Intensity

- Everything so far has looked at 2nd order properties of a point pattern: pairs of points
- Equivalent to variances and covariances for quantitative data
- What about 1st order properties, equivalent to mean
- That is the intensity of the point process, λ

Count / area

$$\lambda(s) = \lim_{dA \to 0} \frac{\#\text{events in area dA centered at } s}{dA}$$

- Homogeneous Poisson process (CSR):
  - P[event at s] independent of presence/absence of other events
  - $\lambda(s)$ constant
- Inhomogeneous Poisson process:
  - $\lambda(s)$ not constant

Indep

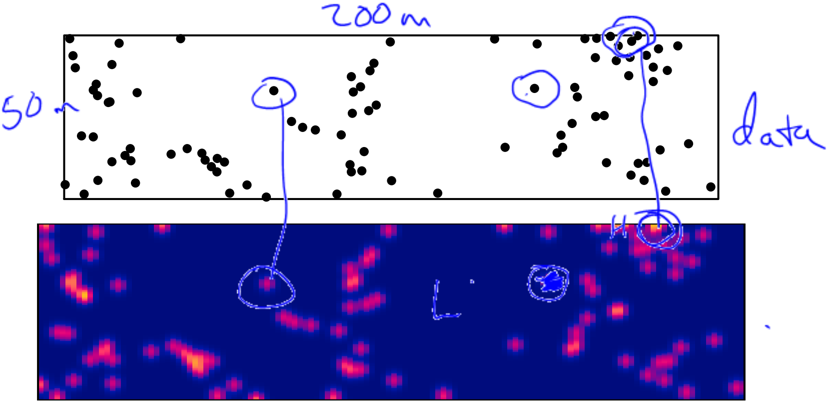
Estimating intensity

big boxes ⇒ precise estimate
biased

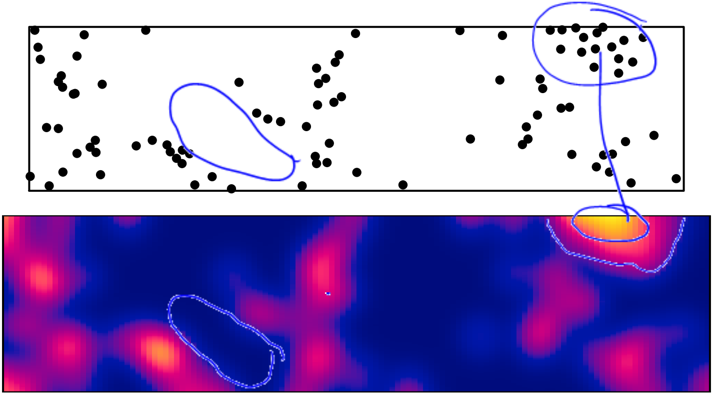
low boxes ⇒ less/no bias
high variability

- Goal: estimate $\lambda(s)$ at a set of s locations (e.g. a grid)?
- use kernel smoothing, as we did to estimate $\hat{g}(x)$
- bandwidth of the kernel controls smoothness of the map
  - large bandwidth ⇒ smoother map
  - small bandwidth ⇒ rougher (bumpier) map
- illustrate with $\sigma = 1.5$, $\sigma = 4.5$, and $\sigma = 15$ plots
- Also have to deal with edge effects

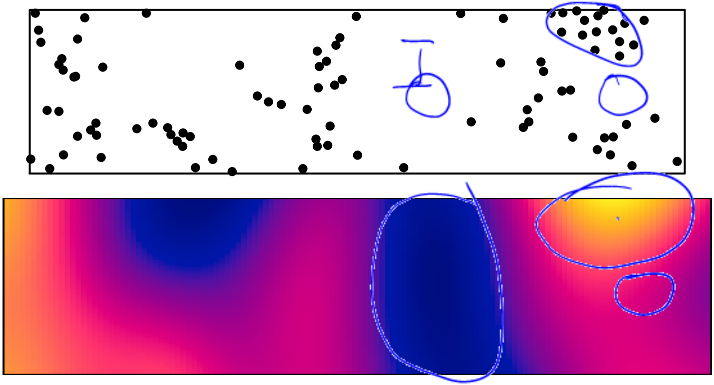
BW: 1.5m



BW: 4.5m



BW: 15m



How to choose σ?

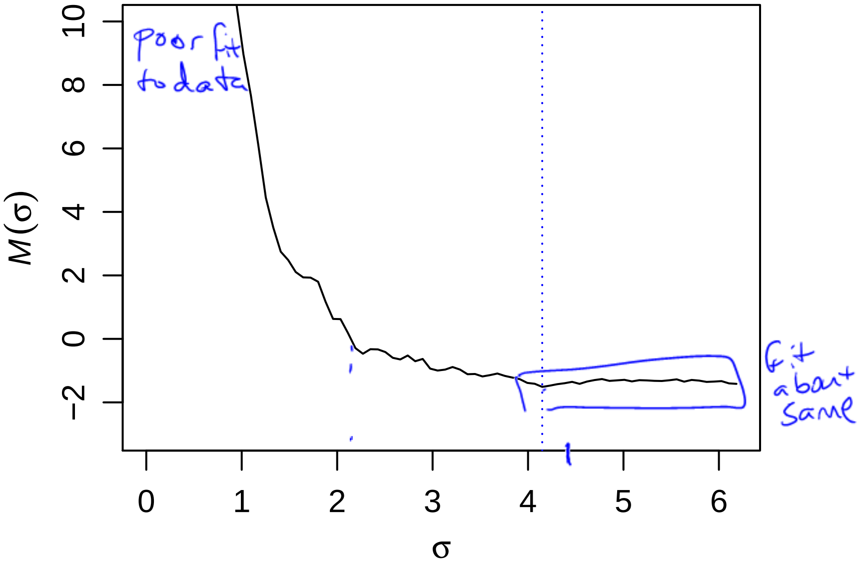
sample size

- What looks good?
- Simple data-based rules: Scott's rule, 25% percentile of interpoint distances
- Cross-validation, concept:
  - omit a point, estimate $\lambda(s)$ there, want $\lambda(s)$ to be large
  - location without a point, want $\lambda(s)$ to be small
- Two versions of cross-validation, both *a-priori* reasonable
  - Minimize mean-square error (Diggle-Berman criterion)
  - Maximize data log-likelihood

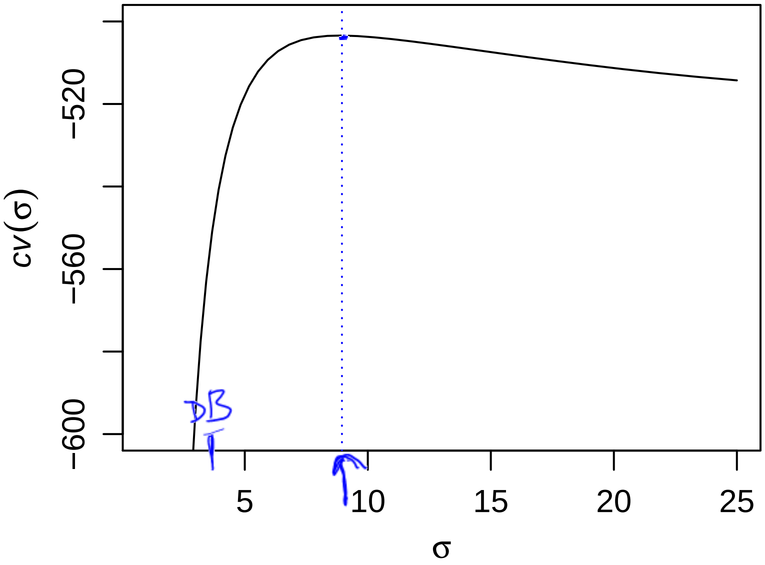
Cornie-VanLieshout: Same as lnL . much faster

- choose σ that does this the best
- My experience: Diggle-Berman undersmooths

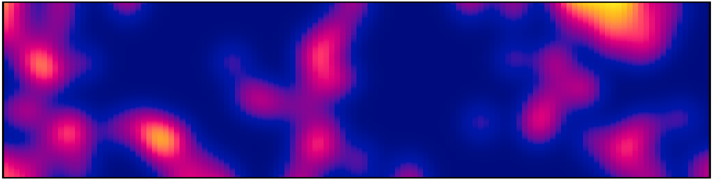
Diggle-Berman



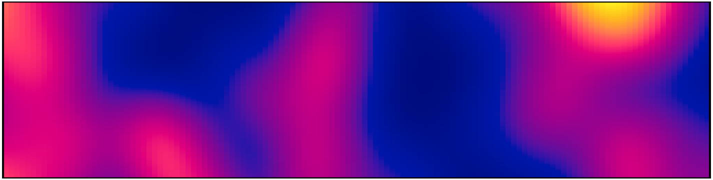
Likelihood



Cypress: Diggle, likelihood



DB
4



PPL
9

Modeling $\lambda(s)$ as a function of covariates

- imagine have $X(s)$ at every possible location s
- examples of potential $X(s)$:

fine

  - geographic coordinates (x,y)
  - distance to field edge or hazardous waste site
  - elevation from DEM
  - areal data
- $\lambda(s) \geq 0$, so a plausible model is $\lambda(s) = \exp(\boldsymbol{X}\beta)$
  - e.g. $\lambda(s) = \exp(\beta_0 + \beta_1 elev(s))$
  - or $\log \lambda(s) = \beta_0 + \beta_1 elev(s)$

approximate

- kriged surface based on geostat data
  - but this is estimated and is not the "true" $X(s)$ ignored
  - creates complicated issues (error in covariates problem)

Modeling $\lambda(s)$ as a function of covariates

- Data are locations of events
  - anticipate $\lambda(s)$ larger at those locations than elsewhere
- To get started, imagine 10 1x1 quadrats:
  - observe an event in 2 of them and not in 8 of them X each box.
  - Use maximum likelihood to estimate $\lambda(s_i)$ for each quadrat
  - Model $Y_i \sim$ Poiss $(\lambda(s_i))$ log λ = Xβ

Count

$$f\left(Y_i \mid \lambda(s_i)\right) = \frac{e^{-\lambda(s_i)}\lambda(s_i)^{Y_i}}{Y_i!}$$

log likelihood

$$\log L\left(\lambda(s_i) \mid Y_i\right) = -\lambda(s_i) + Y_i \log\left(\lambda(s_i)\right) - \log Y_i!$$

  - event quadrats $(Y_i = 1)$: $\log \mathrm{L} = -\lambda(s_i) + \log\left(\lambda(s_i)\right) - 0$
  - non-event quadrats $(Y_i = 0)$: $\log \mathrm{L} = -\lambda(s_i) + 0 - 0$
  - So, $\log \mathrm{L} = \sum_{events} \log\left(\lambda(s_i)\right) - \sum_{all} \lambda(s_i)$ quadrats
- Include covariates by modeling $\lambda(s_i)$
  - $\lambda(s_i) \geq 0$, so model $\log \lambda(s_i) = \boldsymbol{X}_i \boldsymbol{\beta}$

Modeling $\lambda(s)$ as a function of covariates

- 10 quadrats: $\log \mathrm{L} = \sum_{events} \log(\lambda(s_i)) - \sum_{all} \lambda(s_i)$
- Now: make quadrats smaller and smaller.
  - Still 2 event locations, Many "all" locations
  - Event locations still a sum (event is a point)
  - All locations become an integral $\sum_{all} \lambda(s_i) \Rightarrow \int_A \lambda(u)\mathrm{d}u$
- log likelihood for an inhomogeneous Poisson process

events area

$$\log L = \sum_{i=1}^{n} \log \lambda(s_i) - \int_A \lambda(u)\mathrm{d}u$$

- When $\lambda(s)$ depends on elevation,
  $\lambda(s_i) = \exp(\beta_0 + \beta_1\, elev(s_i))$, n events

$$\log L = \sum_{i=1}^{n} [\beta_0 + \beta_1 elev(s_i)] - \int_A \exp\left(\beta_0 + \beta_1 elev(u)\right) \mathrm{d}u$$

Modeling $\lambda(s)$ as a function of covariates

- Estimate log intensity function by finding the parameter values that maximize the log likelihood
- When $\lambda(s)$ is constant (CSR, homogeneous Poisson process):

events   × at event   all points

$$
\begin{aligned}
\log L &= \sum_{i=1}^{n} \log \lambda - \int_A \lambda \mathrm{d}u \\
&= n \log \lambda - \lambda \|A\| \\
\frac{\mathrm{d}\ \log L}{\mathrm{d}\ \lambda} &= \frac{n}{\lambda} - \|A\| = 0 \\
\hat{\lambda} &= \frac{n}{\|A\|}
\end{aligned}
$$

$\frac{\#\text{ events}}{\text{area}}$

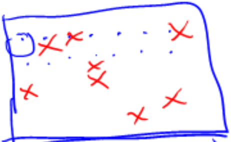
- "obvious" estimator of λ for HPP, $n/\|A\|$, is an ML estimator
- Maximizing $\log \lambda(s) = \beta_0 + \beta_1 elev(s)$ requires numeric maximization, no analytical solution

Modeling $\lambda(s)$ as a function of covariates

- Is this useful?
  - Yes - likelihood is the most common estimator / test method, when you move away from normal distributions
  - Many of the "usual" methods are ML or refinements of ML
  - Discovering that "obvious" estimator of λ for HPP, $n/\|A\|$, is an ML estimator tells you a lot:
  - All the general properties of ML estimators apply:
    - Estimates are consistent (get closer to true values as sample size increases)
    - Asymptotic normal (have normal sampling distributions for suitably large sample sizes)
    - Variance from Fisher or observed information (so can easily compute Var $\hat{\beta}_1$
    - log L is the foundation for model selection statistics: AIC, AICc, BIC

Modeling $\lambda(s)$ as a function of covariates

- Is this useful?
- Notice the practical problem: Need a lot of covariate information,
- both:
  - Covariate values (e.g., elevation) at the event locations
    - Very commonly have $X(s)$ at event locations
  - *AND* covariate values everywhere else in the study area
- No problem when λ is a function of coordinates (e.g., trend surface)
- Otherwise looks intractable: $X(s)$ at every s?
- Actually only need to estimate / approximate $\int_A \exp\left(\beta_0 + \beta_1 X(u)\right) \mathrm{d}u$
  - Often approximate by values at a grid: $\sum_{grid} \|gridcell\| exp\left(\beta_0 + \beta_1 X(u)\right)$ $\|a\ cell\|$
  - Or by values at a simple random sample of locations: $\|A\| \sum_{sample} \left[exp\left(\beta_0 + \beta_1 X(u)\right)\right] / n_{sample}$



$\exp(\beta_0 + \beta_1 X)$

An aside: MAXENT modeling of species distributions

- MAXENT is a very popular algorithm / software program for modeling species distributions
  - Given GIS images with elevation, precipitation, ...
  - and location records, where a species has been found
  - predict P[species occurs at a new location | covariates]
- Often called niche or species distribution modeling
- Phillips et al., 2006, Ecol. Model. 190:231-259
- Developed from maximum entropy principles (machine learning technique)
- Very popular because does not require explicit samples of absences
- Data collection for usual logistic regression:
  - Random sample of locations
  - Visit an d observe whether species present or absent
  - Simple statistical model, practically impossible

An aside: MAXENT modeling of species distributions

- Wharton and Shepherd (2010, Ann. Appl. Statistics 4:1383-1402) showed that the quantity maximized by MAXENT is the IPP log likelihood
  - Provided immediate answers to difficult questions about MAXENT, such as role of "pseudo-absences"
- However, appropriate use of MAXENT demands specific sort of data
  - random sample of presences
  - good estimate of background prevalence

$$\sum_{\text{event}} \log \lambda - \int \lambda \, d\text{area}$$

- Critical review of assumptions:
  - Royle, J.A. et al. 2012, Methods in Ecology and Evolution, 3(3):545-554
- And data are often not "the right sort"
  - Review of many applications of MAXENT
    Yackulik, C. et al. 2013, MEE 4(3):246-243.

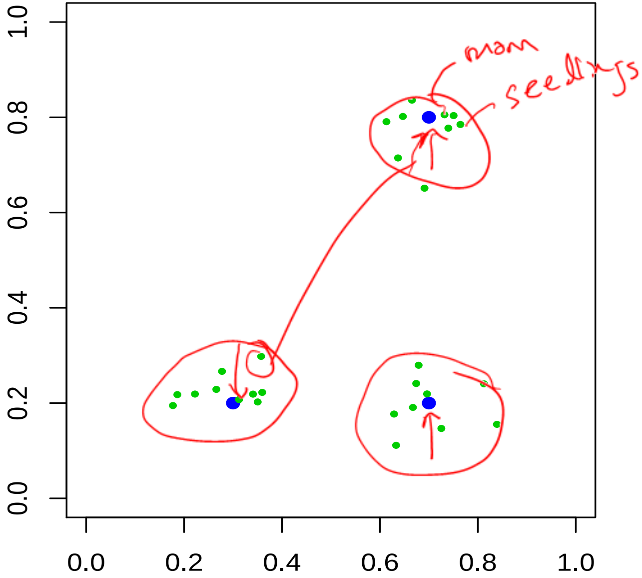
Modeling spatial patterns

Modern

- Historically (through 2000 or later)
  - Classify pattern as clustered, random, segregated
- Current best practice, more insightful:
  - model the spatial pattern,
  - learn more about the characteristics of the clusters or the regularity
  - not just clustering: yes/no?, regular: yes/no?
- Many different models for spatial point patterns
  - I will only talk about two to illustrate what can be done.
  - Chapter 6 of Diggle's spatial point pattern book describes many more.

Estimating seed dispersal distance

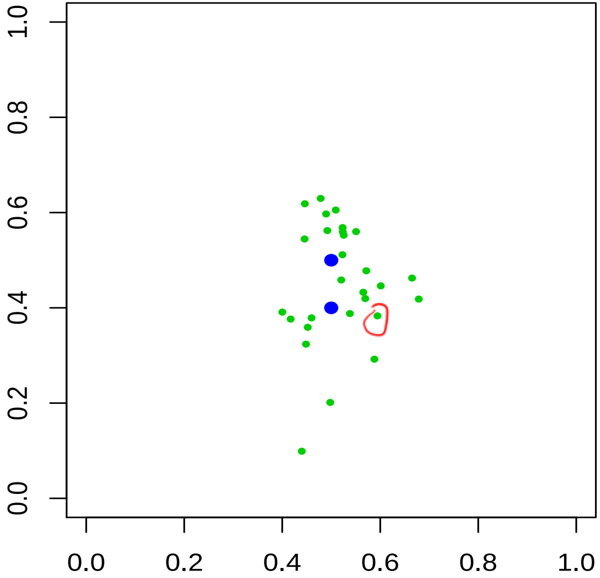
- How far are seeds moved away from mom?
  - Plant produces seeds
  - In most plants, those seeds are dispersed away from mom.
  - Higher survival/growth if not really close to mom.
  - How far do they move?
- very very difficult to measure directly
- If mom's are widely spaced, and you know the location of mom, can look at locations of seedlings to estimate directly (picture on next slide)

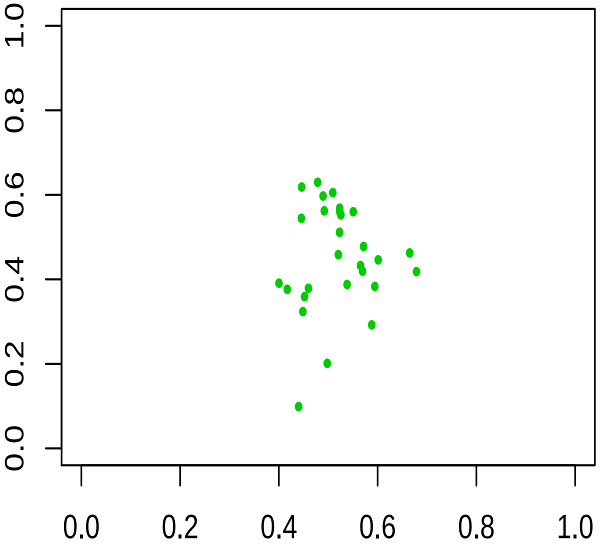


0.0
0.2
0.4
0.6
0.8
1.0
mom
seedlings

The problem gets harder

- In previous plot, seedling distribution “looks” like short-distance dispersal.
  reasonable to assume points around a mom all came from that mom
- What about next plot?
- Which seedlings belong to each mom?
  - not clear
  - genetic markers sometimes help, but expensive
- And what if the plant is an annual, so when you can see the seedlings, you don’t know where mom was? (2nd plot)


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

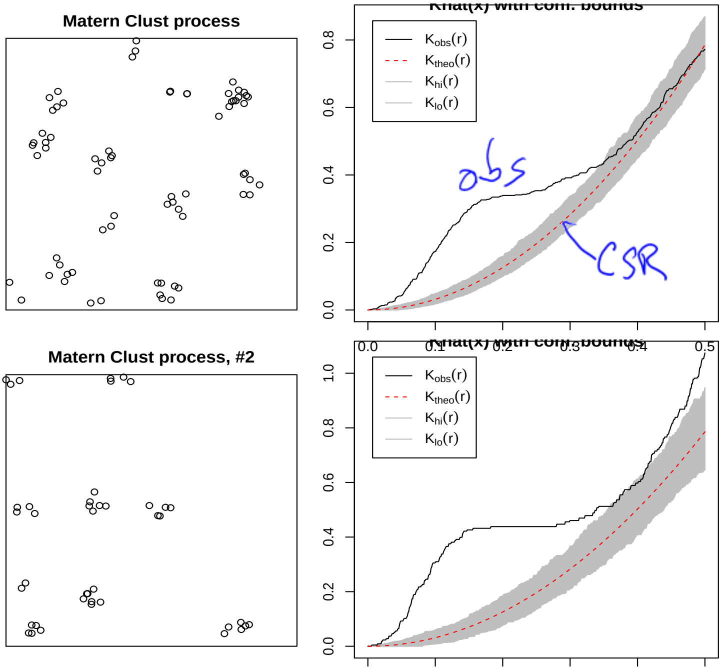

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Neyman-Scott process for clustered events

Jerzy : Neyman - Pearson
Elizabeth

- a very general model
  - mothers are CSR with an intensity k
  - daughters have a specified distribution of distance from mom
    often bivariate normal $(0, \sigma^2)$ = Thomas process
    another common choice:
    uniform w/i disk of radius r = Matern cluster process
  - with a Poisson # of daughters per mom, with mean μ
  - only observe daughter locations, not mom
  - parameters are k, σ^2, and μ, or k, r, and μ
- Pictures and K functions on next slide

Thomas   Matern


Matern Clust process
Khat(x) with conf. bounds
$K_{obs}(r)$
$K_{theo}(r)$
$K_{hi}(r)$
$K_{lo}(r)$
obs
CSR
0.0 0.1 0.2 0.3 0.4 0.5
0.0 0.2 0.4 0.6 0.8
Matern Clust process, #2
Khat(x) with conf. bounds
$K_{obs}(r)$
$K_{theo}(r)$
$K_{hi}(r)$
$K_{lo}(r)$
0.0 0.2 0.4 0.6 0.8 1.0

Estimation

$NS = \pi x^2 + \text{stuff}$
> 0

- key parameter in seed dispersal question is σ^2 (or r)
- For many N-S-type processes, can calculate (or look up) theoretical $K(x|k, \sigma^2,$ and $\mu)$

  CSR: $K = \pi x^2$
- So, estimate k, σ^2, and μ by finding the theoretical $K(x)$ that is closest to the $\hat{K}(x)$ computed from the events
- How to determine "closest"?
  - Commonly use least-squares estimation: "minimum contrast" estimation
  - i.e. minimize $\sum_{\text{all } x} \left[\hat{K}(x) - K(x \mid k, \sigma^2, \mu) \right]^2$
  - problem here is that Var $\hat{K}(x)$ is not constant
  - so LS would "pay more attention" to distances x with large variances because LS assumes all distances have the same variance

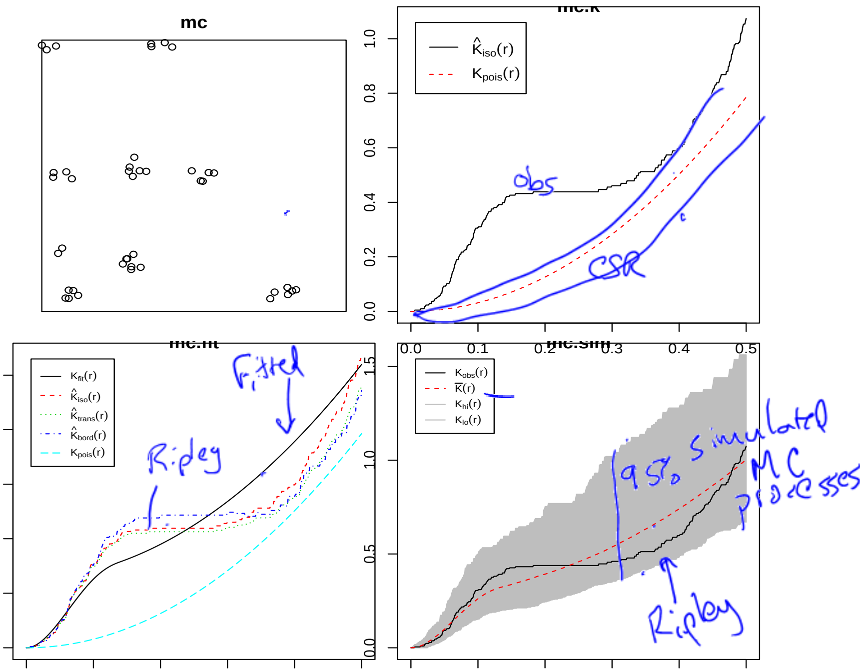
Estimation

- Dealing with unequal Var $\hat{K}(x)$: Diggle and Gratton suggest

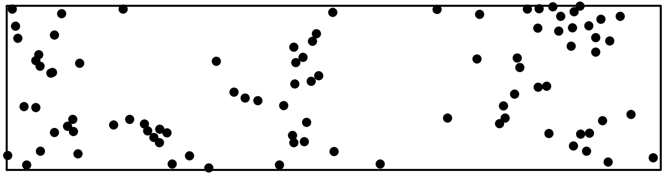
$$\left(\Sigma \left[\mid \hat{K}(x)^{1/4} - K(x \mid k, \sigma^2, \mu)^{1/4} \mid\right]^2\right)$$

- This is like the Cressie-Hawkins variogram estimator,
  - Using 1/4 power to control the variance
  - Don't need the C-H denominator because comparing two functions.
- Calculating theoretical $K(x)$ usually requires integration
- Q: What if you can't do that integration analytically?
- A: calculate a Monte-Carlo approximation to that integral
  - simulate process $\mid k, \sigma^2, \mu$
  - calculate $\hat{K}(x)$
  - repeat above 2 steps many times (1000?) and average to estimate $K(x \mid k, \sigma^2, \mu)$

$K(x) = \int$ distances


mc
mc.K
$\hat{K}_{iso}(r)$
$K_{pois}(r)$
obs
CSR
mc.fit
$K_{fit}(r)$
$\hat{K}_{iso}(r)$
$\hat{K}_{trans}(r)$
$\hat{K}_{bord}(r)$
$K_{pois}(r)$
Ripley
Fitted
mc.sim
$K_{obs}(r)$
$\overline{K}(r)$
$K_{hi}(r)$
$K_{lo}(r)$
95% simulated MC processes
Ripley

Cypress trees in Savannah River Swamp



Modeling clustering of cypress trees

- Matern process: N-S process in which daughters are randomly distributed within a disk with radius R
- Theoretical $K(x)$ for a Matern process

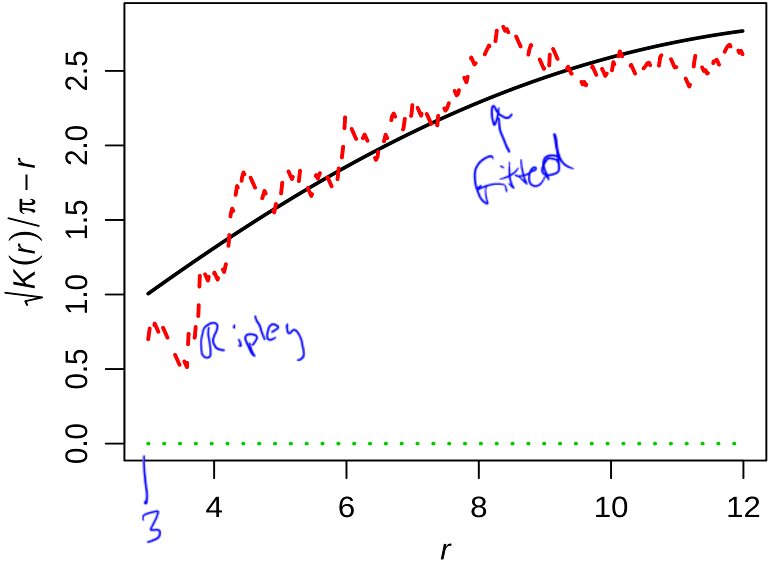
$$K(x) = \pi x^2 + \frac{h(x/2R)}{k},$$

stuff

where $h()$ is a known function, details not important

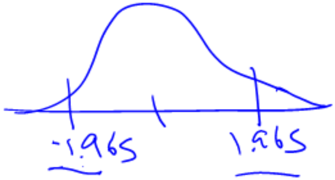
- fitting this model to the cypress locations gives:
  - $\hat{k} = 0.0024$ m o m's = # events/10,000 m^2
  - $\hat{R} = 12.01$ 12 m
  - $\hat{\mu} = 3.70$ 3.7 trees.
- Interpretation:
  - a total of 24 = 0.0024*area = 0.0024*50*200 clusters
  - each with radius 12m and containing 3.7 trees

Modeling clustering of cypress trees



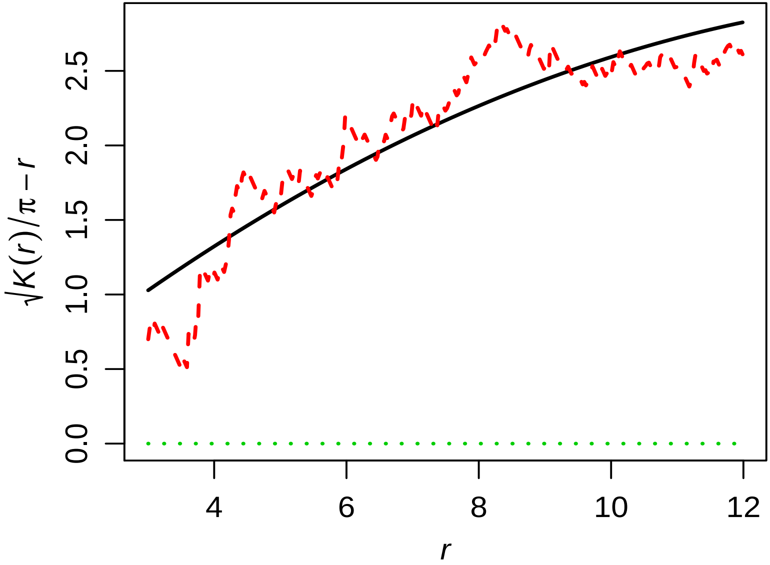
Modeling clustering of cypress trees

- Many different N-S processes, differing in the distribution of daughters around mom
- Thomas process: daughters (what you see) $\sim N(0, \sigma^2)$ around unseen mom
- Estimates are similar:
  - $\hat{k} = 0.0027$
  - $\hat{\sigma}^2 = 36.09$
  - $\hat{\mu} = 3.40$



- Interpretation:
  - a total of 27 = 0.0027*area = 0.0027*50*200 clusters
  - clusters containing 3.4 trees on average
  - (36.09 have sd of 6.1m, so 95% of trees within $2.45s$ = 15.0m.
  - where does 2.45 come from?
    - cluster is isotropic, so distance$^2/\sigma^2 \sim$ Chi(2)
    - 0.95 quantile of Chi(2) = 5.99.
    - $\sqrt{5.99} = 2.45$
    - Approximate calculation, ignores uncertainty in s^2

Modeling clustering of cypress trees



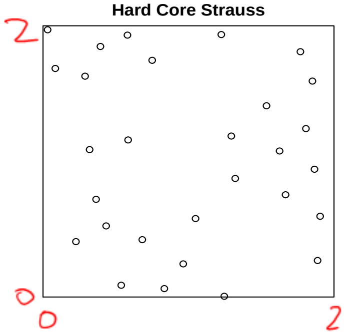
A process with inhibition

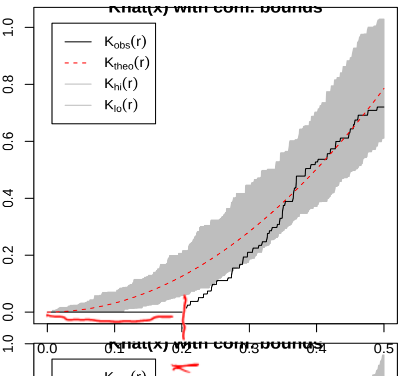
- Point patterns that tend to be regular
- there are many of these - I will use Strauss process as an illustration
- consider sequentially simulating points
- remember def'n of a Poisson process: P[event in dA] does not depend on presence / absence of any other events independence
- to get inhibition, a nearby point **reduces** P[event in dA]
- Strauss process with interaction radius of r
    - generate the tentative location of an event using a Poisson process, with intensity λ
    - draw a circle of radius r around the tentative location
    - count number of already existing events in that circle
    - if $n = 0$, keep the event (intensity is λ)
    - if $n > 0$, keep the event with probability γ^n (intensity | other events is $\lambda\gamma^n$.

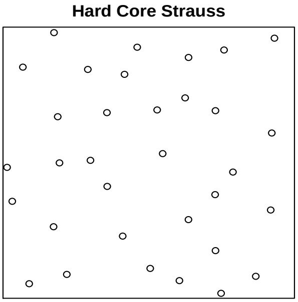
gamma

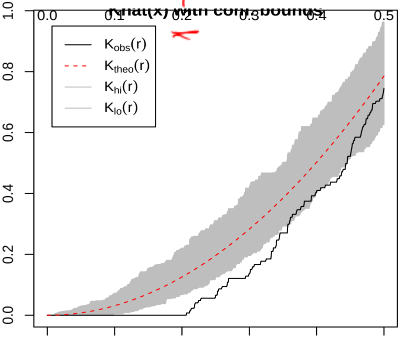
Interpreting a Strauss process

- Characteristics of the process depend on r and especially γ radius gamma
  - $\gamma = 0$: hard-core process. No event allowed w/i distance r of another
  - $0 < \gamma < 1$: soft-core process. Events w/i distance r are less likely.
  - $\gamma = 1$: no inhibition, Poisson process
- Three parameters in this model:
  - r: radius of interaction
  - γ: strength of inhibitation
  - β: related to overall intensity (# of events)


Hard Core Strauss
2
0
0
2


Khat(x) with conf. bounds
$K_{obs}(r)$
$K_{theo}(r)$
$K_{hi}(r)$
$K_{lo}(r)$
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5


Hard Core Strauss


Khat(x) with conf. bounds
$K_{obs}(r)$
$K_{theo}(r)$
$K_{hi}(r)$
$K_{lo}(r)$
0.0
0.2
0.4
0.6
0.8
1.0

Estimating parameters in an inhibition model

$\sum_{\text{events}} \log \lambda$

- Can write down an approximation to $K(x)$
  - use as we did for a cluster process
- Or use likelihood:
- Likelihood for CSR or inhomogeneous Poisson process is easy to write down
- lnL is a sum because points are independent
  - Hard to write down $\log L$ for processes with inhibition
    - Need joint distribution of all events, not sum of independent pieces
  - And even harder to maximize

Estimating param. of a process with inhibition

- Two issues:
  - log Likelihood is not a sum of independent pieces
  - hard to find maximum for some parameters
- Solutions (as of now, not the final word):
- 1) pseudolikelihood
  - Approximate the joint distribution:

other locations

$$f(Y_1, Y_2, \cdots, Y_n \mid \theta) \approx \underbrace{f(Y_1|Y_{-1}, \theta)}\underbrace{f(Y_2|Y_{-2}, \theta)} \cdots f(Y_n|Y_{-n}, \theta)$$

    where Y_{-i} means without event Y_i
  - resulting lnL is:

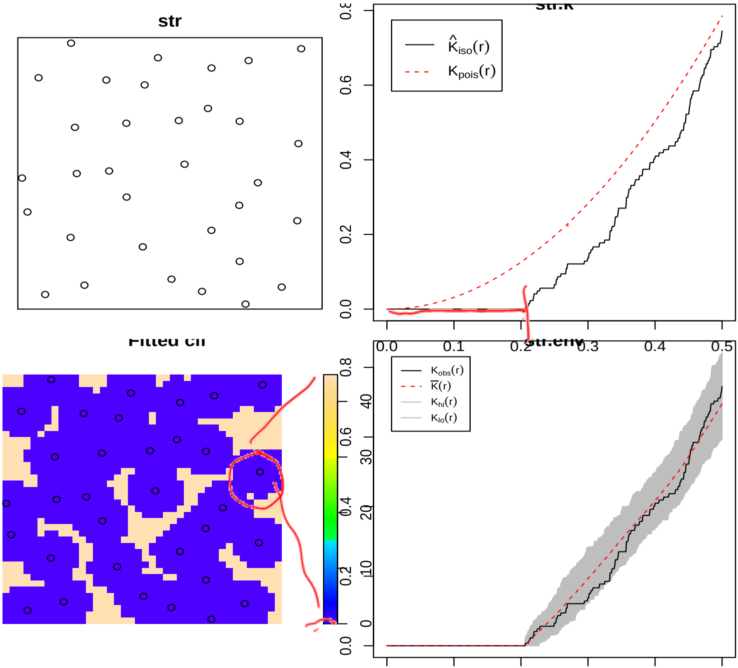
pseudo likelihood

$$\log L(\theta \mid Y_1, Y_2, \cdots, Y_n) = \Sigma_{i=1}^{n} \log L(\theta, Y_i \mid Y_{-i})$$

  - leads to good estimates but Var θ badly estimated
  - so bad tests, confidence intervals

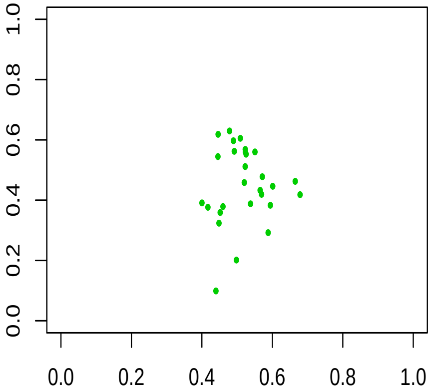
Estimating param. of a process with inhibition

- 2) profiling over r
  - no analytical equations for MLE's of a Strauss process
  - have to numerically maximize
  - turns out to be easy to maximize lnL for γ and β, not for r
  - r is called an irregular parameter. very hard to find a maximum, even numerically
  - Solution: profile likelihood
  - pick a value of r, find $\hat{\gamma}$ and $\hat{\beta}$ that maximize $\log L(\gamma, \beta \mid Y, r)$, i.e. fixed value of r
  - repeat for various values of r
  - find the "best" value of r (at least approximately).
  - that is $\hat{r}$, use corresponding $\hat{\gamma}$ and $\hat{\beta}$.


str
str.K
$\hat{K}_{iso}(r)$
$K_{pois}(r)$
Fitted cif
str.env
$K_{obs}(r)$
$\overline{K}(r)$
$K_{hi}(r)$
$K_{lo}(r)$

Combining pattern and trend

- Example point pattern:



Combining pattern and trend

- Two possible interpretations
  - Events are independent, intensity varies
  - Intensity is constant, events are clustered
- Remember geostats: trend + spatial correlation
  - No unique decomposition based on the data alone
- Same thing with a point pattern
- Can construct two processes with exactly the same $K(x)$ function
  - One is varying intensity, independent events
  - One is constant intensity, clustered events

Combining patterns and trend

- Usual solution: relies on covariates
  - Trend is something you can predict from covariates
  - Pattern is what is left over
- Examining pattern when intensity not constant

$$\frac{\#}{\|A\|}$$

$\Sigma \frac{\# \text{ events}}{\text{edge corr.}}$

* (- Adjust estimator - most common

  - "Inhomogeneous" $K(x)$:

test

examine pattern

plot K( )

$$\hat{K}_I(x) = \frac{1}{\|A\|} \sum_{i \neq j} \frac{I(d_{ij} < x)}{w_{ij} \lambda(s_i) \lambda(s_j)}$$

- model to estimate λ

model driven by covariates

kernel smooth

large bandwidth.

  - Note: when $\lambda(s)$ constant $= n/\|A\|$ get usual $\hat{K}(x)$

$$\hat{K}(x) = \frac{\|A\|}{n^2} \sum_{i \neq j} \frac{I(d_{ij} < x)}{w_{ij}}$$

only test.

powerful test.

- Adjust expectation - Null hypothesis
  - Fit trend model $\hat{\lambda}(s)$,
  - simulate inhomogeneous Poisson process with that $\lambda(s)$ surface
  - Compute $K(x)$, repeat

Combining patterns and trend

- Modeling patterns and trend simultaneously
  - Inhibition / segregation — large scale
    - Pseudolikelihood: Easy to include trend and inhibitation
  - Clustering — small scale
    - Not settled: current usual practice is to estimate $\lambda(s)$ as function of covariates
    - Use the inhomogenous K(x) estimator with that $\lambda(s)$
    - using minimum contrast to fit the clustering process

doesn't work well