

File: /General/MLAB-Text/Papers/voronoi/voronoi.tex

Spatial Statistics(Voronoi Diagrams): An MLAB Example

Gary D. Knott, Ph.D.
Civilized Software, Inc.
12109 Heritage Park Circle
Silver Spring, MD 20906
Tel. (301) 962-3711
Email: csi@civilized.com
URL: <http://www.civilized.com>

Our problem is to decide if two bivariate random vectors have the same bivariate distribution functions, except for possibly different isotropic scalings, given two sets of samples: $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_n\} \subset \mathcal{R}^2$.

One approach to this problem is based on the so-called Voronoi diagram (a Voronoi diagram is also known as a Dirichlet tessellation) of each point set. The Voronoi diagram of a point set $\{x_1, \dots, x_n\}$ in the plane is the collection of closed regions V_1, \dots, V_n which are the “natural” neighborhoods of the data points x_1, \dots, x_n in the plane; the Voronoi regions V_1, \dots, V_n satisfy the following conditions.

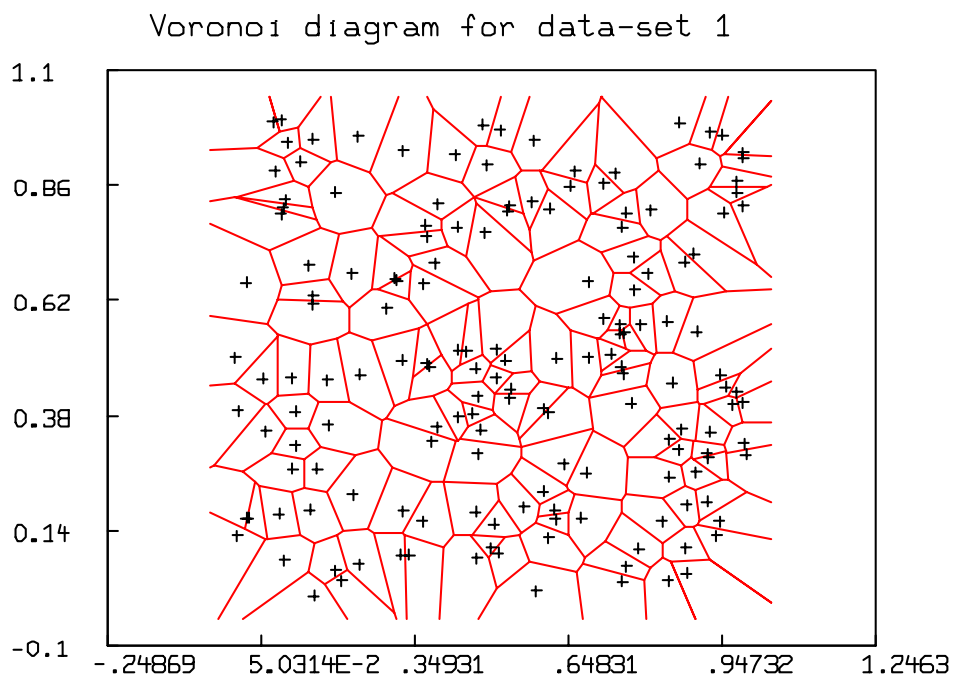
1. $x_i \in V_i$ for $i = 1, \dots, n$
2. $V_i \cap V_j = \partial V_i \cap \partial V_j$ for $i \neq j$. When $V_i \cap V_j$ is not empty, it is a segment on the bisector line of the line-segment between x_i and x_j .
3. $V_1 \cup V_2 \cup \dots \cup V_n = \mathcal{R}^2$.
4. $V_i = \{x \mid |x - x_i|_2 \leq |x - x_j|_2 \text{ for } j = 1, \dots, n\}$.

If the two spatial samples come from the same bivariate population (with possible isotropic scale differences), then the suitably-scaled areas of the bounded Voronoi regions will also have the same univariate distribution. Thus we may compute the two empirical cumulative distribution functions of the two sets of values corresponding to the scaled areas of the bounded Voronoi regions of the two spatial samples.

We may then use the Kolomogorov-Smirnov test to decide if these two cumulative distribution functions are plausibly based on the same underlying population.

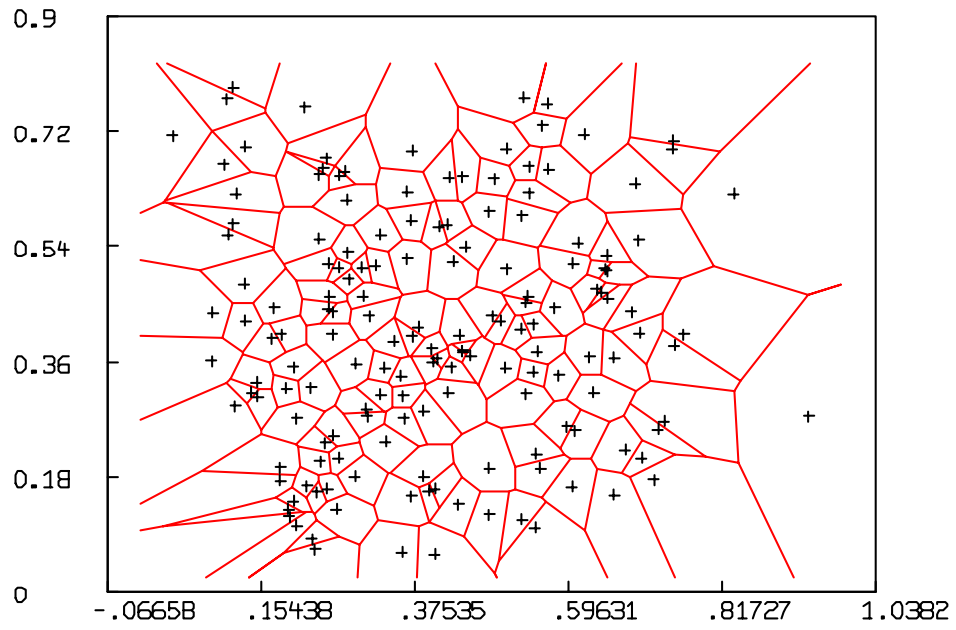
This process is carried-out in MLAB as follows.

```
* d1 = read(d1,200,2)
* draw d1 pt crosspt ptsize .01 lt none
* draw vorcurve(d1) lt marker color red
* window adjust wmatch
* top title "Voronoi diagram for data-set 1"
* view
```



```
* d2 = read(d2,200,2)
* draw d2 pt crosspt ptsize .01 lt none
* draw vorcurve(d2) lt marker color red
* window adjust wmatch
* top title "Voronoi diagram for data-set 2"
* view
```

Voronoi diagram for data-set 2

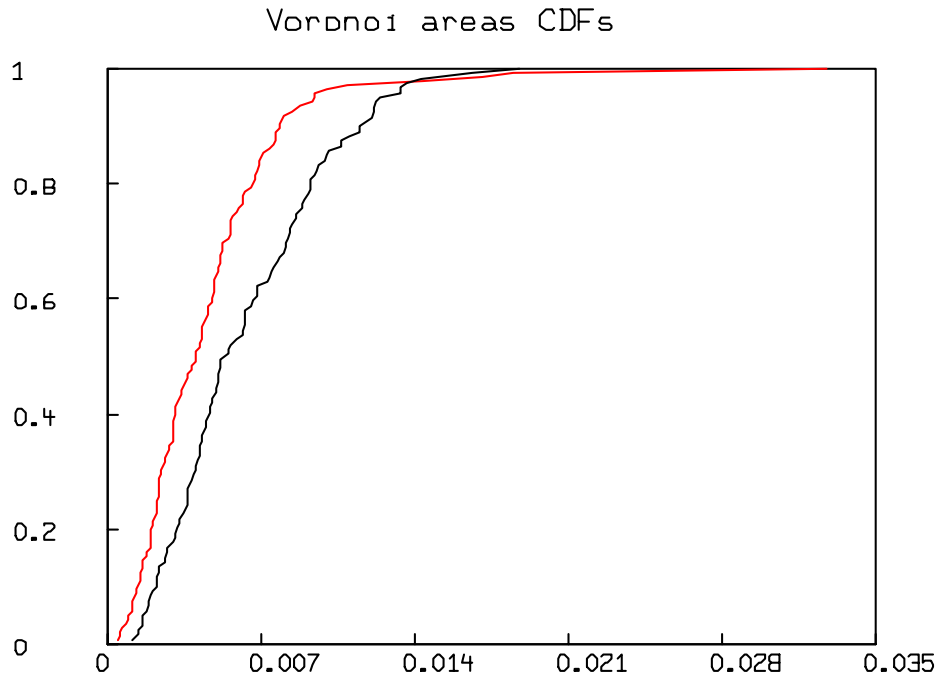


```
* v1 = vorstat(d1) col 2
* v1 = compress(v1)
* xd = maxv(d1 col 1) - minv(d1 col 1)
* yd = maxv(d1 col 2) - minv(d1 col 2)
* az1 = xd*yd
* v1 = v1/az1
* z1 = cdf(v1)
*
* v2 = vorstat(d2) col 2
* v2 = compress(v2)
* xd = maxv(d2 col 1) - minv(d2 col 1)
* yd = maxv(d2 col 2) - minv(d2 col 2)
* az2 = xd*yd
* v2 = v2/az2
* z2 = cdf(v2)
*
* draw z1
```

```

* draw z2 color red
* top title "Voronoi areas CDFs"
* view

```



```

* ksf2t(z1,z2)
[K-S-test: are M1 and M2 empirical cdf estimates of the same distribution?]
null hypothesis H0: M1 and M2 are cdf estimates of the same distribution.
Then the scaled maximum deviation between the cdf(M1) and the cdf(M2) is
distributed according to the Kolmogorov-Smirnov K statistic.
The K-statistic value = 1.991649

```

The probability $P(K > 1.991649) = 0.000717$
This means that a value of K larger than 1.991649 arises
about 0.071718 percent of the time, given H0.

```

: a 5 by 1 matrix

```

1: 1.99164923
2: 0.25
3: 0.25
4: 7.17178243E-4
5: 6.19572536E-3