# Spatial Point Processes

Eric Feigelson
Penn State University

Harvard-Smithsonian Center for Astrophysics, Jan 2014

# Background on Spatial Point Processes

Study of spatial point processes is a part of the field of spatial statistics that includes: graph, map, network data; lattice data (e.g. images); spatio-temporal processes; mathematical morphology; and more.

The statistical characterization of point process in 2-3 dimensions has been dominated by applications in geostatistics & ecology (e.g. forestry). But the methods are often related to those of multivariate statistics and density estimation.

In astronomy spatial point processes most naturally arise in the 2D and 3D distribution of stars and galaxy in space. But the methods also can be applied to any low-dimensional parameter space. Most work has concentrated on characterizing the unusual hierarchical, anisotropic clustering of galaxies in space: virialized groups & clusters, unequilibrated filaments and Great Walls, voids.

# Concepts of spatial point processes

**Stationarity:** Process properties are invariant under spatial translation

**Poisson process:** When the expected number of points is Poisson-distributed

**Complete Spatial Randomness:** a stationary Poisson process. Alternative is a **inhomogeneous process.**

**Mark variables:** covariates measured at each point

For stationary processes, several global statistics are used to characterize the **spatial autocorrelation** and **clustering.**

For inhomogeneous processes, spatial **interpolation** can map the intensity or mark variable distributions.

**Spatial modeling** characterizes deterministic and/or stochastic relationships of the intensity, mark variable and/or temporal behavior. Examples: (non)linear trends, Gaussian and Markov random fields.

# Spatial autocorrelation

**Moran's I** and **Geary's G** (1940-50s), with their **correlograms** and **variograms**, extend the Pearson bivariate linear correlation coefficient to the spatial domain, evaluating the autocorrelation between bins on different spatial scales.  These functions can be evaluated globally (stationary) or locally (inhomogeneous).

# Spatial interpolation

**Kernel density estimators, k-nearest neighbor estimators, Inverse Distance Weighting, tessellations,** and **kriging** (for mark variables) are commonly used interpolation methods.  Kriging, a least squares local regression technique, has been extensively developed in geostatistics.

# Spatial clustering

***Cumulative second-order moment measures for stationary processes***

G(d): distribution of nearest-neighbor distances from data points

F(d): distribution of nearest-neighbor distances from random locations

Ripley's K(d): average number of points within *d* of data points

*Ripley's K is the integral of the astronomers' two-point correlation function. K does not require choice of bin size.*

L*(d): a scaled K(d) to stabilize the variances

J(d): a ratio involving G and F

*Baddeley's J has reduced dependence on edge effects*

These measures have known means & variances under assumptions of CSR. Hence, under the Central Limit Theorem, probabilities of hypotheses involving CSR can be obtained for a large data set.