0.42850 1.63529
6	0.00000 0.00000	0.00000 0.00000	0.31314 1.76919	0.44269 1.69624	0.51802 1.64108	0.53925 1.34606	0.43345 1.41527	0.50724 1.36222	0.51940 1.39106
8	0.06000 0.00000	0.00000 0.00000	0.60000 0.00000	0.54096 1.72840	0.61969 1.65701	0.55304 1.54171	0.70884 1.49601	0.69344 1.46935	0.54924 1.36270
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.63269 1.42996	0.66492 1.34735	0.59671 1.35747	0.66827 1.27890	0.65285 1.29470
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.67977 1.33401	0.72450 1.34062	0.64926 1.26297	0.58969 1.22572
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.75756 1.27605	0.74028 1.25845	0.79142 1.23666
30	0.00000 0.00000	0.77789 1.18936	0.79470 1.23467						
40	0.00000 0.00000	0.81551 1.20399							

	ALFA=0.005		P=0.250		Q=0.250		SIMULACIONES= 200		VALORES DEL INDICE A		
	2	4	6	8	10	15	20	30	40		
2	0.03496 2.27278	0.03613 1.21872	0.03074 1.26053	0.01007 1.14826	0.03795 1.16697	0.06460 1.11950	0.02114 0.95060	0.03628 1.01701	0.02923 1.05795		
4	0.00000 0.00000	0.33475 1.35616	0.46378 1.35115	0.37147 1.20282	0.49033 1.18951	0.36142 1.15413	0.38504 1.08087	0.38020 1.07188	0.57495 1.11413		
6	0.00000 0.00000	0.00000 0.00000	0.40729 1.12586	0.33338 0.96860	0.39889 1.09058	0.34192 1.02210	0.36833 0.95766	0.33927 0.91653	0.40707 0.88490		
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.41144 1.03298	0.42830 1.07960	0.53641 1.01857	0.46000 0.95138	0.47460 0.90490	0.50000 0.97437		
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.47292 0.95667	0.45476 0.92648	0.40155 0.93831	0.47436 0.89242	0.48176 0.90709		
15	0.00000 0.00000	0.00000 1.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.43103 0.89988	0.46046 0.89254	0.52522 0.87673	0.45912 0.86003		
20	0.00000 0.00000	0.00000 1.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.56335 0.88495	0.51660 0.87312	0.50377 0.86060		
30	0.00000 0.00000	0.50000 0.85875	0.50343 0.82300								
40	0.00000 0.00000	0.00000 1.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.57496 0.85811		

ALFA=0.025

P=0.250

q=0.250

SIMULACIONES= 200

INDICE DE DISPERSION

	2	4	6	8	10	15	20	30	40
2	0.07975 1.55383	0.12621 1.52211	0.11769 1.31036	0.10465 1.39493	0.08917 1.32274	0.12389 1.35497	0.10489 1.30319	0.18158 1.32004	0.16484 1.40258
4	0.00000 0.00000	0.45070 2.19548	0.50544 1.61690	0.59376 1.65262	0.61391 1.57851	0.63446 1.60401	0.64942 1.63442	0.58267 1.54354	0.62051 1.58361
6	0.00000 0.00000	0.00000 0.00000	0.53209 1.35590	0.56559 1.47737	0.58855 1.41401	0.61173 1.28245	0.64451 1.32483	0.63771 1.24322	0.61918 1.36620
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.68720 1.44159	0.69165 1.44828	0.69717 1.34427	0.76390 1.37644	0.73678 1.37867	0.71319 1.30152
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.65616 1.36472	0.71240 1.27395	0.68054 1.31602	0.73393 1.26012	0.72429 1.25508
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.72670 1.20894	0.74968 1.23035	0.71164 1.19493	0.75075 1.16909
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.80018 1.25513	0.78478 1.22772	0.83712 1.20597
30	0.00000 0.00000	0.81361 1.18312	0.84672 1.17076						
40	0.00000 0.00000	0.86000 0.00000	0.84776 1.20016						

ALFA=0.025

P=0.250

Q=0.250

SIMULACIONES= 200

VALORES DEL INDICE A

	2	4	6	8	10	15	20	30	40
2	0.10574 1.90070	0.09063 1.07705	0.09991 1.16718	0.05215 1.04376	0.07299 1.08202	0.09012 0.99925	0.07361 0.87871	0.10084 0.88106	0.14180 0.93231
4	0.00000 0.00000	0.39716 1.29647	0.54261 1.25195	0.46365 1.12541	0.54307 1.12858	0.47743 1.09068	0.50266 1.04082	0.46932 1.04048	0.45969 1.06755
6	0.00000 0.00000	0.00000 0.00000	0.48355 1.09668	0.40667 0.91771	0.43849 1.03208	0.47997 0.95198	0.41028 0.89477	0.42939 0.89434	0.45024 0.87592
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.46288 0.98953	0.56451 0.97775	0.57501 0.97890	0.54872 0.93395	0.57174 0.92365	0.56134 0.92292
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.50655 0.94374	0.52726 0.87125	0.50639 0.84955	0.52297 0.86891	0.51532 0.86899
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.50394 0.87023	0.49820 0.84926	0.55011 0.84527	0.51183 0.83312
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.58207 0.84730	0.59988 0.85802	0.57764 0.82655
30	0.00000 0.00000	0.60196 0.83566	0.58240 0.79606						
40	0.00000 0.00000	0.64935 0.82620							

	ALFA=0.050	P=0.250	Q=0.250	SIMULACIONES= 200			VALORES DEL INDICE A		
	2	4	6	8	10	15	20	30	40
2	0.17807 1.77967	0.11828 1.01410	0.12221 0.11015	0.08764 0.93933	0.09110 0.98340	0.10102 0.91553	0.09517 0.84982	0.12596 0.85671	0.12235 0.83581
4	0.00000 0.00000	0.49445 1.24279	0.58808 1.20878	0.53462 1.08594	0.58545 1.09863	0.52742 1.05578	0.55275 1.01290	0.51932 1.02574	0.49400 1.03939
6	0.00000 0.00000	0.00000 0.00000	0.50028 1.06427	0.44968 0.89217	0.48030 0.96047	0.49750 0.90872	0.47468 0.86147	0.47360 0.85877	0.47750 0.84691
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.49238 0.96358	0.59852 0.96570	0.60296 0.94819	0.58038 0.90219	0.61164 0.90198	0.58410 0.90913
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.53903 0.92348	0.54212 0.85892	0.55050 0.82539	0.54666 0.84505	0.54453 0.85475
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.52035 0.84084	0.52788 0.80892	0.57455 0.82429	0.54392 0.80093
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.60131 0.83199	0.61910 0.84897	0.60207 0.81550
30	0.00000 0.00000	0.62981 0.82083	0.60384 0.79061						
40	0.00000 0.00000	0.60000 0.00000	0.63190 0.80427						

	ALFA=0.050	P=0.250	Q=0.250	SIMULACIONES= 200			INDICE DE DISPERSION		
	2	4	6	8	10	15	20	30	40
2	0.12007 1.11507	0.19029 1.35157	0.19577 1.24111	0.15306 1.17567	0.12992 1.25194	0.14921 1.24598	0.17413 1.25277	0.21191 1.23639	0.18921 1.23769
4	0.00000 0.00000	0.55763 1.79297	0.58528 1.50827	0.65731 1.57424	0.64351 1.54627	0.68282 1.48183	0.70589 1.53586	0.68000 1.48319	0.71980 1.54725
6	0.00000 0.00000	0.00000 0.00000	0.56154 1.39129	0.58600 1.36893	0.64606 1.29341	0.64650 1.25254	0.69800 1.25873	0.66929 1.21462	0.67573 1.25324
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.70399 1.37040	0.72117 1.33600	0.73320 1.28008	0.77744 1.33851	0.77453 1.33559	0.78488 1.25843
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.68601 1.31089	0.75055 1.23425	0.72620 1.25658	0.77193 1.21210	0.76385 1.22405
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.74183 1.14872	0.77533 1.17147	0.77596 1.16790	0.77377 1.14313
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.82821 1.20046	0.79886 1.19309	0.85543 1.17370
30	0.00000 0.00000	0.82638 1.11717	0.85727 1.13832						
40	0.00000 0.00000	0.00000 0.00000	0.86541 1.16161						

	ALFA=0.005	P=0.500	q=0.750		SIMULACIONES= 200		INDICE DE DISPERSION		
	2	4	6	8	10	15	20	30	40
2	0.25052 *****	0.21487 3.23100	0.36105 3.82359	0.21467 3.17823	0.17198 3.31992	0.29615 2.96948	0.28436 2.77021	0.36309 2.81499	0.17171 2.76578
4	0.00000 0.00000	0.36211 3.23008	0.64721 2.32976	0.59088 2.16203	0.40565 2.04328	0.49121 1.94942	0.56628 1.91526	0.58844 1.90027	0.52205 1.78608
6	0.00000 0.00000	0.70700 0.00000	0.57514 2.14090	0.57984 2.04164	0.67517 1.95908	0.68098 1.89093	0.60011 1.63238	0.65579 1.70682	0.71129 1.66418
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.48970 1.70348	0.57917 1.81000	0.60499 1.86796	0.69772 1.60104	0.74928 1.51180	0.59314 1.54724
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.68341 1.73555	0.68053 1.58988	0.71633 1.50784	0.70246 1.56311	0.63384 1.44773
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.79857 1.57631	0.79283 1.54428	0.82737 1.50080	0.80074 1.41642
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.75037 1.47824	0.82604 1.43525	0.80308 1.44259
30	0.00000 0.00000	0.85793 1.39923	0.82653 1.49032						
40	0.00000 0.00000	0.00000 0.00000	0.82216 1.34797						

ALFA=0.005

P=0.500

Q=0.750

SIMULACIONES= 200

VALORES DEL INDICE A

	2	4	6	8	10	15	20	30	40
2	0.01598 2.22737	0.24385 2.02289	0.36562 1.71672	0.25791 1.80231	0.26061 1.68400	0.35815 1.66598	0.25579 1.66678	0.45453 1.61640	0.24251 1.59101
4	0.00000 0.00000	0.35696 1.48163	0.40314 1.32859	0.54295 1.33565	0.51153 1.27901	0.52350 1.22646	0.65164 1.24153	0.58606 1.22182	0.57110 1.20820
6	0.00000 0.00000	0.06000 0.00000	0.57785 1.24130	0.49066 1.23461	0.55916 1.15465	0.67277 1.16035	0.64064 1.14679	0.72346 1.13220	0.63290 1.14264
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.52843 1.19343	0.59075 1.13069	0.65152 1.11986	0.70673 1.11738	0.64662 1.08003	0.65151 1.08798
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.61391 1.09267	0.66295 1.08748	0.74999 1.11437	0.66406 1.05838	0.74338 1.07562
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.69459 1.06873	0.76380 1.07596	0.77489 1.03495	0.75681 1.04757
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.79062 1.04078	0.78011 1.01706	0.78782 1.02442
30	0.00000 0.00000	0.74447 1.01914	0.82215 0.99905						
40	0.00000 0.00000	0.79412 1.00256							

ALFA=0.025

P=0.500

Q=0.750

SIMULACIONES= 200

INDICE DE DISPERSION

	2	4	6	9	10	15	20	30	40
2	0.45949 9.83044	0.38105 2.58538	0.26175 3.07419	0.43921 2.74304	0.44014 2.88623	0.50046 2.76254	0.54415 2.50470	0.53476 2.58679	0.42667 2.31955
4	0.00000 0.00000	0.48912 0.85156	0.71173 2.00742	0.69714 1.86104	0.59521 1.82956	0.65402 1.85821	0.60176 1.72678	0.73985 1.73279	0.72532 1.62809
6	0.00000 0.00000	0.00000 0.00000	0.78538 1.94546	0.72602 1.80032	0.76758 1.79098	0.70824 1.66171	0.71464 1.56621	0.77168 1.59891	0.81609 1.61773
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.70949 1.66599	0.74478 1.58461	0.84639 1.66910	0.79399 1.51955	0.82724 1.48692	0.77621 1.47976
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.77569 1.62740	0.84651 1.54188	0.84944 1.45630	0.85906 1.48685	0.85741 1.43269
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.87170 1.52621	0.84103 1.37499	0.87166 1.37264	0.85053 1.36647
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.84834 1.42235	0.86723 1.37032	0.87925 1.35071
30	0.00000 0.00000	0.95576 1.35423	0.91592 1.31007						
40	0.00000 0.00000	0.94158 1.26004							

	ALFA=0.025	P=0.500	Q=0.750	SIMULACIONES= 200			VALORES DEL INDICE A		
	2	4	6	8	10	15	20	30	40
2	0.11753 0.22729	0.49001 1.44951	0.51986 0.68014	0.61187 0.95177	0.43109 0.84947	0.54944 1.57504	0.57734 1.63148	0.69176 1.54965	0.47977 1.54875
4	0.00000 0.00000	0.58555 1.48083	0.54860 0.52235	0.72578 0.34318	0.66715 0.23767	0.65791 1.21712	0.72195 1.22957	0.68807 1.20829	0.75541 1.19718
6	0.00000 0.00000	0.00000 0.00000	0.67563 1.19866	0.70491 0.24035	0.67765 1.14559	0.73341 1.12428	0.72109 1.15313	0.77166 1.12252	0.75857 1.11235
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.60182 0.48130	0.67482 1.11629	0.72644 1.09624	0.74201 1.11170	0.79441 1.07441	0.77529 1.07420
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.65869 1.08811	0.67901 1.07532	0.79806 1.08363	0.77265 1.04644	0.81169 1.06134
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.76104 1.05668	0.82140 1.06254	0.81795 1.03043	0.81246 1.03463
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.80212 1.03268	0.80745 1.00602	0.81446 1.00835
30	0.00000 0.00000	0.78981 0.99935	0.84277 0.99318						
40	0.00000 0.00000	0.00000 0.00000	0.83124 0.98723						

ALFA=0.050

P=0.500

q=0.750

SIMULACIONES= 200

INDICE DE DISPERSION

	2	4	6	8	10	15	20	30	40
2	0.47230 8.38840	0.47990 2.35276	0.61004 2.82307	0.56177 2.43955	0.56656 2.63896	0.58428 2.34358	0.65912 2.32370	0.61791 2.43924	0.62430 2.26769
4	0.00000 0.00000	0.62889 1.75645	0.74455 1.95256	0.82557 1.76782	0.70246 1.78055	0.70356 1.78897	0.74714 1.67386	0.80134 1.63530	0.82142 1.57506
6	0.00000 0.00000	0.00000 0.00000	0.82169 1.89674	0.79443 1.59854	0.82118 1.69787	0.80824 1.60374	0.79840 1.51805	0.83286 1.50836	0.85743 1.50886
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.79299 1.53239	0.82363 1.52153	0.88255 1.55991	0.84883 1.48009	0.86050 1.41894	0.84162 1.42958
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.80941 1.52675	0.89385 1.44856	0.88792 1.40387	0.91153 1.38980	0.91317 1.38146
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.91155 1.46220	0.87971 1.32931	0.90122 1.35159	0.93466 1.30751
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.88730 1.35056	0.91848 1.33941	0.93465 1.31194
30	0.00000 0.00000	0.94851 1.32146	0.94698 1.27989						
40	0.00000 0.00000	0.95529 1.26429							

	ALFA=0.050		P=0.500		Q=0.750		SIMULACIONES= 200		VALORES DEL INDICE A		
	2	4	6	8	10	15	20	30	40		
2	0.14517 1.83651	0.61578 1.85832	0.54663 1.65099	0.99620 1.86210	0.56985 1.60028	0.67203 1.54923	0.67833 1.61171	0.72071 1.52966	0.67753 1.52160		
4	0.00000 0.00000	0.68116 1.39930	0.64969 1.25502	0.77589 1.29020	0.72394 1.22677	0.72937 1.20026	0.82738 1.22197	0.78230 1.18211	0.82977 1.18339		
6	0.00000 0.00000	0.00000 0.00000	0.70231 1.17706	0.79228 1.20120	0.71656 1.13744	0.74939 1.10632	0.76499 1.11787	0.81426 1.09329	0.79281 1.10235		
8	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.74696 1.14911	0.74015 1.10036	0.76408 1.07598	0.80794 1.09530	0.81463 1.06698	0.80974 1.06184		
10	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.71175 1.06726	0.73433 1.05721	0.81949 1.06397	0.80890 1.03684	0.84160 1.05742		
15	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.78421 1.03863	0.84727 1.05752	0.85282 1.02674	0.86852 1.02991		
20	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.00000 0.00000	0.82548 1.02560	0.82811 1.00031	0.83542 1.00282		
30	0.00000 0.00000	0.82656 0.99265	0.85260 0.98890								
40	0.00000 0.00000	0.84638 0.97635									

## AGRADECIMIENTOS

El tema de la tesis fue sugerido por Yoko Sugiura durante un curso en la Escuela Nacional de Antropología e Historia en enero de 1983; sirva este resultado como mínimo reconocimiento a su interés por el desarrollo del proyecto. A partir de mayo de ese año el director de tesis fue Guillermo Espinosa, quien facilitó el trabajo enormemente con sus críticas y sugerencias, sobre todo acerca de la forma en que esta tesis podía ser enfocada en su conjunto. En este mismo sentido quisiera agradecer a Margarita Chávez Cano sus comentarios a esta tesis, y, principalmente, su amabilidad.

Evidentemente, el Instituto de Investigaciones Antropológicas de la UNAM hizo posible este trabajo al proporcionar un ambiente de trabajo (casi) óptimo; en particular, agradezco al Dr. Jaime Litvak el apoyo necesario para que un proyecto de matemáticas pudiera realizarse en un instituto de antropología y por sus comentarios acerca de ese trabajo. Asimismo, debo reconocer la influencia que tuvo en la tesis la convivencia generada a partir del trabajo con los compañeros del Instituto.

Por otra parte, agradezco su ayuda durante y después de la carrera a la mayoría de los profesores que tuve en la Facultad de Ciencias y a todos mis compañeros de generación; al personal de la Sección Escolar de la Facultad y al de la Biblioteca del IIMAS por su poco sentido burocrático; a mis alumnos de la ENAH por aguantarme en las primeras veces que he dado clase; a don Fernando Botas, Moisés Aguirre y Víctor Ortiz por los dibujos de esta tesis; a Patricia Martel por sus comentarios a la redacción final; a Anita, Manuel y su abuelita (que si tuviera ruedas a lo mejor sería bicicleta); a Polo Tzinacantlácatl (aunque no devuelva el poster);

a Erik por las esferas; y, por último, pero no al último, a mi familia, que en una u otra forma siempre me ha apoyado.

Ciudad Universitaria,

23 de octubre de 1984

## REFERENCIAS

- Abler, R., J.S. Adams y P. Gould (1971), Spatial organization. The geographer's view of the world, Prentice-Hall, Englewood Cliffs, Nueva Jersey.
- Alden, J.R. (1979), "A reconstruction of Toltec period political units in the valley of Mexico", en Renfrew, C. y K.L. Cooke (eds.) Transformations. Mathematical approaches to culture change, Academic Press, Nueva York.
- Aldenderfer, M.S. (1981), "Computer simulation for archaeology: an introductory essay", en Sabloff, J. A. (ed.) Simulations in Archaeology, University of New Mexico Press, Albuquerque.
- Anderberg, M. (1973), Cluster Analysis for applications, Academic Press, Nueva York.
- Aranda, F. y Rufiz-Velasco, S. (1984), SAREG. Un programa interactivo para simulación y análisis de problemas de regresión, Comunicaciones Técnicas, serie azul num. 79, IIMAS/UNAM.
- Atkinson, A.C. (1979), "Computer generation of Poisson random variables", JRSS-G28, pp. 329-357.
- Bartlett, M.S. (1974), "The statistical analysis of point pattern", AAP 6, pp. 336-358.
- Ibid. (1975), The statistical analysis of point pattern, Chapman and Hall, Londres.
- Ibid. (1978), "Nearest neighbour models in the analysis of field experiments", JRSS-B40, pp. 147-174.
- Bell, J.A. (1981), "Scientific method and the formulation of testable computer simulation models", en Sabloff, J.A. (ed.), Simulations in Archaeology, University of New Mexico Press, Albuquerque.
- Buck, W. (1977), Advanced Calculus, McGraw-Hill, Tokio.

- Chadwick, A.J. (1978), "A computer simulation of Mycenaean settlement", en Hodder, I. (ed.), Simulation studies in archaeology, Cambridge University Press, Cambridge.
- Ibid. (1979), "Settlement simulation", en Renfrew, C. y K.L. Cooke (eds.), Transformations. Mathematical approaches to culture change, Academic Press, Nueva York.
- Chisholm, M. (1979), Human geography, Pelican Books, Londres.
- Clark, P.J. y F.C. Evans (1954), "Distance to nearest neighbor as a measure of spatial relationships in populations", Ecology 35, pp. 445-453.
- Ibid. (1979), "Generalization of a nearest neighbor measure of dispersion for use in k dimensions", Ecology 60, pp. 316-317.
- Clarke, D.L., (1972) (ed.) Models in archaeology, Methuen, Londres.
- Ibid. (1977) (ed.) Spatial archaeology, Academic Press, Londres.
- Ibid. (1977) "Spatial information in archaeology", en Clarke, D.L. (ed.) Spatial archaeology, Academic Press, Londres.
- Cox, D. y D.V. Hinkley (1978) Theoretical Statistics, Chapman and Hall, Londres.
- Cox, T.F. y T. Lewis (1976), "A conditioned distance ratio method for analysing spatial patterns", Biometrika 40, pp. 483-491.
- Curry, L. (1964), "The random spatial economy: an exploration in settlement theory", AAAG 54, pp. 138-146.
- Dacey, M.F. (1960), "A note on the derivation of nearest-neighbor distances", Journal of Regional Science vol. 2, pp. 81-87.
- Ibid. (1964), "Modified Poisson probability law for point pattern more regular than random", AAAG 54, pp. 559-565.
- David, J.C. (1973), Statistics and data analysis in geology, John Wiley, Nueva York.
- Diggle, P.J., J. Besag y J.T. Gleaves (1976), "Statistical analysis

of spatial point patterns by means of distance methods",  
Biometrics 32, pp. 659-667.

Donnelly, K.P. (1978), "Simulation to determine the variance and edge effect of total nearest-neighbour distance", en Hodder, I. (ed.), Simulation studies in archaeology, Cambridge University Press, Cambridge.

Doran, J. y F.R. Hodson (1975), Mathematics and Computers in Archaeology, Edinburgh University Press, Edinburgo.

Fishman, G.S. (1973), Concepts and methods in discrete event digital simulation, Wiley-Interscience, Nueva York.

Garner, B. (1967), "Models of urban geography and settlement location" en Haggett, P. y R.J. Chorley (eds.), Models in geography, Methuen, Londres.

Getis, A. (1964), "Temporal land-use patterns analysis with the use of nearest-neighbour and quadrats methods", AAAG 54, pp. 391-399.

Getis, A. y B. Boots (1978), Models of spatial processes, Cambridge University Press, Cambridge.

Gómez, M.T. (1980), Algunos modelos de patrón de asentamiento y su aplicación en arqueología, tesis de matemático, Facultad de Ciencias, UNAM.

Greig-Smith, P. (1964), Quantitative Plant Ecology, Butterworths, Londres.

Hacking, I. (1965), The logic of statistical inference, Cambridge University Press, Cambridge.

Haggett, P. y R.J. Chorley (eds.) (1967), Models in Geography, Methuen, Londres.

Haggett, P., A.D. Cliff y A. Frey (1977), Locational models in human geography, Edward Arnold, Londres.

- Hammersley, J.M. y D.C. Handscomb (1979), Monte Carlo methods, Chapman and Hall, Londres.
- Hammond, R. y P.S. McCullagh (1978), Quantitative techniques in geography, Oxford University Press, Oxford.
- Hansen, R.M. y E.E. Remmenga (1961), "Nearest neighbor concept applied to pocket gopher populations", Ecology 42, pp. 812-814.
- Harvey, D. (1973), Explanation in geography, Edward Arnold, Londres.
- Hodder, I. (1972), "Locational models and the study of Romano-British settlements", en Clarke, D.L. (ed.), Models in archaeology, Methuen, Londres.
- Ibid. (1977), "Some new directions in the spatial analysis of archaeological data at the regional scale (macro)", en Clarke, D.L. (ed.), Spatial Archaeology, Academic Press, Londres.
- Hodder I. y E. Okell (1978), "An index for assesing the association between distributions of points in archaeology", en Hodder I. (ed.), Simulation studies in archaeology, Cambridge University Press, Cambridge.
- Hodder, I. y C. Orton (1976), Spatial analysis in archaeology, Cambridge University Press, Cambridge.
- Hoel, P., S. Port y C. Stone (1971), Introduction to probability theory, Houghton-Mifflin, Boston.
- Hogg, R.V. y E.A. Tanis (1977), Probability and statistical inference, Collier-MacMillan, Nueva York.
- Hogg, R.V. y A.T. Craig, (1978), Introduction to mathematical statistics, Collier-MacMillan, Nueva York.
- Holgate, P. (1964), "The efficiency of nearest-neighbour estimators", Biometrics 20, pp. 647-649.
- Ibid. (1965), "Tests for randomness based on distance methods", Biometrika 52, pp. 345-353.

- Ibid. (1972), "The use of distance methods for the analysis of spatial distributions of points", en Lewis, P.A.W. (ed.), Stochastic point processes, John Wiley, Nueva York.
- House, P. y J. MacLeod (1977), Large-scale models for policy evaluation, Wiley-Interscience, Nueva York.
- King, L.J. (1968), "A quantitative expression of the pattern of urban settlement in selected areas of the United States", en Marble, D.F. y B.J.L. Berry (eds.), Spatial analysis: a reader in statistical geography, Prentice-Hall, Englewood Cliffs, Nueva Jersey.
- Ibid. (1969), Statistical analysis in geography, Prentice-Hall, Englewood Cliffs, Nueva Jersey.
- Kleijnen, J.P.C. (1974), Statistical techniques in simulation, Marcel Dekker, Nueva York.
- Larson, H.J. (1978), Introduction to probability theory and statistical inference, Wiley series in probability and mathematical statistics, Nueva York.
- Lewis, P.A.W. (1972), (ed), Stochastic point processes, John Wiley, Nueva York.
- Lindgren, B.W., (1970), Statistical theory, MacMillan, Nueva York.
- Loomis, C. y H. Sternberg, (1968), Advanced Calculus, Addison-Wesley, Nueva York.
- Marble, D.F y B.J.L. Berry (eds.) (1968), Spatial analysis: a reader in statistical geography, Prentice-Hall, Englewood Cliffs, Nueva Jersey.
- McNutt, C.H. (1981), "Nearest neighbors, boundary effects and the old flag trick: a general solution", Am.Ant. 46, pp. 571-591.
- Miller, R.C. y J.S. Kahn (1962), Statistical analysis in the geological sciences, John Wiley, Nueva York.

- Miller, R.S. y W.J.D. Stephen (1966), "Spatial relationships in flocks of Sandhill Cranes (*Grus Canadensis*)", Ecology 47, pp. 323-327.
- Milnor, J.W. (1979), Topología desde un punto de vista diferencial, comunicación interna num. 3, Facultad de Ciencias, UNAM.
- Ord, K. (1975), "Estimation methods for models of spatial interaction", JASA 70, pp. 120-126.
- Papoulis, A., (1972), Probability, random variables and stochastic processes, McGraw-Hill, Tokio.
- Pielou, E.C. (1959), "The use of point-to-plant distances in the study of the pattern of plant populations", JE 47, pp. 607-613.
- Ibid. (1961), "Segregation and symmetry in two species populations as studied by nearest neighbour methods", JE 49, pp. 255-269.
- Ibid. (1962), "The use of plant-to-neighbour distances for the detection of competition", JE 50, pp. 357-367.
- Ibid. (1969), An introduction to mathematical ecology, John Wiley, Nueva York,
- Ibid. (1977), Mathematical Ecology, Wiley-Interscience, Nueva York.
- Pinder, D., I. Shimada y D. Gregory (1979), "The nearest neighbor statistics: archaeological applications and new developments", Am.Ant. 44, pp. 430-444.
- Popper, K.R. (1975), The logic of scientific discovery, Hutchinson, Londres.
- Renfrew, C. y K.L. Cooke (eds.) (1979), Transformations. Mathematical approaches to culture change, Academic Press, Nueva York.
- Ripley, B.D. (1977), "Modelling spatial patterns", JRSS-B39, pp. 172-212.
- Ibid. (1979), "Tests for 'randomness' for spatial point patterns", JRSS-B41, pp. 368-374.

- Ibid. (1982), Spatial statistics, John Wiley series in probability and mathematical statistics, Nueva York.
- Ross, S. M. (1970), Applied probability models with optimization applications, Holden-Day, San Francisco.
- Sabloff, J.A. (ed.) (1981), Simulations in archaeology, University of New Mexico Press, Alburquerque.
- Stanislawski, M.B. (1973), "review: Longacre, W., 'Archaeology as Anthropology'", Am.Ant. 38, pp. 117-122.
- Stark, B. y D.L. Young (1981), "Linear nearest neighbor analysis", Am.Ant. 46, pp. 284-300.
- Sugiura, Y y M. Cortina (en prensa), "Epiclásico en la cuenca del río Lerma; aplicación de dos modelos estadísticos: el análisis de vecino más cercano y la estadística de Pielou", Revista mexicana de estudios antropológicos.
- Thompson, H.R. (1956), "Distribution of distance to n-th neighbour in a population of randomly distributed individuals", Ecology 37, pp. 391-394.
- Tocher, K.D. (1974), The art of simulation, The English Universities Press, Londres.
- Washburn, D.K. (1974), "Nearest neighbor analysis of Pueblo I-III settlement patterns along the Río Puerco of the East, New Mexico", Am.Ant. 39, pp. 315-335.
- Whallon, R. (1974), "Spatial analysis of occupation floors II: the application of nearest neighbor analysis", Am.Ant. 39, pp. 16-34.
- Young, D.L. (1982), "The linear nearest neighbour statistics", Biometrika 69, pp. 477-480.

Zubrow, E.Z.B. y H.W. Harbaugh (1978), "Archaeological prospecting: kriging and simulations", en Hodder, I. (ed.), Simulation studies in archaeology, Cambridge. University Press, Cambridge.

## ADDENDA

Lehmann, M. (1959), Testing Statistical Hypothesis, Wiley Publications in Statistics, Nueva York.

Silvey, D.S. (1975), Statistical Inference, Chapman and Hall, Londres.

## Abreviaturas usadas:

AAAG	Annals of the Association of American Geographers
AmAnt.	American Antiquity
JASA	Journal of the American Statistical Association
JE	Journal of Ecology
JRSS	Journal of the Royal Statistical Society