

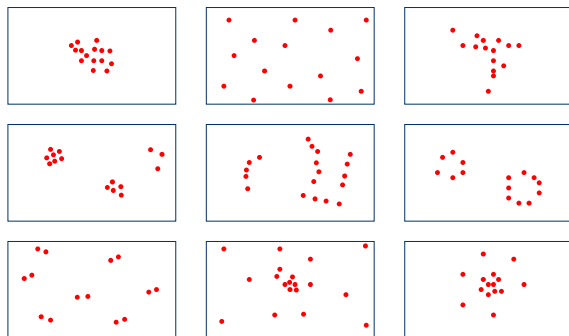
VI. 空間解析

VI. 空間解析

VI-1 空間解析とは何か

空間中のオブジェクト分布や現象においてパターン、秩序、規則性を見いだす（あるいはその不在を確認するための数理的手法）

空間解析によって見いだされるパターンは、その背後にある構造・原因を明らかにするのを助ける



VI. 空間解析

パターンというのは、目で見てわかるような明らかなものから、一見してもわからない複雑な構造を持つものまで様々である。

空間解析では、（どちらかと言えば）前者よりも後者に重きを置く。特に、そのパターンが「偶然」起こったものなのか、「特別な」ものなのかを調べるために、統計的検定をよく用いる。

VI. 空間解析

都市における商業施設の分布

銀行、バー、酒屋、コンビニエンスストア、ファミリーレストラン、...

⇒ 横浜市の事例

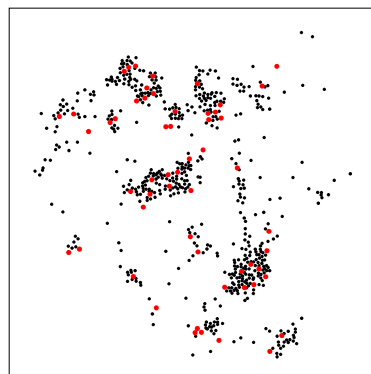
VI. 空間解析

一体、これらの商業施設は、なぜこのような分布パターンをとっているのだろうか。

パターン形成の背景にある行動原理は何か。

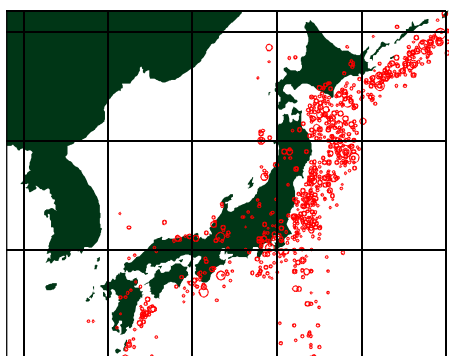


Cases of Childhood Cancer
in the West Midlands Health Authority Region from 1980 to 1984.



- Lung cancer
- Larynx cancer

Lung and larynx cancer cases in the Chorley-Ribble area



Spatial distribution of large earthquakes
1895-1980 in Japan and its vicinity

VI. 空間解析

空間解析手法の3つの分類基準

1) 対象とする空間オブジェクトの種類

点分布

線分布

面分布

連続分布 (標高などのスカラー場, ベクトル場など)

流れ分布 (交通流などのflow)

VI. 空間解析

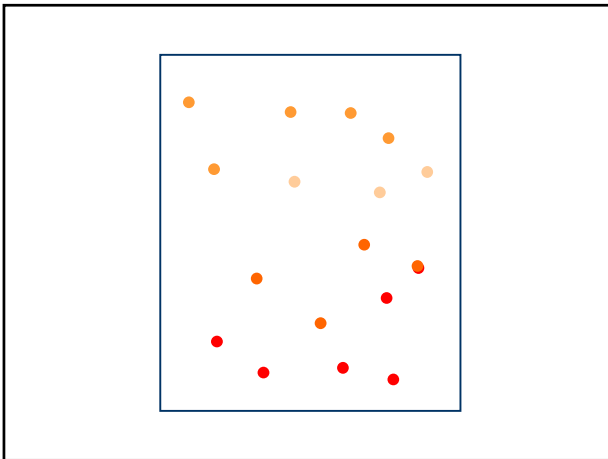
2) 対象とするデータの種類の

空間データのみ

…各オブジェクトが全て同質である場合

空間データ+属性データ

…各オブジェクトの性質がそれぞれ異なる場合



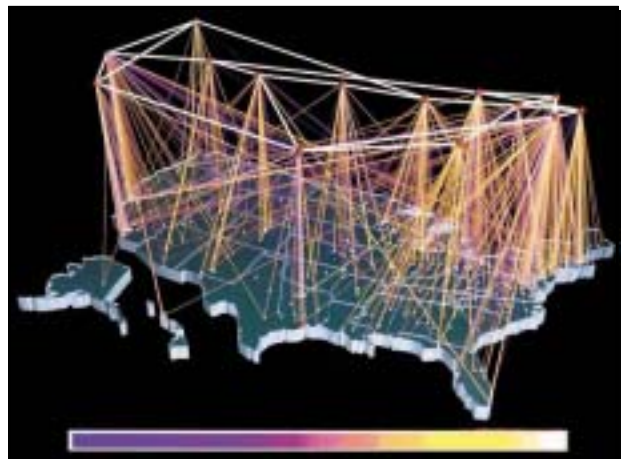
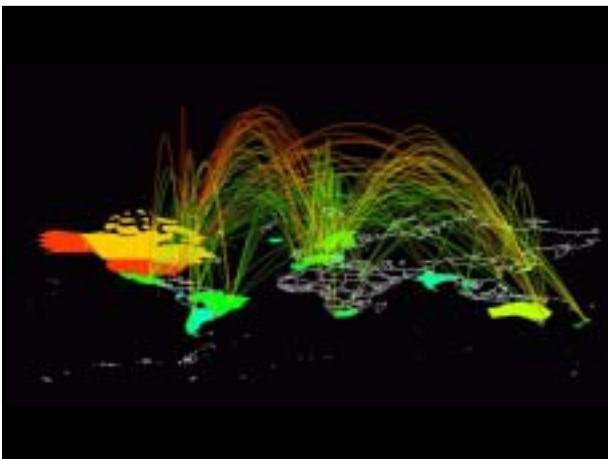
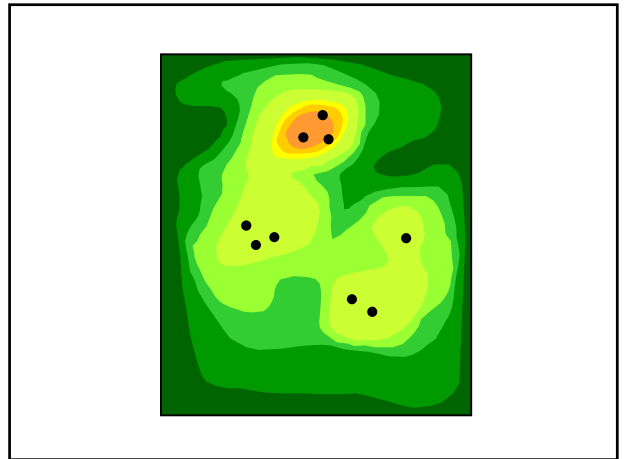
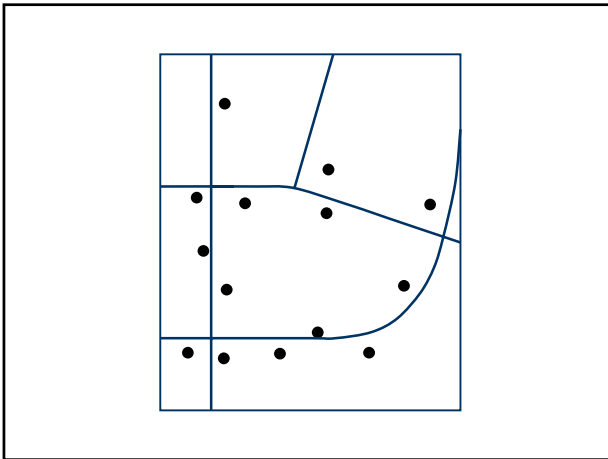
3) 対象とする空間オブジェクト分布の数

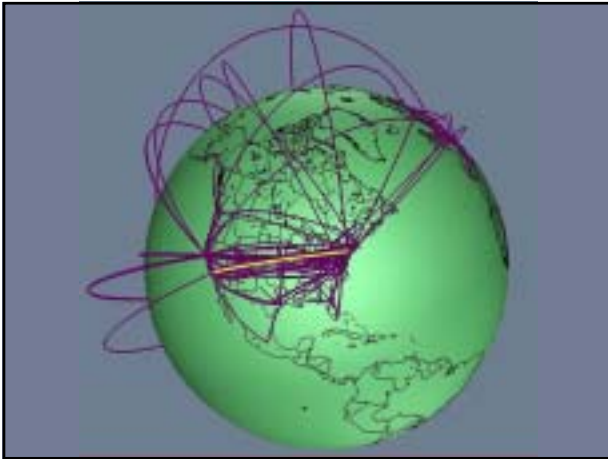
単一の分布を対象とする

…一つの分布の中でのパターン

複数の分布間の関係を扱う

…複数の分布間の関係におけるパターン





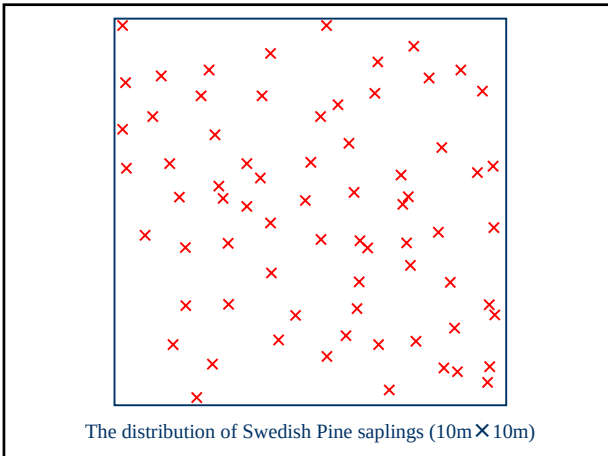
VI. 空間解析

VI-2 単一の点分布を分析する方法

空間に分布する点オブジェクトの分布パターンを分析する方法

「点」と見なすことができるオブジェクトであればこの分析方法を適用することができる。

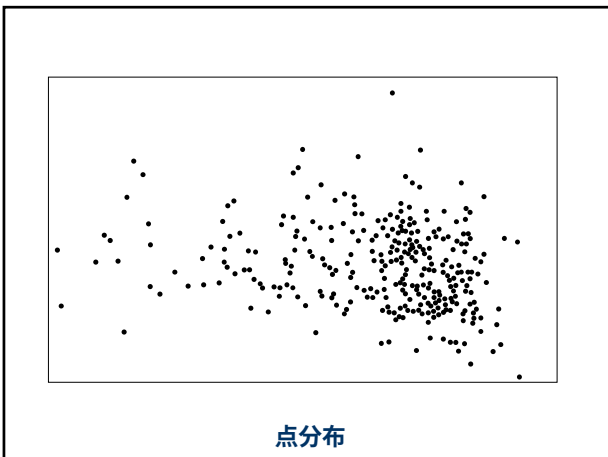
ここでは、点の属性は考慮しない(全ての点は等質であるとする)



VI. 空間解析

VI-2.1 可視化

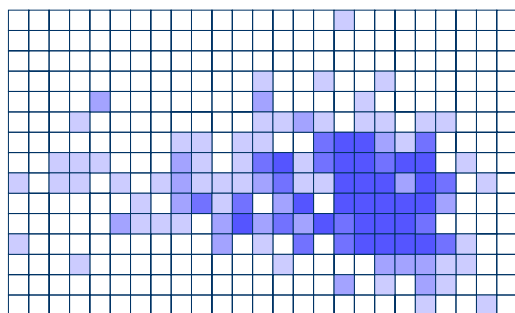
点の分布を視覚的に分析する



VI. 空間解析

VI-2.1.1 区画法

点の分布をラスターデータとして表現する方法



区画法による点分布の可視化

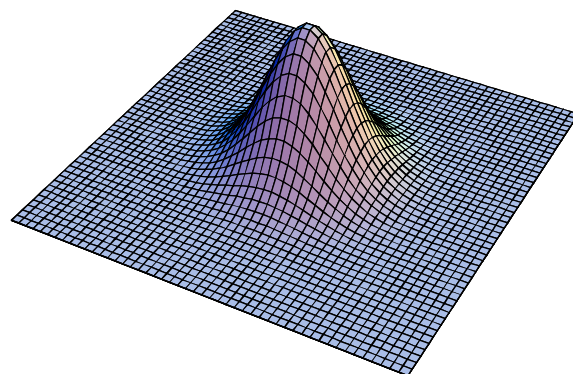
問題点

- a) 分布が滑らかではなく、場合によっては不自然
- b) 区画の大きさによって結果が異なる

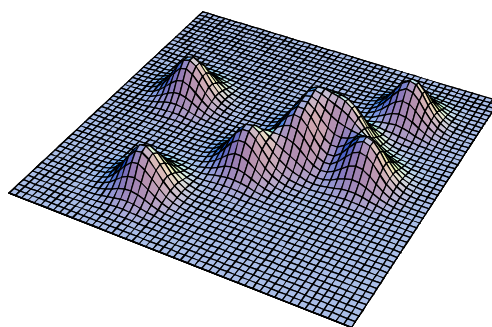
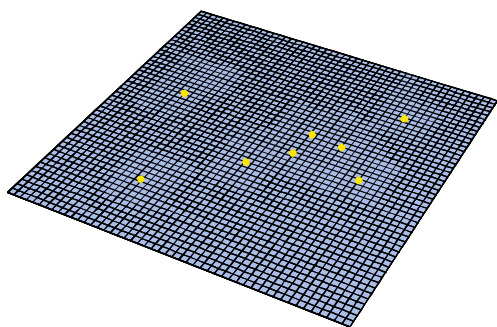
VI-2.1.2 カーネル法

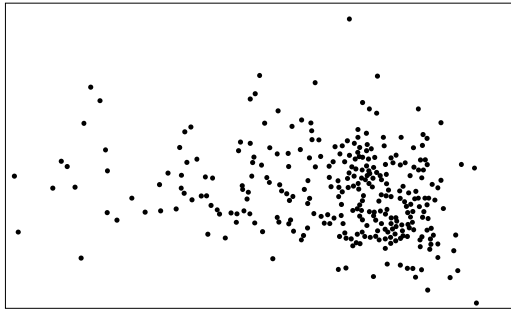
区画法の問題点のうち、「滑らかではない」点を改善した方法

各点を中心に、「カーネル」と呼ばれる「山」を置いていき、その合計を色などで表現する。

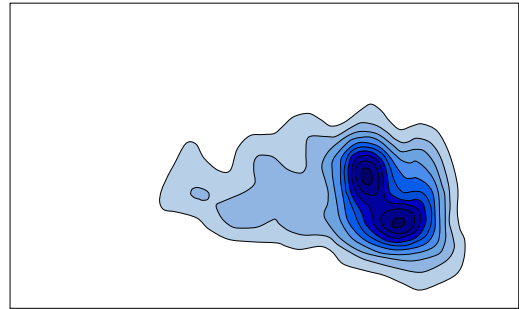


Gaussian kernel





点分布

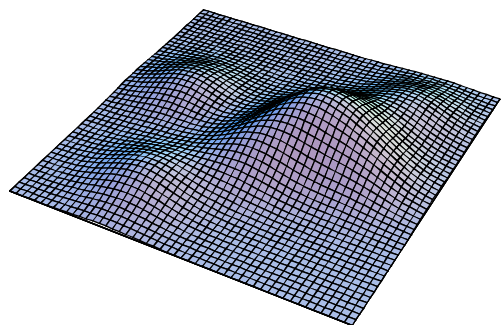
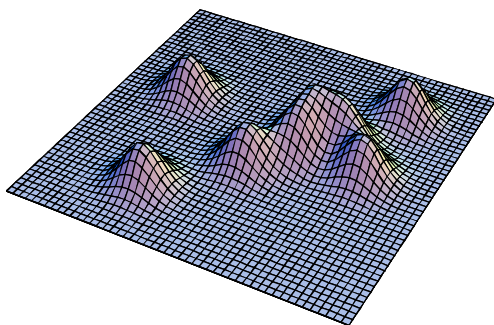
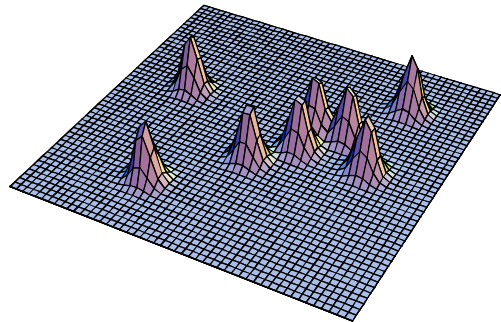


カーネル法による点分布の可視化

VI. 空間解析

問題点

カーネルの形状, 特にその傾きによって結果が異なる



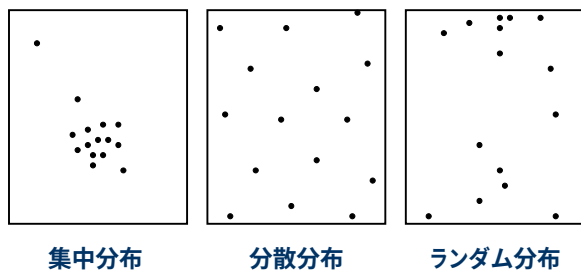
VI-2.2 数値・関数による情報の集約

可視化した点分布には多くの情報が含まれているが、却って情報量が多すぎて全体の様子が分かりにくいことがある。

そこで、情報を数値や関数で集約的に表現するという操作が行われる。

VI-2.2.1 最近隣距離法

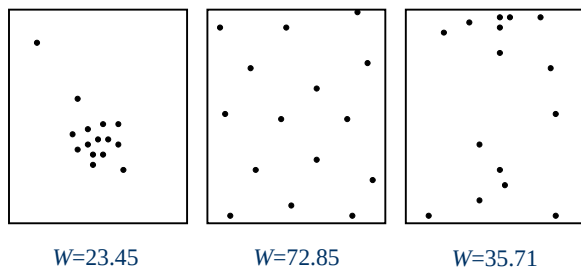
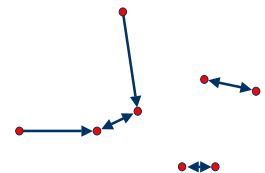
点の分布を「分散型」「集中型」という観点から分類するために用いられる



平均最近隣距離とは、各点から最も近い点までの距離の平均値である。

d_i : 点*i*から最近隣点までの距離
 n : 点の個数

$$W = \frac{1}{n} \sum_{i=1}^n d_i$$



ここで、各分布への分類基準として、点が面積*S*の平面上でランダムに分布している（一様ポアソン分布に従っている）場合を考える。このときの平均最近隣距離*W*の期待値は、

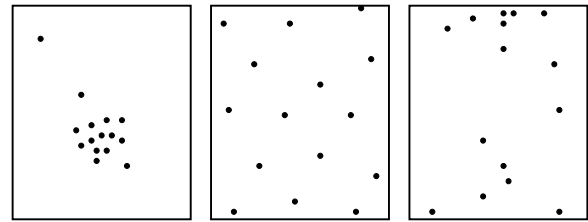
$$E[W] \approx \frac{1}{2\sqrt{n/S}}$$

である。そして、

VI. 空間解析

$W \ll E[W]$: 点は集中している
 $W \approx E[W]$: 点はランダムに分布している
 $W \gg E[W]$: 点は分散している

と判断すればよい。



$W=23.45$

$W=72.85$

$W=35.71$

ランダム分布における W の期待値=39.53

VI. 空間解析

なお、 W や $E[W]$ の値は点の個数や密度に依存する。
 そのため、点の個数や密度の異なる場合を比較する
 には、

$$w = \frac{W}{E[W]}$$

と基準化した値を用いる。この場合、

VI. 空間解析

$w \ll 1$: 点は集中している
 $w \approx 1$: 点はランダムに分布している
 $w \gg 1$: 点は分散している

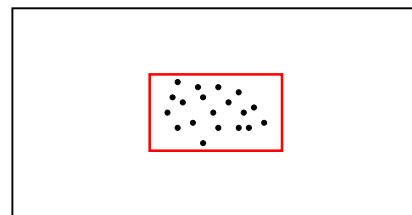
と判断する。

VI. 空間解析

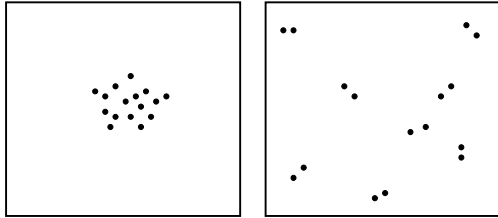
最近隣距離法の問題点

VI. 空間解析

a) 領域の設定によって結果が異なる



b) 最近隣距離だけでは識別のできない分布がある

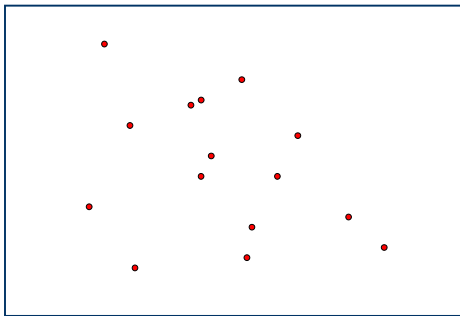


VI-2.2.2 K-関数法

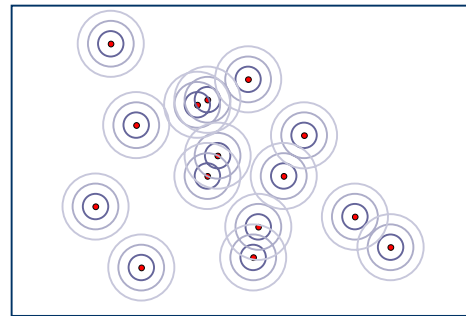
最近隣距離法では判別のつかない分布を識別するために開発された方法

$$K(h) = \frac{\text{点から距離}h\text{以内に存在する点の個数}}{n\lambda}$$

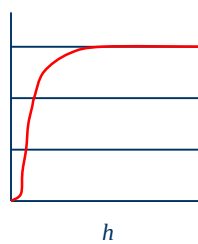
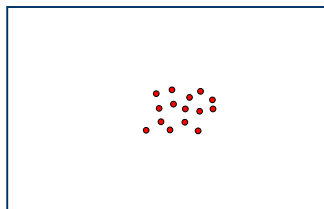
$\lambda: n/S$ (点の密度)



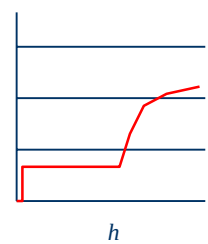
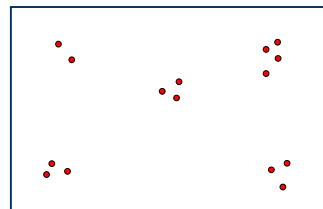
点分布



K-関数の構成法



h



h

K -関数を用いると、どのくらいのスケール (空間的な範囲) で点が集中・分散しているのかを判断することができる。点がランダムに分布している場合、 K -関数の期待値は、

$$E[K(h)] = \pi h^2$$

である。従って、

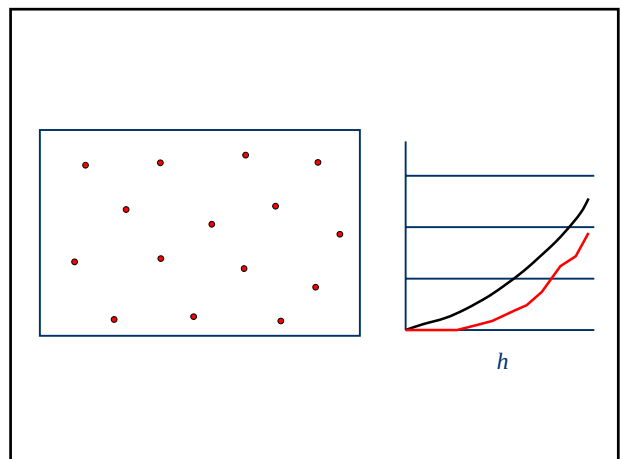
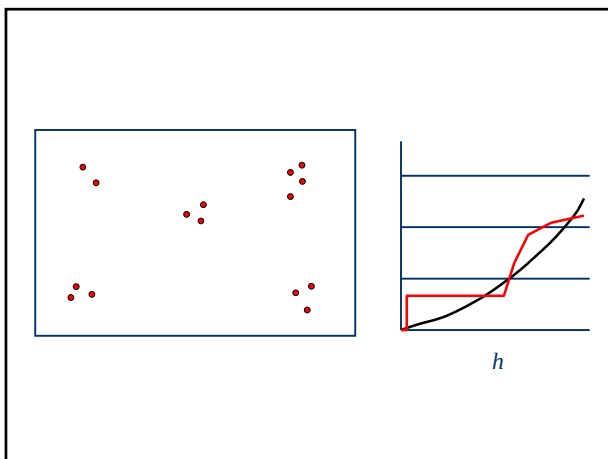
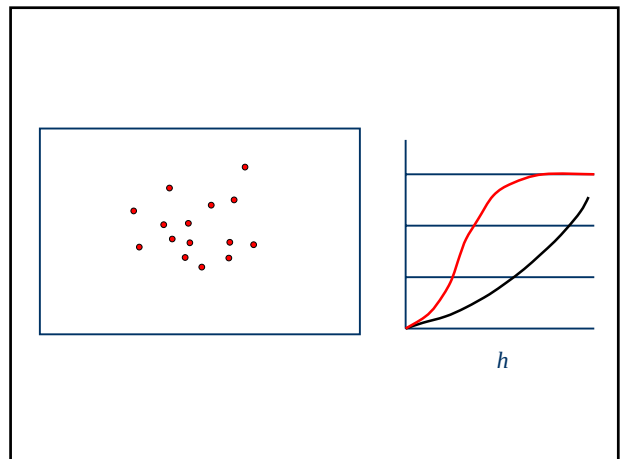
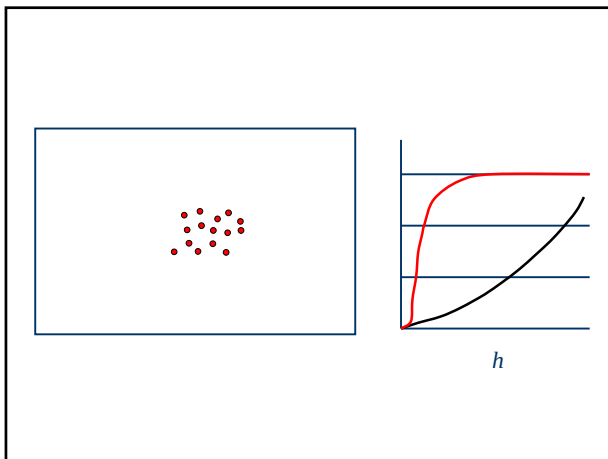
半径 h の円という程度のスケールにおいて、

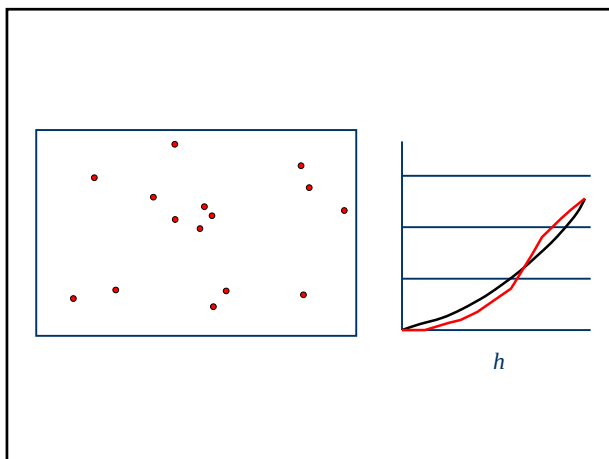
$K[h] \gg \pi h^2$: 点は集中している

$K[h] \approx \pi h^2$: 点はランダムに分布している

$K[h] \ll \pi h^2$: 点は分散している

と判断する。



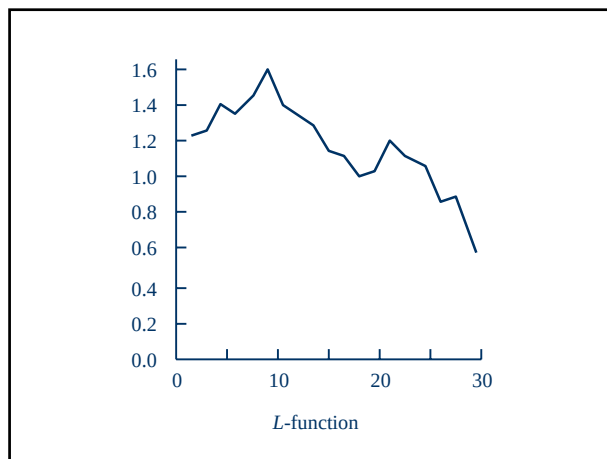
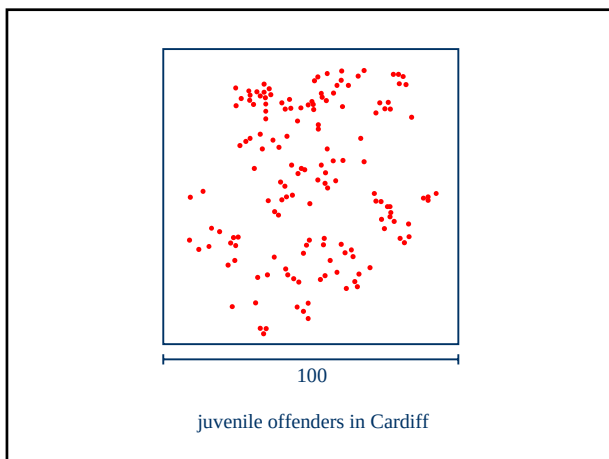


VI. 空間解析

なお, K -関数も平均最近隣距離と同様, 点の個数や密度の影響を取り除くために基準化されることがある. それは通常, L -関数と呼ばれ,

$$L(h) = \sqrt{\frac{K(h)}{\pi}} - h$$

と定義される. L -関数の場合, 点がランダムに分布していれば0, 集中していれば正, 分散していれば負の値をとる.



VI. 空間解析

VI-2.3 統計的検定

点分布の傾向 (集中・分散) をより明確に分類するための方法. 即ち, 平均最近隣距離が「珍しいほどに」小さいのか, 「ほどほどに」小さいのかを明確に区別する.

これによって, パターンが「偶然」発生したものか, あるいは, 何らかの原因によって発生したものかを (確率的判断として) 知ることができる.

VI. 空間解析

VI-2.3.1 最近隣距離法

点の分布が (一様) ランダム分布に従うときの, 平均最近隣距離の確率分布を用いて, 統計的検定を行う.

但し, この分布形を解析的に表すことは難しいため, 普通は近似計算を用いる.

以下, 点の個数を n とする.

1) 点の個数が十分に多い (数100個) の場合

n 個の点の中からランダムに m 個を選び、それらについて平均最近隣距離を計算する.

$$W = \frac{1}{m} \sum_{i=1}^m d_i$$

d_i : 点 i から最近隣点までの距離

点がランダムに分布する場合, この距離 W は近似的に以下の正規分布に従う.

$$N\left(\frac{1}{2\sqrt{\lambda}}, \frac{4-\pi}{4\pi n\lambda}\right)$$

従って, $\lambda=n/S$ とし,

$$Z = \frac{\frac{1}{m} \sum_{i=1}^m d_i - \frac{1}{2\sqrt{\lambda}}}{\sqrt{\frac{4-\pi}{4\pi n\lambda}}}$$

を計算すれば, 標準正規分布による検定ができる.

この場合,

帰無仮説: 点はランダムに分布している

対立仮説: 点はランダムに分布していない (集中あるいは分散分布している)

2) 点の個数があまり多くない場合

この場合, 全ての点について平均最近隣距離を計算する.

$$W = \frac{1}{n} \sum_{i=1}^n d_i$$

点がランダムに分布する場合, この距離 W は近似的に以下の正規分布に従う.

$$N\left(0.5\sqrt{\frac{S}{n}} + 0.051\frac{L}{n} + 0.041\frac{L}{n\sqrt{n}}, 0.070\frac{S}{n^2} + 0.037\sqrt{\frac{S}{n^5}}\right)$$

ここで, S は対象領域の面積, L は周長である.

VI-2.3.2 K-関数法

K -関数を用いても, 平均最近隣距離とほぼ同様の方法で検定を行うことができる.

1) 点の個数が十分に多い (数100個) の場合

n 個の点の中からランダムに m 個を選び、それらについて K -関数を構成する。すると、点がランダムに分布する場合には、 K -関数は以下の正規分布に従う。

$$N\left(\pi h^2, \frac{\pi h^2}{m\lambda}\right)$$

従って、平均最近隣距離と同様の検定が可能である。

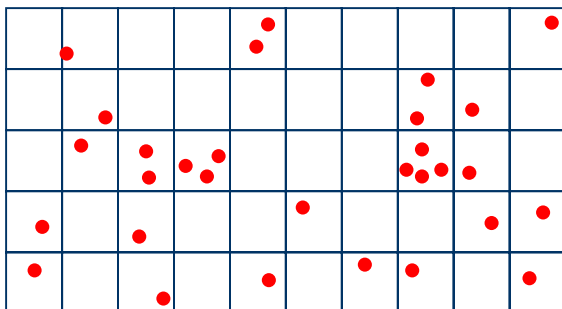
2) 点の個数があまり多くない場合

平均最近隣距離の場合と異なり、この場合の K -関数の分布形を表す良い近似式は見つかっていない。そのため、与えられている分析領域と点の個数を用いてモンテカルロ・シミュレーションを行うことにより分布形を計算する。即ち、所与の領域に所与の個数の点をランダムに分布させ、その都度、 K -関数を計算するという作業を10000回程度繰り返す。

VI-2.3.3 区画法

平均最近隣距離と K -関数は、いずれも点間の距離に基づく検定方法である。それに対し、可視化と同様、区画法を検定に用いることもできる。

いま、分析領域を合同な M 個の領域 (通常は長方形) に分割する。そして、区画 i に含まれる点の個数を x_i とする。



0	1	0	0	2	0	0	0	0	1
0	1	0	0	0	0	0	2	1	0
0	1	2	3	0	0	0	4	1	0
1	0	1	0	0	1	0	0	1	1
1	0	1	0	1	0	1	1	0	1

すると,

$$\chi^2 = \frac{\sum_{i=1}^M (x_i - \bar{x})^2}{\bar{x}}$$

を統計量として用いることが可能である. この値は, 点が集中していれば大きく, 分散していれば小さくなる. また, 点が分散して分布する場合には, 近似的に自由度 $M-1$ の χ^2 分布に従う (χ^2 検定).

ところで, χ^2 検定とは?

いま, M 通りの結果をとりうる離散的現象を考える (さいころ振りなど). そして, この現象を n 回繰り返し, 事象 i の起こった回数 (例えば, 1の目の出た回数) を x_i とする.

一方, 事象 i の起こる理論的確率 (さいころの例でいえば $1/6$) を p_i とする. χ^2 検定とは, 現象が理論と合致しているか, 具体的にいえば, x_i が np_i と等しいと考えて良いかどうかを調べる方法である.

χ^2 検定では,

$$\chi^2 = \frac{\sum_{i=1}^M (x_i - np_i)^2}{np_i}$$

を統計量として用いる. 現象が理論と合致していれば, この値は小さくなり, そうでなければ大きくなる. 従って, 帰無仮説を「理論と現象が一致している」とした検定が可能である. この場合, χ^2 値は自由度 $M-1$ の χ^2 分布に従う.

従って, χ^2 検定における「理論値」を, 点が一様ランダムに分布する場合に各区画に点の入る確率 $1/M$ として考えれば, 点の分布に χ^2 検定に適用することができる.

但し, χ^2 検定は分布が一様であるかどうかの検定であり, ランダムであるかどうかの検定ではないということに注意する必要がある. 即ち,

帰無仮説: 点は一様に分布している

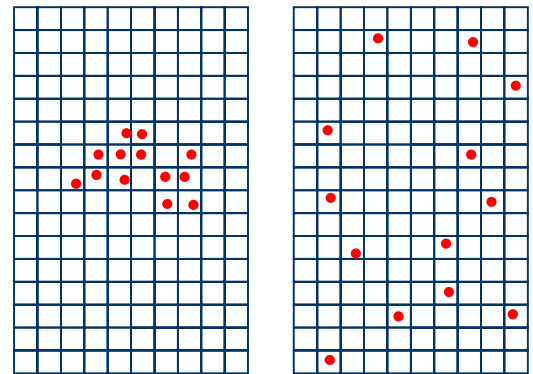
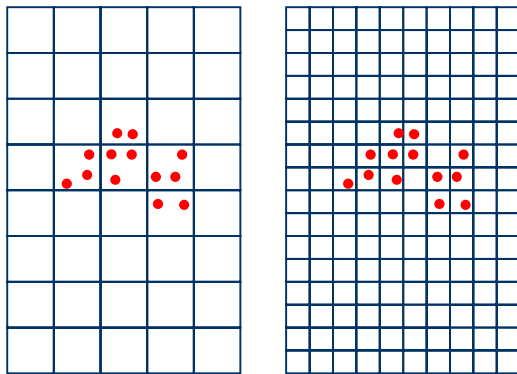
対立仮説: 点は一様に分布していない (偏りがある)

区画法の利点

様々な「理論的」分布に対する検定ができること。これは、もともと χ^2 検定が様々な理論分布への適合性を調べるための方法であることによる。一様分布だけでなく、何らかの偏りのある分布を考え、データがそれに適合しているかどうかを調べることも可能である。

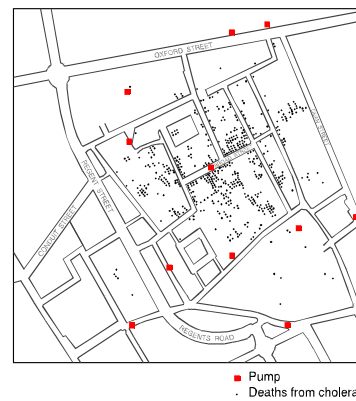
区画法の問題点

- a) 領域の設定, セルの大きさによって結果が異なる
- b) 点の数がセルの数と比べて十分多い必要がある

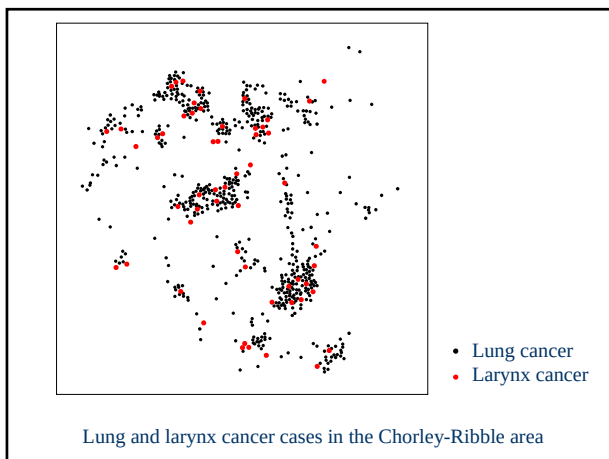


VI-3 異なる2つの点分布の関係を分析する方法

異なる2種類の点オブジェクトが、同一地域内に分布している場合、それらの関係が問題になることがある。

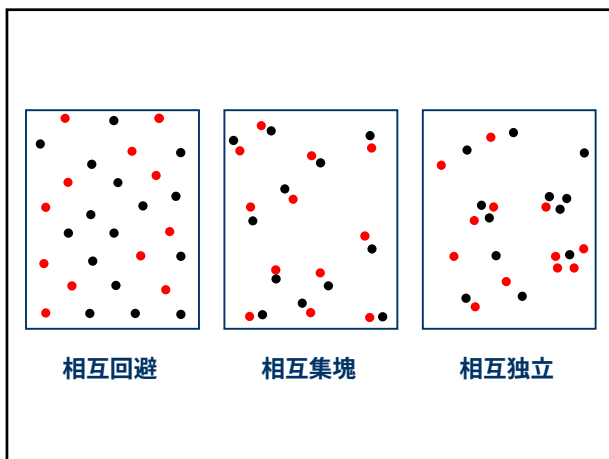


Deaths from cholera in London, 1854



VI. 空間解析

二つの分布間の関係といっても様々なものがあり得るが、ここでは、二つの分布が互いに接近しているのか、互いに避けあっているのか、そのどちらでもないのか、ということに焦点を当てる。



VI. 空間解析

二つの分布の類似性は、

- 1) 相互の因果関係 (一方が他方を引き起こす)
例: コレラ患者と井戸の分布
- 2) 同一の原因を持つ (どちらも同じ原因による)
例: 肺ガンと咽頭ガンの患者の分布

を暗に物語る。

VI. 空間解析

他にも、
産業廃棄物焼却施設と咽頭ガン患者の分布
犯罪発生地点と交番の分布
プランクトンとその捕食者の分布
鉄道駅と商業施設の分布
...

VI. 空間解析

VI-3.1 区画法

いま、分析領域内にA, Bという2種類の点分布が存在する。ここで、分析領域を同数のK個の区画に分割し、区画を以下のように分類して数える。

	A種を含む	A種を含まない	合計
B種を含む	k_{AB}	k_{0B}	K_B
B種を含まない	k_{A0}	k_{00}	$K-K_B$
合計	K_A	$K-K_A$	K

ここで、もし二つの分布間に関係がないとすると、表の左列と中列、上段と中段はそれぞれ相似形になるはずである。このことを検定するには、以下の統計量を計算すればよい。

$$\chi^2 = \frac{(k_{AB} - K_A K_B / K)^2}{K_A K_B / K} + \frac{\{k_{0B} - (K - K_A) K_B / K\}^2}{(K - K_A) K_B / K} + \frac{\{k_{A0} - K_A (K - K_B) / K\}^2}{K_A (K - K_B) / K} + \frac{\{k_{00} - (K - K_A) (K - K_B) / K\}^2}{(K - K_A) (K - K_B) / K}$$

この統計量は、二つの分布が独立である場合、自由度1の χ^2 分布に従う。

問題点

- a) 区画の大きさによって結果が異なる
- b) 空間的な位置に関する情報を生かしていない

VI-3.2 相互最近隣距離法

単一点分布における最近隣距離法に対応

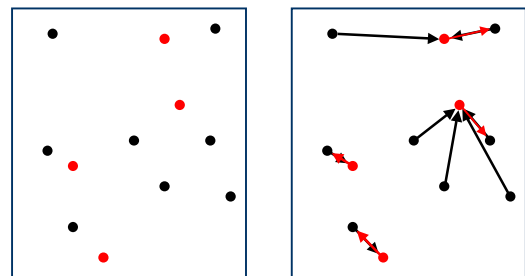
いま、分析領域内にA、Bという2種類の点分布が存在し、それぞれの点の個数を n_a 、 n_b 、密度を λ_a 、 λ_b とする。

それぞれの種類の全ての点について、他種の最近隣点までの距離の平均値（平均相互最近隣距離）を計算する。

$$D = \frac{1}{n_a + n_b} \left(\sum_{i=1}^{n_a} d_{ai} + \sum_{i=1}^{n_b} d_{bi} \right)$$

d_{ai} : Aの点*i*から、Bの最寄りの点までの距離

d_{bi} : Bの点*i*から、Aの最寄りの点までの距離



二種類の点が、いずれも無限平面上でランダムに分布する場合、平均相互最近隣距離 D の期待値は簡単な式で表される。

$$E[D] = \frac{1}{n_a + n_b} \left(\frac{n_a}{2\sqrt{\lambda_b}} + \frac{n_b}{2\sqrt{\lambda_a}} \right)$$

従って、 D と $E[D]$ を比較することにより、二種類の点が互いに接近しているのか、互いに避け合っているかを判断することができる。

検定の方法は大きく2つに分かれる。点の数が十分多い場合には、それぞれの点分布から点を m_a , m_b 個、それぞれ n_a , n_b に比例するようにランダムに抽出し、それらについて平均相互最近隣距離 D を計算する。

すると、点分布がランダムな場合の D の確率分布は、

$$N \left(\frac{1}{m_a + m_b} \left(\frac{m_a}{2\sqrt{\lambda_b}} + \frac{m_b}{2\sqrt{\lambda_a}} \right), \frac{m_a \lambda_a (4 - \pi m_a) + m_b \lambda_b (4 - \pi m_b) + 2\pi m_a m_b \sqrt{\lambda_a \lambda_b}}{4\pi \lambda_a \lambda_b (m_a - m_b)^2} \right)$$

となり、正規分布に基づく検定ができる。

点の数が少ない場合には、通常の最近隣距離法と異なり、平均相互最近隣距離 D の確率分布が明らかではない。そのため、検定はモンテカルロシミュレーションによらざるを得ない。

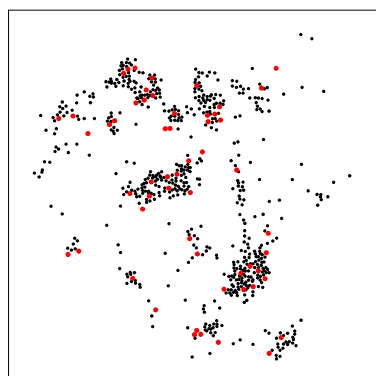
VI-3.3 相互 K -関数法 (2つの分布が対等な場合)

2つの点分布を分析する場合、点分布間の立場は大きく2通りに分かれる。

- 2つの分布が対等な場合
- 2つの分布が対等ではない場合

- 2つの分布が対等な場合
 - 相互に影響を与えあう
 - 同一の原因を持つ (どちらも同じ原因による)

例：八百屋と肉屋の分布
肺ガンと咽頭ガンの患者の分布
白人と黒人による犯罪発生地点の分布



• Lung cancer
• Larynx cancer

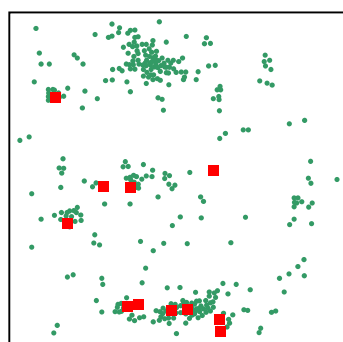
Lung and larynx cancer cases in the Chorley-Ribble area

- 2つの分布が対等ではない場合
一方が他方に影響を与える (因果関係が想定される)

例: コレラ患者と井戸の分布

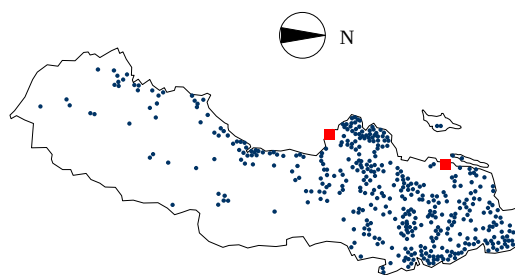
産業廃棄物処理場と白血病患者の分布

原子力発電所と小児急性白血病患者の分布



■ Hazardous waste sites
• Leukaemia cases

Leukaemia cases in upstate New York, 1978-82



■ Nuclear power plants
• Leukaemia cases

Childhood acute lymphocytic leukaemia cases,
ages 0-15 years for 1980-1990 in Sweden

最近隣距離法の拡張が相互最近隣距離法であるが、同様に、 K -関数法の拡張が相互 K -関数法である。

いま、分析領域内にA, Bという2種類の点分布が存在し、それぞれの点の個数を n_a , n_b とする。

このとき、Bから見たAの相互 K -関数は次のように定義される。

$$K_{AB}'(h) = \frac{\text{Bの点から距離}h\text{以内に存在するAの点の個数}}{n_b \lambda_a}$$

Aから見たBの相互 K -関数 $K_{BA}'(h)$ も同様に定義される。

そして、AB間の関係はこれらの加重平均

$$K_{AB}(h) = \frac{n_b K_{AB}'(h) + n_a K_{BA}'(h)}{n_a + n_b}$$

で表される。但し、通常はこの値を基準化した相互L-関数

$$L_{AB}(h) = \sqrt{\frac{K_{AB}(h)}{\pi}} - h$$

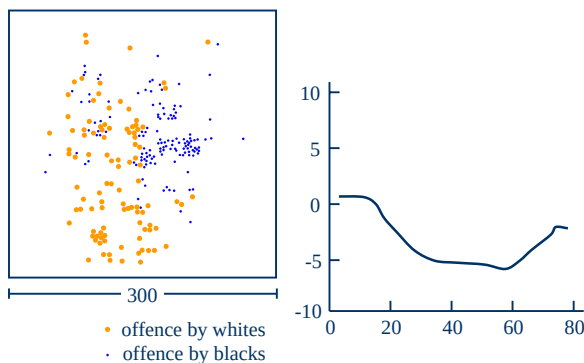
が用いられることが多い。

なお、相互K-関数、相互L-関数の統計的検定は、いずれもモンテカルロ・シミュレーションによる。

VI-3.4 相互K-関数法 (2つの分布が対等ではない場合)

二つの分布が対等ではない場合、相互K-関数法は適当ではない。

なぜならば、この方法における統計的検定は、二つの点分布がそれぞれ自由に (ランダムに) 分布できると仮定しているからである。この仮定では、「二つの分布が無関係である」か、「二つの分布は互いに影響し合っている」ということしかわからない。



Oklahoma City offences by whites and blacks and the cross L -function

そこでこの場合、相互K-関数の定義において用いられる関数の一方 $K_{AB}'(h)$ あるいは $K_{BA}'(h)$ を利用する。 $K_{AB}'(h)$ を用いれば、Bから見たAの様子がわかる。従って、例えば産業廃棄物焼却施設と白血病患者の分布を比較する場合、前者をB、後者をAとして $K_{AB}'(h)$ を計算すればよい。なおこの場合にも、統計的検定はモンテカルロシミュレーションによる。