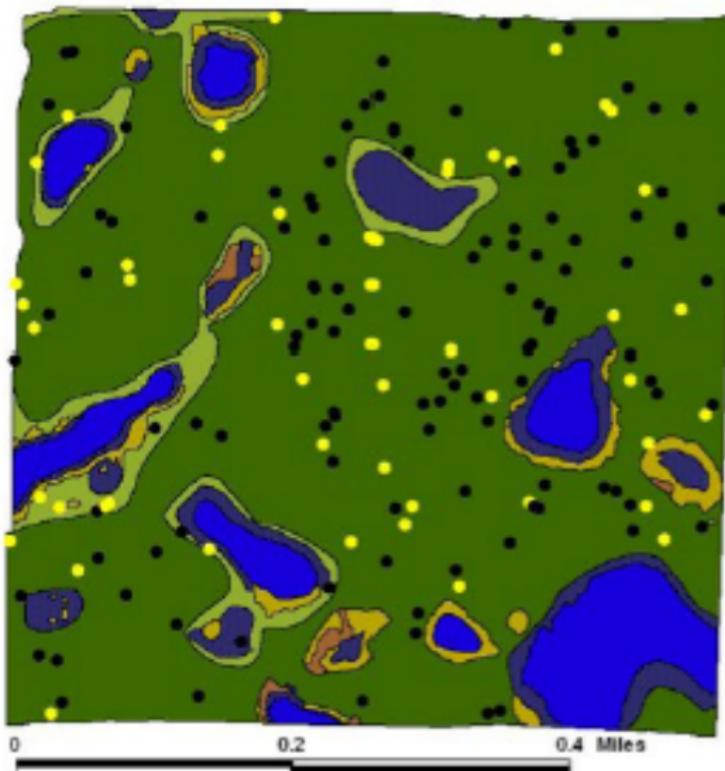


Spatial Point Patterns

- Observed locations of events: datum is the location
 - Major shift in interest from previous material!
 - Up to now, location has been fixed point or fixed area,
 - Location arbitrary or happenstance, often controlled by the investigator (where to take point samples)
 - Random quantity has been the value at each location
- Random quantity is now the location of an event
- May record additional information at each location \Rightarrow marked point process
 - Sometimes small # of classes
 - Examples: species of tree, live / dead plant, successful / unsuccessful bird nest, disease case / not diseased person
 - Or, may be continuous quantity
 - Examples: diameter of tree, angle of a crystal
- But that addn info only exists when there is an event at that location

Duck nests in a 1/4 section of ND



Potential questions

- ① is intensity (# events / unit area) const. or vary. over the study area
- ② how does intensity vary as function of potential covariates
 - EX: does intensity of duck nests decline with distance to wetland?
- ③ are events randomly scattered, clustered, or regular
 - EX: are duck nests independently located in space, or do they cluster near other duck nests, or do they avoid being near other nests?
- ④ how can we describe pattern at multiple scales?
- ⑤ how can we describe rel. between two (or more) types of points?
 - EX: do depredated nests tend to occur near other depredated nests?
- ⑥ how can we describe the cor. between marks as a function of distance?
 - when mark is a continuous value
 - Historically, 3) was most important Q
 - Now, moving beyond to all the other Q.
 - We'll begin with 3 and 4, then 5 and 6, end w/intensity

Application areas

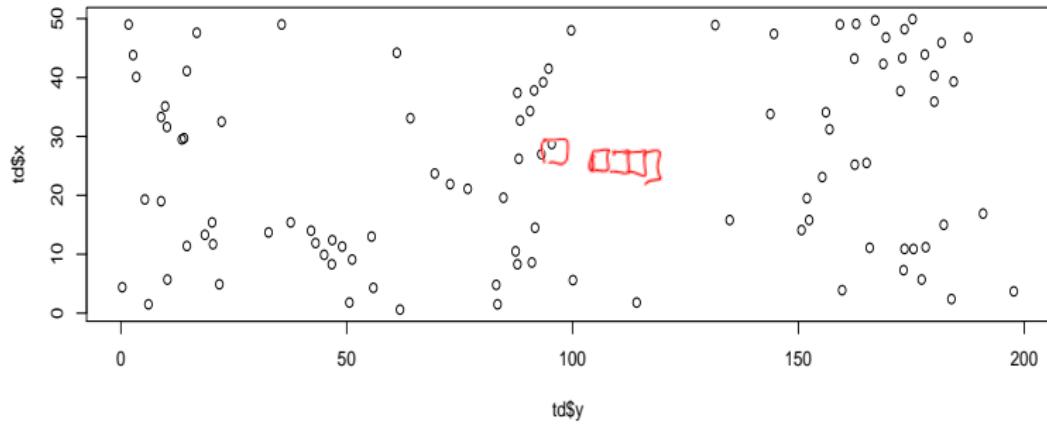
- Many, including:
- Ecology: historically important field of application, many different applications, including:
 - spatial pattern (random / clustered / avoidance) of a single species
 - patterns of mortality (clustered or not?)
- transportation: locations of accidents
- neurology: locations of neurons
- geology: locations of earthquakes (space, or space/time)
- geography: do similar types of stores tend to cluster near each other?
- epidemiology: do cases of a particular disease cluster?
 - If so, suggests contagious disease or single spatial cause

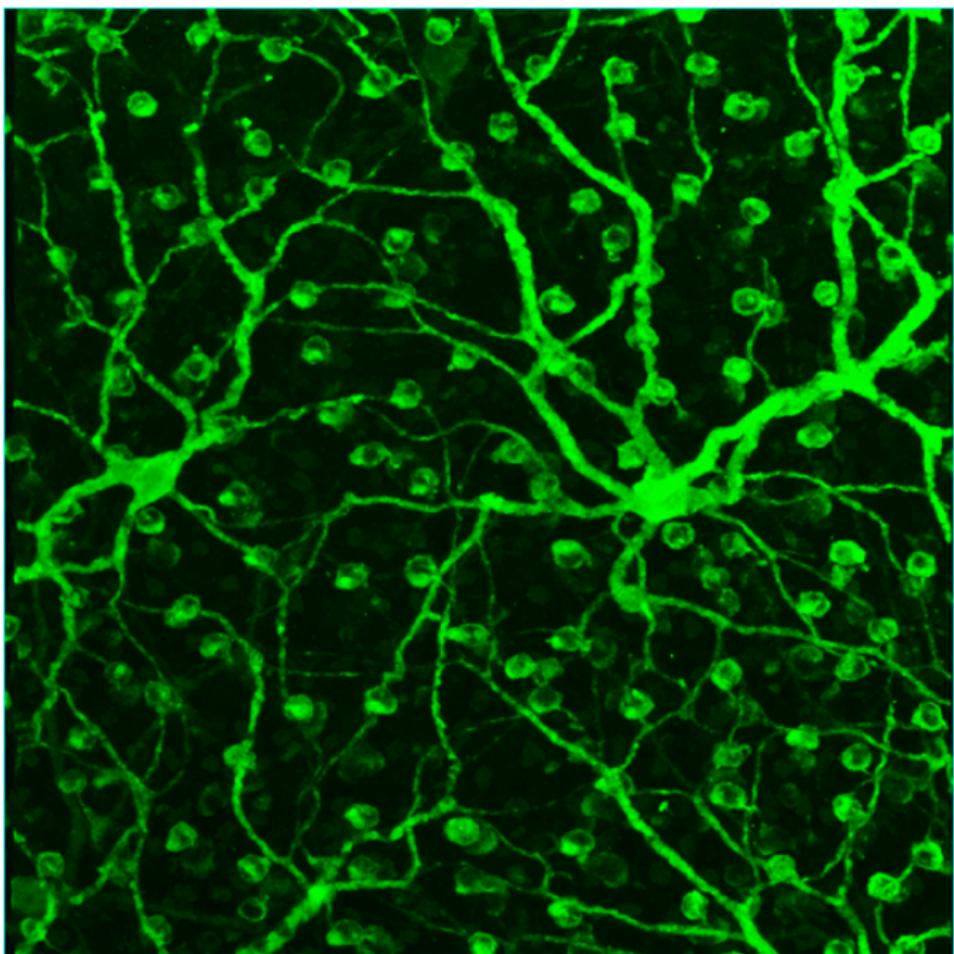
Examples in pictures:



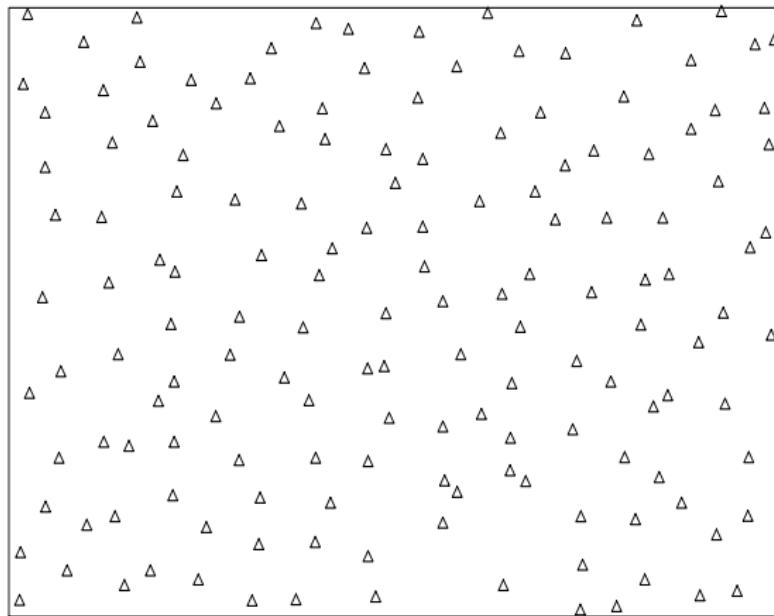


Cypress trees in Good plot 1





Amacrine cells, on



Homogeneous Poisson Process = Complete Spatial Randomness

- imagine a very small area, dA , with $P[\text{event occurs in } dA] = \lambda dA$
- dA small enough that:
 - at most 1 event in dA
 - most areas have 0 events
- $\lambda = \text{expected } \# \text{ events / unit area}$
- λ is the intensity of the spatial process
- Two assumptions that give HPP = CSR \rightarrow random pattern.
 - λ constant over study area
 - the outcome (0/1) in dA_1 is independent of the outcome in non-overlapping area dA_2

tiny tiny area

intensity

not d time

λdA

• λ

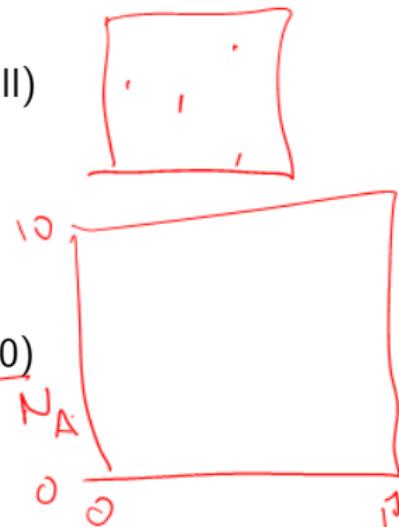
random pattern.

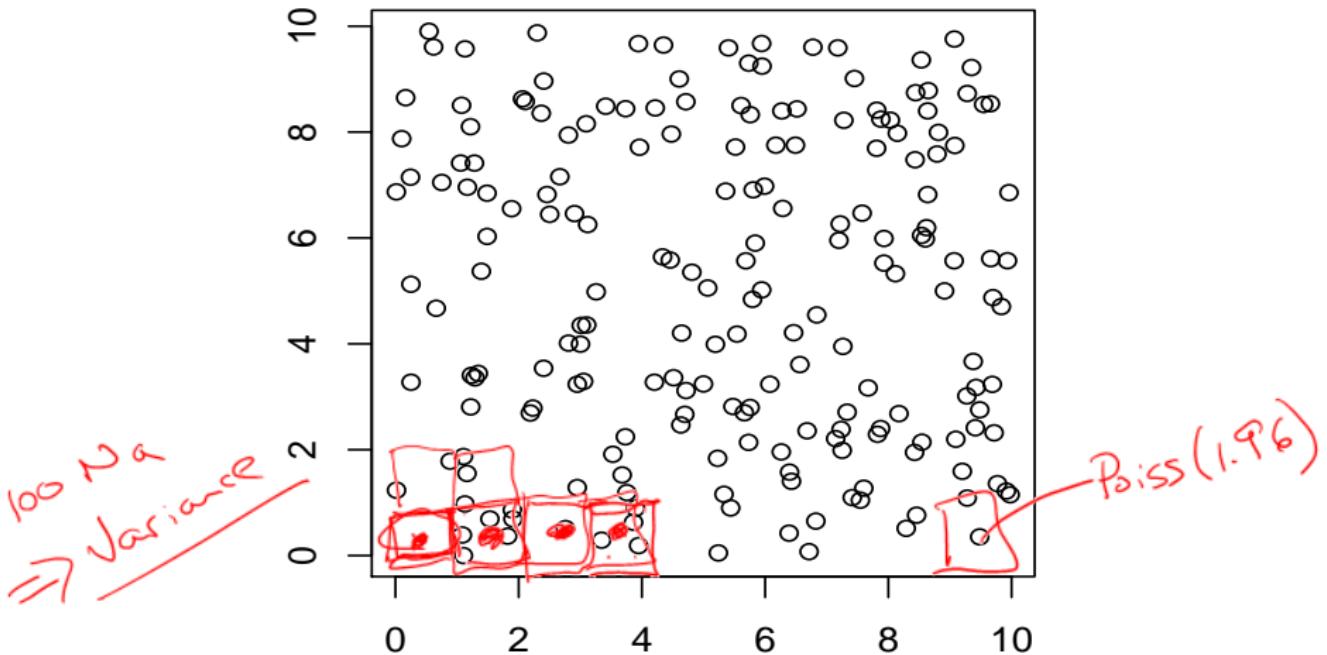
□ □

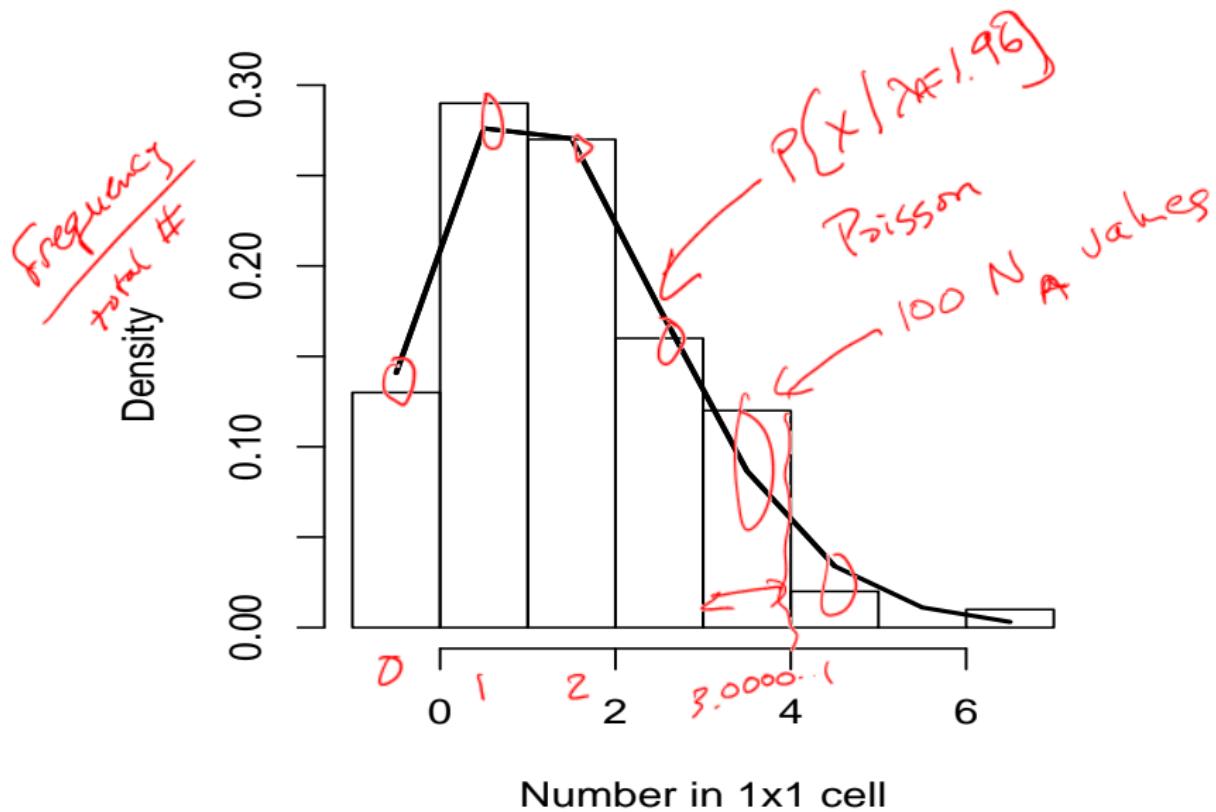
Homogeneous Poisson Processes

Some mathematical results:

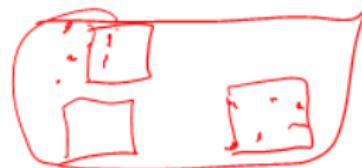
- Define $N_A = \#$ points in area A (no longer small)
- $N_A \sim \text{Poiss}(\lambda A)$ *When HPP*
 - mean #: λA *mean \propto area*
 - var #: λA
 - pmf $P[X = \lambda A] = \frac{e^{-\lambda A} (\lambda A)^{\lambda A}}{X!}$
- examples: CSR, observe 196 obs on $(0,10), (0,10)$
- Look at individual 1×1 quadrats \rightarrow 100 values N_A
 - mean count per 1×1 quadrats = 1.96
 - Var count = 1.90 *sampling variation*
 - Histogram close to theoretical pmf







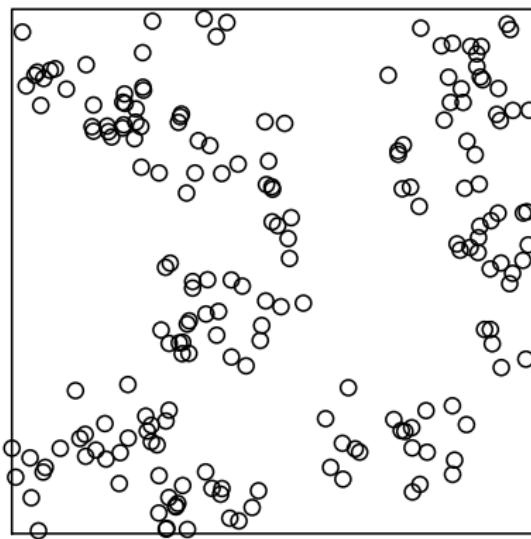
Clustering

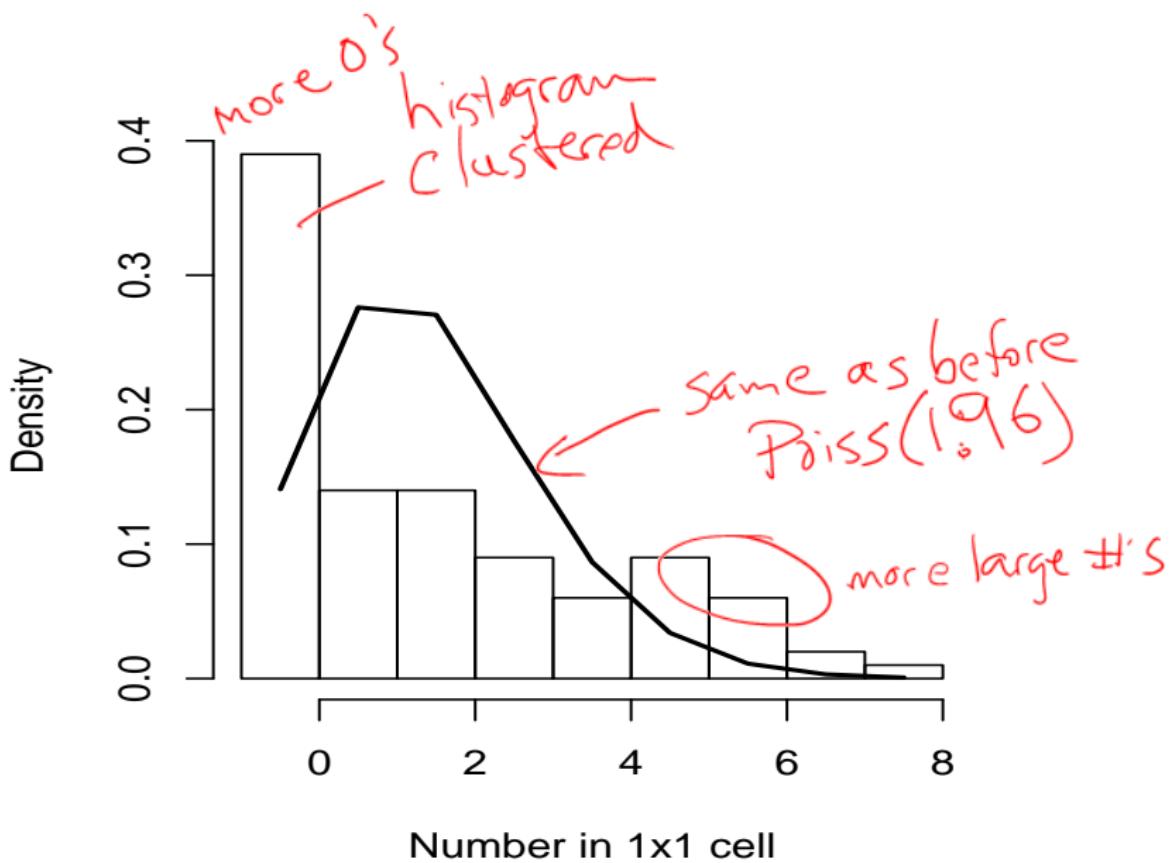


- Clustered processes/pattern: points more likely to occur near other points.
- For quadrats, means that:
 - some quadrats contain a cluster, have more points than expected
 - other quadrats have no points
- same mean, larger variance
- for clustered process with 196 points on next two slides:
 - mean = 1.96, variance = 4.70



A clustered process





Regular processes

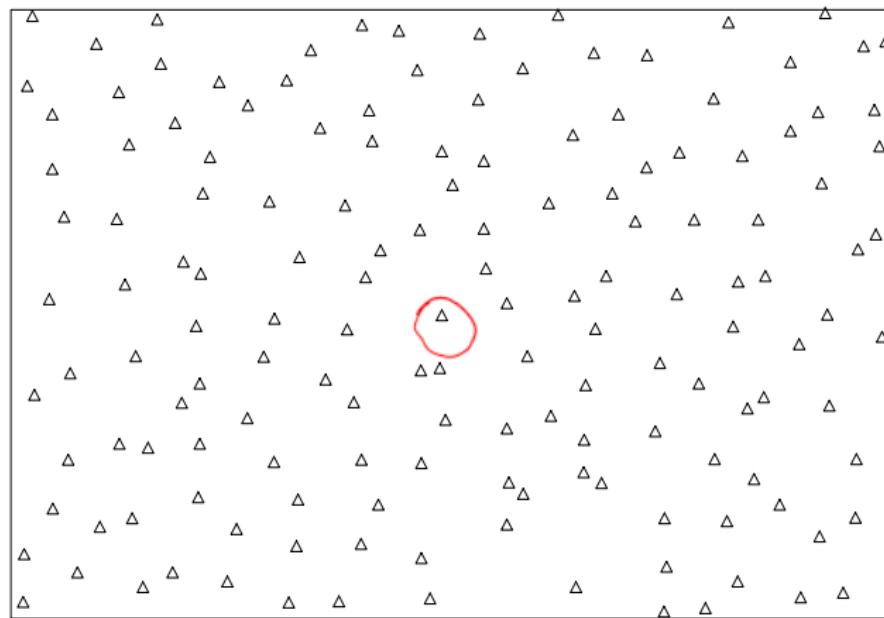


HP: new point
equally likely
anywhere

- $P[\text{event in } dA]$ lower if dA close to another point
- Tends to “space out” points

point \rightarrow constant
 $\Rightarrow P(dA)$ doesn't
 depend on
 previous event

Amacrine cells, on



106 even
mean 1.96
variance
small
var < mean

A better approach

- Historically: Quadrats used extensively
- But, very limited. Restricted to one specific scale (size of quadrat)
- Better approach
 - record locations of events, not just count in a box
 - usually all events in a predefined area *complete enumeration*
 - can be random sample of events
 - But, hard to take a simple random sample
- Can convert to quadrat counts, but can do a LOT more with (x,y) data

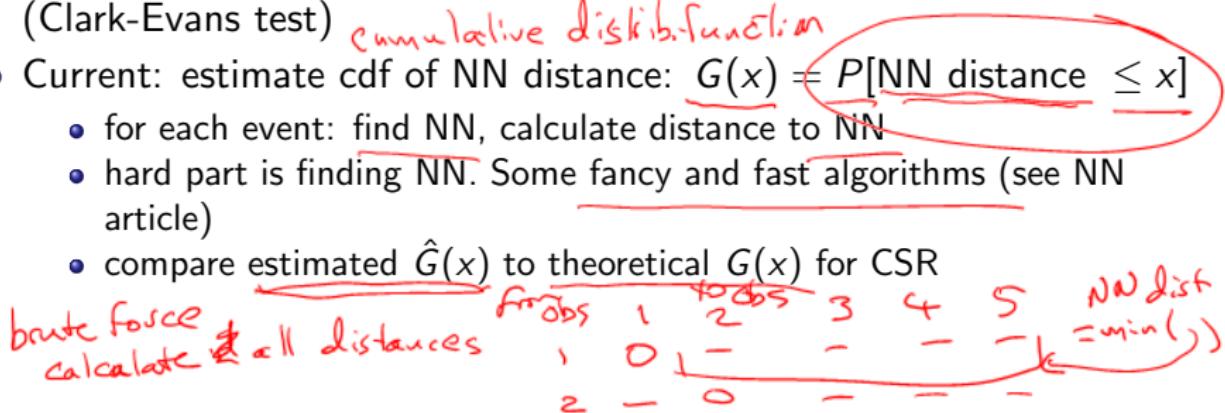
Summary functions

- Concept:
 - measure something as a function of distance
- Various choices of summary
 - Distance to nearest neighbor (event - event distance)
 - Distance to nearest point (point - event distance)
 - Combination of these two
 - Ripley's K function — *last 20 yrs*
better
 - pair correlation function
- Each has uses

Distance to nearest neighbor

- How close is each obs. to its nearest neighbor?
 - clustering: NN distances tend to be small
 - random (CSR): intermediate
 - regular: NN distances tend to be large
- Historical: calculate mean NN distance, compare to theoretical value (Clark-Evans test)
- Current: estimate cdf of NN distance: $\hat{G}(x) = P[\text{NN distance} \leq x]$
 - for each event: find NN, calculate distance to NN
 - hard part is finding NN. Some fancy and fast algorithms (see NN article)
 - compare estimated $\hat{G}(x)$ to theoretical $G(x)$ for CSR

$$\hat{G}(x) = \frac{\#\text{NNd} \leq x}{\text{total \#}}$$

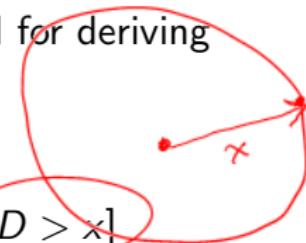


Theoretical CDF on NN distance

CSR $\bullet P[D < x] = 1 - e^{-\lambda \pi x^2}$ 2D
area of circle, radius x

- x is distance of concern,
- λ is intensity (events per unit area)
- πx^2 is area of circle, radius x

- (For the statisticians). Nice ex. of CDF method for deriving transforming a random variable
- Define D = distance to NN dist.



$$G(x) = P[D \leq x] = 1 - P[D > x]$$

P[no obs in circle of radius x]

CSR $N_A \sim \text{Pois}(\lambda A)$, so:

$$P[0] = \frac{e^{-\lambda A} (\lambda A)^0}{0!}$$

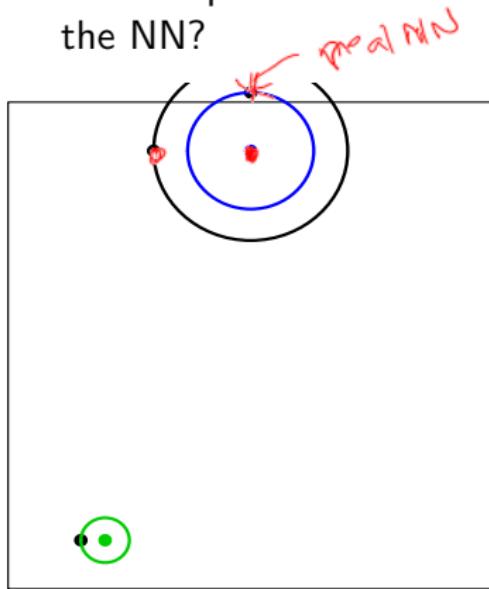
$A = \pi x^2$

$$P[D \leq x] = 1 - e^{-\lambda \pi x^2}$$

- Similar ideas, different formula for 1D or 3D.

Edge effects

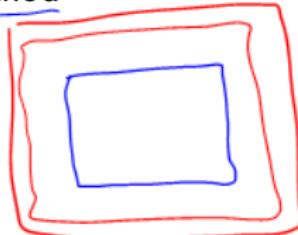
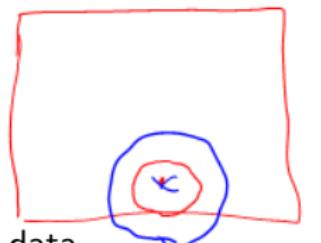
- Above assumes infinite plane
- Real study areas have edges
- When a point is close to edge of mapped area, what is the distance to the NN?

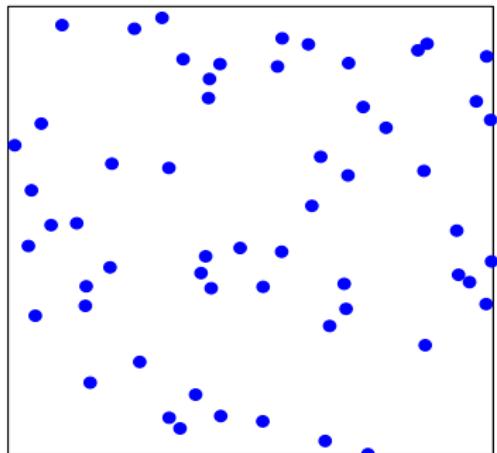


- overestimate D .
- true NN may be just over the boundary (close to event)
- observed NN (inside study area) is larger than it "should be"
- So underestimate $G(x)$ especially for large x

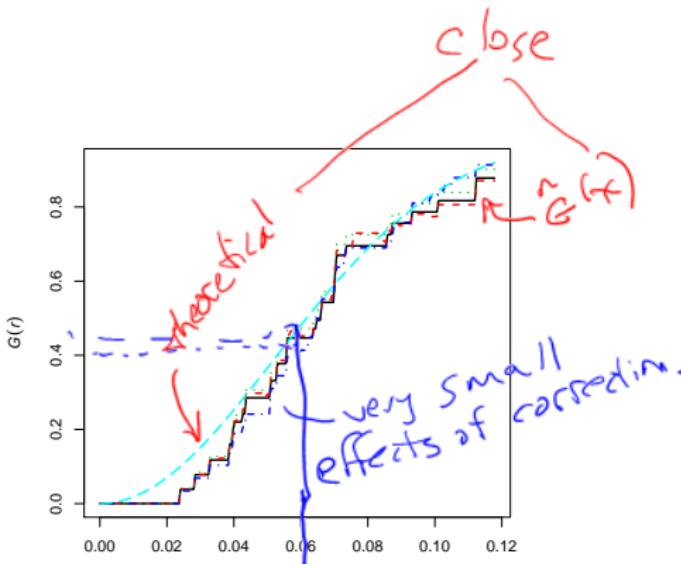
Edge effects

- Three approaches to edge effects
- 1) Ignore problem.
 - Study area edge is a real edge (e.g. lake shoreline)
 - really care about distance to nearest valid event
- 2) Traditional: adjust estimator
 - “edge-corrected” estimator of $\hat{G}(x)$
 - Usual: use the Kaplan-Meier estimator for censored data
 - others have been proposed, avoid reduced sample method
 - but, bias correction increases $\text{Var } \hat{G}(x)$
- 3) Radical: adjust expectation
 - Use uncorrected estimator
 - Change theoretical $G(x)$ to account for edge effects
- If goal is to est. $G(x)$, 2) much better
- If goal is to test CSR (or some other process), 3) has higher power
- Estimate theoretical $G(x)$ by simulation



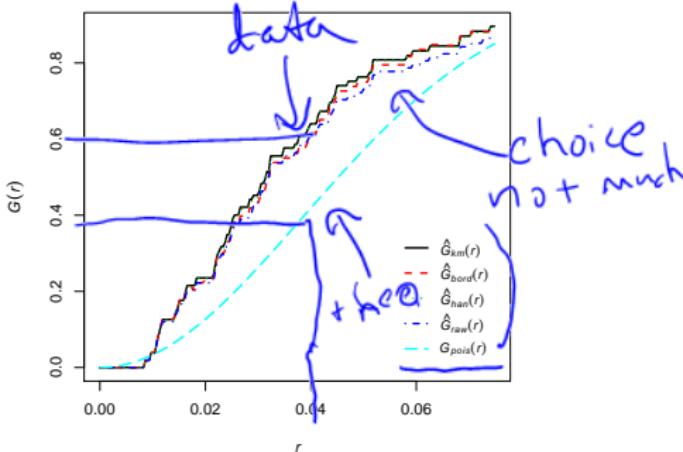
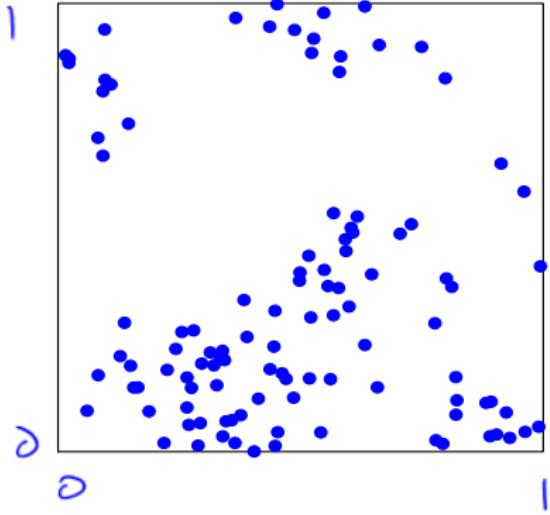


Point patterns
cluster/regular / random = CSR km method.
 Summary functions
 $G(x) \approx P[\text{NN} \leq x]$



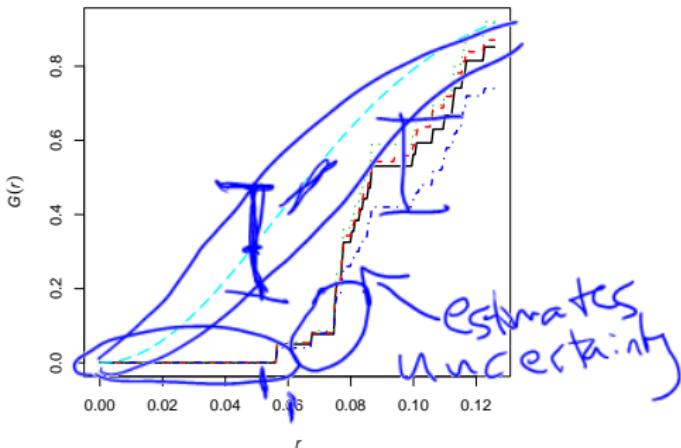
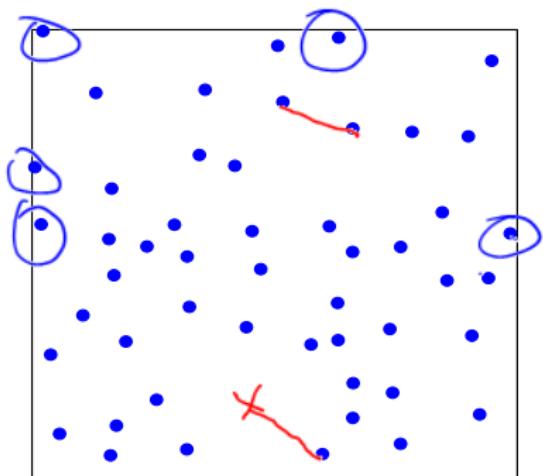
clustering

- More points close: shorter NN distance



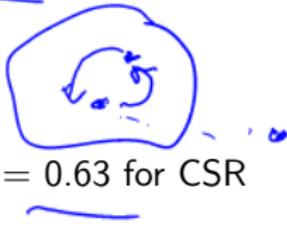
inhibition / regularity

- Fewer points close: longer NN distance
- Hard core process: no points within a minimum distance



Sampling variance of $\hat{G}(x)$

- How to calculate $\text{Var } \hat{G}(x)$?
- Quite a hard problem: 1) edge effects
- 2) "Reflexive NN's": pair of points
 - B is A's NN, A is B's NN
 - same NN distance
 - surprisingly common: $P[\text{reflexive}] = 0.63$ for CSR
 - increases $\text{Var } \hat{G}(x)$ —
- $\text{Var } \hat{G}(x)$ has been derived under CSR, ignoring edge effects
- Now, almost always computed by simulation
 - Simulate a realization of null hypothesis process (e.g. CSR)
 - Estimate $\hat{G}(x)$
 - Repeat simulate/estimate 99 or 999 times
 - Calculate $\text{Var } \hat{G}(x)$ at various x
- Or go straight to a confidence interval
 - Calculate 0.025 and 0.975 quantiles of $\hat{G}(x)$ at a specified x value
 - repeat for various x 's

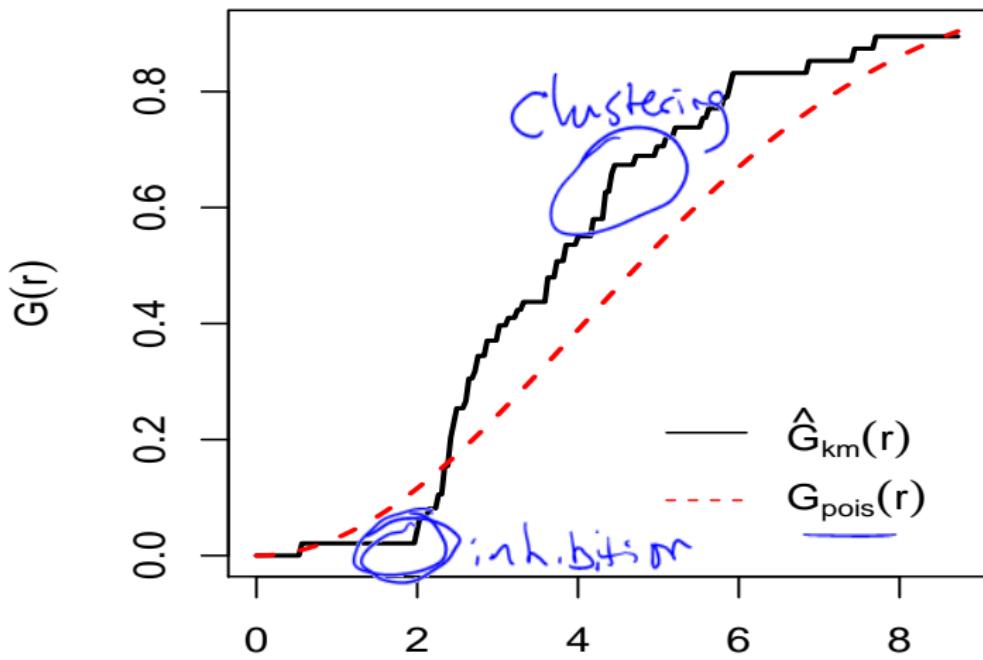


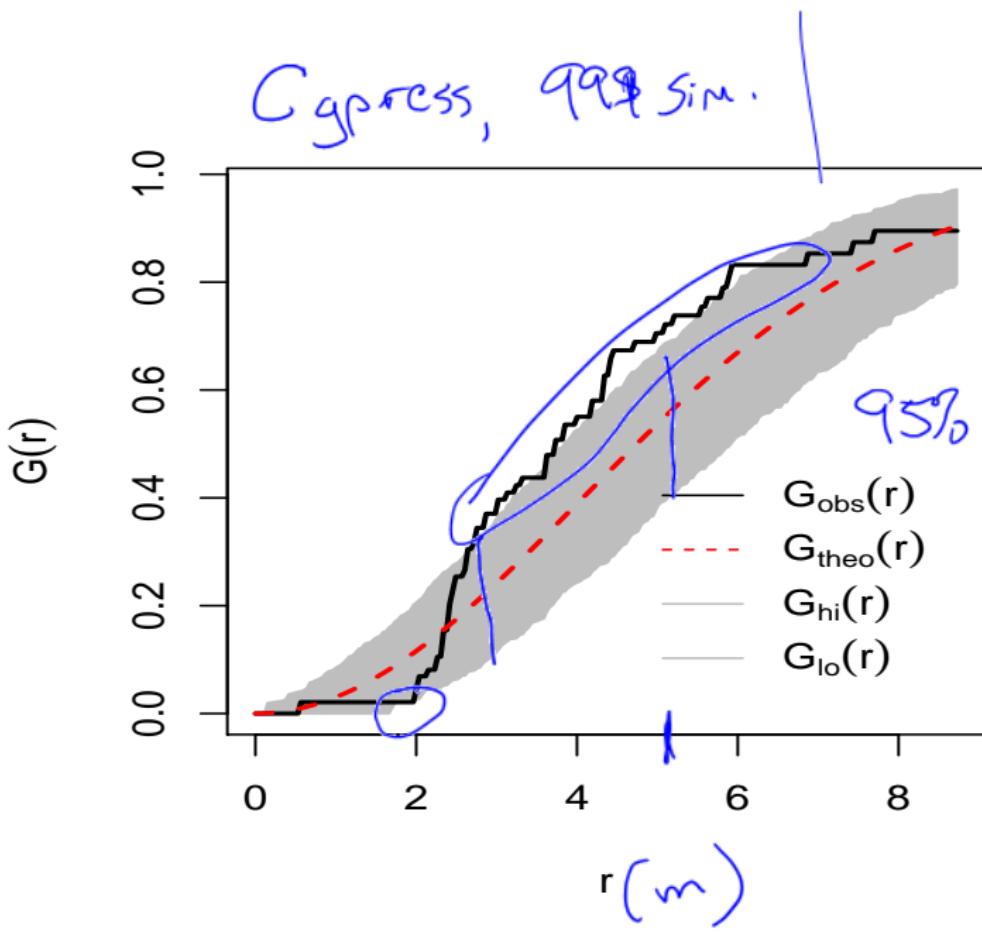
$\text{Var } \hat{G}$

$\text{Var } \hat{G}$ given CSR

n_{siml}

Cypress trees



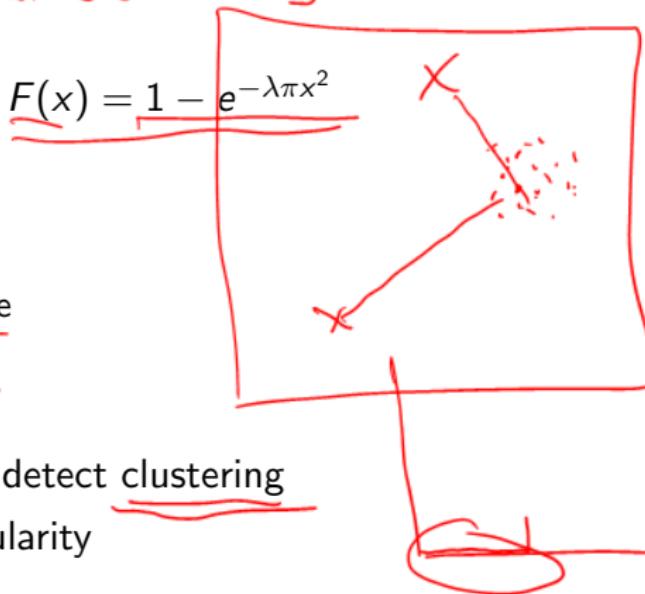


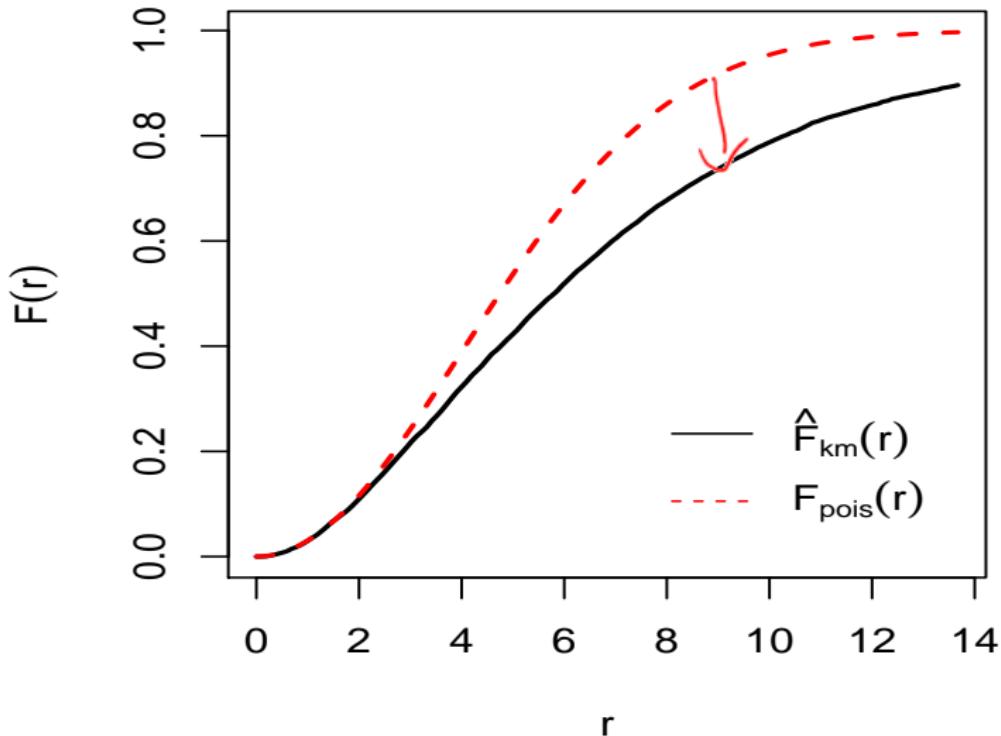
Point-event distance = empty-space function

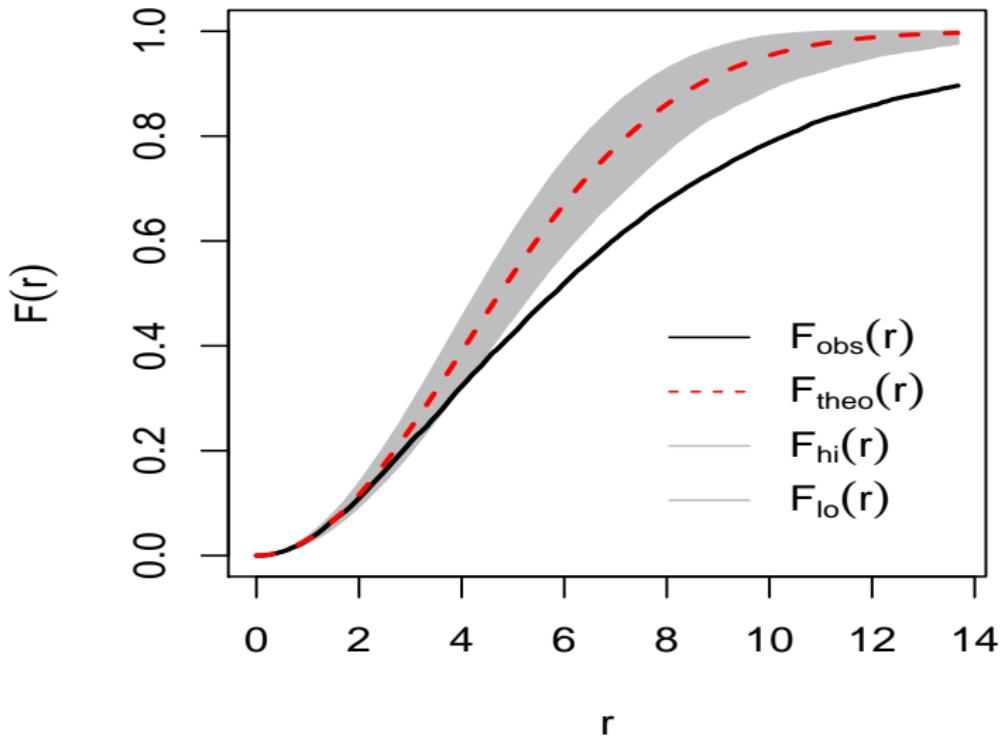
$G(x)$: point-point

- cdf of distance from randomly chosen point (not an event) to nearest event
- Usually denoted $F(x)$
- Under CSR, ignoring edge effects: $F(x) = 1 - e^{-\lambda\pi x^2}$
 - same derivation as for $G(x)$
- But now:
 - large distances \Rightarrow clustering, because big areas of empty space
 - small distances \Rightarrow regularity
- Evaluated in same way as $G(x)$
- $F(x)$ more powerful than $G(x)$ to detect clustering
- $G(x)$ more powerful to detect regularity

$P\{\text{point event} < x\}$





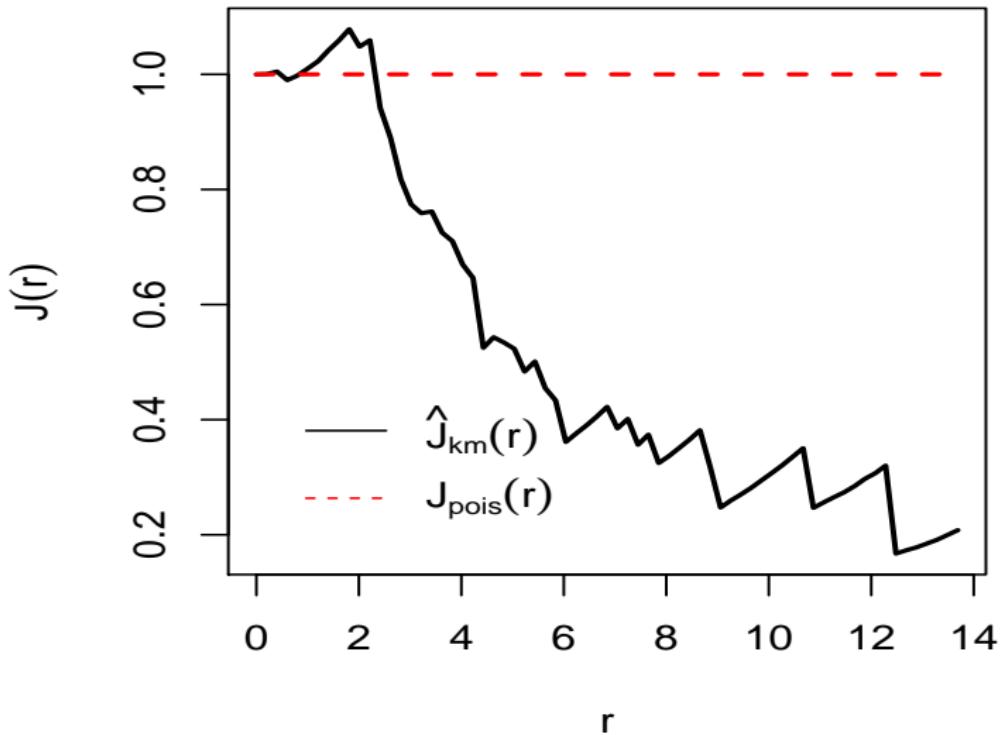


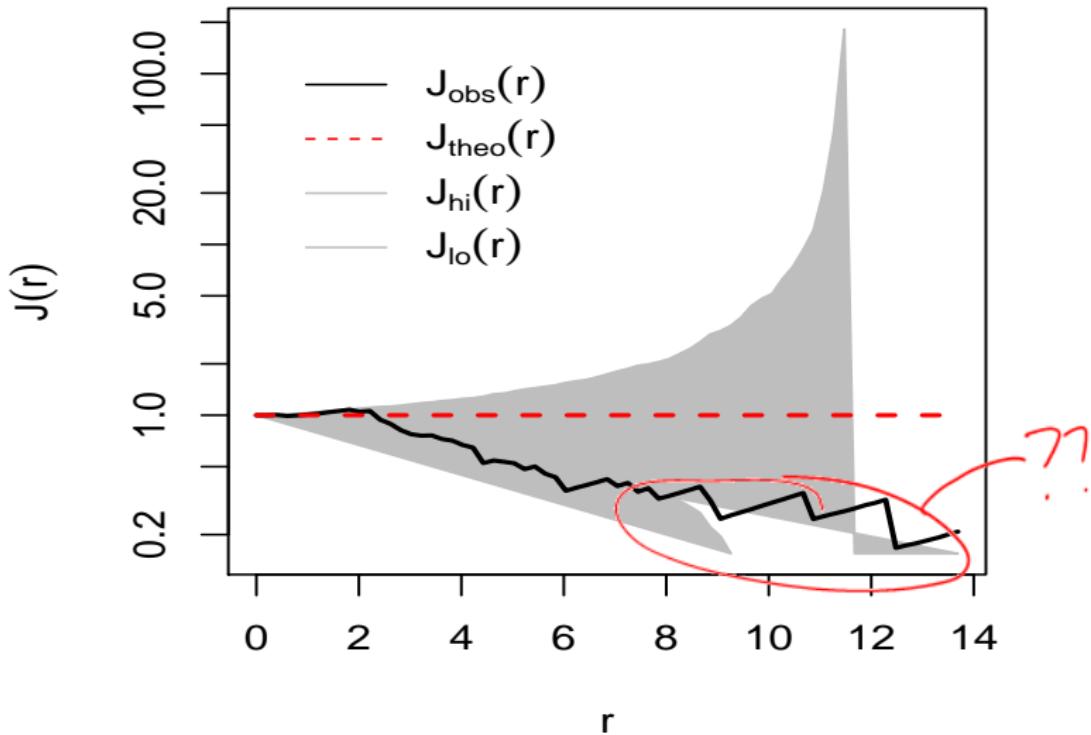
Baddeley's J function

- Can combine $F(x)$ and $G(x)$

$$J(x) = \frac{1 - G(x)}{1 - F(x)}$$

- Interpretation:
 - clustering: $J(x) < 1$
 - CSR: $J(x) = 1$
 - ~~regularity~~: $J(x) > 1$
- Much newer than $F(x)$ or $G(x)$: 1996 paper
- Few have much experience with it





Looking at multiple scales simultaneously

$F(x)$ nearest count
 $G(x)$

- I've emphasized nearest neighbor (of an event, of a point).
- closest event (to the event, to the point)

①

- Straightforward extension to 2nd NN (next closest), 3rd NN, ...
 - Gets harder to interpret
 - and you have a separate plot for each NN rank

②

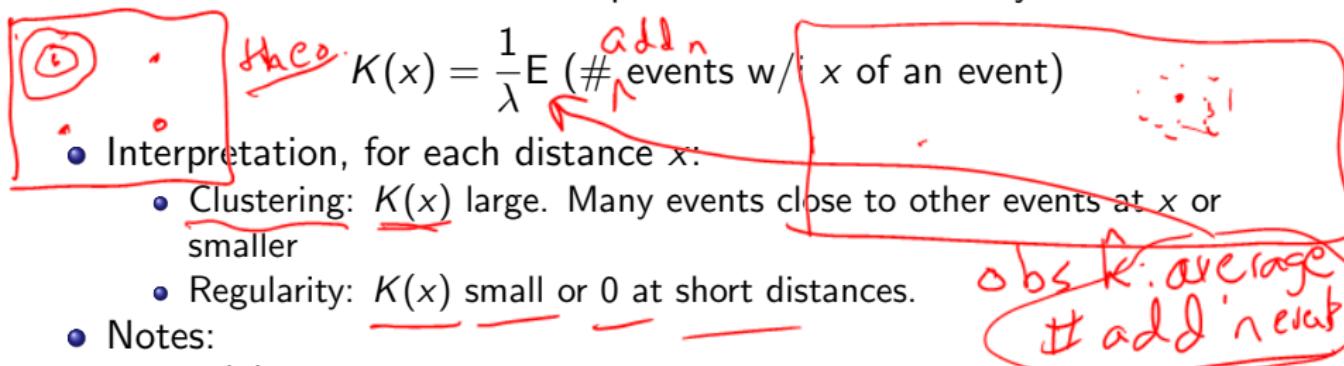
- Rethink how to compute the summary
 - Instead of "how far to closest point"
 - think of how many points within a specified distance?
- leads to Ripley's K statistic (Ripley 1976, *J. Appl. Prob*)



do not include center

Ripley's K function

- Like $F(x)$, $G(x)$, and $J(x)$, looks at 2nd order characteristics of a point pattern
- Now, the most commonly used point pattern analysis function
- Provides information at multiple scales simultaneously



- Notes:
 - $K(x)$ can detect clusters of regularly spaced points
 - i.e., different patterns at different scales
 - but it is cumulative (number of points *within* distance x)
 - we'll see a refinement, the pair correlation function, that looks at points at distance x
 - which simplifies (greatly) inferring the scale of a pattern

$K(x)$ under CSR

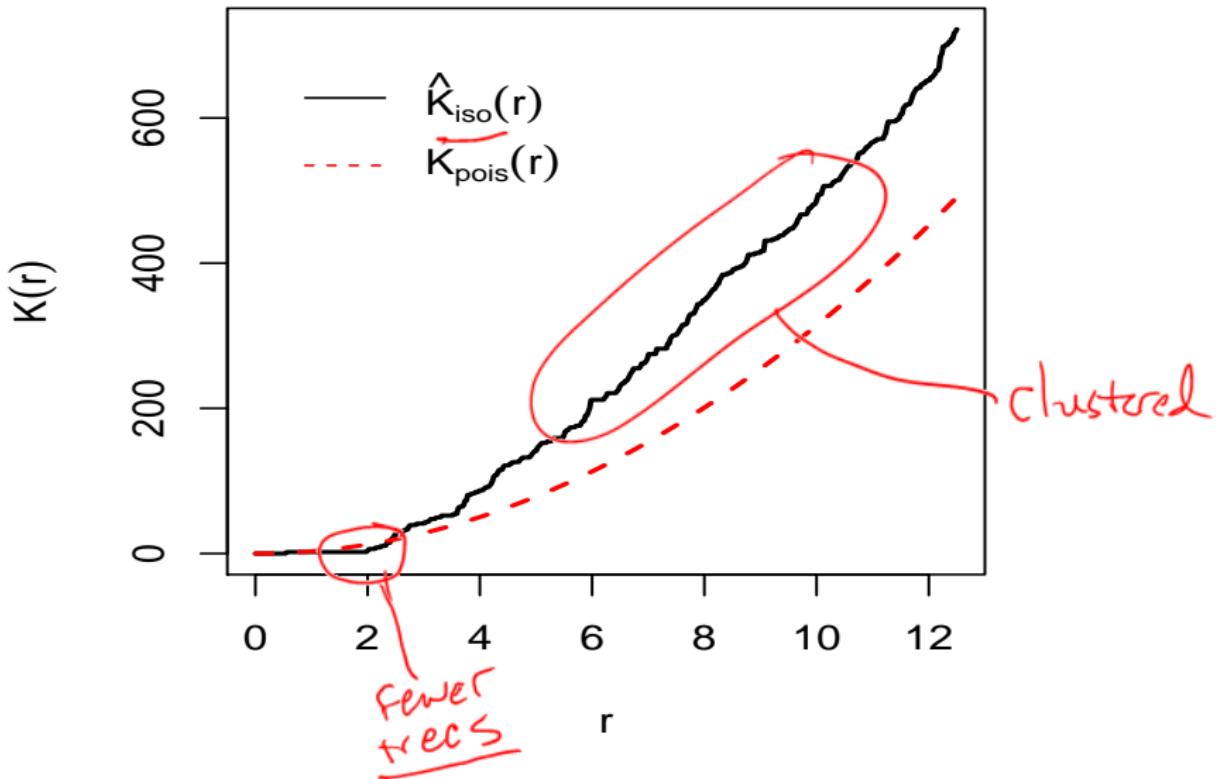
- Expected value, $E \hat{K}(x)$:

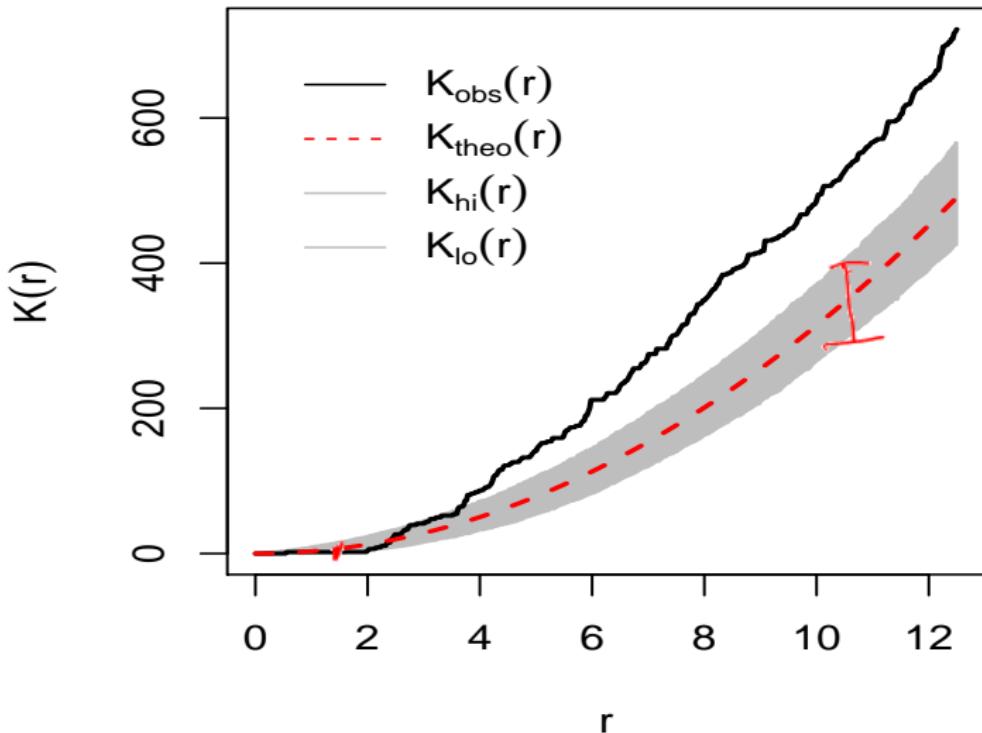
- Under CSR, events are independent, $E \# \text{ in area } A = \lambda A$
- $E \hat{K}(x) = E \# \text{ in area } \pi x^2 / \lambda = \lambda \pi x^2 / \lambda = \pi x^2$

- Variance, $\text{Var } \hat{K}(x)$:

- $\text{Var } \# \text{ in area } A = \lambda A$
- so, $\text{Var } \hat{K}(x) = \frac{1}{\lambda^2} \text{Var } \# \text{ in } \pi x^2$
- $= \pi x^2 / \lambda$
- smaller with more expected points (larger λ)
- increases with distance, x

any intensity
CSR $\hat{K} = \pi x^2$





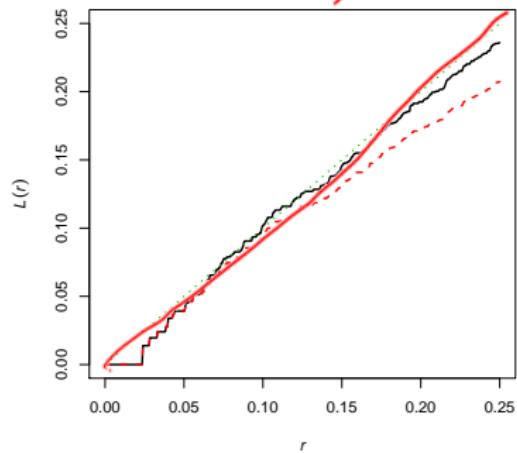
Besag's L function

- $\underline{L(x)}$ is a variance equalizing transformation
- if $\underline{Y} \sim \text{Poiss}(\underline{X}\beta)$, then $\underline{\sqrt{Y}}$ has constant variance
- Besag's original version

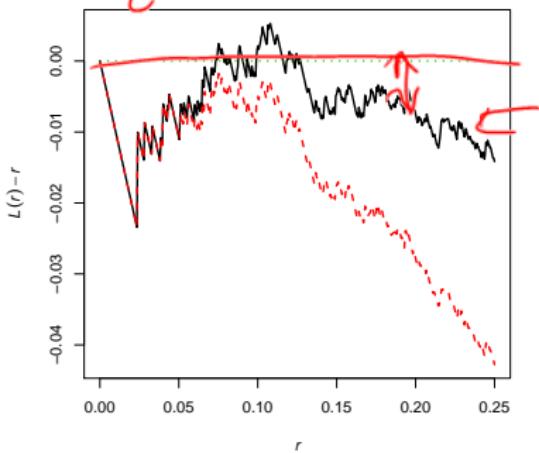
$$\underline{\underline{L(x)}} = \sqrt{\underline{\underline{K(x)}}/\pi} \quad K = \pi x^2$$

- Wiegand and Moloney (2014) call this $\underline{\underline{L_1(x)}}$
- Under CSR: $L(x) = x$, $\text{Var } \hat{L}(x)$ approx constant.
- I prefer $\underline{\underline{L^*(x)}} = \underline{\underline{L(x)}} - \underline{\underline{x}}$
 - Wiegand and Moloney (2014) call this $\underline{\underline{L_2(x)}}$
- Nice feature of $L^* = L_2(x)$: under CSR, $L^*(x) = 0$
- I believe plots of L^* are much clearer (but you decide which you prefer)

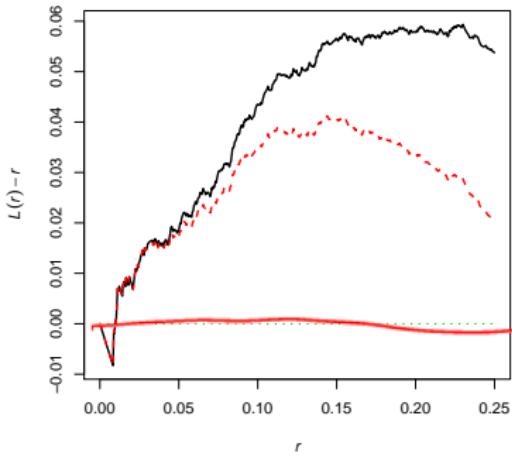
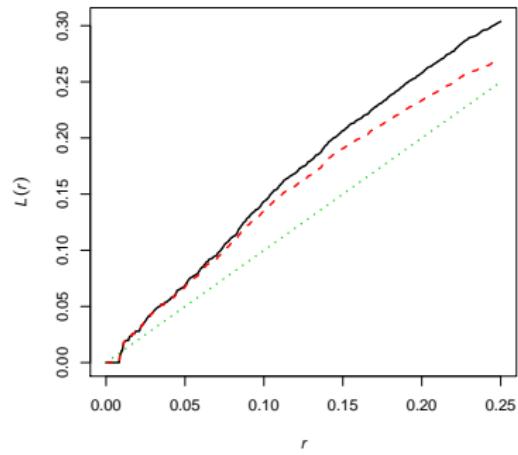
$$L(\gamma) = L_1(\gamma)$$



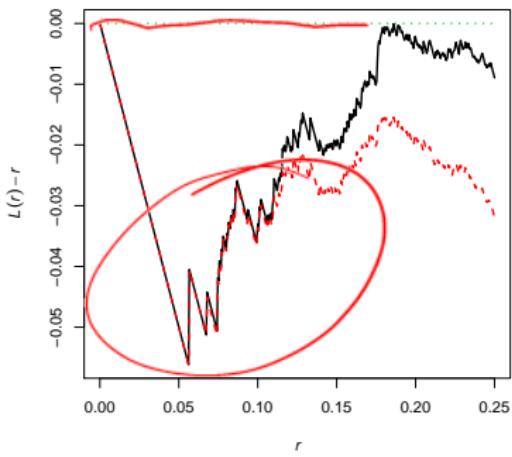
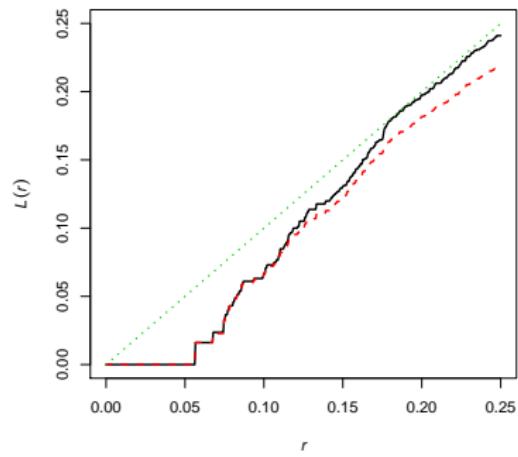
$$L_2(\gamma) = \frac{L(\gamma)}{\gamma} - \gamma$$

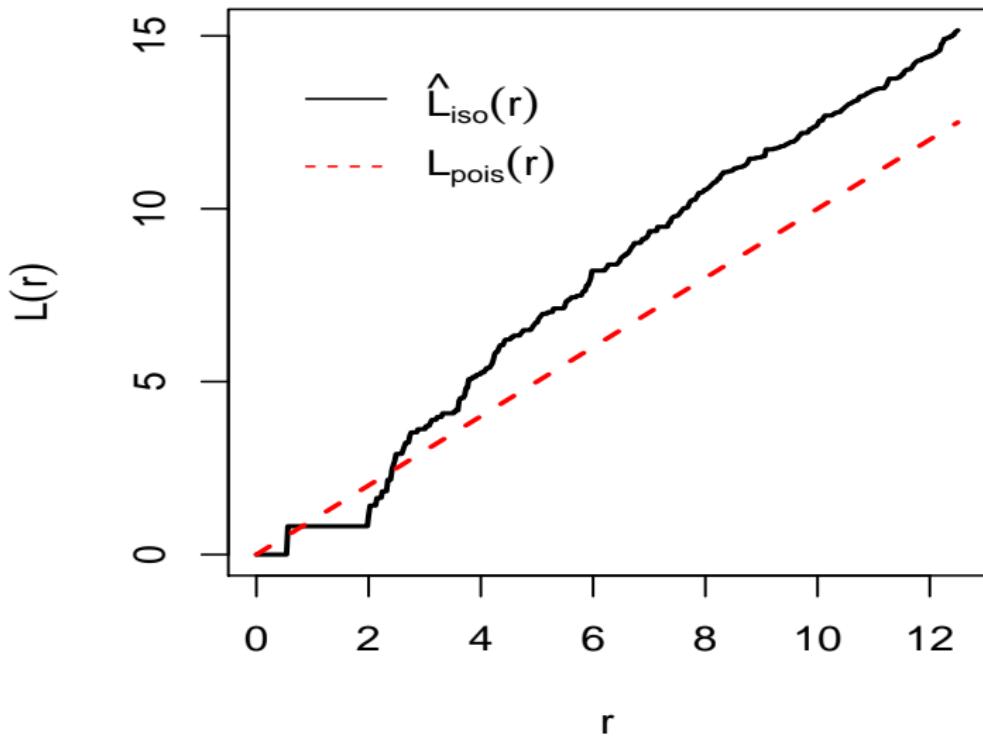


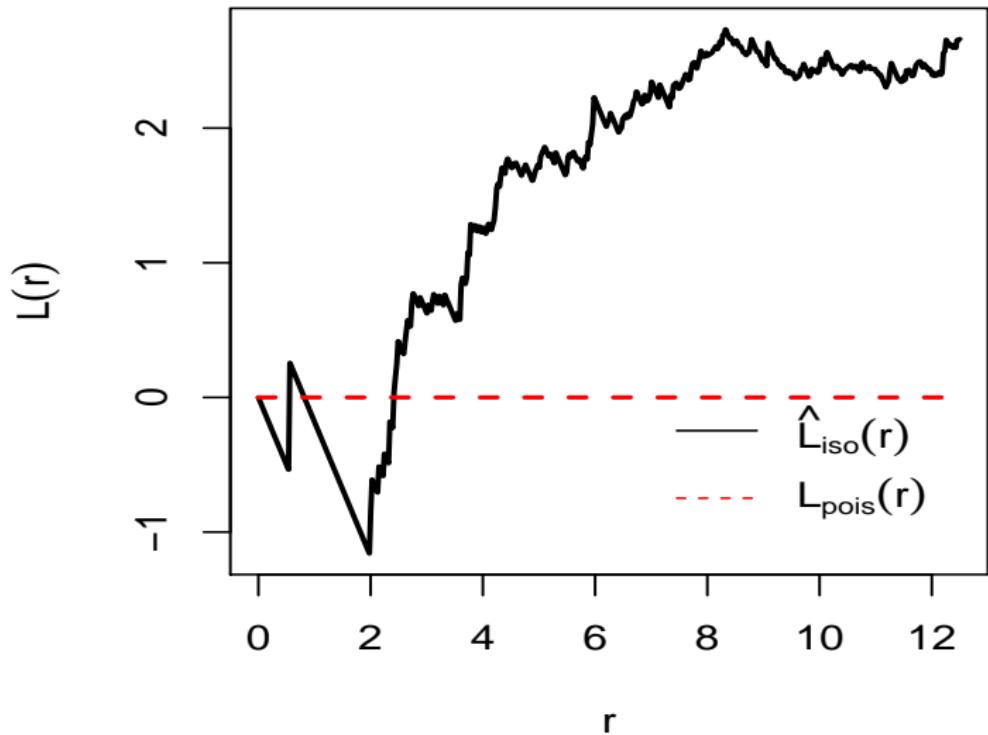
Clustered

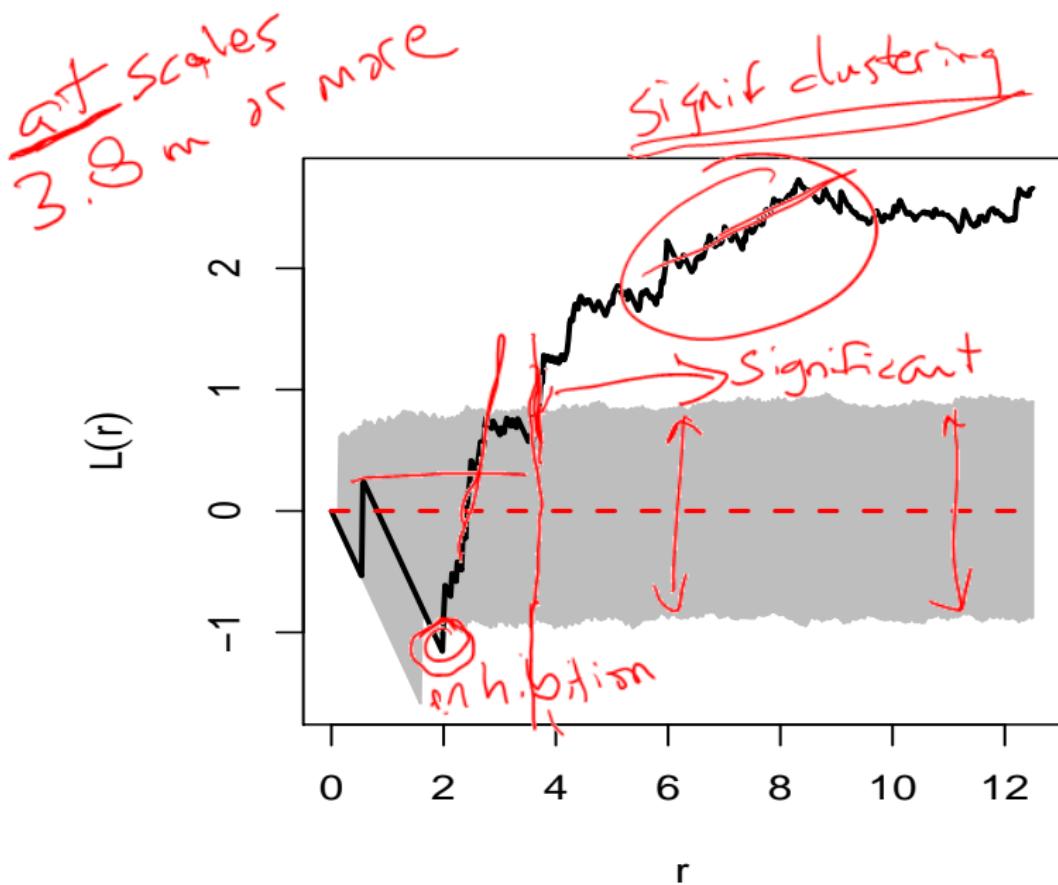


inhibition









Pair correlation functions

- $K(x)$ and $L(x)$ are cumulative measures
 - Based on number of events *within* x of another event
- What if you want to describe association *at* distance x ?
 - Closer to intuition about spatial scale \equiv
 - Can untangle multiple processes
 - inhibition at short distances
 - clustering at large distances
- pair-correlation function, $g(x)$ or $\rho(x)$

$$g(x) = \frac{1}{2\pi x} \frac{dK(x)}{dx}$$

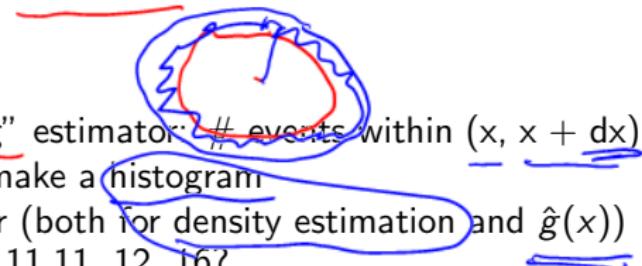
- under CSR ($K(x) = \pi x^2$), $\frac{dK(x)}{dx} = 2\pi x$, and $g(x) = 1$
- $g(x) > 1$ \Rightarrow events more likely AT distance x than under CSR \Rightarrow clustering at a scale of x
- $g(x) < 1$ \Rightarrow events less likely AT distance x than under CSR \Rightarrow repulsion at a scale of x
- range is $(0, \infty)$ with 1 as the neutral point
- so often log transform: evaluate $\log g(x)$

CSR

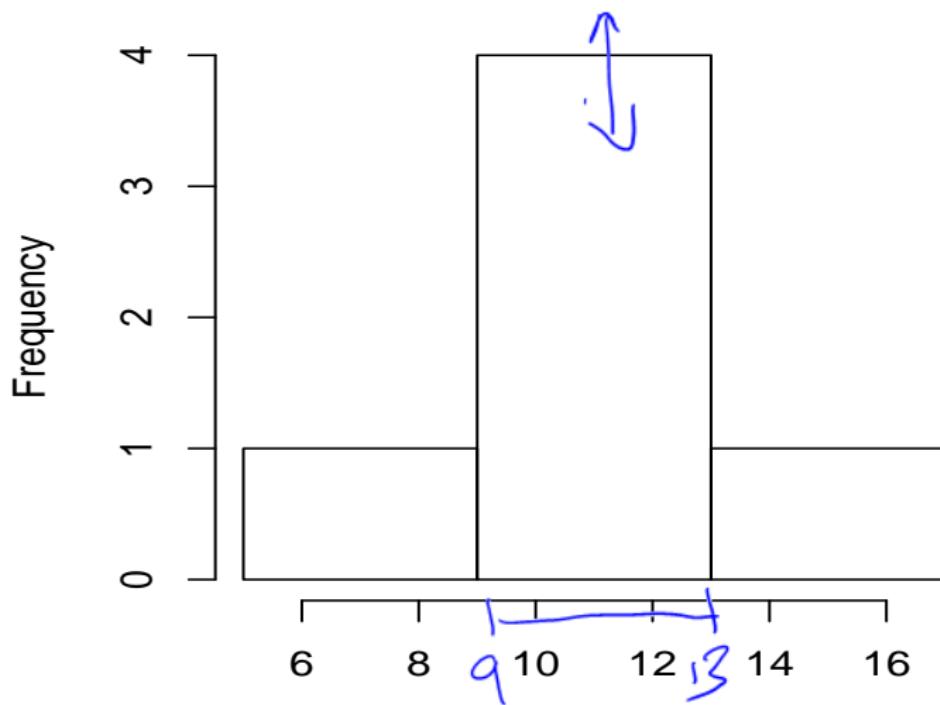
0 \Rightarrow CSR
0 \Rightarrow inhibition
0 \Rightarrow clustering

Estimators of $g(x)$

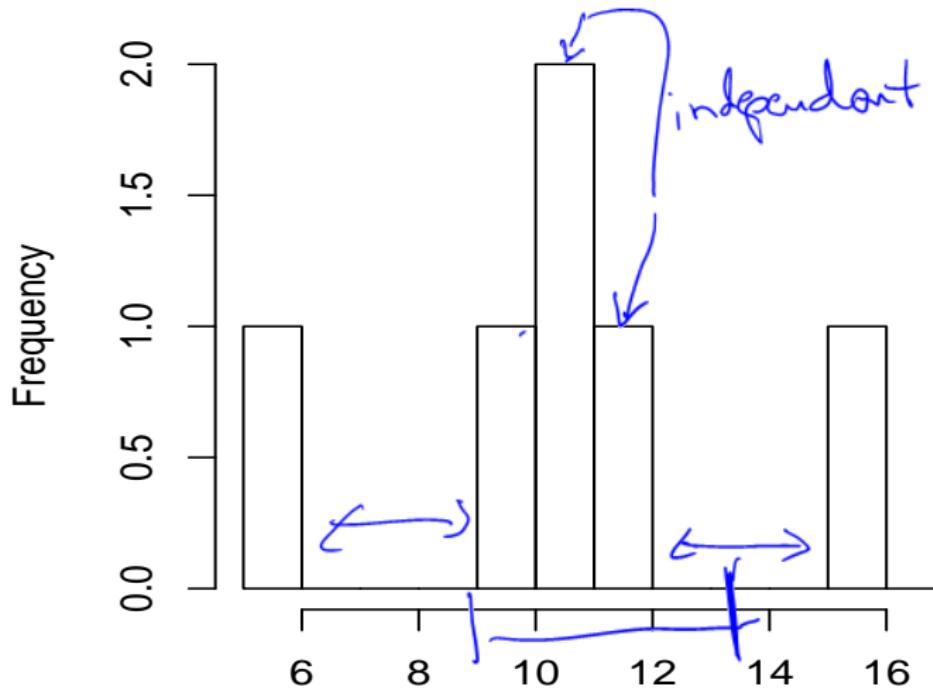
- $g(x)$ is much harder to estimate than $K(x)$
 - $\hat{K}(x)$ is a sum (# events within distance $< x$)
 - $\hat{g}(x)$ depends on 0/1 variable: is there an event at distance $= x$ or not
 - parallel to the issue that a cdf: $P[X < x]$ is easier to estimate than a pdf: $f[X = x]$
- Two proposed estimators:
 - Wiegand and Moloney “O-ring” estimator: # events within $(x, x + dx)$ equivalent to binning obs. to make a histogram
 - kernel smoothing: much better (both for density estimation and $\hat{g}(x)$)
 - What is the histogram of 5,10,11,11, 12, 16?
 - choice of “bin width” really matters
 - see next two slides



Histogram with wide categories



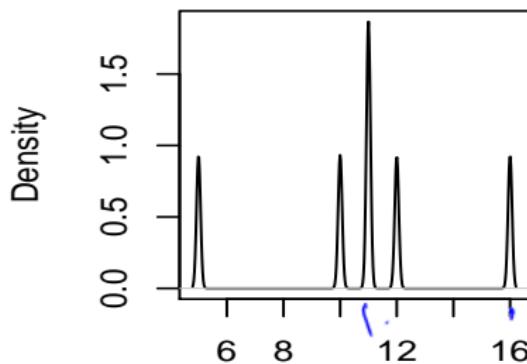
Histogram with narrow categories



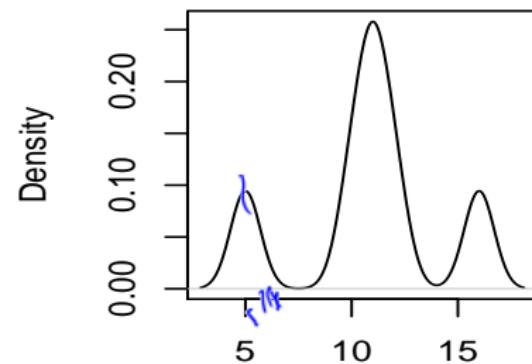
Kernel smoothing

- Histograms estimate probability density for a range of X using only the values in that range
- the variance in the estimated probability depends on the number of obs in the bin
 - wide bin: many points, low variance, but biased estimate (one number for many X values)
 - narrow bin: low bias (small range of X values), but large variance (few obs in bin)
- Density estimation partially avoids this tradeoff and is less dependent on the breaks between categories *each obs \Rightarrow many small bins*
- Concept: superimpose little “bumps” of probability around each obs. Add up the probability to estimate $f(x)$
- result depends on sd of each “bump”
- sd called “bandwidth”
- result also depends on “kernel”, i.e. the shape of the bump
- notice that range of density estimate is wider than data range.

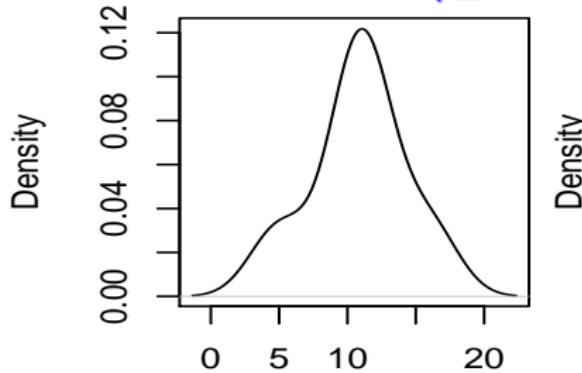




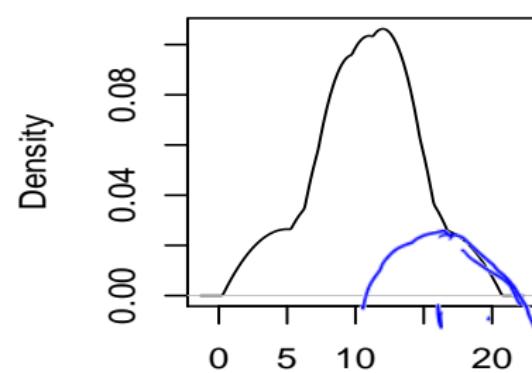
$N = 6$ Bandwidth = 0.07



$N = 6$ Bandwidth = 0.70



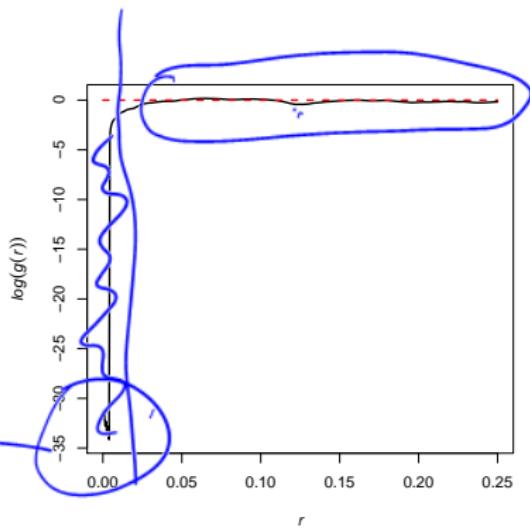
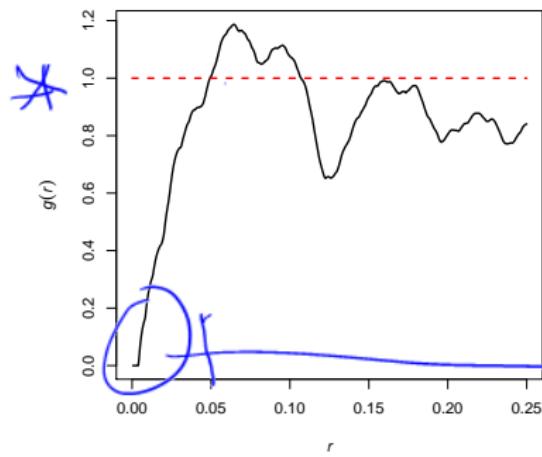
$N = 6$ Bandwidth = 2.11



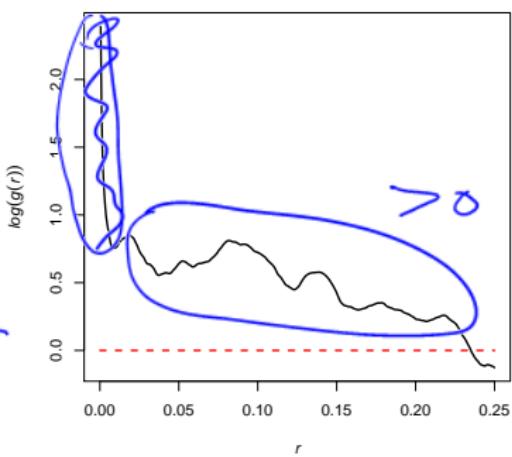
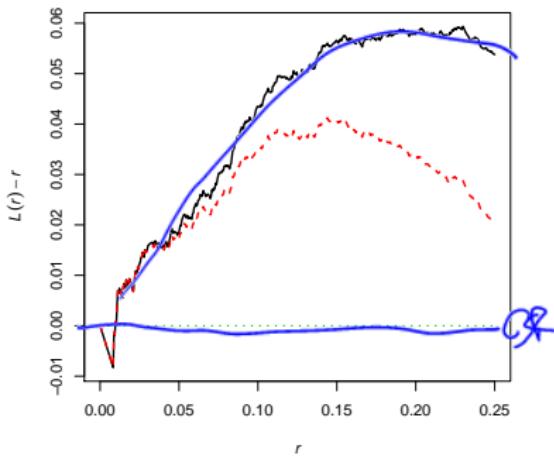
$N = 6$ Bandwidth = 2.11

CSR

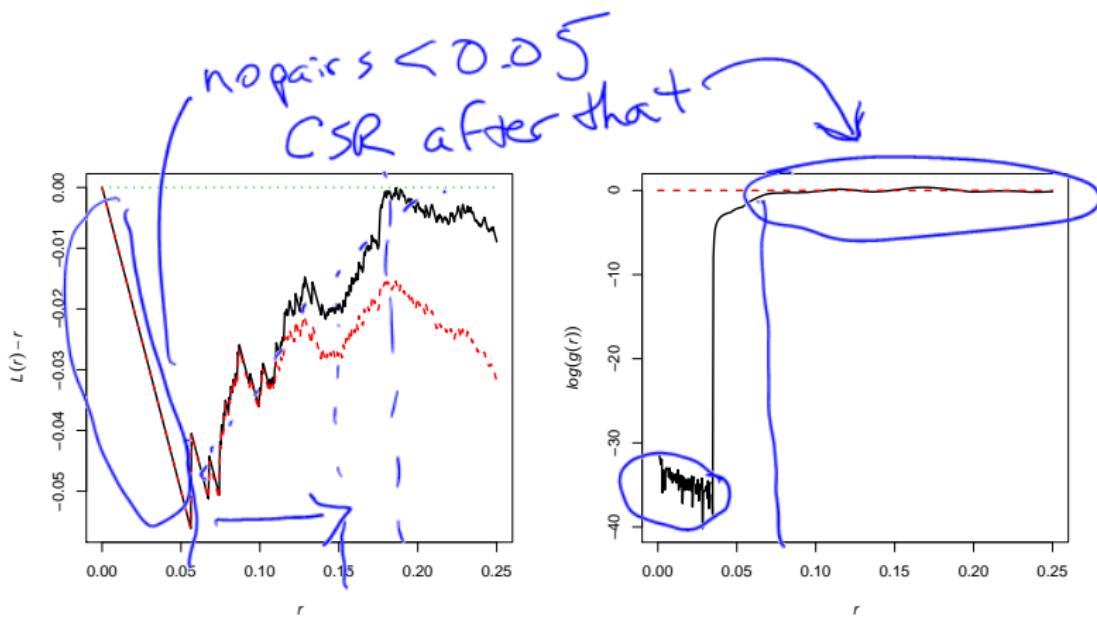
all pairs of distance

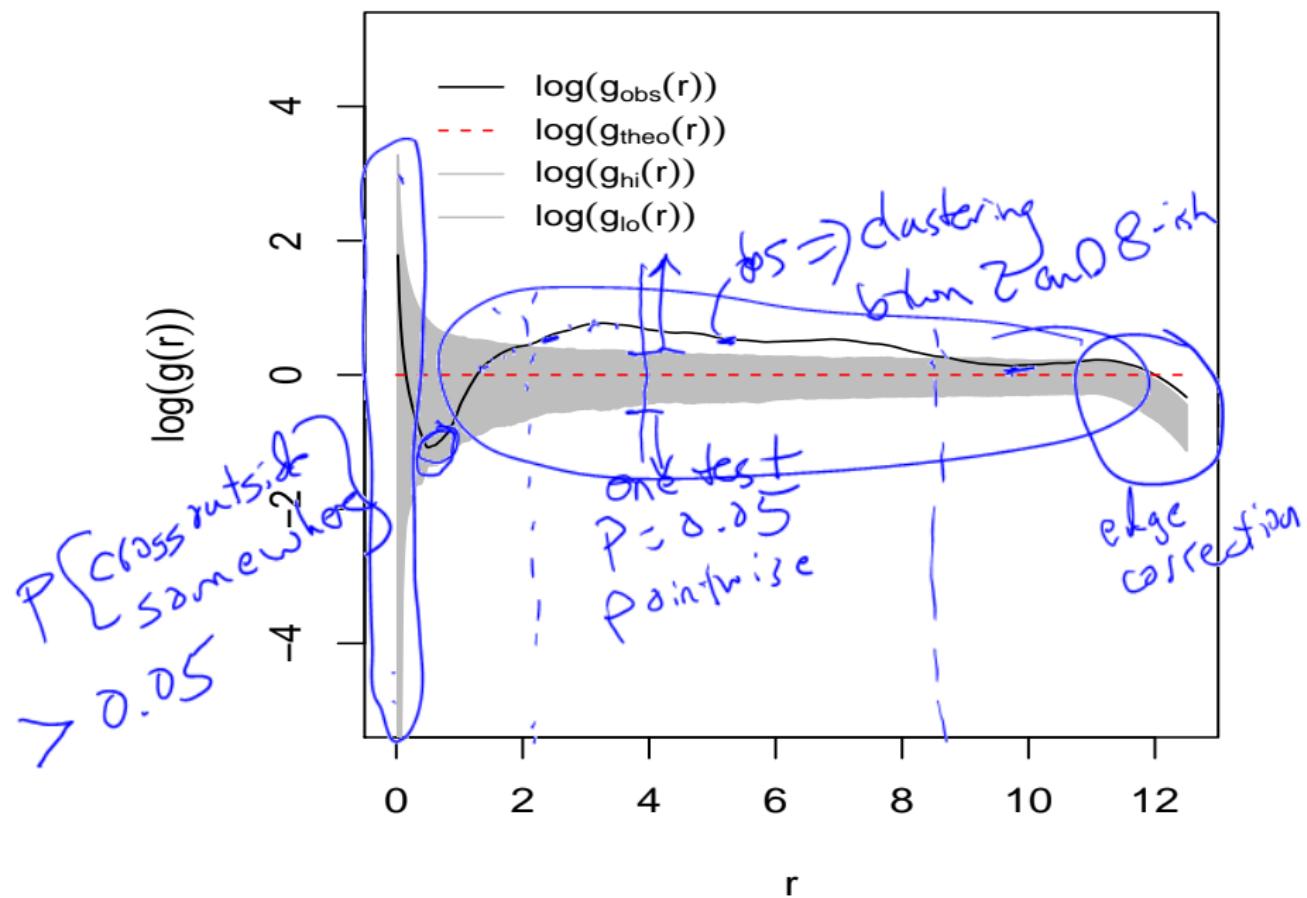


Clustered



inhibition





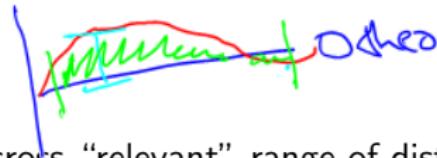
Interpreting / using measures

multiple testing .

- 1) to test CSR: pointwise tests

- estimate $\hat{L}(x)$ at a range of distances, x
- use simulation to calculate point-wise quantiles of $\hat{L}(x)$
- plot $\hat{L}(x)$ and simulation envelope
- interpret deviations above and below expected
- consider distance x_1 , then distance x_2
- called pointwise-tests.
- Type I error rate, α level, correct for one test
- One issue (serious): multiple testing
 - doing many tests, one at each distance
 - $P[\text{reject} \text{ — CSR at any distance}]$ is much larger than $P[\text{reject} \text{ — CSR}]$
 - especially for cumulative summary functions, $K(x)$ and $L(x)$
 - Quite hard to do a true level α test
 - usual approaches don't work well because $\hat{L}(x_1)$ and $\hat{L}(x_2)$ are correlated

Interpreting / using measures



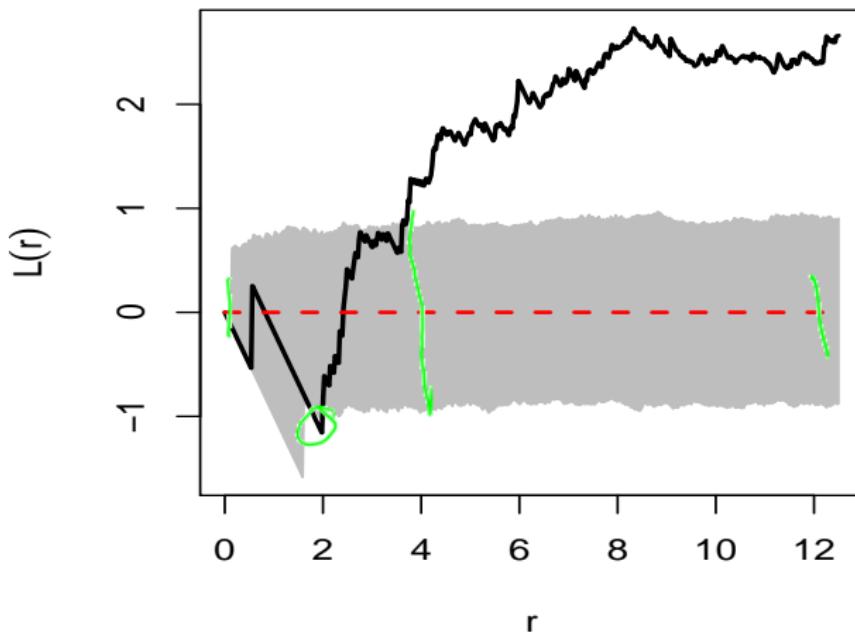
- 2) summary tests of CSR L_2

- Calculate a summary statistic across “relevant” range of distances
- Two common choices, using $L(x)$ as example:
 - $s = \max_x [|\hat{L}(x) - L(x)|]$ (maximum statistic, Maximum Absolute Deviation) $\frac{\text{obs}}{\text{theo}}$
 - $s = \int_x [\hat{L}(x) - L(x)]^2$ (integral statistic, Loosmore and Ford test)
 - Both computed by evaluating “interesting” set of x , finding max or sum
- $L(x)$ can be theoretical expectation ($K(x) = \pi x^2$, $L_2(x) = 0$)
- Or, $L(x)$ computed as average of n simulations (see below)
 - accounts for bias due to edge corrections
- Integral better when consistent but small deviations above expected curve
 - more commonly used
- Max better when large excursion from theoretical value for a small range of distances

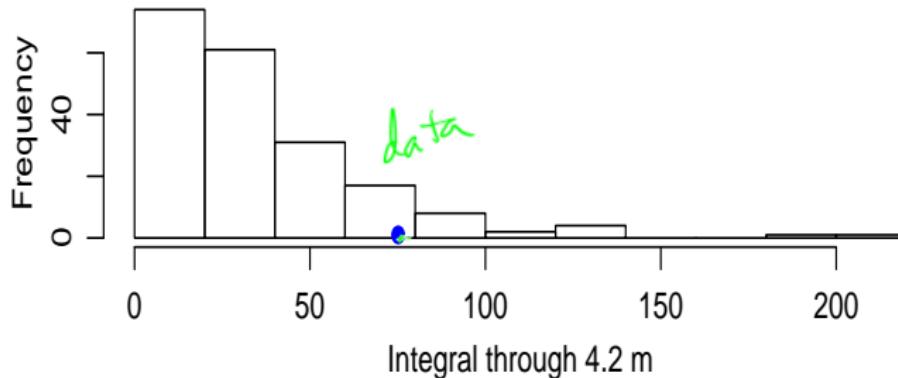
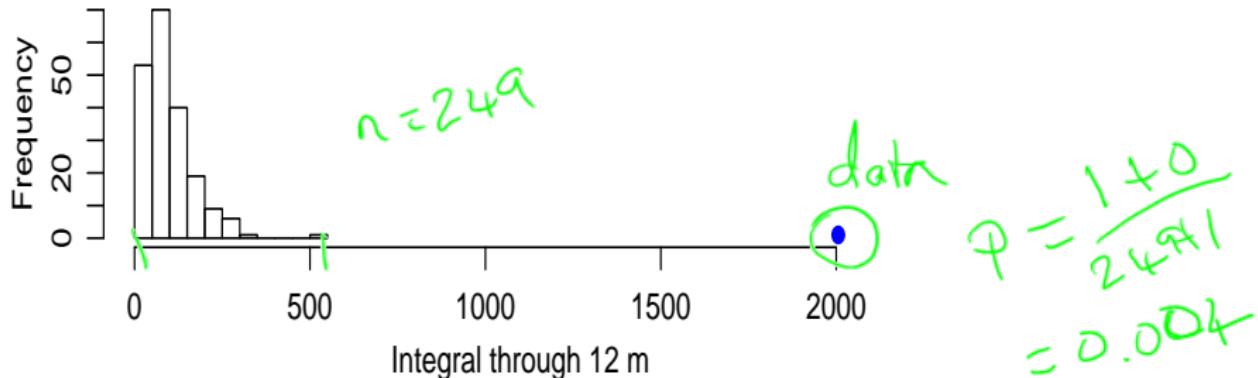
Summary tests

- Turning into a hypothesis test
 - Have s_{obs} from the observed pattern
 - Simulate many (39, 99, 999) random patterns under H_0 (e.g., CSR)
 - Calculate summary statistic for observed data and each simulated data set
 - Calculate $P[\text{as or more extreme summary statistic}] = p\text{-value}$
 - Usually one-sided definition of more extreme (only care about large s)
- This avoids multiple testing issues and gives valid p-value

Summary tests: Cypress pointwise tests

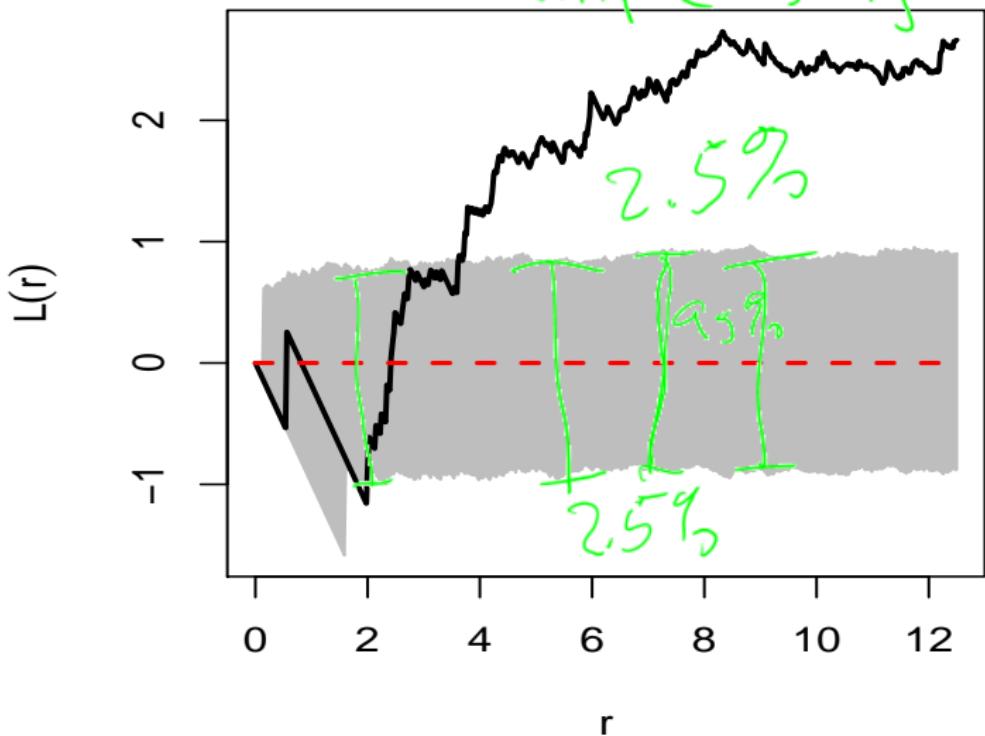


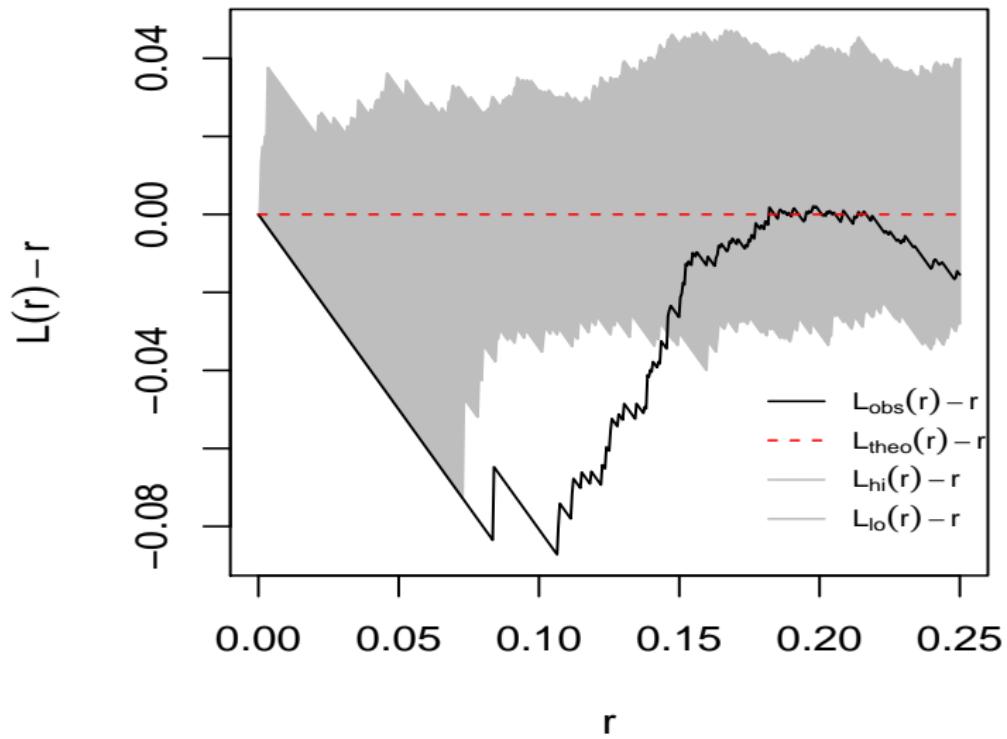
Summary tests: Cypress summary tests



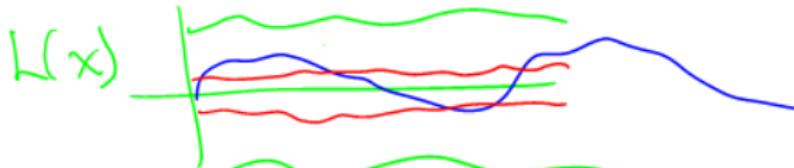
- Need to choose upper and lower distances
 - best when chosen to be biologically relevant.
 - DON'T look for the most significant region
- most commonly used to test CSR
 - But you specify the null hypothesis
 - same approach can be used for any point process model (examples coming soon)
- The integral and especially the maximum statistic assume $\text{Var } \hat{L}(x)$ approximately constant
 - Don't use $K(x)$: because $\text{Var } K(x)$ is definitely not constant
 - Use $L(x)$ instead, approximately constant variance
 - But not perfect (see next two plots)
 - Can transform $G(x)$ or $F(x)$ (both proportions), e.g. $\sin^{-1} \sqrt{G(\hat{x})}$
- There are studentized summary statistics, if unequal Var is bad

any deviation \neq Summary
= multiple testing





A poor use of a test of CSR



- Until a few years ago, it was fashionable to
 - map locations of all things in an area. Usually trees or other plants, could be animal nests
 - usually many species
 - for each species, test CSR (usually using K/L functions) ✓
 - tabulate # species that are clustered, # random, # regular
- then make ecological conclusions about the community
- do you see the issue here? $> 95\%$ $< 95\%$ in 95% level.

no evidence of a diff \neq no diff

accept H_0
depends on sample size

H_0 is true

A poor use of a test of CSR

- If you don't reject H0: CSR for a species, do you know that species is randomly distributed?
- What if you only had 10 individuals for that species?
 - Statistical power to detect 'not-CSR' is really small
- failure to reject H0 does not \Rightarrow H0 is true
- In my experience (mostly with trees)
 - large # events: detect clustering, sometimes regularity not random
 - small # events: accept H0
- If you expect intensity to vary over a study area, that introduces clustering.
- If you believe that non-random spatial patterns are the norm, the hypothesis test is really telling you only whether you have a sufficiently large sample size to detect that non-random pattern.

Interpreting / using measures

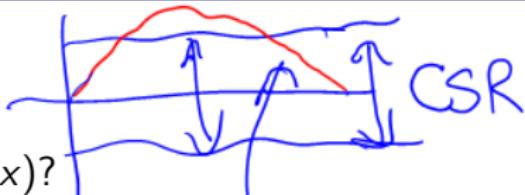
$$\lambda \hat{K}(x)$$

- 3) Estimating # "excess events"
 - if pattern is CSR, expect $\lambda \pi x^2$ events within radius x of another event
 - data says an average of $\lambda \hat{K}(x)$ events within radius x of another event
 - $\lambda \hat{K}(x) - \lambda \pi x^2$ is average "excess events"
 - describes magnitude of clustering in subject-matter terms
 - less frequently used is $\frac{\hat{K}(x)}{\pi x^2} - 1$ $\Rightarrow k = \frac{\hat{K}(x)}{\pi x^2} - 1$ proportion of excess events at distance x
- Cypress tree illustration
 - $\hat{\lambda} = 98 / (50 \times 200) = 0.0098$
 - at distance of 10m, $\hat{K}(10) = 488.6 \Rightarrow$ ave. of 4.7 cypress trees within 10m of another cypress tree
 - $\pi 10^2 = 314.7 \Rightarrow 1.7$ excess cypress trees within 10m of another.
 - Or, $\frac{488.6}{314.7} - 1 = 0.55 \Rightarrow 55\%$ more cypress trees within 10m of another.

Interpreting / using measures

- 4) Describing spatial scale $g(x) = \text{pair correct}$
 - “scale” is a tricky concept. Various definitions
 - Here, scale = distance(s) at which events repulse each other or attract each other
 - A distance-specific concept
 - Many studies have used $\hat{K}(x)$ or $\hat{L}(x)$ to estimate scale, e.g. find x where $L(x)$ is most different from theoretical value
- Increasingly understood to be wrong
 - Both $\hat{K}(x)$ and $\hat{L}(x)$ are cumulative functions: # points within circle of radius x
 - Small # at distance x may be because repulsion (fewer pts.) at distances $< x$, even if strong clustering at x
 - Really want to know what is going on AT distance x , not $\leq x$
- Use pair-correlation function

Interpreting / using measures



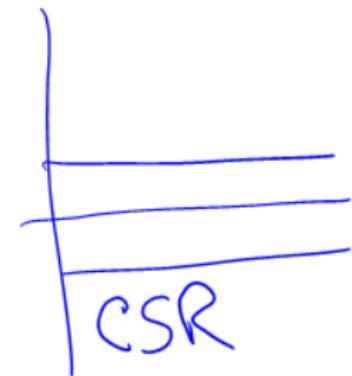
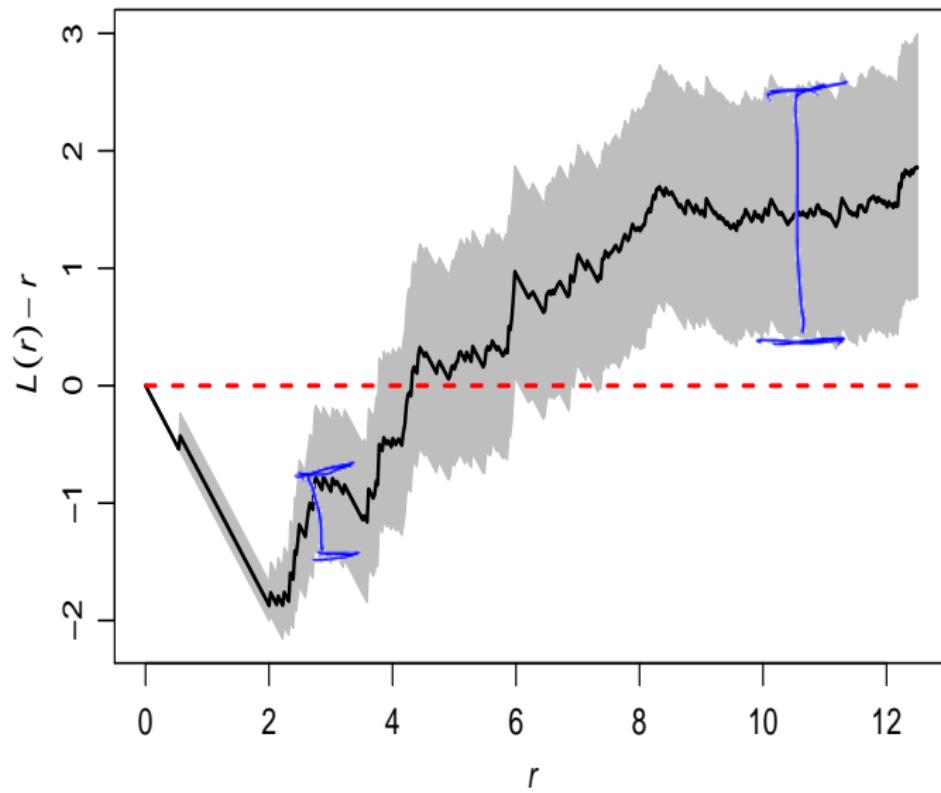
- 5) How precise is $\hat{L}(x)$ or $\hat{g}(x)$?
- Not the width of the Null hypothesis envelopes
- Precision of $\hat{L}(x)$ or $\hat{g}(x)$
 - Certainly depends on $N = \#$ points
 - But also on the spatial pattern
 - $\hat{K}(x)$ more variable for clustered patterns
- If you know the true spatial pattern, simulate from that pattern and calculate envelope *use estimated mode* \Rightarrow data $\Rightarrow c_i$
- If you don't know the true pattern, use a bootstrap \Rightarrow data $\Rightarrow c_i$
- Point pattern bootstrap proposed by Loh, 2008

not CSR
different variance
not simulation
envelope

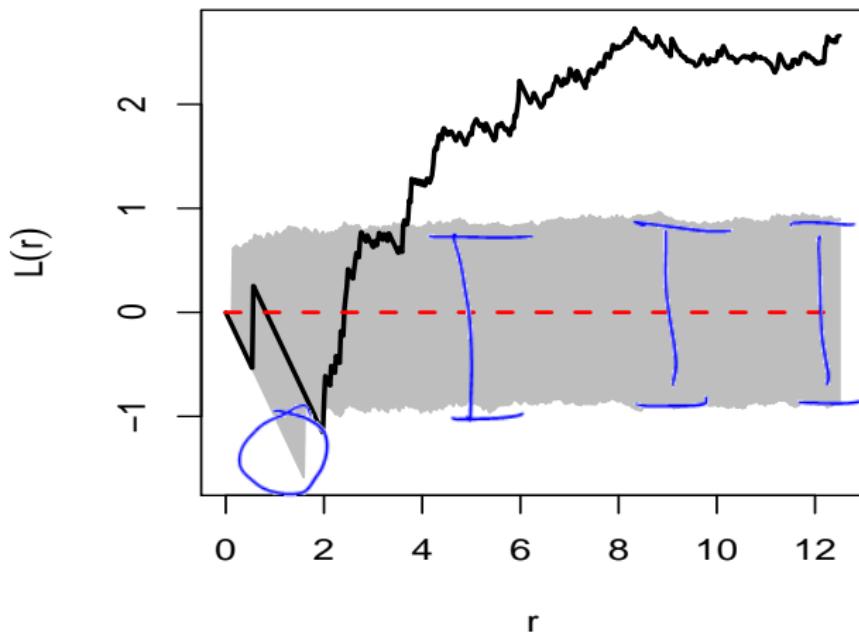
The bootstrap

- General method for estimating precision of a statistic
- Uses resampling the data to approximate the unknown sampling distribution of a statistic ~~X~~
- Gives you the se of a statistic or a confidence interval for a statistic
- CI much more common
- Not the same as a randomization test or a null hypothesis test
 - Hypothesis test: simulate / resample assuming H_0 (CSR, no diff. in means)
 - Bootstrap: simulate / resample assuming H_a (arbitrary pattern, non-zero diff)
- Extremely useful tool for “difficult” problems boot library
- Usual forms of bootstrap don’t work for point pattern data
- Problem is that one point contributes to many $L(x)$
- Loh devised something that (so far) is acceptable sometimes
 - resample contributions to $\hat{g}(x)$ or $\hat{L}(x)$
- Issues when bootstrap average curve not same as data curve (see below)

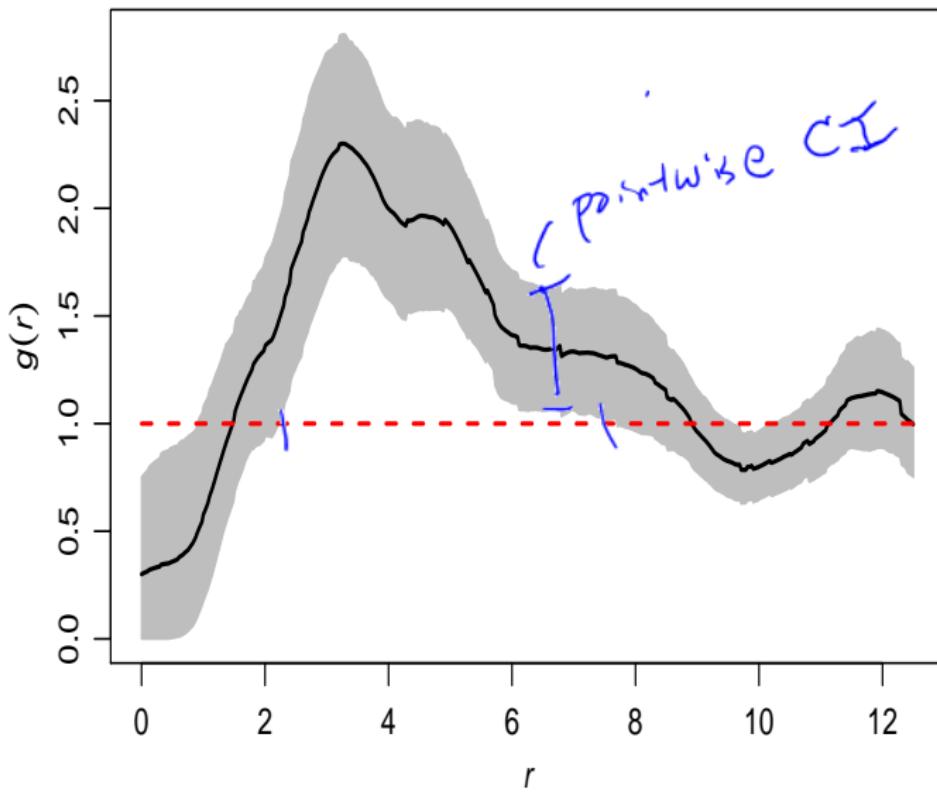
Cypress $L(x)$ bootstrap



Cypress L(x) CSR (null) envelope



Cypress $g(x)$ bootstrap



Summary functions
descriptive
evaluating
clustered
apparently random
inhibition
General issues
multiple test
interpretation
bootstrap \Rightarrow CI

Cypress $g(x)$ CSR (null) envelope

