

**КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ  
УНИВЕРСИТЕТ**

---

**Институт экологии и природопользования**

**ТЕОРИЯ ПРОСТРАНСТВЕННЫХ  
ТОЧЕЧНЫХ ПРОЦЕССОВ  
В ЗАДАЧАХ ЭКОЛОГИИ  
И ПРИРОДОПОЛЬЗОВАНИЯ**

**(с применением пакета R)**

**Учебное пособие**



**КАЗАНЬ  
2014**

**УДК 519.2, 004.43**

**ББК 22.17**

**Т33**

*Печатается по решению Редакционно-издательского совета ФГАОУВПО  
«Казанский (Приволжский) федеральный университет»*

*Методической комиссии Института экологии и природопользования  
(протокол № 10 от 11 ноября 2014 г.)*

*Кафедры моделирования экологических систем  
(протокол № 4 от 27 октября 2014 г.)*

**Авторы-составители:**

доктор биологических наук, профессор **А.А. Савельев**,  
кандидат биологических наук, доцент **С.С. Мухарамова**,  
кандидат биологических наук, доцент **Н.А. Чижикова**,  
старший преподаватель **А.Г. Пилюгин**

**Научный редактор**

доктор физико-математических наук, профессор **Э.В. Скворцов**

**Рецензенты:**

доктор физико-математических наук, профессор **А.М. Елизаров**  
кандидат технических наук, специалист ООО «НекстГИС» **Д.А. Колесов**

**Т33 Теория пространственных точечных процессов в задачах экологии и природопользования (с применением пакета R): учебное пособие / сост.: А.А. Савельев, С.С. Мухарамова, Н.А. Чижикова, А.Г. Пилюгин. – Казань: Изд-во Казан. ун-та, 2014. – 146 с.**

**ISBN 978-5-00019-311-2**

Учебное пособие предназначено для студентов естественнонаучных факультетов, и может быть использовано при проведении курсов, посвященных изучению методов численного описания природных пространственных явлений. В пособии рассматриваются основные понятия теории пространственных точечных процессов, методы анализа точечных образов. Приводятся примеры реализации изучаемых методов с использованием среды статистического программирования R.

**ISBN 978-5-00019-311-2**

**УДК 519.2, 004.43**

**ББК 22.17**

© Издательство Казанского университета, 2014

# СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	5
ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ.....	10
СВОЙСТВА ПРОСТРАНСТВЕННОГО ТОЧЕЧНОГО ПРОЦЕССА... 16	
Свойства первого и второго порядка.....	16
Интенсивность .....	16
Интенсивность второго порядка .....	17
Парная корреляционная функция .....	19
Функции $K(r)$ , $L(r)$ .....	20
Функции $G(r)$ , $F(r)$ , $J(r)$ .....	22
МОДЕЛИ ПРОСТРАНСТВЕННЫХ ТОЧЕЧНЫХ ПРОЦЕССОВ..... 25	
Однородный пуассоновский процесс .....	29
Неоднородный пуассоновский процесс.....	33
Процессы Кокса .....	35
Смешанный пуассоновский процесс .....	36
Кластерные процессы: процессы Неймана-Скотта .....	38
Кластерный процесс Матерна.....	40
Процесс Томаса .....	40
Процессы ингибирования (угнетения).....	42
Процесс Матерна I.....	42
Процесс Матерна II.....	43
Процессы Гиббса, марковские точечные процессы .....	44
Процесс твердого ядра.....	49
Процесс Штрауса .....	51
Другие модели парного взаимодействия .....	53
АНАЛИЗ ТОЧЕЧНЫХ ОБРАЗОВ..... 57	
Установка программы R.....	60
Подготовительный этап .....	61
Визуализация точечного образа .....	65
Оценка и анализ интенсивности.....	69
Метод квадратов.....	70

Ядерная аппроксимация .....	75
Проверка гипотезы о CSR .....	86
Критерий хи-квадрат на основе «квадратов» .....	88
Проверка гипотезы о неоднородном пуассоновском процессе ...	93
Замечание.....	98
Оценка и анализ взаимодействий.....	100
Оценки $\hat{K}(r)$ , $\hat{L}(r)$ , $\hat{g}(r)$ .....	100
Оценки $\hat{G}(r)$ , $\hat{F}(r)$ , $\hat{J}(r)$ .....	107
Проверка гипотез .....	110
Индексы на основе «квадратов».....	110
Индексы расстояний до ближайшего события.....	114
Критерии на основе $\hat{G}(r)$ , $\hat{F}(r)$ , $\hat{J}(r)$ . Метод «конверта».....	118
Критерии на основе $\hat{K}(r)$ , $\hat{L}(r)$ , $\hat{g}(r)$ .....	126
ЛИТЕРАТУРА.....	139
ПРИЛОЖЕНИЕ.....	145

## ВВЕДЕНИЕ

Данные с пространственной привязкой (в виде координат) часто используются в экологии, лесоведении, геологии и других науках о Земле, особенно в настоящее время, когда точные географические координаты можно определять с помощью устройств системы глобального геопозиционирования (Global Positioning System, GPS). Примерами таких данных являются местоположения отдельных деревьев в лесном массиве, координаты птичьих гнезд, координаты вспышек заболеваний, эпицентров землетрясений, кратеров вулканов, гейзеров, местоположения археологических находок и т.д.

Общим для всех этих данных является то, что число и размещение зафиксированных исследователем событий или объектов можно рассматривать как результат действия некоторых стохастических (случайных, вероятностных) механизмов. Например, для растений такими стохастическими механизмами, определяющими их появление и существование, могут считаться диссеминация (рассеяние семян), конкуренция, лесные пожары. Термин стохастический в контексте изучения пространственного размещения объектов означает невозможность предугадать точно, сколько именно объектов встретится на исследуемой площади и каковы будут координаты этих объектов – всё это случайные величины.

Цель изучения наблюдаемого размещения объектов или появления событий заключается в определении свойств пространственного точечного процесса, формирующего исследуемый пространственный рисунок. Вот примеры традиционных вопросов:

- Существует ли тенденция к формированию систематического рисунка в размещении объектов на изучаемой территории? Или пространственный рисунок абсолютно случаен, хаотичен?
- Можно ли утверждать, что для пространственного рисунка характерно наличие агрегаций (скоплений) объектов? Или, наоборот, размещение объектов показывает наличие регулярности (разрежений)? Каков масштаб наблюдаемых неоднородностей?
- Сколько в среднем объектов приходится на единицу площади или, выражаясь математическим языком – какова интенсивность точечного процесса?
- Постоянна ли интенсивность точечного процесса или ее величина варьирует в пространстве?
- Существует ли пространственная зависимость между событиями? Связано ли появление одних объектов с наличием других? Проявляется ли эта зависимость как фасилитация или как ингибирование?
- Какая модель адекватно описывает наблюдаемый точечный рисунок?

Помимо пространственных координат событий или объектов, собранные исследователем данные могут содержать дополнительную информацию, такую, как диаметр ствола дерева, число яиц в гнезде, степень тяжести заболевания и т.п. В этом случае исследователя может интересовать вопрос, влияет ли взаиморасположение деревьев / гнезд / инфекционных заболеваний на диаметр стволов / число яиц / степень тяжести заболевания?

С позиции пространственных зависимостей может быть исследована межвидовая и внутривидовая конкуренция, изучаемая экологией растений и животных. Много примеров изучения пространственных взаимодействий в медицине, особенно в эпидемиологии. Так, при наличии пространственной информации об инфекционных заболеваниях исследуется зависимость случаев появления инфекционных заболеваний от других пространственно распределенных событий или объектов, таких, как места размножения и агрегаций насекомых. Можно привести аналогичные примеры из геологии, когда изучаются тенденции появления событий или объектов вдоль некоторых линий, например, возникновение землетрясений и гейзеров в зоне разломов или вдоль линеаментов.

Размещение изучаемых событий и объектов в пространстве абстрактно может быть представлено в виде пространственного размещения точек, которое называется точечным образом (паттерном, рисунком). *В данном пособии представлены методы, предназначенные для анализа точечных образов и моделирования пространственных точечных процессов, формирующих эти образы.*

**ВАЖНО:** не каждые пространственные данные могут быть изучены методами анализа точечных образов. Эти методы применяются при изучении процессов, создающих размещение дискретных объектов или событий. Измерения свойств таких объектов могут быть получены только в некоторых точках, которые предопределены самим процессом. Например, диаметр дерева может быть измерен только в том месте, где есть дерево, число яиц может быть подсчитано только там, где имеется

гнездо, но не в любом местоположении по желанию исследователя. Только данные такой природы могут исследоваться методами анализа точечных образов.

В этой связи следует понимать различие между пространственным точечным процессом и пространственной случайной функцией (или случайным полем). Случайное поле – это модель явления, непрерывно распределенного в пространстве, например – высота рельефа, поле температур, содержание химических элементов в почве. При этом, хотя явление пространственно непрерывно, полевые данные по его изучению обычно представлены в виде набора точек, регулярно или случайно размещенных в пространстве. Примером таких данных могут быть точки, где была измерена высота рельефа, температура, определена концентрация химического элемента в почве. Число этих точек и их расположение определяется человеком. Исследователь сам выбирает местоположения, где заложить тестовые площадки и измерить высоту рельефа, температуру, отобрать пробы (рис. 1). Данные, полученные таким способом, отражают закономерности явлений, непрерывно распределенных в пространстве, и изучаются методами геостатистики, но не методами анализа точечных образов и точечных процессов.

Тем не менее, методы анализа точечных образов могут быть полезны и при изучении случайных полей, а именно на этапе планирования отбора проб или анализа пространственного дизайна имеющейся выборки. Для некоторых видов обработки данных и статистических выводов о закономерностях явлений, непрерывно распределенных на изучаемой территории, требуется полная случайность расположения в пространстве точек

измерения. Поэтому необходимо не только уметь тестировать такую случайность и проверять, что отсутствуют значительные сгущения пробных площадок (кластеризация), но также и уметь генерировать случайное размещение точек для планирования мест отбора.

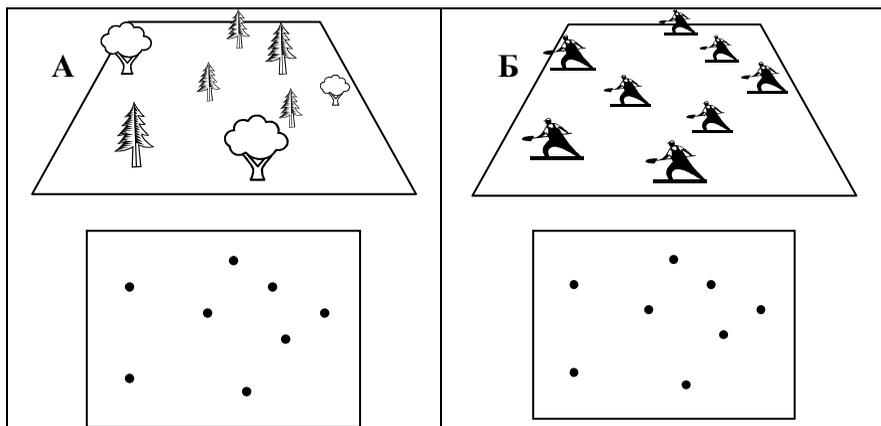


Рис. 1. А: Местоположения дискретных объектов определяются точечным процессом; в ходе обследования установлены координаты всех объектов на изучаемой территории. В этом случае для исследования наблюдаемого размещения объектов могут быть использованы методы анализа точечных образов.

Б: Данные получены в результате отбора проб в точках пространства, выбранных исследователем. Полученный рисунок точек выборки может быть изучен методами анализа точечных образов, но свойства исследуемого явления должны изучаться методами геостатистики.

## ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

В теории пространственных точечных процессов наличие объекта в некоторой точке исследуемой области пространства называется *событием*. Каждое событие можно рассматривать как точку в пространстве. События могут представлять деревья, гнезда, эпицентры землетрясений, очаги возникновения новых случаев гриппа и т.д.

Если набор данных отражает информацию о размещении в пространстве площадных объектов, например, лесных или ландшафтных выделов, окон возобновления в лесном покрове, то часто и эти данные могут быть представлены точками, как если бы площадные объекты были рассмотрены в более генерализованном масштабе (например, в масштабе всего ландшафта)<sup>1</sup>.

События могут располагаться на плоскости (двумерное пространство, 2D), на земной поверхности, в 3D объеме. Это могут быть точки в пространственно-временных осях, например, координаты и время возникновения эпицентра землетрясения. Однако в данном пособии не будут затрагиваться временная составляющая точечных процессов или трехмерные случаи, - ограничимся двумерным пространством (местоположения точек

---

<sup>1</sup> Заметим, что понятия мелкого и крупного масштаба в биологии часто трактуются иначе, чем в географии. С биологической точки зрения: чем меньше размер и больше детальность изучаемых вариаций, тем меньше масштаб, в котором они наблюдаются, и наоборот, большие, протяженные элементы структуры называют крупномасштабными. Географическая же точка зрения на понятия мелкого и крупного масштаба обратна той, что принята в биологии.

фиксируются в двумерной системе координат). Обозначим через  $S$  область исследований,  $S \subseteq \mathbf{R}^2$ ,  $\mathbf{R}^2$  - двумерное евклидовое пространство (плоскость).

*Пространственный точечный процесс  $\mathbf{X}$*  в области  $S$  – это случайное счетное подмножество событий (точек) в области  $S$ . Число событий и их местонахождения случайны и зависят от свойств процесса. Это понятие используется как математическая (вероятностная, стохастическая) модель, представляющая реально существующий пространственный феномен, который рассматривается как случайный.

*Точечный образ  $\mathbf{x}$*  (часто его называют *точечный паттерн*, point pattern) – это реализация точечного процесса или, другими словами, набор событий  $(x_1, x_2, \dots, x_n)$ , зафиксированных в исследуемой области  $S$ . Обозначение  $x_i$  соответствует местоположению (точке)  $i$ -ого события и содержит информацию о его координатах:  $x_i = (X_i, Y_i)$ . То есть  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  – фиксированная комбинация точек, где  $x_i \in S$ ,  $i=1, \dots, n$ .

В данном пособии мы всегда будем предполагать, что область  $S$  ограничена, а любая реализация процесса  $\mathbf{X}$  конечна. Такой точечный процесс называется *конечным*.

Иногда исследователи обладают дополнительной информацией, относящейся к каждому событию. В этом случае набор событий описывается *маркированным* точечным образом. Он представляется в виде набора векторов  $\{(X_i, Y_i; \mathbf{m}_i)\}$ , определяющих местоположение  $x_i = (X_i, Y_i)$  в области наблюдения и *метку*  $\mathbf{m}_i$  события. Например, в исследованиях лесных экосистем местоположение  $x_i$  может представлять координаты ствола дерева, а метки могут характеризовать вид, пол, возрастную группу, степень поражения дерева вредителем.

Метки часто бывают категориальными (номинальными, рис. 2 А), ординальными, но могут быть и скалярными (например, возраст или размер объекта, рис. 2 Б). Метки могут быть многомерными, например, при изучении деревьев часто указывается и диаметр, и порода.

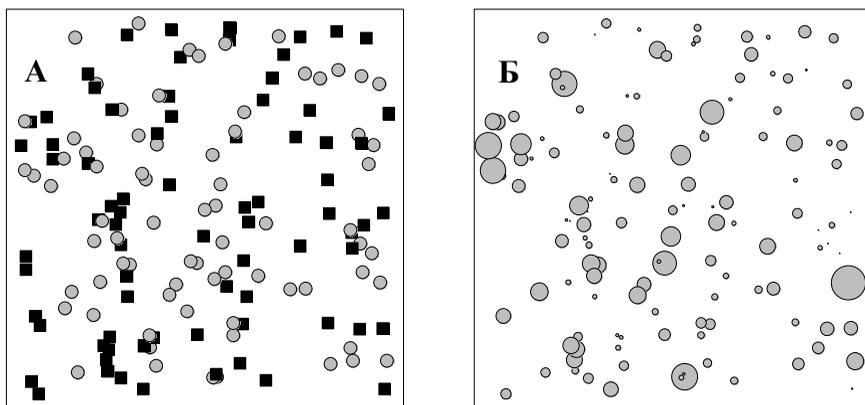


Рис. 2. Иллюстрация маркированного точечного образа.

А: Точечный образ размещения двух пород деревьев, «метка» является категориальной переменной, на рисунке показана разными значками в соответствии с породой дерева;

Б: Точечный образ размещения деревьев, «метка» является скалярной переменной, показана размером значка в соответствии с диаметром ствола.

Необходимо понимать отличие точечных процессов и точечных образов. Процесс – это теоретическая стохастическая модель (число событий и их координаты являются случайными переменными, неизвестно точно, где и сколько появится событий), тогда как наблюдаемый точечный образ – это одна из возможных реализаций процесса (каждая случайная переменная приняла одно конкретное значение: получено конкретное число событий и координаты их известны).

При изучении точечного образа предполагается, что в области  $S$  существует некий пространственный точечный процесс  $X$ , а наблюдаемый точечный образ  $x$  является одной из его реализаций.

Так, в случае с деревьями модель точечного процесса описывает природные явления, порождающие размещение деревьев - рассеяние семян, конкурентные взаимоотношения между деревьями. В случае с гнездовьями птиц – это явления миграции и размножения птиц в природе и т.д. Реализацией точечного процесса, точечным паттерном в данных примерах будет карта деревьев реально существующего леса или карта гнездовых. Число событий (деревьев, гнездовых) и их местоположения зависят от свойств точечного процесса.

Каким же образом эти свойства могут быть описаны, или, другими словами, как может быть охарактеризовано вероятностное поведение пространственного точечного процесса?

Рассмотрим случайную величину  $N_X(B)$  – число событий, порождаемых процессом  $X$  в произвольной подобласти  $B$ , принадлежащей области  $S$  ( $B \subseteq S$ ) (рис. 3). Вероятностное поведение пространственного точечного процесса (его *закон распределения*) описывается множеством многомерных функций распределения наборов случайных величин  $\{N_X(B_1), \dots, N_X(B_k)\}$  для любого положительного целого  $k$  и для любой пространственной конфигурации  $B_1, \dots, B_k$ , где все  $B_i \subseteq S$ .

# X

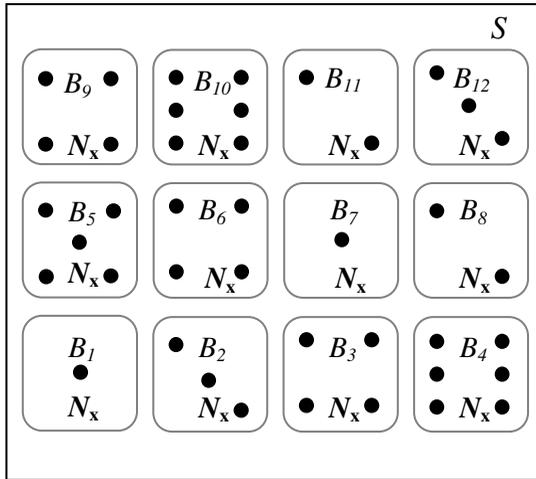


Рис. 3. Приведенные на рисунке числа событий  $N_x$  в каждой подобласти  $B \subseteq S$  являются реализациями случайных величин  $N_x(B)$ .

В отличие от классической статистики, для изучаемого пространственного феномена невозможно получить повторные реализации - наблюдаемый точечный образ уникален. Действительность уникальна, и изучение должно быть проведено на основе одной единственной реализации. Это важное ограничение диктует необходимость стационарности и эргодичности пространственного точечного процесса. Пространственный процесс *стационарен* (в строгом смысле), если его закон распределения инвариантен относительно сдвига. То есть все свойства процесса одинаковы во всех подобластях изучаемого пространства. Если закон распределения стационарного процесса инвариантен и относительно поворота, то такой процесс называется *изотропным*. Пространственный процесс называется *эргодичным*, если среднее по всем

возможным реализациям равно среднему отдельной безграничной в пространстве реализации (Cressie, 1993), которое на практике заменяется средним по исследуемой области.

Стационарность и эргодичность случайного точечного процесса позволяют оценивать его свойства и получать достоверные статистические выводы на основе анализа наблюдаемого точечного образа.

На практике, прежде всего, в силу ограниченности объема данных трудно сделать надежные статистические выводы о законе распределения точечного процесса. Поэтому обычно ограничиваются описанием поведения исследуемого процесса в терминах его *свойств первого порядка* (изменение средних в пространстве) и *свойств второго порядка* (пространственная корреляция процесса).

# СВОЙСТВА ПРОСТРАНСТВЕННОГО ТОЧЕЧНОГО ПРОЦЕССА

## *Свойства первого и второго порядка*

Начальный курс статистики вводит понятия моментов первого (математическое ожидание) и второго порядков (дисперсия, ковариация). В пространственной статистике также рассматривают свойства первого и второго порядка. Свойства первого порядка связаны с ожидаемым числом событий  $E[N_X(B)]$ , где  $N_X(B)$  – случайная величина, число событий в произвольной подобласти  $B$ . Свойства второго порядка описываются ковариацией случайных величин:  $COV[N_X(B_1), N_X(B_2)]$ , и показывают, связано ли число  $N_X(B_1)$  событий в области  $B_1$  с числом  $N_X(B_2)$  событий в области  $B_2$ . Так как математическое ожидание числа событий и ковариация очевидно зависят от размера подобластей  $B$ , свойства первого и второго порядка точечного процесса характеризуют на основе предельного поведения этих величин, отнесенных к единице площади.

## *Интенсивность*

Свойства первого порядка описываются *интенсивностью*  $\lambda(x)$ , которая определена как ожидаемое число событий процесса  $X$  в окрестности точки  $x \in S$ :

$$\lambda(x) = \lim_{A_{dx} \rightarrow 0} \frac{E[N_X(dx)]}{A_{dx}}.$$

Здесь  $dx$  – это бесконечно малая окрестность точки  $x$ ,  $A_{dx}$  – площадь этой окрестности,  $E[\cdot]$  – математическое ожидание,  $N_X(dx)$  – случайное число событий в  $dx$ .

Интенсивность может быть константой или изменяться в пространстве. Если интенсивность константа, то есть ее значение не зависит от положения точки  $x$  и постоянно во всей изучаемой области  $S$ :  $\lambda(x) \equiv \lambda$ , тогда пространственный точечный процесс  $X$  называется *однородным* (homogeneous), иначе – *неоднородным* (inhomogeneous) (рис. 4). Однородность процесса - это один из типов стационарности, указывающий на постоянство в пространстве свойств первого порядка.

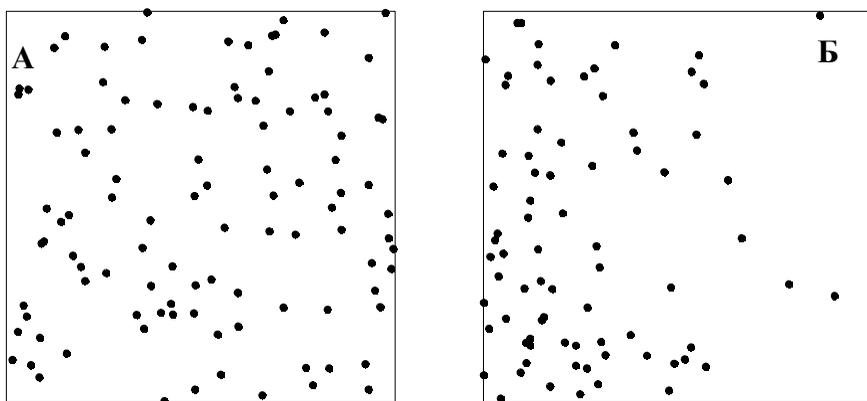


Рис. 4. А: Реализация однородного пространственного точечного процесса, интенсивность постоянна в изучаемой области. Б: Реализация неоднородного пространственного точечного процесса, интенсивность варьирует в изучаемой области.

### ***Интенсивность второго порядка***

Свойства второго порядка отражают пространственные зависимости процесса, взаимодействие между событиями в области  $S$ . Одной из характеристик этих свойств является *интенсивность второго порядка*  $\lambda_2(u, v)$ :

$$\lambda_2(u, v) = \lim_{A_{du}, A_{dv} \rightarrow 0} \frac{E[N_X(du)N_X(dv)]}{A_{du}A_{dv}}.$$

Здесь  $du, dv$  – это бесконечно малые окрестности точек  $u, v \in S$  (принадлежащих области  $S$ ), а  $A_{du}, A_{dv}$  – площади этих окрестностей. Интенсивность второго порядка характеризует возможность встретить пару событий - в окрестности точки  $u$  и в окрестности точки  $v$  (поскольку окрестности точек настолько малы, что  $N_X$  может принимать значение либо 0, либо 1).

Интенсивность второго порядка подобна коэффициенту ковариации между двумя переменными, используемому в многомерном анализе данных, хотя в отличие от коэффициента ковариации она не может принимать отрицательные значения. Коэффициент ковариации характеризует связь двух случайных величин, а интенсивность второго порядка отражает взаимодействие событий в окрестностях двух точек пространства.

Пространственное парное взаимодействие событий может быть позитивным, негативным и нейтральным. Взаимодействие будет нейтральным (взаимодействие отсутствует), если события возникают независимо друг от друга. В этом случае  $\lambda_2(u, v) = \lambda(u) \cdot \lambda(v)$ . Взаимодействие позитивно, если события в окрестностях точек  $u$  и  $v$  чаще возникают одновременно. Тогда интенсивность второго порядка  $\lambda_2(u, v)$  будет принимать высокие значения. Взаимодействие негативно, если при возникновении события в окрестности одной точки событие в окрестности второй точки чаще не происходит. Тогда значение  $\lambda_2(u, v)$  будет близко к нулю.

Пространственный точечный процесс обладает *стационарностью второго порядка*, если выполняются следующие условия:

- интенсивность процесса постоянна во всей изучаемой области  $S$ :  $\lambda(x) \equiv \lambda, x \in S$ ;
- интенсивность второго порядка  $\lambda_2(u, v)$  не зависит от местоположения точек  $u$  и  $v$ , а зависит только от длины и направления вектора  $r$ , разделяющего точки  $u$  и  $v$ :  $\lambda_2(u, v) = \lambda_2(u - v) = \lambda_2(r)$ .

Если  $\lambda_2(u, v)$  зависит только от длины разделяющего вектора  $r$  и не зависит от его направления -  $\lambda_2(u, v) = \lambda_2(\|u - v\|) = \lambda_2(\|r\|) = \lambda_2(r)$ , то свойства второго порядка этого процесса будут не только стационарны, но и *изотропны*. В противном случае процесс будет анизотропным. На рис. 5 представлена иллюстрация анизотропного процесса, для которого свойства процесса зависят от направления в пространстве.

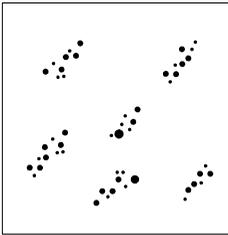


Рис. 5. Анизотропия пространственного процесса, в некотором направлении возможность соседства особей выше.

### ***Парная корреляционная функция***

Помимо  $\lambda_2(u, v)$  существуют и другие функции, характеризующие пространственные зависимости точечного процесса. Так, *парная корреляционная функция*  $g(u, v)$  – это отношение интенсивности второго порядка  $\lambda_2(u, v)$  к произведению интенсивностей в точках  $u$  и  $v$ :

$$g(u, v) = \frac{\lambda_2(u, v)}{\lambda(u)\lambda(v)},$$

(аналогом в многомерной статистике является переход от коэффициента ковариации к коэффициенту корреляции с помощью деления ковариации на дисперсии переменных). Если взаимодействие отсутствует, то  $g(u, v) = 1$  (в окрестностях точек  $u$  и  $v$  события возникают независимо друг от друга). Если  $g(u, v) > 1$ , то это указывает на позитивное взаимодействие: в местоположениях  $u$  и  $v$  события более вероятно произойдут одновременно. А если  $g(u, v) < 1$ , то наоборот, имеет место негативное взаимодействие: появление события в одном из местоположений  $u$  или  $v$  более вероятно приведет к тому, что во втором местоположении события не произойдет.

Если процесс стационарен, то функция  $g$  инвариантна относительно сдвига:  $g(u, v) = g(u - v) = g(r)$ . Для изотропного процесса значение  $g$  зависит только от длины разделяющего вектора  $r$  между точками  $u$  и  $v$ :  $g(u, v) = g(r)$ . Заметим, что возможны случаи, когда  $g$  инвариантна относительно сдвига и поворота, но стационарности нет – процесс неоднородный (интенсивность непостоянна).

### **Функции $K(r)$ , $L(r)$**

Для стационарного изотропного процесса Рипли (1976) показал, что свойства второго порядка могут быть описаны *K-функцией*, иначе называемой *«редуцированной функцией второго момента»* (reduced second moment function). Значение этой функции пропорционально ожиданию числа событий  $N_X$  в произвольной подобласти  $B_r$ , имеющей форму круга радиуса  $r$ :

$$K(r) = \frac{E[N_X(B_r)]}{\lambda},$$

здесь  $\lambda$  – интенсивность процесса.

Для неоднородных процессов (но со стационарной парной корреляционной функцией  $g$ , инвариантной относительно сдвига)  $K$ -функция была обобщена как:

$$K(r) = \frac{1}{A_S} E \left[ \sum_u \sum_{v \neq u} \frac{I(\|u-v\| < r)}{\lambda(u)\lambda(v)} \right],$$

где  $A_S$  – площадь области  $S$ ,  $I(\cdot)$  - индикаторная функция (она равна 1, если расстояние между событиями  $u$  и  $v$  меньше  $r$ , иначе – равна 0).

Бесаг (1977) предложил анализировать измененную версию  $K$ -функции, которая обладает более удобными свойствами. Часто используют ее модифицированное определение:

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

Если парная корреляционная функция  $g$  изотропна (инвариантна относительно поворота), то она связана с  $K$ -функцией следующими соотношениями:

$$K(r) = \int_0^r g(t) 2\pi t dt \quad \text{и} \quad g(r) = \frac{K'(r)}{2\pi r},$$

здесь  $2\pi r$  – это длина окружности с радиусом  $r$  (производная от площади круга с радиусом  $r$ ).

Недостатком функций  $K(r)$  и  $L(r)$  является их кумулятивность: значение функции для расстояния  $r$  зависит и от ее поведения на расстояниях, меньших  $r$ . В отличие от них парная корреляционная функция  $g(r)$  более проста в интерпретации, хотя для нее сложнее получить оценку. В общем случае, по крайней мере, для небольших  $r$ , значение  $g(r) > 1$  указывает на позитивное взаимодействие событий, находящихся

друг от друга на расстоянии  $r$ . Тогда в порождаемых процессом точечных образах будут наблюдаться *агрегации* (кластеры, скопления) событий, разделенных таким расстоянием. Значение  $g(r) < 1$  указывает на негативное взаимодействие событий на этом расстоянии, что порождает *регулярность* (разреженность с расстоянием  $r$ ) в их размещении. Такие типы поведения могут быть обусловлены некими латентными механизмами, вызывающими либо притяжение, либо отталкивание точек. В свою очередь, по крайней мере, для небольших  $r$ , значение  $L(r) > 0$  говорит о наличии агрегаций точек, находящихся друг от друга на расстояниях, меньших  $r$ , а  $L(r) < 0$  – о регулярности с расстояниями, меньшими  $r$ .

### **Функции $G(r)$ , $F(r)$ , $J(r)$**

Рассмотренные выше функции характеризуют свойства точечного процесса в различных масштабах, тогда как для получения информации о свойствах второго порядка в *локальном* масштабе для стационарного изотропного процесса могут быть использованы функции, основанные на распределении расстояний до ближайшего события. Одна из таких функций – кумулятивная функция распределения расстояний до ближайшего события  $G(r)$  (*nearest neighbour event-event distance distribution function*). Обозначим  $W$  - расстояние между произвольным событием и событием, являющимся его ближайшим соседом (рис. 6 А), тогда функция  $G(r)$  равна вероятности встречи ближайшего события на расстоянии, не более  $r$  от произвольного события:

$$G(r) = P(W < r).$$

К примеру, значение  $G(20)$  – это вероятность того, что на расстоянии менее 20 единиц от любого события можно встретить еще одно событие, которое будет его ближайшим соседом. Так, значение  $G(20) = 0$  означает, что практически ни одно событие не имеет соседа в радиусе 20.

Можно также определить расстояние  $U$  от любой случайно выбранной точки пространства до ближайшего события (рис. 6 Б). Функция распределения таких расстояний обозначается через  $F(r)$  (*nearest neighbour point-event distance distribution function; empty space function*):

$$F(r) = P(U < r).$$

Она может быть полезной при исследовании размеров пустот (окон) в точечных образах. Например,  $F(20)$  - это вероятность встретить событие на расстоянии менее 20 единиц от случайно выбранной точки в пространстве.

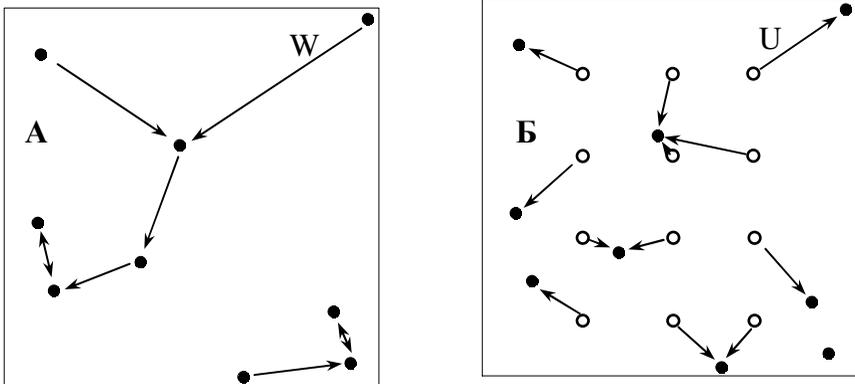


Рис. 6. Расстояния до ближайшего события. Закрашенные значки обозначают события, полые значки соответствуют местоположениям, не содержащим событий.

А: Для функции  $G(r)$ ,  $W$  – расстояние от события до ближайшего события.

Б: Для функции  $F(r)$ ,  $U$  – расстояние от случайно выбранной точки, не содержащей событие, до ближайшего события.

Чтобы численно сравнивать распределения  $G(r)$  и  $F(r)$ , была предложена специальная функция  $J(r)$ :

$$J(r) = \frac{1 - G(r)}{1 - F(r)} \text{ для } F(r) < 1.$$

Более подробно об этих функциях можно прочитать в работе Cressie (1993). В общем случае, для небольших  $r$  соотношение  $G(r) > F(r)$  (или  $J(r) < 1$ ) указывает на агрегацию событий - события будут чаще встречаться вокруг других событий, чем вокруг случайно выбранной точки пространства. Если  $G(r) < F(r)$  (или  $J(r) > 1$ ), то это указывает на регулярность в локальном масштабе - между событиями существует некоторая «непреодолимая» дистанция.

## МОДЕЛИ ПРОСТРАНСТВЕННЫХ ТОЧЕЧНЫХ ПРОЦЕССОВ

Ключевой вопрос изучения наблюдаемого пространственного размещения объектов или событий состоит в определении его пространственной структуры, а именно пространственных вероятностных закономерностей стохастического процесса, формирующего исследуемый рисунок. Полное отсутствие пространственной структуры называют полной пространственной случайностью (*complete spatial randomness, CSR*). Это понятие близко к термину «белый шум» в теории временных рядов и теории обработки сигналов. Такой процесс является однородным и независимым, он предполагает, что вероятность появления события в любом местоположении (location) области  $S$  одинакова, и события возникают независимо друг от друга (нет взаимодействий, no interaction). Простейшей математической моделью полностью случайного пространственного процесса является *однородный пуассоновский процесс* (homogeneous Poisson process).

Отклонение от полной пространственной случайности может проявляться либо наличием *агрегаций* (кластеров, скоплений событий в каких-либо подобластях изучаемой области), либо наличием некоторой пространственной *регулярности* в размещении событий (события «отталкиваются» друг от друга; крайний случай – полная регулярность, регулярная сетка) (рис.7).

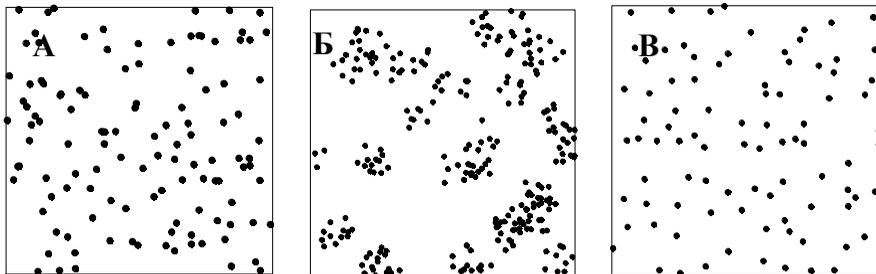


Рис. 7. Основные типы точечных процессов, примеры реализаций.  
 А: полная пространственная случайность.  
 Б: агрегационный процесс, события образуют кластеры.  
 В: регуляризованный процесс, события «стремятся отдалиться» друг от друга на некоторое расстояние.

Различают различные виды агрегирования. Во-первых, агрегации могут порождаться неоднородностью точечного процесса, когда интенсивность изменяется в пространстве. Тогда скопления событий будут появляться в тех подобластях изучаемой области, где интенсивность процесса выше. Этот механизм обычно обусловлен влиянием внешних факторов, приводящих к пространственной вариации интенсивности. Например, в подобластях повышенной влажности лучше всхожесть семян, что приведет к повышенной плотности проростков. В другом случае, агрегации могут порождаться процессом с позитивным взаимодействием событий, которое «притягивает» события друг к другу: появление одного события увеличит вероятность появления в его окрестности других. Скажем, если речь идет о размещении особей в популяциях, то такой механизм может быть связан с биолого-экологическими свойствами вида. Например, появление первой, родительской особи может привести к появлению вокруг нее особей-потомков и формированию агрегации. Стоит отметить, что разные

процессы способны порождать сходные точечные образы, а образование агрегаций может быть вызвано несколькими механизмами, которые сложно отделить друг от друга. Для математического представления агрегационных процессов используются *процессы Кокса* (Cox process). Это обширный класс моделей, которые способны описать разнообразные механизмы образования агрегаций. Если скопления событий образуются за счет неоднородности процесса и при этом взаимодействия между событиями отсутствуют, то как основная модель рассматривается *неоднородный пуассоновский процесс* (inhomogeneous Poisson process). Если образование кластеров связано с позитивным пространственным взаимодействием событий, то для моделирования используются *процессы Неймана-Скотта* (Neyman-Scott process), например, кластерный процесс Матерна (Matern cluster process), процесс Томаса (Thomas process).

Пространственная регулярность в размещении событий также может порождаться различными причинами. В основном ее проявление обусловлено процессами с негативным взаимодействием, которое «отталкивает» события друг от друга: конкурентные отношения, аменсализм. В точечном образе регулярность часто проявляется наличием вокруг события некоего «ядра», которое покрывает его и препятствует появлению других событий в близлежащих местоположениях. Для математического представления такого поведения используется класс моделей, названных моделями *угнетения* (ингибирования), в которых пуассоновский процесс трансформируется в новый процесс с помощью операции прореживания (*модели прореживания* (thinning process), *модели твердого ядра* (hard core process)). Часто точечные процессы с негативным

взаимодействием событий моделируют с помощью *марковских точечных процессов* (или *процессов Гиббса*). В общем случае, процессы Гиббса позволяют моделировать не только регулярность в размещении событий, но и агрегацию, а также различные типы более сложного поведения.

<b>Что наблюдается</b>	<b>Какие закономерности лежат в основе</b>	<b>Класс моделей</b>	<b>Частные случаи</b>
Полная пространственная случайность	Интенсивность постоянна, между событиями нет взаимодействий.	Однородный пуассоновский процесс	
Агрегации событий	Интенсивность меняется по территории, между событиями нет взаимодействий. Агрегации образуются вследствие неоднородности распределения внешних факторов, ресурсов.	Неоднородный пуассоновский процесс	
	Между событиями есть позитивное взаимодействие. Агрегации образуются событиями-потомками, размещающимися вокруг родительских событий.	Процессы Кокса  Процессы Неймана-Скотта	Процесс Матерна, процесс Томаса и др.
Разреженность (регулярность) событий	Между событиями есть негативное взаимодействие. События отдаляются друг от друга вследствие конкурентных отношений или ингибирования.	Процессы ингибирования  Марковские (Гиббса) процессы	Модель Матерна I, модель Матерна II  Процесс Штрауса и др.
Сложное поведение	Эффекты любых порядков.	Марковские (Гиббса) процессы	

## ***Однородный пуассоновский процесс***

Моделью полной пространственной случайности (CSR) служит однородный пуассоновский процесс, задаваемый следующим образом:

1). Интенсивность процесса является константой:  $\lambda(x) \equiv \lambda$ ;

2).  $N_X(B)$  – случайное число событий, порождаемых процессом в произвольной подобласти  $B$  ( $B \subseteq S$ ), подчиняется распределению Пуассона со средним  $\mu(B) = \lambda \cdot A_B$ , где  $\lambda$  – интенсивность процесса,  $A_B$  – площадь подобласти  $B$ . То есть вероятность того, что число событий в подобласти  $B$  примет значение равное  $k$ , определяется как:

$$P(N_X(B) = k) = \frac{e^{-\mu(B)} \mu^k(B)}{k!}. \quad (1)$$

3). Для непересекающихся подобластей  $B_1, \dots, B_k \subseteq S$  случайные величины  $N_X(B_1), \dots, N_X(B_k)$  независимы.

Эти свойства означают, что события имеют одинаковую вероятность появления в любом местоположении изучаемой области  $S$ , а также не взаимодействуют друг с другом – размещаются в пространстве независимо, появление одних событий не увеличивает и не уменьшает вероятность появления других.

Однородный пуассоновский процесс является стационарным и изотропным. Для него функция Рипли  $K(r) = \pi r^2$ , т.е. равна площади круга радиусом  $r$ . Чем больше площадь круга, тем больше событий можно встретить, а произведение  $\lambda \pi r^2$  даст ожидаемое число событий в круге радиуса  $r$ . Также легко показать, что функция  $L(r) = 0$  для

любого радиуса. Для этого процесса интенсивность второго порядка  $\lambda_2(r) = \lambda^2$ , а парная корреляционная функция  $g(r) = 1$  для любых радиусов  $r$ .

На этих свойствах однородного пуассоновского процесса основаны тесты, используемые для проверки гипотезы о полной пространственной случайности точечного процесса. При этом значения  $K(r) > \pi r^2$ ,  $L(r) > 0$ ,  $g(r) > 1$  указывают на наличие агрегаций, а  $K(r) < \pi r^2$ ,  $L(r) < 0$ ,  $g(r) < 1$  – на пространственную регулярность событий (рис. 8).

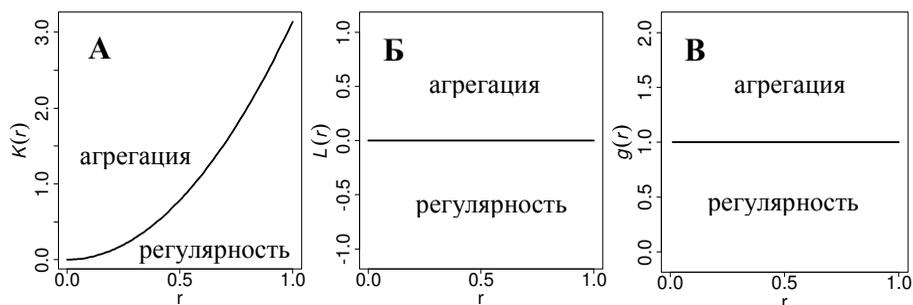


Рис. 8. Характеристики свойств второго порядка однородного пуассоновского процесса. А:  $K(r)$ , Б:  $L(r)$ , В:  $g(r)$ .

Отклонение этих функций от поведения, соответствующего однородному пуассоновскому процессу, будет свидетельствовать о наличии взаимодействия между событиями, либо позитивного, приводящего к их агрегированию, либо негативного, приводящего к регулярности в их размещении.

Так как в случае однородного пуассоновского процесса взаимодействие между событиями отсутствует, распределение расстояний между ближайшими событиями и распределение расстояний до ближайших событий от случайно выбранных точек пространства одинаковы, т.е. функции  $G(r)$  и  $F(r)$  равны, и  $G(r) = F(r) = 1 - e^{-\lambda \cdot 2\pi r^2}$ . Отсюда следует, что, функция  $J(r) = 1$

для любого радиуса. Эти свойства также часто используют для проверки гипотезы о полной пространственной случайности. Если  $G(r) > F(r)$  или  $J(r) < 1$ , это будет означать, что события чаще встречаются вокруг других событий, чем вокруг случайно выбранной точки пространства, поэтому можно констатировать, что события агрегируются. Если  $G(r) < F(r)$  или  $J(r) > 1$ , то события, скорее всего, показывают наличие некоторой регулярности распределения в пространстве, к примеру, расстояние между событиями почти всегда не меньше определенной величины (рис. 9).

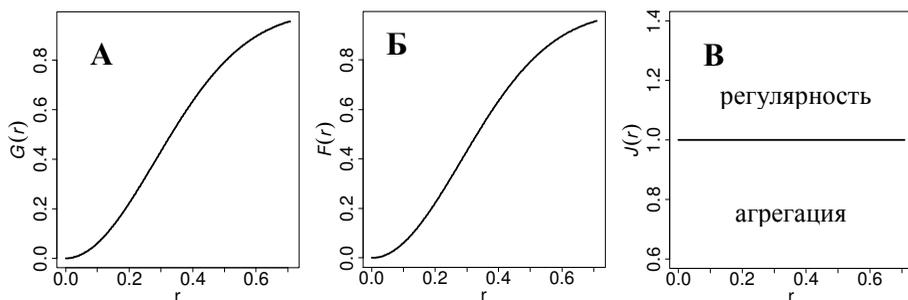


Рис. 9. Характеристики свойств второго порядка одномерного пуассоновского процесса с интенсивностью  $\lambda=1$ .

А:  $G(r)$ , Б:  $F(r)$ , В:  $J(r)$ .

Отклонение этих функций от поведения, соответствующего однородному пуассоновскому процессу, будет свидетельствовать о наличии взаимодействия между событиями, либо позитивного, приводящих к их агрегированию, либо негативного, приводящего к регулярности в их размещении.

В качестве иллюстрации свойств CSR рассмотрим следующий пример. Пусть в области  $S$  имеет место однородный пуассоновский процесс с интенсивностью  $\lambda = 1$ . Выделим в области  $S$  произвольную подобласть  $B$  с площадью  $A_B = 10$ . Число событий  $N_X$  в подобласти  $B$  - случайная величина, заранее

неизвестно, чему точно будет равно ее значение в каждой реализации. Известно лишь то, что число событий будет подчиняться пуассоновскому закону со средним

$$\mu(B) = \lambda \cdot A_B = 1 \cdot 10 = 10.$$

Тогда вероятность того, что в такой подобласти  $B$  не будет ни одного события, равна:

$$P(N_x = 0) = \frac{e^{-\mu(B)} \mu^k(B)}{k!} = \frac{e^{-10} 10^0}{0!} = \frac{1}{e^{10}} = 0.000045 \approx 0.$$

Вероятность, что в  $B$  встретится 5 событий, равна:

$$P(N_x = 5) = \frac{e^{-10} 10^5}{5!} = 0.0378.$$

Вероятность встречи числа событий, равного 10:

$$P(N_x = 10) = \frac{e^{-10} 10^{10}}{10!} = 0.1251.$$

Вероятность встречи числа событий, равного 20:

$$P(N_x = 20) = \frac{e^{-10} 10^{20}}{20!} \approx 0.002$$

и т.д. Наибольшую вероятность имеет число событий, соответствующее среднему,  $P(N_x = 10) = 0.1251$ . В соответствии с этим, если рассмотреть различные реализации однородного пуассоновского процесса с  $\lambda=1$ , то чаще всего в любой подобласти с площадью, равной 10, можно будет встретить 10 событий, реже – 5, 15 и т.д. Число событий, к примеру, равное 0 или 20, практически не будет встречаться, так как вероятность таких значений весьма мала.

На рис. 10 в качестве примера приведен вид распределения Пуассона, соответствующего различным значениям параметра  $\mu$ .

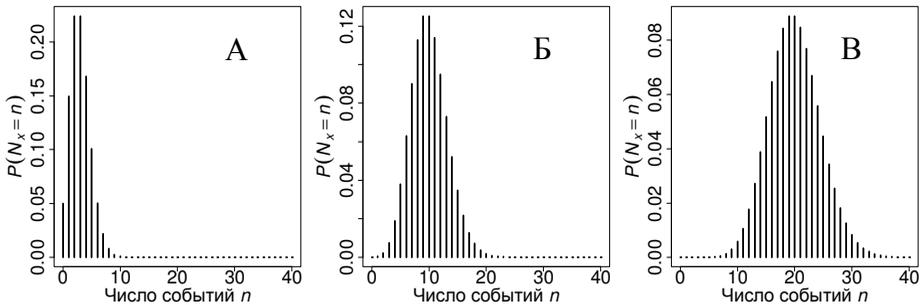


Рис. 10. Распределение Пуассона - А:  $\mu=3$ , Б:  $\mu=10$ , В:  $\mu=20$ .

### ***Неоднородный пуассоновский процесс***

Для *неоднородного* пуассоновского процесса интенсивность не является константой, а варьирует на изучаемой территории (является функцией от  $x$ ,  $x \in S$ ). Это самая простая модель нестационарного процесса, являющегося одной из альтернатив CSR. Варьирующая интенсивность может приводить к повышенной плотности событий в областях высоких значений  $\lambda(x)$ , визуально проявляющейся как агрегации. При этом появление событий остается независимым друг от друга, между событиями нет взаимодействий.

Неоднородный пуассоновский процесс задается следующим образом:

1).  $N_x(B)$  – случайное число событий, порождаемых процессом в произвольной подобласти  $B$  ( $B \subseteq S$ ), подчиняется распределению Пуассона со средним

$$\mu(B) = E[N_x(B)] = \int_B \lambda(x) dx,$$

где  $\lambda(x)$  – интенсивность процесса,  $\mu(B)$  - ожидаемое число событий, попадающих в подобласть  $B$ . Вероятность того, что число событий в подобласти  $B$  примет значение, равное  $k$ , определяется как:

$$P(N_x(B) = k) = \frac{e^{-\mu(B)} \mu^k(B)}{k!}.$$

2). Для непересекающихся подобластей  $B_1, \dots, B_k \subseteq S$  случайные величины  $N_x(B_1), \dots, N_x(B_k)$  независимы.

В отличие от однородного пуассоновского процесса интенсивность процесса  $\lambda(x)$  не является константой (меняется по территории), и, следовательно, ожидание числа событий  $\mu(B)$  в произвольной подобласти  $B$  зависит не только от размера этой подобласти, но и от ее локализации на изучаемой территории.

Зависимость интенсивности появления событий от местоположения  $x$  часто может объясняться влиянием каких-то внешних факторов, что можно включить в математическую модель. А именно: интенсивность процесса  $\lambda(x)$  может рассматриваться как функция от значений этих факторов в точках  $x \in S$ :

$$\lambda(x) = f\{z_1(x), z_2(x), \dots, z_p(x)\}.$$

Здесь  $z_1, z_2, \dots, z_p$  - внешние факторы, способные влиять на интенсивность процесса (и, следовательно, на генерируемый им точечный образ). Такими факторами могут быть, например, влажность, почвенное плодородие, доступность пищевых ресурсов, высота над уровнем моря, рН и др. Вид функции  $f$ , связывающей интенсивность с факторами, может быть различным. Например, часто для моделирования используют экспоненциальную функцию вида:

$$\lambda(x) = e^{\alpha + \beta z(x)},$$

где  $z(x)$  - фактор, оказывающий влияние на интенсивность, значение фактора зависит от точки  $x$ ;  $\alpha$  и  $\beta$  - коэффициенты

линейного уравнения. В качестве непрямых факторов в модели могут участвовать сами координаты точки, неявно связанные с прямыми факторами, которые не были измерены (рис. 11).

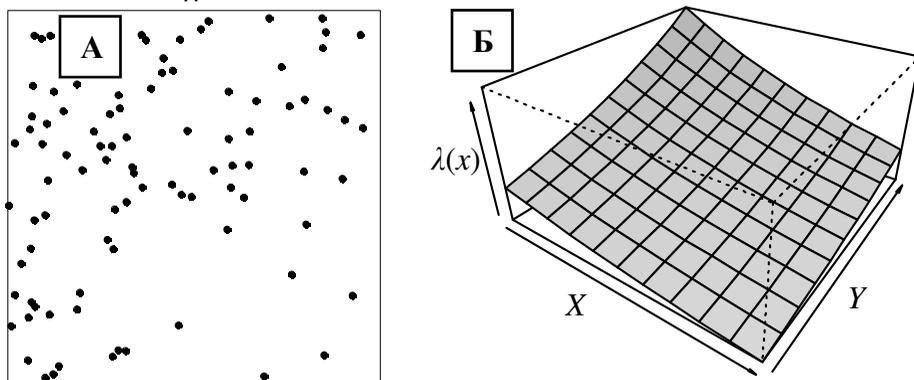


Рис. 11. Пример реализации точечного процесса (А), интенсивность которого является функцией положения в пространстве:  $\lambda(x) = e^{-X+Y}$  (Б).

Для неоднородного пуассоновского процесса, так же как для однородного, взаимодействие между событиями отсутствует, и парная корреляционная функция  $g(r) = 1$  для любых радиусов  $r$ .

### ***Процессы Кокса***

Процессы Кокса являются расширением пуассоновского процесса, а именно: интенсивность пуассоновского процесса рассматривается как реализация *случайной* функции  $\Lambda(x)$ . То есть, это «дважды стохастический процесс»: случайно не только появление событий в тех или иных точках пространства, но и интенсивность процесса в каждой точке  $x$  принимает случайное значение в соответствии законом распределения случайного поля  $\Lambda(x)$ . Такая модель была впервые предложена английским статистиком Дэвидом Коксом (1955).

## Смешанный пуассоновский процесс

Простейшим примером процесса Кокса является смешанный пуассоновский процесс (mixed Poisson process), когда интенсивность является случайной величиной  $\lambda$  (не зависит от  $x$ ), принимающей неотрицательные значения в соответствии с некоторым законом распределения вероятностей (к примеру, гамма, логнормальным, экспоненциальным и др.). Реализации этого процесса ведут себя как реализации однородного пуассоновского процесса, но однородный пуассоновский процесс генерирует реализации с постоянной интенсивностью, а в случае смешанного пуассоновского процесса интенсивность для каждой реализации может быть различной.

В качестве иллюстрации рассмотрим следующий пример. Пусть в изучаемой области  $S$  имеет место смешанный пуассоновский процесс  $X$  (частный случай процесса Кокса), интенсивность которого – случайная величина, распределенная по экспоненциальному закону с параметром  $a=0.01$ . Пусть для трех реализаций этого процесса интенсивность случайным образом приняла значения  $\lambda$ , равные 5, 3 и 2, соответственно.

Рассмотрим подобласть  $B$ , принадлежащую  $S$  и имеющую площадь 4 единицы. В каждой из реализаций  $x$  процесса  $X$  число событий  $N_x(B)$  в подобласти  $B$  является реализацией случайной величины  $N_X(B)$ , имеющей закон распределения Пуассона с матожиданием числа событий  $\mu(B) = \lambda \cdot A_B$ . Для рассматриваемых трех реализаций значение  $\mu(B)$  в нашей подобласти  $B$  будет равно 20, 12 и 8 соответственно. На рис. 12 показаны эти три реализации смешанного пуассоновского процесса, где число событий  $N_x(B)$  случайным образом приняло значения 23, 11 и 10.

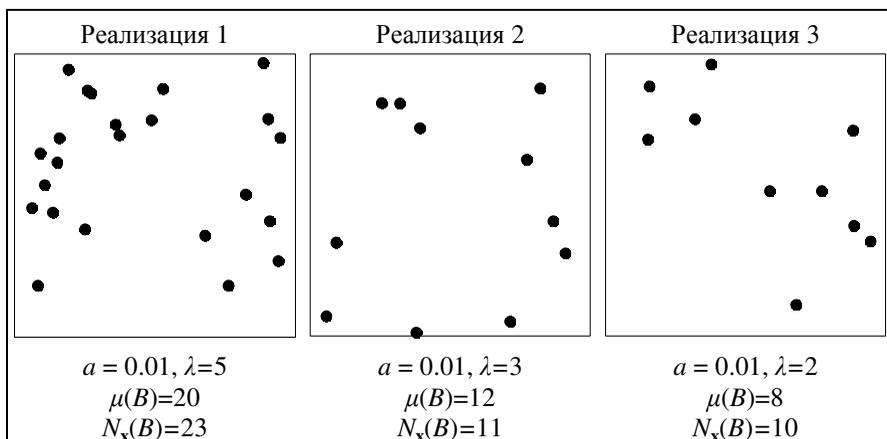


Рис. 12. Пример трех реализаций смешанного пуассоновского процесса.

Если дана только одна реализация, то невозможно отличить смешанный пуассоновский процесс от однородного пуассоновского процесса. Аналогичная ситуация и для процессов Кокса в общем случае: имея только одну реализацию, невозможно отличить его от неоднородного пуассоновского процесса. Решение о том, какая из двух моделей является наиболее подходящей, зависит от априорных знаний об изучаемом явлении. Процессы Кокса позволяют моделировать агрегацию или кластеризацию событий, связанную с влиянием внешних факторов, но при этом внешние факторы, влияющие на интенсивность процесса, предполагаются *стохастическими* по своей природе.

Для большинства моделей класса процессов Кокса парная корреляционная функция  $g \geq 1$ .

## ***Кластерные процессы: процессы Неймана-Скотта***

Кластерные процессы - один из наиболее важных и широко применяемых классов моделей точечных процессов. Здесь события в области  $S$  появляются в агрегациях, скоплениях.

Основная идея этих моделей состоит в том, что существует некоторый случайный точечный процесс, порождающий размещение «родительских» событий (центров кластеров), вокруг которых согласно своим независимым точечным процессам случайно размещаются так называемые «дочерние» события (или события-«потомки»), образуя кластеры. Таким образом, кластерный точечный процесс есть суперпозиция точечных процессов «потомков».

В качестве примера природного процесса, который может быть описан с помощью данной модели, можно назвать размещение деревьев в лесу. Побегов могут создавать видимые агрегации, размещаясь вокруг погибших родительских особей.

Процессы Неймана-Скотта (их определение впервые дано Нейманом (1939)) являются частным случаем кластерных процессов. Их иногда называют кластерными пуассоновскими процессами (Poisson cluster process). Здесь:

- «Родительские» события размещаются в области  $S$  в соответствии с однородным пуассоновским процессом с интенсивностью  $\lambda$ .
- Каждое «родительское событие» генерирует случайное число  $K$  событий-«потомков». Все дискретные случайные величины  $K$  (число потомков каждого «родительского» события)

независимы и имеют одинаковый закон распределения вероятностей<sup>2</sup>  $P(K=k)$ ,  $k=0, 1, 2, \dots$

- Позиции «потомков» относительно их «родительских» событий (центров кластеров) независимы и имеют одинаковый закон распределения с некоторой двумерной функцией плотности вероятности  $f(\cdot)$ .
- Результирующий точечный процесс состоит только из событий-«потомков». «Родительские» события отсутствуют либо неотличимы от событий-«потомков» (рис. 13).

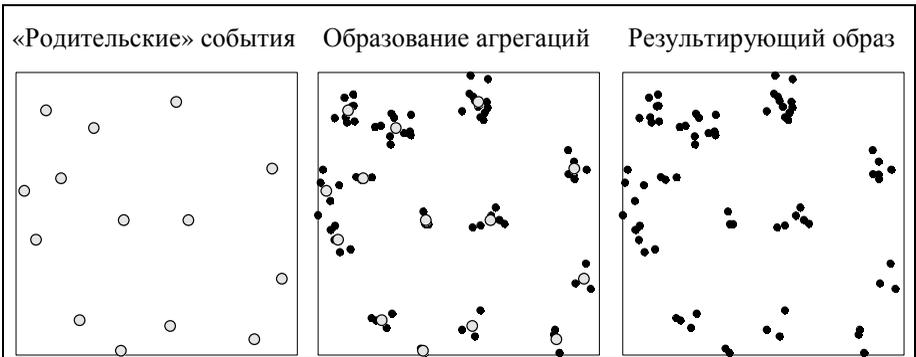


Рис. 13. Результирующий точечный образ состоит только из событий-«потомков», появившихся вокруг событий-«родителей».

Этот класс процессов используется для моделирования отклонения от полной пространственной случайности, связанного с позитивным пространственным взаимодействием событий (эффекты второго порядка). Оно проявляется образованием агрегаций событий: вероятность встретить одно событие рядом с другим повышается, так как предполагается, что «родительское» событие «дало жизнь» нескольким событиям-«потомкам».

<sup>2</sup> В англоязычной литературе употребляют сокращение и термин iid, independently and identically distributed.

Процессы Неймана-Скотта являются однородными по определению, а парная корреляционная функция  $g > 1$ .

Определение процесса Неймана-Скотта не конкретизирует закон образования агрегаций. Ниже приведены примеры (частные случаи) этих процессов.

### **Кластерный процесс Матерна**

В этой модели (рис. 14 А) каждое «родительское» событие создает  $K$  «потомков», число которых случайно и подчиняется распределению Пуассона с матожиданием  $\mu$ :  $P(K = k) = \frac{e^{-\mu} \mu^k}{k!}$ .

«Потомки» размещаются независимо друг от друга в круге радиуса  $R$  вокруг «родительской» особи в соответствии с двумерным равномерным законом распределения, функция плотности вероятности которого равна:

$$f(X, Y) = \frac{1}{\pi R^2}.$$

### **Процесс Томаса**

В этой модели (рис. 14 Б) каждое «родительское» событие также создает случайное число  $K$  «потомков» согласно распределению Пуассона с матожиданием  $\mu$ . И «потомки» размещаются независимо друг от друга вокруг «родительского» события, но в соответствии с двумерным гауссовым законом распределения с матожиданием 0 и ковариационной матрицей  $\sigma^2 \mathbf{I}$ , функция плотности вероятности которого равна:

$$f(X, Y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(X^2+Y^2)}{2\sigma^2}}.$$

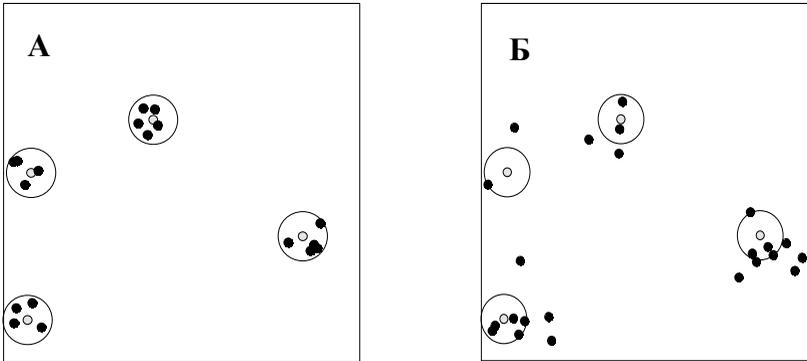


Рис. 14. Иллюстрации кластерных пуассоновских процессов. Обе закартированные области имеют единичную площадь. «Родительское» событие показано белой точкой, вокруг которой очерчена окружность радиуса  $R=0.07$ .

А: Реализация процесса Матерна; число «дочерних» событий (черные точки) подчиняется распределению Пуассона с  $\mu = 5$ ; радиус круга, в котором появляются «потомки»,  $R=0.07$ .

Б: Реализация процесса Томаса; число «дочерних» событий (черные точки) также подчиняется распределению Пуассона с  $\mu = 5$ ; «дочерние» события рассеиваются относительно «родительского» в соответствии с двумерной нормальной (гауссовой) функцией распределения с  $\sigma=0.07$ , центрированной в «родительском» событии.

В литературе описаны расширения этого класса процессов, а именно, неоднородные процессы Неймана-Скотта, например, когда в качестве «родительского» рассматривается неоднородный пуассоновский процесс с интенсивностью  $\lambda(x)$ .

Заметим, что нет строгой границы между кластерными процессами и процессами Кокса, точнее, существует большой класс точечных процессов, которые одновременно являются и кластерными процессами, и процессами Кокса. Если случайное число точек в кластере имеет распределение Пуассона (как, например, в случаях модели Матерна или модели Томаса), тогда кластерный процесс будет также процессом Кокса.

## ***Процессы ингибирования (угнетения)***

Как говорилось выше, другим отклонением от полной пространственной случайности является наличие некоторой регулярности в размещении событий. Подобное поведение может быть связано с негативным пространственным взаимодействием событий, с наличием механизмов, мешающих появлению событий вблизи других событий. Для описания таких точечных процессов может применяться класс моделей ингибирования, которые лимитируют минимальное расстояние между событиями некоторой константой («твердое ядро»). Эти модели могут быть использованы, например, для описания пространственного размещения деревьев в старовозрастном лесу, где расстояние между деревьями ограничено их кронами, или для описания других конкурентных взаимодействий.

Рассмотрим две модели ингибирования, предложенные Матерном (1960) и использующие процедуру прореживания (thinning).

### **Процесс Матерна I**

Этот процесс задается следующим образом: предполагается, что в области  $S$  есть некоторый однородный пуассоновский процесс с интенсивностью  $\rho$ , в реализациях которого удаляются все пары событий, находящиеся друг от друга на расстоянии менее заданного порога  $\delta$  (минимальное допустимое расстояние). Оставшиеся после прореживания события и составляют реализацию процесса Матерна I с результирующей интенсивностью  $\lambda$ , которая зависит от значения  $\rho$  и порога ингибирования  $\delta$ :

$$\lambda = \lambda(\rho, \delta) = \rho \cdot e^{-\rho \cdot \pi \delta^2}.$$

## Процесс Матерна II

Модель также задается на основе однородного пуассоновского процесса в области  $S$  с интенсивностью  $\rho$ . Каждому  $i$ -тому событию в его реализации присваивается случайная метка - «момент наступления»  $t_i$ , которая может принимать значения в некоем числовом интервале, следуя, обычно, равномерному закону распределения вероятностей. Далее в каждой паре событий, находящихся друг от друга ближе, чем порог ингибирования  $\delta$ , удаляется то событие, которое имеет более поздний «момент наступления». Так как в этой модели удаляются не оба события в паре, то интенсивность процесса Матерна II выше по сравнению с интенсивностью в модели Матерна I (рис. 15) при одинаковых  $\rho$ . В качестве примера природного процесса, который может быть описан этой моделью, можно привести взаимоотношения деревьев в лесу, когда крупные деревья угнетают рост более молодых деревьев.

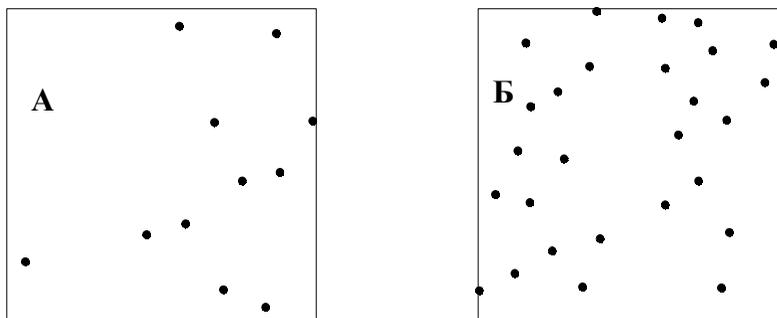


Рис. 15. Иллюстрации процессов ингибирования. Обе области имеют единичную площадь. Порог ингибирования  $\delta = 0.1$ .

А: Реализация процесса Матерна I. Б: Реализация процесса Матерна II. Исходные интенсивности  $\rho$  одинаковы, результирующие интенсивности  $\lambda$  различаются; в каждой реализации отсутствуют пары объектов, находящихся на расстоянии меньшем, чем  $0.1$ .

## *Процессы Гиббса, марковские точечные процессы*

Пространственную регулярность в размещении событий (репульсивное поведение, «отталкивание» событий) часто моделируют с помощью марковских точечных процессов (или процессов Гиббса). Это большое семейство моделей, позволяющих описывать всевозможные взаимодействия между событиями (не только негативные и не только парные), моделирующих не только регулярные точечные образы или точечные образы с агрегацией событий, но и комбинации этих типов размещений.

Процессы Гиббса были введены в статистической физике для описания систем взаимодействующих частиц. Американский ученый Д. Уиллард Гиббс (1839–1903), заложивший основы статистической механики, изучал термодинамику химических и физических процессов в системах, состоящих из частиц. Можно представить, что каждая частица несет некий заряд, и взаимодействие между частицами является силой, которая отталкивает их друг от друга. Чем ближе две частицы находятся друг к другу, тем выше потенциал отталкивания. Частицы стремятся расположиться в пространстве так, чтобы минимизировать этот потенциал. В соответствии с теорией Гиббса, чем меньшим потенциалом обладает система (конфигурация, пространственное размещение частиц), тем выше вероятность ее существования.

Теория Гиббса описывает поля парных потенциалов, однако с ее помощью можно формализовать и другие процессы, например, пространственное размещение деревьев в лесу. Здесь аналогом «потенциала отталкивания» может являться конкуренция за ресурсы. Логично предположить, что этот

конкурентный потенциал будет уменьшаться по мере увеличения расстояния между деревьями. Также можно допустить, что чем деревья больше, тем сильнее потенциал: чаще можно встретить растущие рядом маленькие деревца, чем большие деревья. Взаимодействия между событиями могут быть симметричными и несимметричными. К примеру, взаимодействие деревьев в лесу может рассматриваться как несимметричное: большие деревья сильнее воздействуют на юные деревца, не испытывая значительного конкурентного давления с их стороны.

Процессы Гиббса задаются с помощью *функции плотности* относительно пуассоновского процесса.

Пусть  $Y_1$  – однородный пуассоновский процесс с интенсивностью  $\lambda = 1$  в области  $S$ . Из его свойств следует, что для любого множества  $F$ , состоящего из конечных конфигураций точек в  $S$ , вероятность того, что этот процесс породит конфигурацию из  $F$ , равна:

$$P(Y_1 \in F) = \sum_{n \geq 0} \frac{e^{-A_S}}{n!} \int_S \dots \int_S I[\mathbf{x} \in F] dx_1 dx_2 \dots dx_n,$$

где  $A_S$  – площадь области  $S$ ,  $I[\cdot]$  – индикаторная функция, которая равна 1, если конфигурация  $\mathbf{x}$  принадлежит множеству  $F$ , иначе равна 0.

Говорят, что конечный точечный процесс  $\mathbf{X}$  в области  $S$  имеет функцию плотности  $f$  относительно процесса  $Y_1$ , если для него подобная вероятность может быть определена как:

$$P(\mathbf{X} \in F) = \sum_{n \geq 0} \frac{e^{-A_S}}{n!} \int_S \dots \int_S I[\mathbf{x} \in F] f(\mathbf{x}) dx_1 dx_2 \dots dx_n.$$

Функция плотности  $f$  задается для любой конечной конфигурации точек  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  из  $S$ .

Для однородного пуассоновского процесса с интенсивностью  $\lambda$ :

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) = \alpha \lambda^n, \text{ где константа } \alpha = e^{(1-\lambda)A_S}.$$

Для неоднородного пуассоновского процесса:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) = \alpha \prod_{i=1}^n \lambda(x_i), \text{ где } \alpha = e^{A_S - \int_S \lambda(u) d(u)}.$$

Процессы Гиббса в ограниченной области  $S$  в общем случае определяют через функцию плотности вида:

$$f(\mathbf{x}) = \frac{1}{c} \cdot e^{-U(\mathbf{x})},$$

где  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  - любая конфигурация  $n$  точек в области  $S$ , функция  $U$  в терминах статистической физики обозначает энергию системы - потенциал пространственного размещения событий  $\mathbf{x}$ ,  $c$  - нормализующая константа («статистическая сумма» в физике, она обеспечивает нужные свойства функции  $f(\mathbf{x})$ ).

В большинстве случаев функция  $f$  не выражается в явном виде из-за сложности определения нормализующей константы, и ее задают вплоть до пропорциональности -  $f(\mathbf{x}) = \frac{1}{c} \cdot h(\mathbf{x})$ , где  $h$  - известная функция. Поэтому основным инструментом при работе с процессами Гиббса является *условная интенсивность*  $\lambda(u, \mathbf{x})$  (в англоязычной литературе - *Papangelou conditional intensity*), которая не зависит от нормализующей константы и определяется как:

$$\lambda(u, \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})},$$

где  $\mathbf{x}$  - любая конечная конфигурация событий,  $u$  - точка из  $S$ , не принадлежащая этой конфигурации  $\mathbf{x}$  ( $u \in S, u \notin \mathbf{x}$ ).  $\lambda(u, \mathbf{x})$  может интерпретироваться как условная вероятность наблюдать событие в бесконечно малой окрестности точки  $u$  при заданной конфигурации точек  $\mathbf{x}$  (более точно, условная вероятность – это  $\lambda(u, \mathbf{x}) du$ ). Очевидно, что условная интенсивность  $\lambda(u, \mathbf{x})$  не зависит от константы  $c$ , а для пуассоновского процесса равна просто интенсивности:  $\lambda(u, \mathbf{x}) = \lambda(u)$ , и не зависит от конфигурации  $\mathbf{x}$ . Если функция плотности обладает свойством наследования ( $f(\mathbf{x}) > 0$  влечет  $f(\mathbf{y}) > 0$  для любых  $\mathbf{y} \subset \mathbf{x}$ ), то  $f(\mathbf{x})$  и  $\lambda(u, \mathbf{x})$  взаимно-однозначно соответствуют друг другу.

Процесс  $\mathbf{X}$  часто называют процессом:

- с притяжением (агрегацией, кластеризацией), если  $\lambda(u, \mathbf{x}) \leq \lambda(u, \mathbf{y})$  всякий раз, когда  $\mathbf{x} \subset \mathbf{y}$ ;
- с отталкиванием (регуляризацией), если  $\lambda(u, \mathbf{x}) \geq \lambda(u, \mathbf{y})$  всякий раз, когда  $\mathbf{x} \subset \mathbf{y}$ .

Вообще говоря, все процессы, для которых существует функция плотности  $f(\mathbf{x})$  и условная интенсивность  $\lambda(u, \mathbf{x})$ , могут быть отнесены к процессам Гиббса. В частности, с помощью этого математического аппарата могут быть формализованы и пуассоновские процессы, и процессы Кокса.

Процесс Гиббса имеет конечный ранг  $R$ , если условная интенсивность удовлетворяет условию

$$\lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x} \cap B(u, R)),$$

где  $B(u, R)$  - круг радиуса  $R$  с центром в точке  $u$ . То есть, вероятность «вставить» точку  $u$  в точечную конфигурацию  $\mathbf{x}$  зависит только от некоторых соседей этой точки. Другими словами, *вероятность произвольной конечной конфигурации*

точек определяется локальным взаимодействием – зависит от взаимодействия с соседями, причем это «соседство» может определяться различными способами, что в более общем случае ведет к определению марковских точечных полей.

Если рассматривается только один сосед, то мы имеем подкласс процессов Гиббса, называемых *моделями парного взаимодействия (pairwise interaction model)*, для которых функция плотности может быть записана как:

$$f(\mathbf{x}) = \frac{1}{c} \prod_{i=1}^n a(x_i) \prod_{i < j} b(x_i, x_j).$$

То есть, энергия  $U(\mathbf{x})$  равна

$$U(\mathbf{x}) = -\sum \log a(x_i) - \sum \log b(x_i, x_j),$$

а условную интенсивность процесса парного взаимодействия можно выписать как:

$$\lambda(u, \mathbf{x}) = a(u) \prod_{i=1}^n b(u, x_i).$$

Здесь функция  $a(u)$  отражает свойства первого порядка (обычно это интенсивность) процесса; если это константа (не зависит от  $u$ ), то процесс будет однородным. Функция  $b(u, v)$  отражает свойства второго порядка и называется *функцией парного взаимодействия*. Для моделей парного взаимодействия эта функция предполагается симметричной, т.е.  $b(u, v) = b(v, u)$ , и обычно изотропной -  $b(u, v) = b(\|u - v\|)$ . В принципе вид функций  $a$  и  $b$  может быть выбран любым, но при этом они должны обеспечивать интегрируемость функции плотности.

Если  $b(v, u) \leq 1$ , тогда процесс  $\mathbf{X}$  является регуляризованным (события «отталкиваются» друг от друга).

Если  $b(v, u) > 1$ , тогда процесс  $\mathbf{X}$  в общем случае не вполне определен (может не существовать).

### Процесс твердого ядра

В простейшем случае, если  $a(u) \equiv a > 0$ , а функция парного взаимодействия имеет следующий вид:

$$b(u, v) = \begin{cases} 1, & \text{если } \|u - v\| > \delta \\ 0, & \text{если } \|u - v\| \leq \delta \end{cases},$$

где  $\|u - v\|$  - расстояние между событиями (точками)  $u$  и  $v$ , тогда плотность примет вид:

$$f(\mathbf{x}) = \begin{cases} c \cdot a^{n(\mathbf{x})}, & \text{если } \|x_i - x_j\| > \delta, \text{ для всех } i \neq j \\ 0, & \text{в иных случаях} \end{cases},$$

где  $n(\mathbf{x}) = n$  - число точек в образе  $\mathbf{x}$ , а условную интенсивность процесса можно записать как:

$$\lambda(u, \mathbf{x}) = \begin{cases} a, & \text{если } \|u - x_i\| > \delta \text{ для всех } i \\ 0, & \text{в иных случаях} \end{cases}.$$

То есть, вероятность появления события в окрестности точки  $u$  не равна нулю (равна  $a \, du$ ), если на расстоянии  $\delta$  от этой точки нет ни одного другого события. Если же не дальше, чем на этом расстоянии от  $u$ , есть хотя бы одно событие, то вероятность появления в этом местоположении нового события равна нулю.

Эта модель известна как процесс твердого ядра. Она создает регулярные, разреженные точечные образы, в которых все наблюдаемые события находятся друг от друга на расстояниях не ближе заданного  $\delta$ . Можно представить, что каждое событие находится в центре ядра, в котором жестко ингибируется появление других событий. На рис. 16 приведен пример

реализации процесса твердого ядра и проиллюстрированы некоторые его свойства.

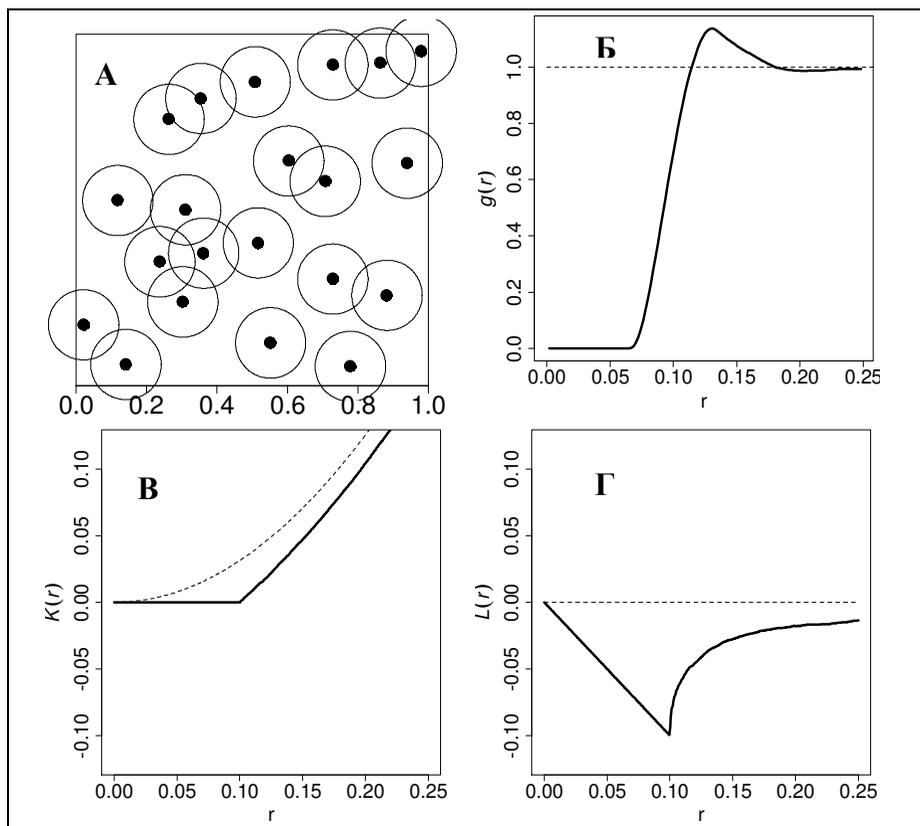


Рис. 16. Иллюстрация процесса твердого ядра.

А: Карта реализации процесса; вокруг каждого события очерчена окружность радиуса  $\delta = 0.1$ – «твердое ядро». Б: Вид парной корреляционной функции. В: Вид  $K$ -функции Рипли. Г: Вид функции  $L(r)$ . Сплошной линией показан пример поведения функций в случае процесса твердого ядра. Пунктиром показано поведение функций, соответствующее CSR.

## Процесс Штрауса

Процессы простого ингибирования и твердого ядра жестко лимитируют минимальное расстояние между событиями. Однако в реальном мире ингибирование не всегда является строгим, и хотя вероятность встречи близких событий может быть низка, она не является нулевой. Такое поведение учтено в модели Штрауса (Strauss process).

Этот процесс, как и предыдущий, является однородным -  $a(u) \equiv a$ , но функция парного взаимодействия, кроме параметра  $\delta$ , зависит еще от одного параметра  $\gamma$ :

$$b(u, v) = \begin{cases} 1, & \text{если } \|u - v\| > \delta \\ \gamma, & \text{если } \|u - v\| \leq \delta \end{cases},$$

отсюда плотность:

$$f(\mathbf{x}) = c \cdot a^{n(\mathbf{x})} \gamma^{s(\mathbf{x}, \delta)},$$

где  $s(\mathbf{x}, \delta)$  – число пар точек из  $\mathbf{x}$ , которые находятся друг от друга на расстоянии не более  $\delta$ , т.е. таких, что  $\|x_i - x_j\| \leq \delta, i \neq j$ .

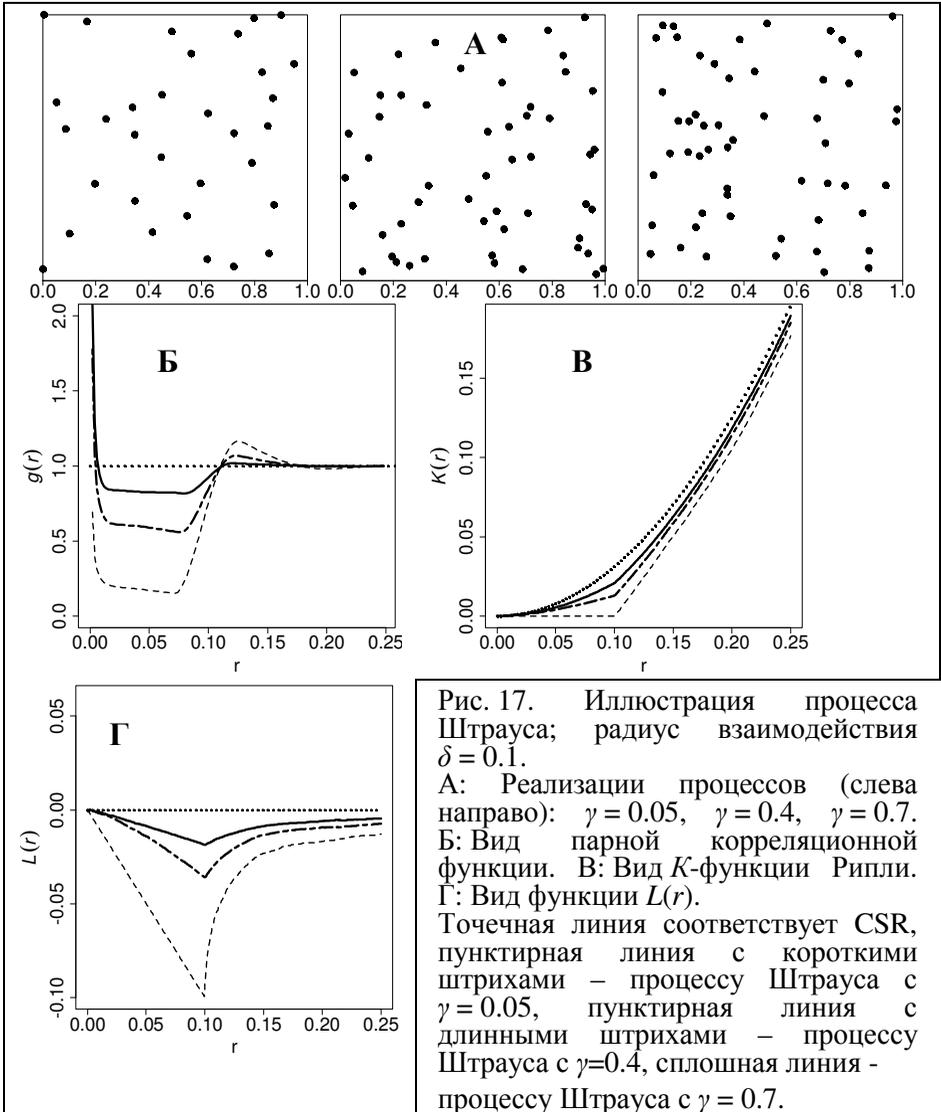
Условная интенсивность процесса Штрауса равна:

$$\lambda(u, \mathbf{x}) = a \cdot \gamma^{t(u, \mathbf{x}, \delta)},$$

где  $t(u, \mathbf{x}, \delta)$  – число точек из  $\mathbf{x}$ , которые находятся от точки  $u$  на расстоянии не более  $\delta$ , т.е. таких, что  $\|u - x_i\| \leq \delta$ .

Параметр  $\gamma$  описывает взаимодействие событий и принимает значения в интервале  $[0; 1]$ . Если  $\gamma = 1$ , тогда ингибирование полностью отсутствует, и процесс становится пуассоновским. Если  $\gamma = 0$ , то наблюдается строгое ингибирование, как в случае процессов твердого ядра. Нестрогое ингибирование соответствует значению  $0 < \gamma < 1$  (рис. 17). В этом случае, чем больше событий наблюдается в окрестности точки  $u$ ,

тем меньше вероятность того, что там может появиться еще одно событие. Параметр  $\delta$  – ранг процесса или радиус взаимодействия между событиями, то есть расстояние, на котором проявляется ингибирование.



## Другие модели парного взаимодействия

Перечислим некоторые другие модели парного взаимодействия, получаемые с помощью различного вида функций взаимодействия.

*Комбинация процесса твердого ядра и процесса Штрауса.* Пусть радиус строгого ингибирования (твердого ядра)  $h > 0$ , а радиус взаимодействия  $\delta > h$ . Функция парного взаимодействия имеет вид:

$$b(u, v) = \begin{cases} 1, & \text{если } \|u - v\| > \delta \\ \gamma, & \text{если } h < \|u - v\| \leq \delta \\ 0, & \text{если } \|u - v\| \leq h \end{cases}$$

То есть, при больших расстояниях между событиями ( $> \delta$ ) влияние их друг на друга отсутствует. На меньших расстояниях имеет место негативное взаимодействие событий, причем на малых расстояниях ( $\leq h$ ) происходит полное ингибирование (рис.18).

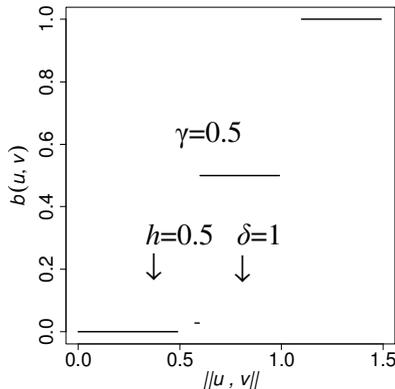


Рис. 18. Функция парного взаимодействия для комбинации процессов твердого ядра и Штрауса. Радиус строгого ингибирования (твердого ядра)  $h = 0.5$ , радиус взаимодействия  $\delta = 1.0$ , степень ингибирования  $\gamma = 0.5$ .

Процесс мягкого ядра (soft core model). Для него функция парного взаимодействия событий, зависящая от двух параметров  $\delta$  и  $k$ , имеет вид:

$$b(u, v) = 1 - \exp \left[ - \left( \frac{\|u - v\|}{\delta} \right)^{\frac{2}{k}} \right],$$

где  $\delta > 0$  – масштаб взаимодействия (scale),  $0 < k < 1$  – коэффициент, который определяет скорость затухания влияния событий друг на друга с возрастанием расстояния между ними. При  $k = 0$  модель сводится к процессу твердого ядра с радиусом  $\delta$ ; при  $k = 1$  наблюдается наиболее плавный переход от сильного ингибирования к отсутствию воздействия (very soft core model) (рис. 19).

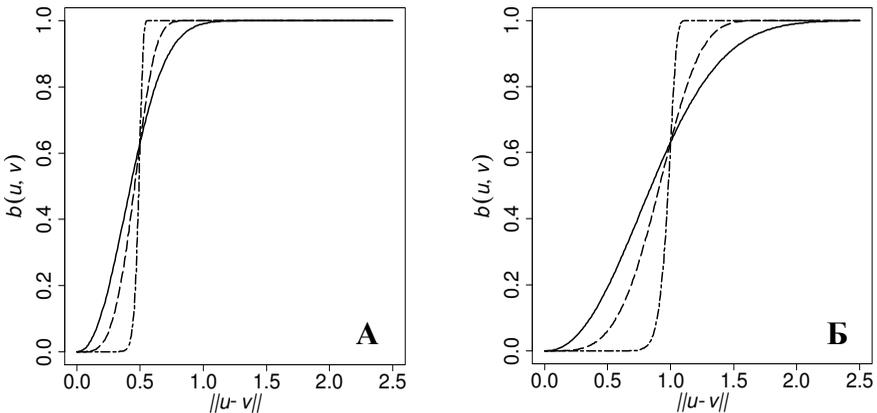


Рис. 19. Функция парного взаимодействия процесса мягкого ядра. По горизонтали откладывается расстояние между точками  $\|u - v\|$ . А: масштаб  $\delta = 0.5$ ; Б: масштаб  $\delta = 1.0$ . Сплошная линия –  $k = 0.9$ , пунктирная линия –  $k = 0.5$ , штрих-пунктирная линия –  $k = 0.1$ .

*Процесс Дигла-Гейтса-Стиббарда* (Diggle-Gates-Stibbard).

Функция парного взаимодействия зависит от радиуса взаимодействия  $\delta$  (рис. 20):

$$b(u, v) = \begin{cases} 1, & \text{если } \|u - v\| > \delta \\ \sin^2\left(\frac{\pi\|u - v\|}{2\delta}\right), & \text{если } \|u - v\| \leq \delta \end{cases}$$

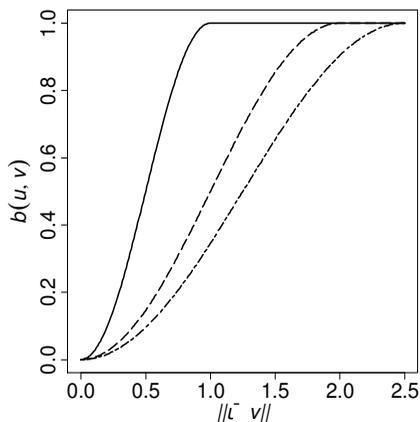


Рис. 20. Функция парного взаимодействия процесса Диггла-Гейтса-Стиббарда (Diggle-Gates-Stibbard). По горизонтали откладывается расстояние между точками  $\|u - v\|$ . Сплошная линия – радиус взаимодействия  $\delta = 1$ , пунктирная линия –  $\delta = 2$ , штрих-пунктирная линия -  $\delta = 2.5$ .

*Процесс Дигла-Граттона* (Diggle-Gratton). Функция парного взаимодействия имеет три параметра: радиус твердого ядра  $h > 0$ , радиус взаимодействия  $\delta > h$  и индекс  $k > 0$ , и записывается как:

$$b(u, v) = \begin{cases} 1, & \text{если } \|u - v\| > \delta \\ \left( \frac{\|u - v\| - h}{\delta - h} \right)^k, & \text{если } h < \|u - v\| \leq \delta \\ 0, & \text{если } \|u - v\| \leq h \end{cases}$$

При  $k=0$  модель становится моделью твердого ядра с радиусом ингибирования, равным  $h$ . При  $k \rightarrow \infty$  модель приближается к модели твердого ядра с радиусом ингибирования, равным  $\delta$  (рис. 21).

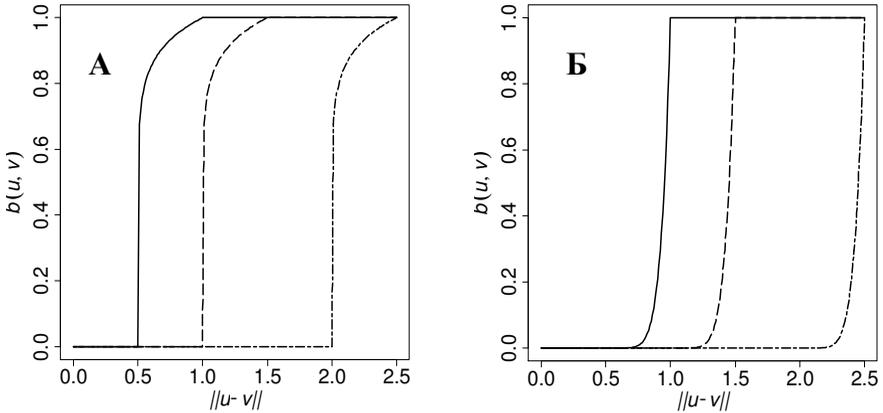


Рис. 21. Функция парного взаимодействия процесса Диггла-Граттона (Diggle-Gratton). По горизонтали откладывается расстояние между точками  $\|u - v\|$ . А: коэффициент  $k = 0.1$ ; Б:  $k = 7$ . Сплошная линия –  $h = 0.5, \delta = 1.0$ ; пунктирная линия –  $h = 1.0, \delta = 1.5$ ; штрих-пунктирная линия  $h = 2.0, \delta = 2.5$ .

## АНАЛИЗ ТОЧЕЧНЫХ ОБРАЗОВ

В данном разделе рассматриваются методы, применяемые для анализа точечных образов и моделирования точечных процессов, способных порождать наблюдаемые образы. В общем случае исследование точечного образа включает в себя следующие этапы:

1. Визуализация точечного образа - построение карт.
2. Оценка по наблюдаемому точечному образу свойств пространственного точечного процесса, порождающего этот образ:
  - оценка интенсивности процесса,
  - оценка парной корреляционной функции  $g(r)$ ,
  - оценка  $K$ -функции и  $L$ -функции,
  - оценка функций  $G(r)$ ,  $F(r)$ ,  $J(r)$ .
3. Проверка гипотезы о полной пространственной случайности (CSR) процесса, лежащего в основе наблюдаемого точечного образа. Классический анализ точечных образов всегда включает проверку этой гипотезы. Многие исследования сфокусированы на установлении того, что данные не согласуются с моделью CSR. Логика здесь следующая: если точечный образ полностью случаен, то он представляет собой «неинтересный» случай, так как местоположения событий совершенно не прогнозируемы, не имеют трендов или связей, сопряженности с чем-либо. Надо заметить, что полная пространственная случайность размещения событий наблюдается не так часто.

4. Если исследуемый образ показывает значимое отклонение от полной пространственной случайности (от CSR), то для дальнейшего исследования привлекают различные модели точечных процессов. Формулируются гипотезы о виде пространственного точечного процесса, способного создать наблюдаемый точечный образ; оцениваются параметры модели; проверяется соответствие точечного образа гипотетическому процессу.

Для проверки гипотез применяют статистические *критерии*, сопоставляющие оценки свойств первого или второго порядка, полученные по наблюдаемому точечному образу, и свойства гипотетического процесса. Рассчитанное значение критерия сравнивается с критическим значением при заданном уровне значимости (вероятности ошибки 1-го рода - отклонить верную нулевую гипотезу; обычно используют уровни 0.01, 0.05). Это сравнение позволяет принять или отвергнуть гипотезу. Критические значения можно найти двумя способами:

1). Если известен закон распределения, которому следует критерий в случае гипотетического процесса, и известны его параметры (обычно: матожидание, дисперсия, число степеней свободы), то, задавшись уровнем значимости, находят соответствующий квантиль этого распределения.

2). С помощью симуляций методом Монте-Карло. Критерий вычисляют на большом числе точечных образов, смоделированных в соответствии с гипотетическим процессом. Квантиль полученного эмпирического распределения статистики, при заданном уровне значимости, даст критическое значение. Заметим, что сами методы симуляции точечных процессов в данном пособии не рассматриваются.

Анализ точечных образов может быть выполнен с использованием пакета **spatstat** (Baddeley, Turner, 2005) среды статистического программирования R (R Core Team, 2014). Этот пакет содержит функции, реализующие большинство методов анализа точечных образов, и постоянно обновляется. Заинтересованные читатели также могут самостоятельно ознакомиться с другими пакетами, которые полезны при изучении пространственных точечных процессов: *spatial*, *MarkedPointProcess*, *spatgraphs*, *splancts*, *ecespa*, *ads*, *spatialkernel*, *Dspat*, *spatialsegregation*. Более полный список пакетов можно найти в обзоре CRAN “Analysis of Spatial Data” по адресу <http://cran.r-project.org/web/views/Spatial.html>.

Для учебных примеров в этом разделе используются:

- Данные о пространственном размещении стволов *Tilia cordata* L. (липы сердцелистной) на площадке размером 50 × 50 м, расположенной на южной границе Раифского участка Волжско-Камского государственного природного биосферного заповедника, Республика Татарстан (Чижикова, 2008). Для каждого ствола указаны координаты на площадке (колонки X и Y) и диаметр ствола (колонка *diameter*).
- Данные о пространственном размещении особей *Adonis vernalis* L. (адониса весеннего) на площадке размером 10 × 10 м в окрестности деревни Николашкино Бавлинского района Республики Татарстан (Фардеева, Саидова, 2010). Для каждой особи указаны координаты на площадке (колонки X и Y), указан код, обозначающий ее онтогенетическое состояние (возраст, колонка *age*), а также указано число цветков (колонка *blossoms*).

## ***Установка программы R***

Язык и среду статистического программирования R можно установить с сайта разработчиков <http://cran.r-project.org/> (The Comprehensive R Archive Network). R – это программная среда вычислений с открытым исходным кодом, разрабатываемая в рамках проекта GNU и распространяемая свободно. Библиотеки R поддерживают широкий спектр методов математической статистики, а также численных методов.

Для компьютеров с операционной системой Windows выберите ссылку «Download R for Windows» и далее ссылку «base». Выберите пункт «Download R 3.1.0 for Windows» – на ваш компьютер будет загружен исполняемый установочный файл R-3.1.0-win.exe. После загрузки файла запустите его и установите программу. Заметим, что на момент написания учебного пособия на странице загрузки доступной была эта версия программы, тогда как сейчас могут быть доступны более новые.

После установки программы потребуется загрузить дополнительные пакеты. Их можно найти, пройдя по ссылке «Packages» (раздел «Software» в левой части страницы). Выберите ссылку «Table of available packages, sorted by name», если необходимо установить пакеты, актуальные для текущей версии R. Если необходимы пакеты для более старых версий, выберите ссылку «bin/windows/contrib» и дальше перейдите в папку, соответствующую требуемой версии программы R.

В нашем случае требуется загрузить пакет **spatstat**, а также вспомогательные пакеты **mgcv**, **deldir**. Пакеты будут сохранены в виде zip архивов на вашем компьютере. Извлекать из архивов их не нужно.

Запустите программу R. Установите пакеты с помощью пункта меню «Пакеты» – «Установить пакеты из локальных zip-файлов».

### *Подготовительный этап*

В среде R все действия над данными выполняются с помощью набора команд в консольном окне «R console» (рис. 22). С основными принципами работы в среде R и командами можно ознакомиться в методическом пособии (Савельев и др., 2007).

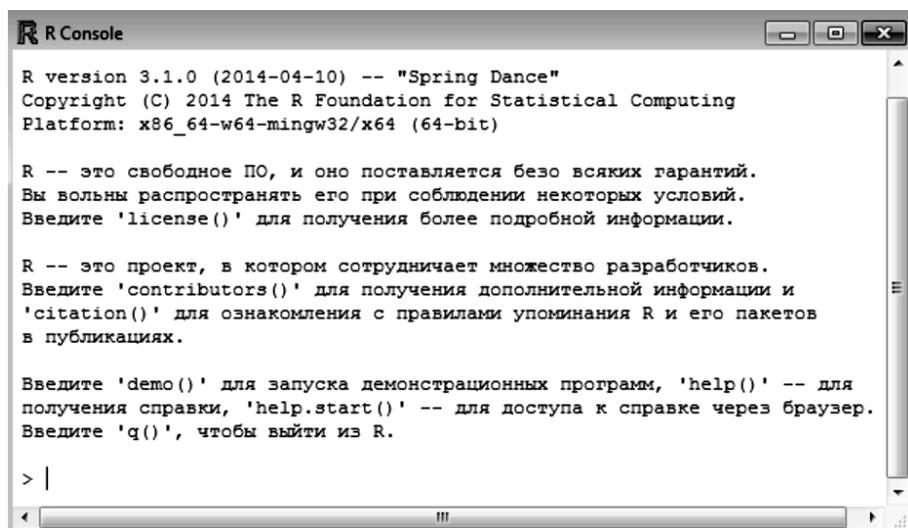


Рис. 22. Вид консольного окна в R.

Перед началом работы необходимо загрузить в память пакет **spatstat** командой:

```
library(spatstat)
```

Дополнительные пакеты, которые требуются для работы **spatstat**, загрузятся автоматически.

При необходимости с помощью команды «?» можно вывести справочную документацию о любой функции. К

примеру, команда:

```
?library
```

вызовет справочную страницу для функции `library`.

Далее для работы потребуется указать путь к рабочей папке, где хранятся данные об изучаемом точечном образе. Допустим, данные находятся в директории `D:\spatstat_course`:

```
setwd("D:/spatstat_course")
```

Загрузим из рабочей папки текстовый файл с данными о размещении стволов липы сердцелистной на исследуемой площадке. Допустим, он называется `plot_4.txt`. Создадим табличную переменную с именем `plot_4`, куда поместим считанные данные:

```
plot_4 <- read.table(file="plot_4.txt",  
                    header = T, sep = "\t", dec = ".")
```

Знак `<-` означает пересылку результата функции `read.table()`, записанной справа от этого знака, в переменную `plot_4`, записанную от него слева. Функция `read.table()` считывает данные из файла. Логическому параметру `header` задано значение `T` (`TRUE`, истина), это означает, что в первой строке файла содержатся заголовки столбцов. Параметр `sep="\t"` означает, что в качестве разделителей между колонками используется табулятор. Параметр `dec = "."` означает, что в качестве разделителя целой и дробной части чисел используется знак «.» – точка.

Просмотрим содержимое созданной табличной переменной `plot_4` во встроенном текстовом редакторе, убедимся, что считывание данных прошло успешно (рис. 23):

```
fix(plot_4)
```

	x	y	diameter	var4	var5	var6	var7
1	0.12	49.9	21				
2	7.02	49.53	1				
3	9.99	49.2	4				
4	11.02	49.16	12				
5	13.15	49.8	19				
6	10.05	46.21	9				
7	14.34	45.71	6				
8	5.91	45.82	6				
9	5.19	46.41	3				
10	2.58	41.69	2				
11	4.29	40.79	1				
12	3.43	40.24	1				
13	3.77	39.97	28				
14	0.61	40.01	2				
15	7.78	41.84	2				
16	7.24	40.09	7				
17	7.9	38.02	3				
18	11.92	40.52	19				
19	11.74	39.95	25				
20	11.96	39.48	25				

Рис. 23. Значения табличной переменной `plot_4` в окне встроенного редактора среды R.

Прежде чем начать анализ точечного образа средствами пакета **spatstat**, нужно конвертировать таблицу с данными в новый тип данных - объект точечного образа `ppp.object`, с которым работают процедуры и функции **spatstat**. Для этого предназначена команда `ppp()`, которая имеет следующие параметры:

x	Вектор координат событий по оси X.
y	Вектор координат событий по оси Y.
marks	Вектор или таблица (объект типа <code>data.frame</code> ), содержащие <b>метки</b> объектов. В нашем случае это таблица из одной колонки: диаметр ствола <i>diameter</i> .
window	Окно наблюдений (область исследования). Объект типа <code>owin</code> ( <code>observation window</code> ). В нашем случае это квадрат с координатами левого нижнего угла (0; 0), правого верхнего - (50; 50).
unitname	Единицы измерения координат, в примере – метры.

Дадим объекту произвольное название `ppp_object1`:

```
ppp_object1 <- ppp(x=plot_4$x, y=plot_4$y,  
                 marks= plot_4$diameter,  
                 window=owin(c(0,50), c(0,50)),  
                 unitname=c("metre", "metres"))
```

Обращение к созданному объекту покажет его основные свойства:

```
ppp_object1
```

```
marked planar point pattern: 306 points  
marks are numeric, of type 'double'  
window: rectangle = [0, 50] x [0, 50] metres
```

Можно видеть, что объект `ppp_object1` состоит из 306 событий (стволов). У каждого события помимо координат имеется числовая «метка». Изучаемая область прямоугольная, размером 50 м × 50 м. Эта информация содержится в различных элементах объекта, и к ней можно обратиться, используя знак `$`. Например, команда:

```
ppp_object1$window
```

```
window: rectangle = [0, 50] x [0, 50] metres
```

выведет сообщение, в котором указаны размеры изучаемой области (окна наблюдений). Полный перечень элементов, содержащихся в типе данных `ppp.object`, можно получить с помощью справки:

```
?ppp.object
```

Загрузим файл (`plot_13.txt`) с данными о размещении особей адониса весеннего на исследуемой площадке, используя табличную переменную с именем `plot_13`:

```
plot_13 <- read.table(file="plot_13.txt",  
                    header = T, sep = "\t", dec = ".")
```

Конвертируем ее в объект точечного образа `ppp.object` с именем `ppp_object2` и покажем его основные свойства:

```
ppp_object2 <- ppp(x=plot_13$x, y=plot_13$y,
  marks=data.frame(age=plot_13$age,
                    blossoms=plot_13$blossoms),
  window=owin(c(0,10), c(0,10)),
  unitname=c("metre", "metres"))
```

**ppp\_object2**

```
marked planar point pattern: 147 points
Mark variables:  age, blossoms
window: rectangle = [0, 10] x [0, 10] metres
```

## ***Визуализация точечного образа***

На начальном этапе исследования изучаемый точечный образ визуализируется в виде карты. Карта является удобным средством предварительного анализа пространственного размещения событий, выявления основных черт наблюдаемого пространственного рисунка. С помощью карт можно ответить на вопрос: наблюдается ли где-либо на изучаемой территории (и если да, то где именно) повышенная концентрация событий, или, наоборот, наблюдаются разреженные участки. Обнаружение таких пространственных особенностей, предваряющее проведение статистического анализа, базируется на визуальном восприятии близости событий, их концентрации, дискретности участков повышенной или пониженной плотности событий.

### ***Пример 1***

Построим карту точечного образа размещения стволов *Tilia cordata* L. для первой исследуемой площадки (рис. 24).

```
plot.ppp(x = ppp_object1, main = "Tilia cordata L.")
```

Функция `plot.ppp()` имеет следующие параметры:

x	Объект типа <code>ppp.object</code> , содержащий точечный образ.
main	Заголовок окна карты.

Добавим к карте оси:

```
axis(side=1, pos = 0, at = seq(0, 50, 5), las = 1)
axis(side=2, pos = 0, at = seq(0, 50, 5), las = 1)
axis(side=3, pos = 50, at = seq(0, 50, 5), labels = F)
axis(side=4, pos = 50, at = seq(0, 50, 5), labels = F)
```

Параметры функции `axis()`:

side	Сторона области построения, где рисовать ось. 1 – нижняя, 2 – левая, 3 – верхняя, 4 – правая.
pos	Координата позиции, где нужно нарисовать ось. Оси должны ограничивать график, поэтому нижняя и левая оси проходят через нулевые координаты, а верхняя и правая оси – через $X=50$ и $Y=50$ , соответственно.
at	Отметки оси, где нужно наносить деления.
las	Код, обозначающий ориентацию подписей отметок оси. 1 – всегда параллельно направлению чтения.
labels	Логический параметр, обозначающий, наносить ли подписи отметок. Для верхней и правой осей отметки отключены, <code>labels = F</code> .

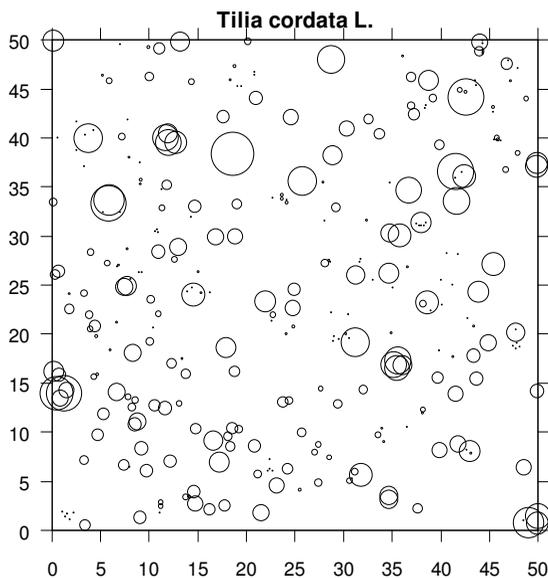


Рис. 24. Карта точечного образа: размещение стволов *Tilia cordata* L. Размер значков пропорционален диаметру стволов.

## Пример 2

Построим карту точечного образа для второй исследуемой площадки. Эти данные о пространственном размещении особей *Adonis vernalis* L. являются маркированным точечным образом. Поэтому при построении карты можно визуализировать не только местоположение особей, но их тематические метки - например, типом или размером значков (рис. 25).

```
plot.ppp(x = ppp_object2, use.marks = T,  
         which.marks = "age", chars = c(19, 24),  
         cols = c("black", "grey"),  
         cex = 1.5, main = "Adonis vernalis L.")
```

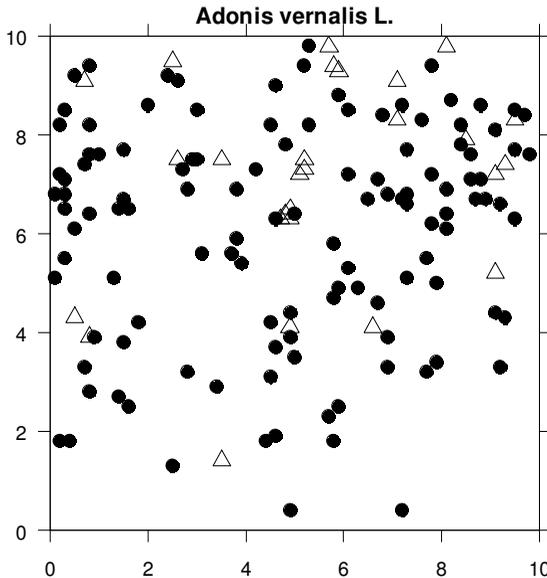


Рис. 25. Карта маркированного точечного образа: размещение особей *Adonis vernalis* L. Черные круги – генеративные особи (код *gene*), полые треугольники – прегенеративные особи (код *pre*).

Дополнительно использованы следующие параметры функции `plot.ppp()`:

use.marks	Логический параметр, который указывает, отображать или нет тематическую информацию (метки) на карте. В данном случае use.marks=T – отображать.
which.marks	Из какой колонки брать тематическую информацию. В данном случае age – категориальный возраст особей.
chars	Вектор, содержащий коды символов для разных категорий. Для генеративных особей ( <i>gene</i> ) указан код 19 – круги, для прегенеративных особей ( <i>pre</i> ) код 24 – полые треугольники.
sex	Размер символов для обозначения событий.

Точечные образы часто являются суперпозицией или наложением нескольких точечных образов, каждый из которых был сформирован под воздействием разных пространственных процессов. Эти процессы могут характеризоваться разной интенсивностью, разными типами взаимодействий между событиями. Например, при изучении древостоев наблюдалось, что для точечного образа, представляющего размещение деревьев верхнего яруса, характерна регулярность как следствие сильных конкурентных отношений. В то же время для молодых деревьев может наблюдаться образование групп, агрегирование. В качестве примеров можно также привести точечные образы размещения особей популяций разных видов или особей разных возрастных групп одного вида в локальном пятне популяции. В подобных случаях может оказаться полезным построение отдельных карт для каждого типа объектов.

### **Пример 3**

Построим отдельные карты для двух возрастных групп особей *Adonis vernalis* L. (рис. 26).

```

plot.ppp(x = ppp_object2[ppp_object2$marks=="gene"],
  which.marks = "age", chars=c(19, 24), cex=1.5,
  main = "Adonis vernalis L.")
plot.ppp(x = ppp_object2[ppp_object2$marks=="pre"],
  which.marks = "age", chars=c(19, 24), cex=1.5,
  main = "Adonis vernalis L.")

```

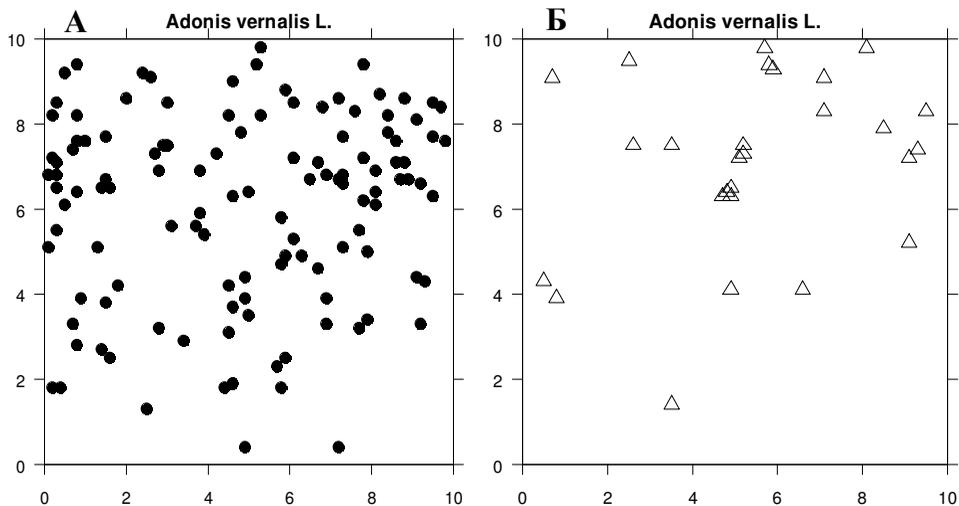


Рис. 26. Карты точечных образов:

А: размещение генеративных особей *Adonis vernalis* L.;

Б: размещение прегенеративных особей *Adonis vernalis* L.

Размещение каждой возрастной группы особей может быть вызвано процессами разной интенсивности.

### ***Оценка и анализ интенсивности***

Оценка интенсивности точечного процесса и визуализация (в виде карты) ее пространственных вариаций должны быть первым шагом анализа изучаемого точечного образа. Карта оценки интенсивности, отражающая локальную плотность событий, может помочь при выборе модели точечного процесса, порождающего наблюдаемый образ. Сама оценка используется как для проверки различных гипотез, так и при анализе и

моделировании зависимости интенсивности процесса от пространственных координат и/или от факторов среды.

В простейшем случае, если *a priory* известно или предполагается, что интенсивность процесса является константой ( $\lambda(x) \equiv \lambda$ ), ее можно оценить как число событий точечного образа ( $n$ ), деленное на площадь изучаемой области  $S (A_S)$ :

$$\hat{\lambda} = n/A_S .$$

В общем случае для получения оценки интенсивности могут применяться рассматриваемые ниже методы: метод квадратов и метод ядерной аппроксимации.

### **Метод квадратов**

Это один из первых методов, разработанных для исследования точечных рисунков и оценки интенсивности. Он рассматривает в изучаемой области *неперекрывающиеся* участки фиксированного размера, где подсчитывается число событий. «Квадраты» могут иметь форму не только квадрата, но и круга, могут быть треугольными, гексагональными или произвольной формы. Какая бы форма ни была выбрана, «квадраты» должны иметь *одинаковую площадь*. Эти участки могут быть заложены случайно (тогда форма участка обычно круг) и не обязательно должны граничить друг с другом; часто они образуют сетку квадратов. Метод очень чувствителен к размеру участков и их расположению. Размер должен быть таким, чтобы обеспечить попадание достаточного числа событий в участки для репрезентативности выборок.

Самый простой путь определить участки и рассмотреть пространственный аспект поведения интенсивности – это поделить изучаемую область сеткой прямоугольных (или квадратных)

ячеек одинаковой площади. Сетка из прямоугольных ячеек также используется, но значительно реже.

Интенсивность оценивается в каждой  $i$ -ой ячейке как локальная плотность событий:

$$\hat{\lambda}_i = n_i / A_Q,$$

где  $n_i$  - число событий точечного образа, попавших в  $i$ -ую ячейку,  $A_Q$  – площадь ячейки. Этот подход позволяет получить простейшую (грубую) оценку распределения интенсивности в исследуемой области.

Размер ячеек сетки выбирается исследователем как компромисс между точностью и пространственной неопределенностью оценки. Ячейки большого размера дадут обобщенную картину, характеризующую широкомасштабные вариации интенсивности, тогда как малый размер ячеек приводит к «пятнистым картам» (рис. 27).

Существует эмпирическое правило (правило «большого пальца», rule of thumb), основанное на свойствах пуассоновского процесса и помогающее выбрать параметры сетки. Размер ячеек сетки выбирается таким, чтобы  $\mu$  - среднее число событий, попавших в ячейку, равнялось 1.6. Для однородного пуассоновского процесса это соответствует тому, что примерно 20% ячеек будут пустыми (при  $\mu=1.6$  согласно формуле (1) вероятность встретить в такой подобласти 0 событий равна  $P(N_X(Q) = 0) \approx 0.20$ ). Можно использовать сетки со средним числом событий в ячейках, равным 1 (40% пустых ячеек), или 4 (2% пустых ячеек) и т.п. Тогда площадь  $A_Q$  ячейки можно найти по следующей формуле:  $A_Q = \mu \cdot A_S / n$ . Если сетка состоит из

квадратных ячеек, то размер стороны  $L_Q$  одной ячейки будет равен:

$$L_Q = \sqrt{\mu \cdot \frac{A_s}{n}}. \quad (2)$$

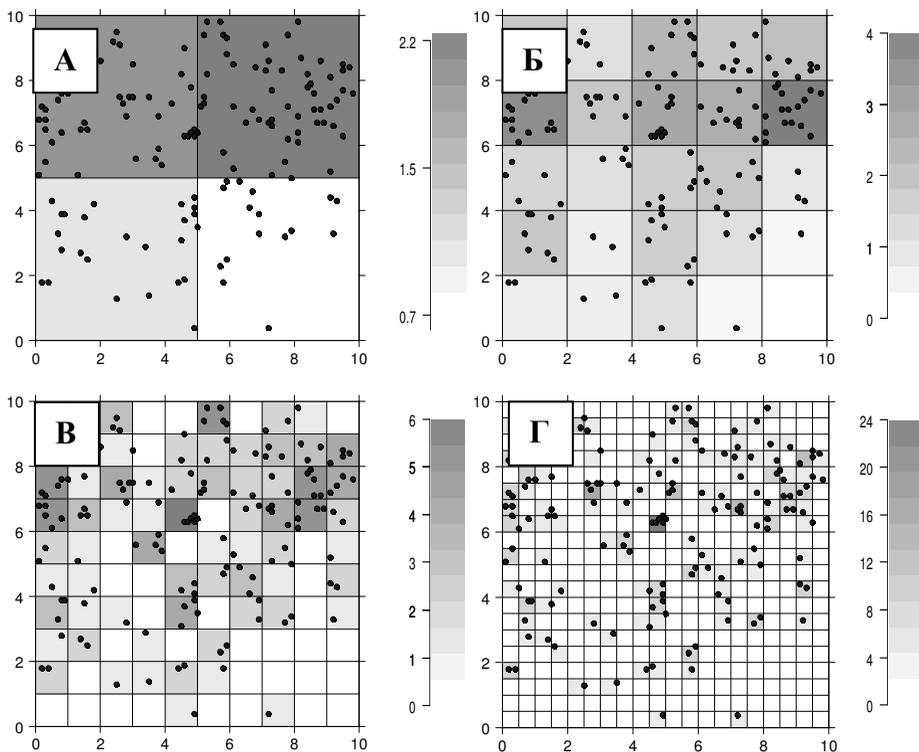


Рис. 27. Оценка интенсивности методом квадратов: карты локальных плотностей, полученные с помощью сеток квадратов разного размера.

А: сетка  $2 \times 2$ , размер ячеек 5 м  $\times$  5 м;

Б: сетка  $5 \times 5$ , размер ячеек 2 м  $\times$  2 м;

В: сетка  $10 \times 10$ , размер ячеек 1 м  $\times$  1 м;

Г: сетка  $20 \times 20$ , размер ячеек 0.5 м  $\times$  0.5 м.

#### Пример 4

С помощью метода квадратов на основе точечного образа размещения особей *Adonis vernalis* L. оценим интенсивность процесса, порождающего этот образ.

1) Зададим параметры сетки: пусть  $\mu = 2$ ; определим рекомендуемый размер ячейки по формуле (2), а также число строк и столбцов сетки:

```
A <- area.owin(ppp_object2$window)
n <- ppp_object2$n
L <- sqrt(2*A/n)
quadnum_x <-
round(as.numeric(dist(ppp_object2$window$xrange))/L)
quadnum_y <-
round(as.numeric(dist(ppp_object2$window$yrange))/L)
quadnum_x
9
quadnum_y
9
L <- as.numeric(dist(ppp_object2$window$xrange)/
quadnum_x)
L
1.111111
```

Функция `area.owin()` вычисляет площадь компоненты `window` объекта `ppp.object`.

Компонента `ppp_object2$n` содержит значение числа событий точечного образа (в нашем случае - число особей).

Переменные `quadnum_x` и `quadnum_y` – число строк и число столбцов сетки, вычисленные как отношение длины соответствующей стороны исследуемой площадки к рекомендуемой длине стороны ячейки. С учетом того, что в нашем случае исследуется квадратная площадка размером  $10 \times 10$ , число строк и столбцов сетки одинаково.

Переменная  $L$  содержит вычисленную длину стороны ячейки сетки.

2) Вычислим  $\hat{\lambda}_i$  - локальные плотности событий в ячейках сетки, и визуализируем пространственное поведение этой оценки интенсивности, построив карту (рис. 28):

```
quadcount <- quadratcount(ppp_object2, nx=quadnum_x,
                           ny=quadnum_y)/L/L
xgrid <- quadrats(ppp_object2, nx = quadnum_x,
                 ny = quadnum_y)$xgrid
ygrid <- quadrats(ppp_object2, nx = quadnum_x,
                 ny = quadnum_y)$ygrid
image(xgrid, ygrid, axes=F,
      t(quadcount[order(1:quadnum_x, decreasing=T)], ),
      col=colorRampPalette(c("white", "grey20"))(15))
```

Функция `quadratcount()` вычисляет число событий, попавших в ячейки сетки (квадраты). Ее параметры  $nx$  и  $ny$  задают размер сетки (число столбцов и строк).

Функция `image()` строит растровый рисунок (карту), где элементами растра являются ячейки сетки объекта `quadcount`. Функция `colorRampPalette()` создает шкалу цветового перехода от белого цвета к серому, состоящую из 15 градаций.

3) Добавим на карту оси и подписи, обозначающие значения  $\hat{\lambda}_i$  в ячейках:

```
axis(2, las=1, pos=0, at = seq(0,10,2), cex.axis=1.5)
axis(1, las=1, pos=0, at = seq(0,10,2), cex.axis=1.5)
axis(3, pos=10, at=seq(0,10,2), labels=F, cex.axis=1.5)
axis(4, pos=10, at=seq(0,10,2), labels=F, cex.axis=1.5)
plot(round(quadcount, 1), add=T, cex=1.5)
```

10	2.4	0	2.4	0	2.4	2.4	0.8	1.6	0
8	2.4	0.8	0.8	0	2.4	1.6	3.2	4	3.2
6	5.7	1.6	4	2.4	2.4	1.6	4	4.9	4
4	2.4	1.6	0.8	1.6	4.9	0.8	0.8	2.4	1.6
2	1.6	0.8	0	0.8	0	3.2	2.4	0.8	0.8
0	2.4	1.6	0	0	4.9	0.8	0.8	0.8	1.6
	1.6	1.6	0.8	0.8	0.8	1.6	1.6	0	0.8
	1.6	0	0.8	1.6	0.8	0.8	0	0	0
	0	0	0	0	0.8	0	0.8	0	0
	0	2	4	6	8	10			

Рис. 28. Оценка интенсивности методом квадратов: карта локальной плотности особей *Adonis vernalis* L., полученная на сетке 9×9.

### Ядерная аппроксимация

Для того, чтобы оценить интенсивность точечного процесса можно воспользоваться непараметрическим *методом ядерной аппроксимации* (kernel method). В общем случае этот метод носит название метода окна Парзена (или метода Парзена-Розенблатта). С помощью ядерной аппроксимации можно получить более сглаженную оценку интенсивности (по сравнению с оценкой, даваемой методом квадратов).

Пусть  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  – наблюдаемый точечный образ. В произвольной точке  $x \in S$  анализируемой области (обычно в узлах достаточно частой сетки) оценка интенсивности пространственного точечного процесса вычисляется как:

$$\hat{\lambda}(x) = \sum_{i=1}^n k_h(x - x_i).$$

Здесь  $k_h(\cdot)$  - специальная *ядерная функция* (функция ядра или окна; kernel function),  $h$  – масштабный параметр, определяющий размер ядра (окна с центром в точке  $x$ ). Вклад каждого события  $x_i$  точечного образа в оценку интенсивности в точке  $x$  зависит от расстояния между точкой  $x$  и этим событием (рис. 29). Как правило, чем расстояние меньше, тем вклад больше, и наоборот. Какая именно будет зависимость - определяется видом ядерной функции и значением ее параметра  $h$ .

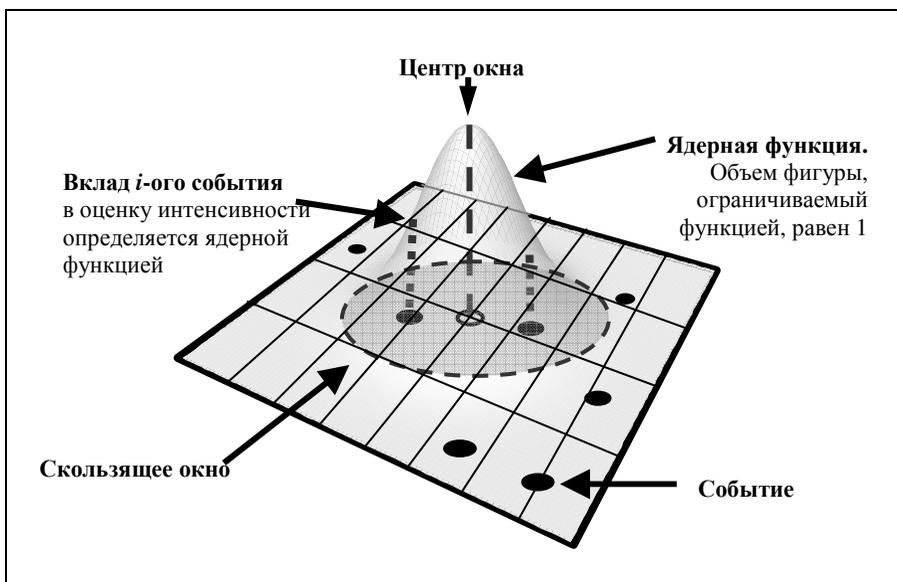


Рис. 29. Иллюстрация ядерной функции. В каждом узле сетки рассчитывается оценка интенсивности  $\hat{\lambda}(x)$ : центр окна (ядра) помещается в текущий узел сетки (точка  $x$ ); вклад в оценку (вес) каждого события определяется ядерной функцией  $k_h(x - x_i)$ . Обычно, чем дальше  $i$ -ое событие от центра окна, тем меньший вклад оно вносит в оценку интенсивности в этой точке.

Для оценки интенсивности могут быть использованы различные ядерные функции. Оптимальной считается ядерная функция Епанечникова (рис. 30 А):

$$k_h(x - x_i) = \begin{cases} \frac{2}{\pi} \left(1 - \frac{\|x - x_i\|^2}{h^2}\right), & \text{если } \|x - x_i\| \leq h \\ 0, & \text{во всех остальных случаях} \end{cases}$$

Широко распространенным является использование ядерной функции на основе Гауссова распределения (рис. 30 Б):

$$k_h(x - x_i) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|x - x_i\|^2}{2\sigma^2}},$$

где  $\sigma$  – это среднеквадратическое отклонение, выступающее в качестве масштабного параметра  $h$ . Сопоставимые значения масштабных параметров для этих двух функций, определяющие размер скользящего окна, которое устанавливает, какие события будут вносить вклад (или основной вклад) при вычислении оценки, даются соотношением  $h \approx 2\sigma$ .

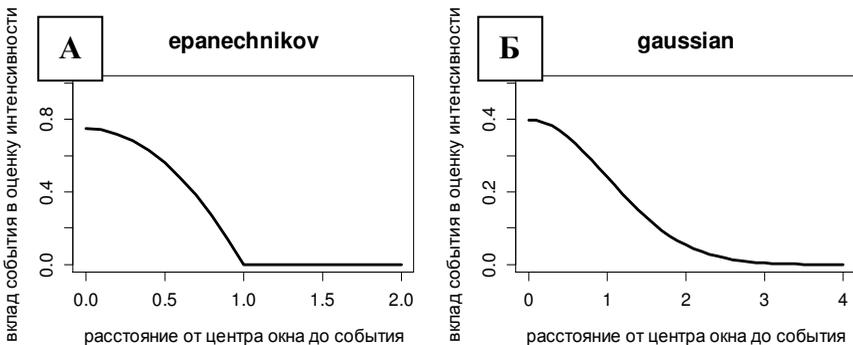


Рис. 30. Ядерные функции «в разрезе».

А: epanechnikov – ядерная функция Епанечникова ( $h=1$ );

Б: gaussian – Гауссова ядерная функция ( $\sigma=1$ ).

Существуют и другие ядерные функции, однако выбор функции  $k(\cdot)$  не столь важен. На рис. 31 представлены оценки

интенсивности, полученные с помощью ядерных функций Епанечникова и Гаусса с сопоставимыми размерами ядер. Видно, что эти карты довольно схожи.

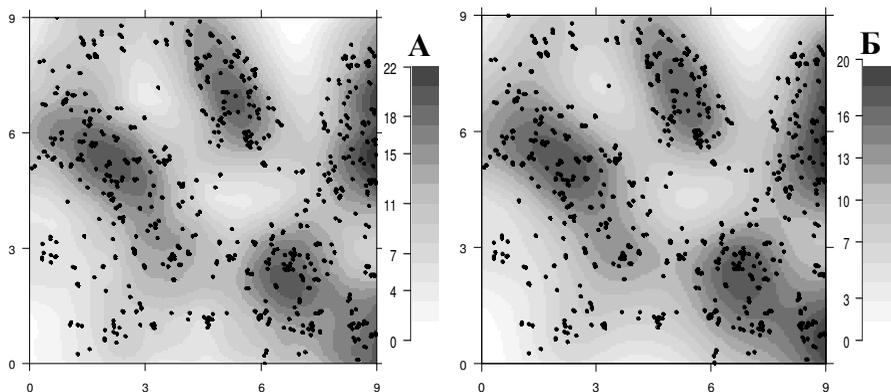


Рис. 31. Оценка интенсивности с помощью ядерных функций. А: Епанечникова ( $h=1.6$  м); Б: Гауссовой ( $\sigma=0.8$  м). Используются данные о проростках ежевики (Hutchings, 1979; Diggle 1983).

Оценка интенсивности гораздо чувствительнее к выбору размера ядра  $h$ , определяющего степень пространственной сглаженности получаемой оценки  $\hat{\lambda}(x)$ . Ядро малого размера будет приводить к занижению оценки во многих точках области исследования, и карта интенсивности будет выглядеть «пятнистой», из-за чего сложно увидеть общие закономерности (рис. 32 А). Ядро большого размера, наоборот, даст сглаженную оценку интенсивности, где будут нивелированы ее мелкие пространственные вариации, которые могут представлять интерес (рис. 32 Б). Таким образом, размер ядра должен выбираться так, чтобы соблюсти компромисс между точностью и генерализацией, позволяя увидеть как глобальный тренд интенсивности, так и ее локальные вариации.

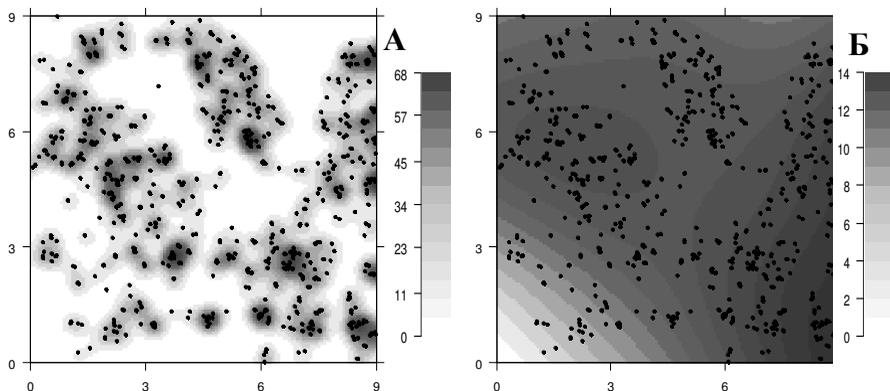


Рис. 32. Оценка интенсивности с помощью Гауссовой ядерной функции. Используются данные о проростках ежевики (Hutchings, 1979; Diggle 1983).

А: выбран недостаточно большой масштабный параметр  $\sigma=0.2$  м, сложно увидеть общую закономерность.

Б: выбранный масштабный параметр  $\sigma=2$  м позволяет проследить глобальный тренд интенсивности, но ее локальные вариации отсутствуют в общей картине.

Существует несколько правил «большого пальца» для определения оптимального размера Гауссова ядра, основанных на предположении о характере распределения интенсивности и обычно применяемых для гладкого поведения:

$$\sigma_{\text{оптим}} = 1.06 \cdot \sigma \cdot n^{-0.2} \quad \text{или} \quad \sigma_{\text{оптим}} = 0.79 \cdot IQR \cdot n^{-0.2} \quad \text{или}$$

$$\sigma_{\text{оптим}} = 0.9 \cdot \min\left(\sigma, \frac{IQR}{1.34}\right) \cdot n^{-0.2}, \quad (3)$$

где  $n$  – число событий точечного образа;  $\sigma$  – выборочное среднеквадратическое отклонение координат событий,  $IQR$  – межквартильный размах (interquartile range);  $\sigma$  и  $IQR$  вычисляются с учетом изменчивости обеих координат,  $X$  и  $Y$ :

$$\sigma = \frac{\sigma(X_{\text{координата}}) + \sigma(Y_{\text{координата}})}{2},$$

$$IQR = \frac{IQR(X_{\text{координата}}) + IQR(Y_{\text{координата}})}{2}.$$

При этом размер ядра, основанный на  $\sigma$ , в случае наличия выбросов может быть завышен, поэтому более надежным является использование  $IQR$ .

В таком же предположении гладкого распределения может быть определен масштабный параметр  $h$  для ядерной функции Епанечникова:

$$h_{\text{оптим}} = 1.76 \cdot \sigma \cdot n^{-\frac{1}{6}}.$$

Также для данного ядра существует вариант, приведенный в книге Cressie (1993):

$$h_{\text{оптим}} = 0.68 \cdot n^{-0.2}.$$

Кроме указанных эмпирических правил для выбора оптимального размера ядра может быть использован метод перекрестной проверки. Например, согласно одной из модификаций этого метода в точке каждого  $i$ -ого события оценка интенсивности с помощью Гауссовой ядерной функции вычисляется двумя способами: с учетом этого события и за исключением этого события. Размер ядра  $\sigma$  подбирается таким, чтобы минимизировать средний квадрат разности этих оценок по всем событиям.

Еще одна особенность, которую необходимо учитывать при оценке интенсивности точечного процесса, это искажение или смещение оценки на границах изучаемой территории, так называемый *граничный эффект* (edge effect). Когда точка  $x$  (центр скользящего окна) находится близко к границе изучаемой области, это окно частично покрывает пространство, где исследования не проводились, и где отсутствует объективная

информация о наличии событий, из-за чего оценка интенсивности занижается. Рассмотрим некоторые из существующих подходов граничной коррекции оценки (рис. 33):

- Использование корректирующего множителя  $p(x)$  при вычислении оценки:

$$\hat{\lambda}(x) = \frac{1}{p(x)} \cdot \sum_i k_n(x - x_i),$$

где

$$p(x) = \int_S k(x - u) du.$$

Значение корректирующего множителя  $p(x)$  равно объему фигуры, ограниченной ядерной функцией и площадью ядра, попадающей на анализируемую территорию  $S$  (рис. 33 А). Если ядро полностью находится внутри  $S$ , то  $p(x) = 1$  – коррекции нет. Иначе, когда ядро пересекается с границей,  $p(x) < 1$ , что приведет к коррекции оценки – ее повышению.

- Построение буферной зоны вдоль границы изучаемой области, в которую не должен попадать центр скользящего окна. Ширина этой зоны определяется в соответствии с масштабным параметром ядерной функции. Оценка  $\hat{\lambda}(x)$  будет вычисляться только для точек, не попадающих в буферную зону (при этом находящиеся в ней события используются при вычислении оценок для точек центральной подобласти). Понятно, что размер эффективной площади, где вычисляется  $\hat{\lambda}(x)$ , уменьшается (рис. 33 Б). Этот подход экстенсивен и редко используется, так как приемлем лишь при работе с большими площадями, когда потеря буферной территории не столь ощутима.

- Тороидальная репликация. Наблюдаемая реализация процесса - точечный образ - реплицируется в горизонтальном и вертикальном направлениях (рис. 33 В). Оценка  $\hat{\lambda}(x)$  вычисляется только для точек  $x$  изучаемой области, однако в расчетах используются и реплицированные события.

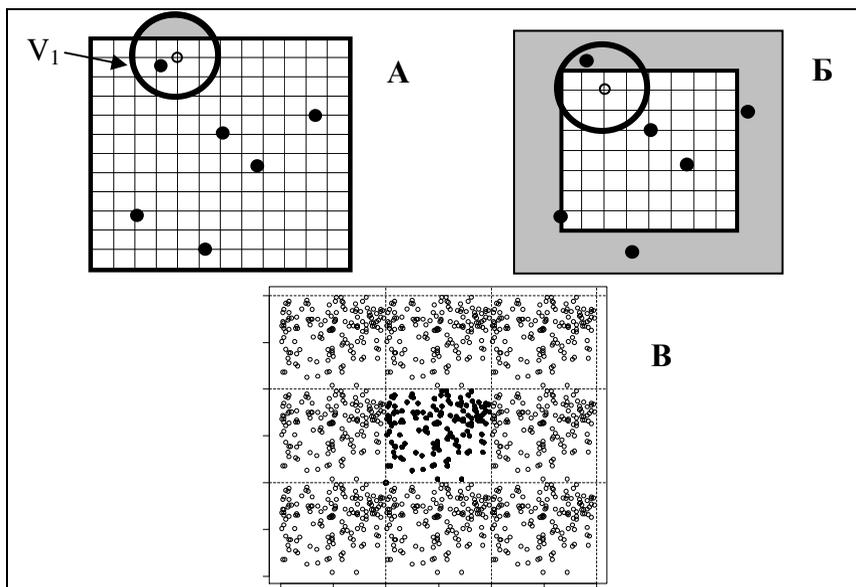


Рис. 33. Методы граничной коррекции при оценке интенсивности.

А: Оценка интенсивности в узлах сетки делится на объем фигуры, ограниченной ядерной функцией и частью  $V_1$  скользящего окна, попадающей в область исследования.

Б: Оценка вычисляется с учетом буферной зоны (серая заливка). События, попавшие в буферную зону, участвуют в оценке интенсивности. Интенсивность оценивается только в узлах сетки, не попадающих в буферную зону.

В: Тороидальная репликация. Черные точки соответствуют наблюдаемой реализации процесса, полые точки соответствуют репликациям. Оценка интенсивности вычисляется только для точек исходной области, при этом учитываются и реплицированные события.

На рис. 34 приведены оценки интенсивности без граничной коррекции и с граничной коррекцией. Можно видеть, что коррекция изменила оценку вблизи границ изучаемой площадки.

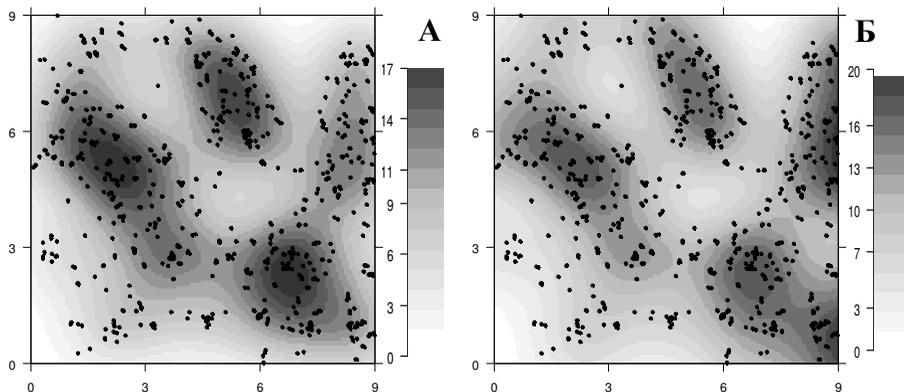


Рис. 34. Оценка интенсивности на основе Гауссовой ядерной функции,  $\sigma=0.8$  м.

А: без граничной коррекции,

Б: с граничной коррекцией.

Использованы данные о проростках ежевики (Hutchings, 1979; Diggle 1983).

### **Пример 5**

С помощью Гауссовой ядерной функции оценим интенсивность пространственного точечного процесса, порождающего точечный образ размещения стволов *Tilia cordata* L. Будем использовать граничную коррекцию с помощью корректирующего множителя.

1) Определим оптимальное значение масштабного параметра  $\sigma_{\text{оптим}}$  по формуле (3):

```
sigma <- (sd(ppp_object1$x) + sd(ppp_object1$y)) / 2
iqr <- (IQR(ppp_object1$x) + IQR(ppp_object1$y)) / 2
sigma_optim1 <- 0.9 * min(sigma, iqr) * ppp_object1$n^(-1/5)
```

Получим:  $\sigma_{\text{оптим1}} = 4.20$  м.

2) Вычислим оценку интенсивности:

```
ppp_density1 <- density.ppp(ppp_object1, sigma_optim1,  
                             edge=TRUE)
```

Параметры функции `density.ppp()`:

<code>ppp_object1</code>	Объект точечного образа типа <code>ppp.object</code> .
<code>sigma_optim1</code>	Значение параметра $\sigma$ в формуле Гауссова ядра.
<code>edge</code>	TRUE – с граничной коррекцией, FALSE - без граничной коррекции.

Эта функция создает объект класса `im.object` (pixel image), который фактически представляет собой большую матрицу, содержащую значения полученной оценки интенсивности в узлах регулярной сетки, покрывающей область исследования.

3) Визуализируем пространственное поведение оценки интенсивности, построив карту:

```
image_color <- colorRampPalette(c("white", "grey",  
                                "grey20"))(15)  
image(ppp_density1, col = image_color)  
plot(ppp_object1, cex=1, add=T, use.marks =F)  
axis(2, las=1, pos=0, at=seq(0, 50, 5))  
axis(1, las=1, pos=0, at=seq(0, 50, 5))  
axis(3, pos=50, at=seq(0, 50, 5), labels=F)  
axis(4, pos=50, at=seq(0, 50, 5), labels=F)
```

Сначала с помощью функции `colorRampPalette()` создается шкала цветового перехода от белого цвета к серому, состоящая из 15 градаций. Затем функция `image()` строит карту, используя созданный объект `ppp_density1`, содержащий оценку интенсивности. Результат представлен на рис. 35.

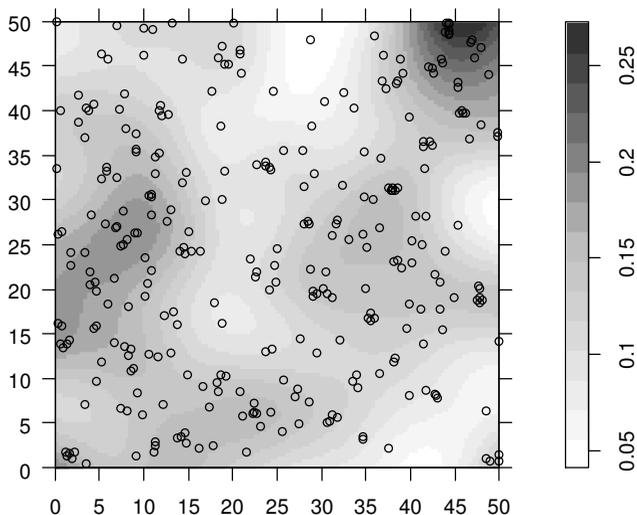


Рис. 35. Оценка интенсивности процесса, порождающего наблюдаемое размещение стволов *Tilia cordata* L. Получена с помощью Гауссовой ядерной функции; размер ядра  $\sigma=4.20$  м.

### Пример 6

Оценим интенсивность точечного процесса с помощью Гауссовой ядерной функции для примера размещения особей *Adonis vernalis* L.

1) Определим оптимальное значение масштабного параметра  $\sigma_{\text{оптим}}$  по формуле (3):

```
sigma <- (sd(ppp_object2$x) + sd(ppp_object2$y)) / 2
iqr <- (IQR(ppp_object2$x) + IQR(ppp_object2$y)) / 2
sigma_optim2 <- 0.9 * min(sigma, iqr) * ppp_object2$n^(-1/5)
```

Получим:  $\text{sigma\_optim2} = 0.86$  м.

2) Вычислим оценку интенсивности:

```
ppp_density2 <- density.ppp(ppp_object2, sigma_optim2,
                             edge=TRUE)
```

3) Построим карту оценки интенсивности (рис. 36):

```
image(ppp_density2, col = image_color)
```

```

plot(ppp_object2, which.marks="age",
     chars=c(19, 24), cex=1.5, add=T)
axis(2, las=1, pos=0, at=seq(0, 10, 2))
axis(1, las=1, pos=0, at=seq(0, 10, 2))
axis(3, pos=10, at=seq(0, 10, 2), labels=F)
axis(4, pos=10, at=seq(0, 10, 2), labels=F)

```

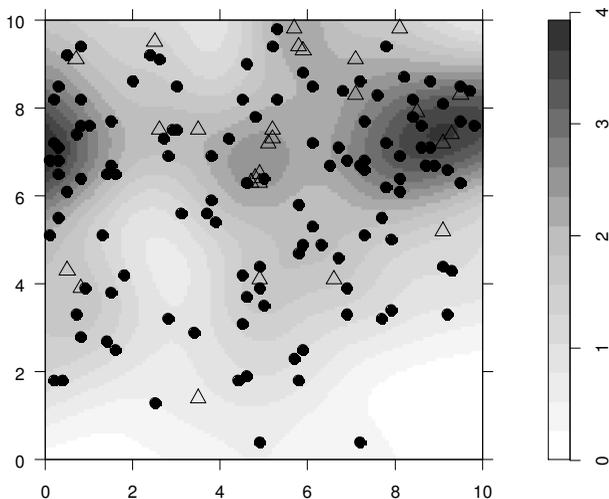


Рис. 36. Оценка интенсивности процесса, порождающего наблюдаемое размещение особей *Adonis vernalis* L. Оценка получена с помощью Гауссовой ядерной функции; размер ядра  $\sigma=0.86$  м. Черные круги – генеративные особи (код *gene*), полые треугольники – прегенеративные особи (код *pre*).

### Проверка гипотезы о CSR

Как уже говорилось, первой проверяемой гипотезой при исследовании точечных рисунков является гипотеза о полной пространственной случайности (CSR) процесса, лежащего в основе наблюдаемого точечного образа. Моделью полной пространственной случайности служит *однородный пуассоновский процесс* (интенсивность процесса постоянна во

всей области исследования, и между событиями нет взаимодействий).

Отклонение от CSR может проявляться:

- наличием *агрегаций* (скоплений) событий в каких-либо подобластях изучаемой области, что в свою очередь может быть следствием:
  - неоднородности точечного процесса, когда его интенсивность изменяется в пространстве,
  - позитивного взаимодействия событий, которое «притягивает» события друг к другу;
- наличием некоторой пространственной *регулярности* размещения событий как следствия негативного взаимодействия событий (события «избегают» друг друга, «отталкиваются» друг от друга; крайний случай – полная регулярность, регулярная сетка).

Заметим, что если анализировать точечный образ только визуально, то даже случайный рисунок, соответствующий однородному пуассоновскому процессу (рис. 7 А), может показывать локальную изменчивость со сгущениями и разрежениями событий. Для надежных статистических выводов при проверке гипотезы о CSR применяют статистические критерии (тесты), основанные на свойствах однородного пуассоновского процесса.

Полученная оценка интенсивности точечного процесса позволяет провести один из простых статистических тестов на CSR, использующий метод квадратов.

### ***Критерий хи-квадрат на основе «квадратов»***

Пусть  $m$  – число квадратов (ячеек) регулярной сетки, покрывающей изучаемую область,  $n_i$  – число событий в  $i$ -ом квадрате ( $i = 1, \dots, m$ ), подсчитанное для изучаемого точечного образа.

В предположении о CSR, числа событий в каждом квадрате – это независимые случайные величины, распределенные по закону Пуассона с одинаковым матожиданием, которое оценивается как

$$\hat{\mu} = n \cdot A_Q / A_S = n / m,$$

где  $n$  – число всех событий точечного образа,  $A_Q$  – площадь ячейки,  $A_S$  – площадь всей области исследования.

Рассматриваемый критерий проверки гипотезы о CSR использует сравнение наблюдаемого ( $n_i$ ) и ожидаемого ( $\hat{\mu}$ ) числа событий в каждой ячейке, для чего применяется статистика, распределенная по закону хи-квадрат:

$$\chi^2 = \sum_{i=1}^m \frac{(n_i - \hat{\mu})^2}{\hat{\mu}} = (m-1) \frac{s^2}{\hat{\mu}}.$$

где  $s^2$  – выборочная дисперсия числа событий в ячейках, она равна:

$$s^2 = \frac{1}{m-1} \sum_{i=1}^m (n_i - \hat{\mu})^2.$$

Если полученное значение  $\chi^2$  превосходит критическую точку распределения хи-квадрат с  $(m-1)$  степенью свободы на заданном уровне значимости  $\alpha$ , то гипотеза о CSR для наблюдаемого точечного образа отвергается. В противном случае гипотеза принимается на уровне значимости  $\alpha$ .

Мощность теста (вероятность ошибки второго рода – принять неправильную нулевую гипотезу) зависит от размера квадратов, и приближается к нулю в случае слишком больших или слишком малых квадратов (вероятность ошибки второго рода высока). Мощность также зависит от альтернативной гипотезы, в частности, от пространственного масштаба отклонений от предполагаемых постоянства интенсивности и независимости событий. Выбор размера квадратов обусловлен этим пространственным масштабом.

Необходимо заметить, что рассматриваемый критерий основан на свойствах первого порядка однородного пуассоновского процесса и фактически проверяет постоянство интенсивности в предположении независимости событий. Если согласно данному критерию гипотеза о CSR принимается, то далее необходимо исследовать свойства второго порядка, отражающие пространственные взаимодействия событий.

Если данный критерий отвергает гипотезу о CSR, то это может свидетельствовать как о неоднородности процесса (интенсивность непостоянна), так и о наличии взаимодействия между событиями, приводящего к неравномерной плотности событий в точечном образе. В этом случае при продолжении исследований привлекаются модели других пространственных точечных процессов.

### **Пример 7**

Проверим гипотезу о CSR размещения стволов *Tilia cordata* L. с помощью критерия хи-квадрат на основе «квадратов». Этот тест реализован в функции `quadrat.test()` пакета **spatstat**.

### Параметры функции `quadrat.test()`:

<code>ppp_object</code>	Объект типа <code>ppp.object</code> , содержащий точечный образ.
<code>nx</code>	Размер сетки квадратов по оси X.
<code>ny</code>	Размер сетки квадратов по оси Y.

Вызовите «Справку», чтобы узнать о других возможных параметрах функции.

Будем использовать сетку 5×5 квадратов.

```
quadttest <- quadrat.test(ppp_object1, nx=5, ny=5)
quadttest
      Chi-squared test of CSR using quadrat counts
      Pearson X2 statistic
data: ppp_object1
X2 = 20.634, df = 24, p-value = 0.6796
alternative hypothesis: two.sided
Quadrats: 5 by 5 grid of tiles
```

Достижимый уровень значимости (p-value) больше 0.05 (обычно задаваемый уровень значимости), следовательно, нулевая гипотеза о соответствии анализируемого точечного образа полной пространственной случайности (CSR) принимается. То есть, данный критерий не обнаружил противоречия тому, что размещение стволов *Tilia cordata* L. на исследуемой площадке является реализацией однородного пуассоновского процесса. И пространственное варьирование оценки интенсивности, которое можно видеть на рис. 35, либо незначимо и отражает выборочную случайность, либо объясняется эффектами второго порядка. Исследование этого точечного образа должно быть продолжено с помощью методов, анализирующих отклонение от CSR, связанное с пространственным взаимодействием событий.

### Пример 8

С помощью критерия хи-квадрат на основе «квадратов» проверим гипотезу о CSR для процесса, порождающего точечный образ размещения особей *Adonis vernalis* L.

Воспользуемся сеткой 9×9 квадратов, построенной в примере 4.

```
quadtest <- quadrat.test(ppp_object2, nx=quadnum_x,
                        ny=quadnum_y)
quadtest
Chi-squared test of CSR using quadrat counts
data: ppp_object2
X2 = 127.9592, df = 80, p-value = 0.001064
Quadrats: 9 by 9 grid of tiles
```

Посмотрим, как изменится результат, если использовать сетку 5×5 квадратов.

```
quadtest <- quadrat.test(ppp_object2, nx=5, ny=5)
quadtest
Chi-squared test of CSR using quadrat counts
data: ppp_object2
X2 = 65.0748, df = 24, p-value = 2.374e-05
Quadrats: 5 by 5 grid of tiles
```

В обоих случаях достигаемый уровень значимости меньше 0.01, следовательно, нулевая гипотеза о соответствии анализируемого точечного образа полной пространственной случайности (CSR) отвергается с надежностью не менее 99%.

Для наглядности построим карту отклонений наблюдаемого числа событий от ожидаемого в случае CSR на сетке 5×5 квадратов (рис. 37 Б).

```
xgrid <- quadrats(ppp_object2, nx = 5, ny = 5)$xgrid
ygrid <- quadrats(ppp_object2, nx = 5, ny = 5)$ygrid
ressiduals2plot <- matrix(quadtest$residuals,
                        ncol=5, nrow=5)
image(xgrid, ygrid, ressiduals2plot[,5:1],
```

```
col= colorRampPalette(c("white", "grey20"))(15),
axes=F, asp=1)
plot(quadtest, add=T)
```

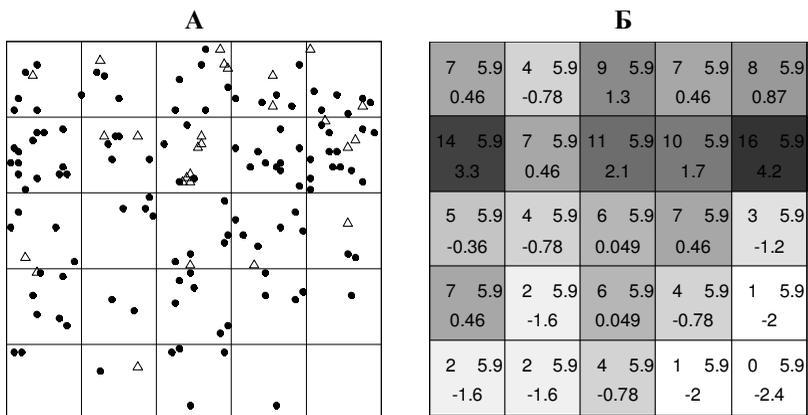


Рис. 37. А: Карта точечного образа размещения особей *Adonis vernalis* L. с сеткой 5×5 квадратов.

Б: В левом углу квадратов показано наблюдаемое число событий (особей)  $n_i$ , в правом углу – ожидаемое число событий в случае CSR  $\hat{\mu}$  ( $\approx 5.9$ ). Во второй строке показано «пирсоновское» отклонение наблюдаемого числа событий от ожидаемого  $(n_i - \hat{\mu}) / \sqrt{\hat{\mu}}$  (сумма квадратов этих значений равна статистике  $\chi^2$ ). Интенсивность заливки квадратов соответствует отклонению.

Таким образом, размещение особей *Adonis vernalis* L. на исследуемой площадке не может считаться реализацией однородного пуассоновского процесса. Причиной этого может быть как наличие значимых агрегаций особей адониса вследствие взаимодействия событий, так и пространственный тренд интенсивности процесса (плотность событий увеличивается к северной границе площадки). Исследование этого образа должно быть продолжено с использованием моделей других процессов.

## Проверка гипотезы о неоднородном пуассоновском процессе

Если при применении критерия хи-квадрат на основе «квадратов» гипотеза о CSR отвергается, то это в частности может свидетельствовать о неоднородности процесса (интенсивность изменяется по территории). И следующей тестируемой гипотезой может быть гипотеза о соответствии точечного образа *неоднородному* пуассоновскому процессу (интенсивность процесса  $\lambda(x)$  варьирует на изучаемой территории, и между событиями нет взаимодействий).

Эта гипотеза формулируется с указанием функции  $\lambda(x)$ . В качестве такой функции может быть использована модельная зависимость интенсивности процесса от каких-либо пространственных переменных (внешних факторов, свойств территории), поведение которых для изучаемой области известно. Например, интенсивность процесса, порождающего размещение деревьев, может зависеть от характеристик рельефа. В простейшем случае в качестве таких переменных в модели могут участвовать сами координаты.

В пакете **spatstat** представлен метод построения параметрической модели зависимости интенсивности точечного процесса от внешних переменных. Интенсивность моделируется как логарифмически линейная функция относительно параметра  $\Theta$ :

$$\log \lambda(x) = \Theta \cdot f(x)$$

где  $f(x)$  - функция от местоположения  $x$ . На практике  $f(x)$  может быть либо функцией от координат, либо пространственной переменной, либо композицией того и другого. Параметры модели определяются методом максимального правдоподобия.

Если построенная модель интенсивности  $\lambda(x)$  адекватна данным, то для проверки гипотезы о соответствии наблюдаемого точечного образа неоднородному пуассоновскому процессу также может быть использован критерий хи-квадрат на основе «квадратов», сравнивающий наблюдаемое число событий в квадратах и ожидаемое. В предположении о неоднородном пуассоновском процессе с интенсивностью  $\lambda(x)$  ожидаемое число событий в каждом  $i$ -ом квадрате будет зависеть от значений  $\lambda(x)$  в этом квадрате:

$$\mu_i = \int_{Q_i} \lambda(x) dx ,$$

и критерий хи-квадрат вычисляется как:

$$\chi^2 = \sum_{i=1}^m \frac{(n_i - \mu_i)^2}{\mu_i} .$$

Если полученное значение  $\chi^2$  превосходит критическую точку распределения хи-квадрат с  $(m-1)$  степенью свободы на заданном уровне значимости  $\alpha$ , то гипотеза для наблюдаемого точечного образа отвергается. Иначе гипотеза принимается на уровне значимости  $\alpha$ .

### **Пример 9**

Для точечного образа размещения особей *Adonis vernalis* L. критерий хи-квадрат на основе «квадратов» показал отклонение от CSR (пример 8). В данном случае причиной этого может быть пространственный тренд интенсивности процесса - плотность событий увеличивается к северной границе площадки (см. карту оценки интенсивности рис. 36).

1) Построим модель зависимости интенсивности пуассоновского процесса от координаты  $Y$  с помощью функции `ppm()`, реализованной в пакете `spatstat`.

Параметры функции `ppm()`:

<code>ppp_object</code>	Объект типа <code>ppp.object</code> , содержащий точечный образ.
<code>trend</code>	R-формула, задающая моделируемую зависимость.

Метод использует логарифмическую функцию связи, то есть формула определяет логарифм интенсивности. Обычно, как уже говорилось, она включает в себя пространственные координаты и/или пространственные переменные, предположительно объясняющие изменчивость интенсивности процесса.

```
intensity_model <- ppm(unmark(ppp_object2),
                       trend = ~y, Poisson(),
                       correction="Ripley")
```

Выведем результат:

```
intensity_model
```

```
Nonstationary Poisson process
```

```
Trend formula: ~y
```

```
Fitted trend coefficients:
```

```
(Intercept)          y
-0.4735872    0.1527306

      Estimate  S.E.      CI95.lo  CI95.hi  Ztest  Zval
(Intercept) -0.4735872 0.2055681 -0.87649339 -0.07068107 * -2.303797
y            0.1527306 0.0302542  0.09343348  0.21202779 *** 5.048245
```

Анализ результатов построения модели указывает на статистически значимую зависимость интенсивности от координаты  $Y$ .

Выпишем построенную модель в явном виде:

$$\lambda(x) = \lambda(Y) = e^{-0.4736+0.1527 \cdot Y} . \quad (4)$$

2) С помощью критерия хи-квадрат на основе «квадратов» проверим гипотезу о том, что изучаемый точечный образ

порождается неоднородным пуассоновским процессом с интенсивностью  $\lambda(x)$ , даваемой формулой (4). В этом случае на вход функции `quadrat.test()` подается переменная, содержащая построенную модель интенсивности пуассоновского процесса, которая также хранит и анализируемый точечный образ:

```
quadtest_n<-quadrat.test(intensity_model,nx=3,ny= 3)  
quadtest_n
```

```
Chi-squared test of fitted Poisson model  
'intensity_model' using quadrat counts  
Pearson X2 statistic  
data: data from intensity_model  
X2 = 10.5148, df = 7, p-value = 0.3225  
alternative hypothesis: two.sided  
Quadrats: 3 by 3 grid of tiles
```

Достигаемый уровень значимости критерия больше 0.1, следовательно, нулевая гипотеза о соответствии анализируемого точечного образа неоднородному пуассоновскому процессу с интенсивностью  $\lambda(x)$  принимается.

3) Построим карту, на которой отобразим полученную модель интенсивности, наблюдаемый точечный образ, сетку квадратов и отклонение наблюдаемого числа событий в «квадратах» от ожидаемого в случае неоднородного пуассоновского процесса с интенсивностью  $\lambda(x)$  (рис. 38).

```
image(predict(intensity_model),theta = 40, horiz = T,  
col=colorRampPalette(c("white","grey20"))(100))  
points(ppp_object2, pch = 21, col="white", cex = 1.2)  
axis(2, las = 1, pos = 0, at = seq(0, 10, 2))  
axis(1, las = 1, pos = 0, at = seq(0, 10, 2))  
axis(3, pos = 10, at = seq(0, 10, 2), labels = F)  
axis(4, pos = 10, at = seq(0, 10, 2), labels = F)  
plot(quadtest_n, add=T, cex=1.3, font=2)
```

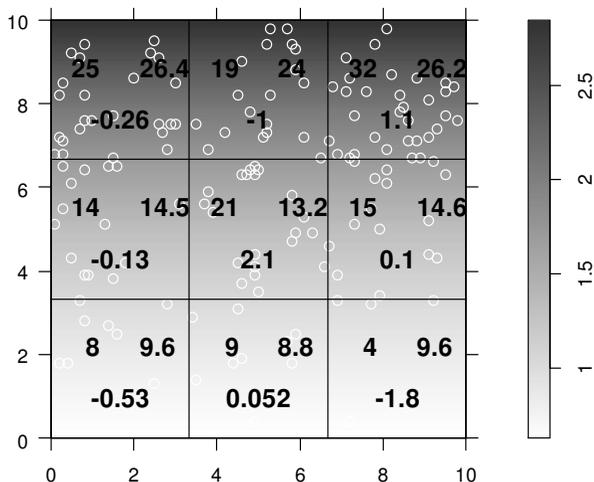


Рис. 38. Карта точечного образа размещения особей *Adonis vernalis* L. с сеткой 3×3 квадратов. Заливкой показана  $\lambda(x)$  - модель интенсивности процесса - формула (4). В левом углу квадратов показано наблюдаемое число событий (особей)  $n_i$ , в правом углу – ожидаемое число событий  $\mu_i$  в случае неоднородного пуассоновского процесса с интенсивностью  $\lambda(x)$ . Во второй строке показаны «пирсоновские» отклонения.

Таким образом, критерий не обнаружил противоречия тому, что размещение особей *Adonis vernalis* L. на исследуемой площадке является реализацией неоднородного пуассоновского процесса. Как уже говорилось, примененный критерий оставляет открытым вопрос о пространственном взаимодействии событий. В силу того, что процесс, порождающий размещение особей *Adonis vernalis* L. на изучаемой площадке, не может считаться стационарным, для дальнейшего исследования пространственных взаимодействий должны привлекаться методы, учитывающие неоднородность процесса.

### Замечание

Оценка и анализ интенсивности пространственного точечного процесса, генерирующего наблюдаемый точечный образ, позволяет сделать вывод об однородности либо неоднородности процесса. От справедливости такого вывода зависит правомерность применения тех или иных методов дальнейшего анализа точечного образа, а также интерпретация получаемых результатов.

Большинство этих методов базируется на предположении, что процесс, формирующий точечный образ, неизменен на всей изучаемой территории (стационарен, или, по крайней мере, однороден) (Cressie, 1993). Однако для природных процессов данное предположение редко выполняется, и игнорирование неоднородности процесса может привести к неправильному истолкованию пространственной структуры.

Если установлена неоднородность точечного процесса, то непостоянство его интенсивности необходимо учитывать, приступая к исследованию взаимодействий между событиями.

Одним из решений может быть выделение внутри исследуемой площади подобластей меньшего размера, где интенсивность может считаться константой, с последующим анализом пространственной структуры в каждой подобласти. Если фактор, ответственный за непостоянство интенсивности процесса известен и закартирован, то процедура выделения однородных подобластей значительно упрощается. Иначе, в предположении, что в основе неоднородности лежат условия среды, выделение однородных подобластей может осуществляться на основе плотности событий, при этом выделяются подобласти высокой и низкой плотности. Однако

такой подход следует использовать с осторожностью, особенно при работе с биологическими объектами, из-за фундаментальной двойственности пространственной неоднородности популяций, которая может формироваться как под влиянием условий среды, так и быть связанной с биолого-экологическими особенностями вида. Оба явления порождают варьирование плотности особей в пространстве и их сложно отличить. По мнению отечественного геоботаника В.И. Василевича (1969) «контагиозные распределения возникают за счет того, что среднее обилие вида варьирует по площади, иначе говоря, вероятность найти особь вида неодинакова в каждой точке или меняется вероятность найти второе растение на площадке после того, как там найдено первое». По сути дела, в первом из этих случаев должна анализироваться неоднородность экологических условий местообитания, а во втором – анализироваться особенности взаимодействия особей вида, внутривидовая конкуренция, биологические особенности вида – тип размножения, обсеменения. Полевые данные не всегда позволяют разделить две эти группы причин.

Если выделение однородных подобластей невозможно, то для дальнейшего исследования привлекаются немногочисленные специальные методы, использующие способы коррекции и уточнения инструментов анализа пространственных взаимодействий для неоднородных процессов.

## **Оценка и анализ взаимодействий**

Пространственные зависимости, взаимодействие между событиями отражают свойства второго порядка точечного процесса. Здесь мы рассмотрим, как по наблюдаемому точечному образу оцениваются и анализируются характеристики этих свойств – парная корреляционная функция  $g(r)$ , функции  $K(r)$ ,  $L(r)$ ,  $G(r)$ ,  $F(r)$ ,  $J(r)$ .

Получить оценки этих характеристик, имея только единственную реализацию пространственного точечного процесса, можно лишь в рамках некоторых ограничений и предположений.

### **Оценки $\hat{K}(r)$ , $\hat{L}(r)$ , $\hat{g}(r)$**

В предположении стационарности и изотропности точечного процесса в изучаемой области  $S$   $K$ -функция Рипли может быть оценена по наблюдаемым данным как:

$$\hat{K}(r) = \frac{1}{\hat{\lambda} \cdot n} \sum_i \sum_{j \neq i} \frac{I(\|x_i - x_j\| < r)}{w(x_i, x_j)},$$

где  $n$  - число событий точечного образа;  $\hat{\lambda}$  - оценка интенсивности процесса (для стационарного процесса эту оценку можно получить как  $\hat{\lambda} = n / A_S$ , где  $A_S$  - площадь изучаемой территории);  $I(\cdot)$  - индикаторная функция (она равна 1, если расстояние между событиями  $x_i$  и  $x_j$  меньше  $r$ , иначе – равна 0). Можно представить, что вокруг каждого  $i$ -ого события очерчивается круг радиусом  $r$  и затем в этом круге подсчитывается число других событий (рис. 39). Подсчет

осуществляется с учетом весов  $w(x_i, x_j)$  - множителей граничной коррекции.

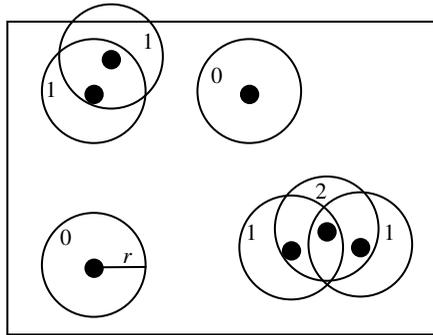


Рис. 39. Иллюстрация вычисления функции  $\hat{K}(r)$ . Для каждого события подсчитывается число других событий, попавших в круг радиуса  $r$  (цифры в каждом круге).

Значение множителя  $w(x_i, x_j)$  для каждой пары точек определяется следующим образом. Рассматривается круг, ограниченный окружностью с центром в точке  $x_i$  и проходящей через точку  $x_j$ . Если этот круг полностью находится на изучаемой территории, то множитель равен 1. Иначе его значение равно отношению площади части круга, находящейся на изучаемой территории, к площади всего круга. Вместо множителя  $w(x_i, x_j)$  могут быть использованы другие способы граничной коррекции, рассмотренные выше: буферная зона или тороидальная репликация.

Следует отметить, что хотя функция  $\hat{K}(r)$  может быть вычислена для любых радиусов  $r$ , при больших  $r$  эта оценка становится смещенной (особенно, если точечный образ состоит из небольшого числа событий) и зависящей от формы изучаемой

области  $S$ . Рекомендуется рассматривать радиусы не более  $1/2$  от наименьшей стороны изучаемой территории (если она прямоугольная), или менее  $\sqrt{A_S / 2}$ , где  $A_S$  – площадь изучаемой территории.

В случае неоднородного процесса, когда интенсивность варьирует на территории, обобщенную функцию Рипли  $K(r)$  можно оценить как:

$$\hat{K}(r) = \frac{1}{A_S} \cdot \sum_i \sum_{j \neq i} \frac{I(\|x_i - x_j\| < r)}{\hat{\lambda}(x_i) \cdot \hat{\lambda}(x_j) \cdot w(x_i, x_j)},$$

где  $\hat{\lambda}(x_i)$  и  $\hat{\lambda}(x_j)$  – оценки интенсивности в соответствующих точках, полученные, например, с помощью ядерной аппроксимации.

Оценку  $L$ -функции, более удобной для интерпретации, чем  $K$ -функция, получают как:

$$\hat{L}(r) = \sqrt{\frac{\hat{K}(r)}{\pi}} - r.$$

Оценка парной корреляционной функции делается в предположении, что она изотропна. В этом случае, как указывалось выше,  $g(r)$  выражается через производную от  $K$ -функции:  $g(r) = \frac{K'(r)}{2\pi r}$ . Но так как довольно часто (для малых выборок)  $\hat{K}(r)$  представляет собой ступенчатую функцию, то вычислить оценку  $K'(r)$  на ее основе непросто. Некоторые исследователи для получения оценки  $K'(r)$  используют сглаженное представление  $\hat{K}(r)$ . Другой подход - оценка парной

корреляционной функции методом ядерной аппроксимации с граничной коррекцией:

$$\hat{g}(r) = \frac{1}{2\pi r A_S} \cdot \sum_i \sum_{j \neq i} \frac{k_h(r - \|x_i - x_j\|)}{\hat{\lambda}(x_i) \cdot \hat{\lambda}(x_j) \cdot w(x_i, x_j)},$$

где  $k_h()$  - одномерная ядерная функция Епанечникова:

$$k_h(t) = \begin{cases} \frac{2}{\pi} \left(1 - \frac{t^2}{h^2}\right), & \text{если } |t| \leq h \\ 0, & \text{в остальных случаях} \end{cases}.$$

Еще раз подчеркнем, что  $K$ -функция и парная корреляционная функция  $g$  отвечают на несколько различные вопросы. Кумулятивная  $K$ -функция может обнаружить агрегацию или регулярность, наблюдаемую вплоть до данного расстояния  $r$ , и, следовательно, смешивает свойства пространственной структуры, наблюдаемые на больших  $r$  со свойствами на малых  $r$ . Она подходит для анализа процессов, которые могут работать вплоть до определенного расстояния  $r$  (например, отрицательный эффект конкуренции), тогда как парная корреляционная функция может обнаружить агрегацию или разрежение именно на данном расстоянии  $r$ . Парную корреляционную функцию легче интерпретировать, чем кумулятивную  $K$ -функцию (так же как легче интерпретировать функцию плотности вероятности по сравнению с кумулятивной функцией распределения вероятности). Поэтому инструменты пространственного анализа должны включать в себя оценки не только  $K$ - и  $L$ -функций, но и оценку парной корреляционной функции  $g$ .

### Пример 10

Оценим функции  $K(r)$ ,  $L(r)$ ,  $g(r)$  по наблюдаемому точечному образу размещения стволов *Tilia cordata* L. и построим графики, показывающие поведение оценок в зависимости от  $r$ .

1) Для получения оценки  $\hat{K}(r)$  в случае стационарного процесса используется функция `Kest()` пакета `spatstat`:

```
my_Kest1 <- Kest(ppp_object1, correction = "Ripley")
plot(my_Kest1, iso ~ r, main="K function", ylab="K(r)")
```

2) Оценку  $\hat{L}(r)$ , более удобную для анализа, можно получить с помощью функции `Lest()` или функции `eval.fv()`:

```
my_Lest1 <- Lest(ppp_object1, correction = "Ripley")
my_Lest1$iso <- my_Lest1$iso - my_Lest1$r
```

или

```
r = my_Kest1$r
my_Lest1 <- eval.fv(sqrt(my_Kest1 / pi) - r)
plot(my_Lest1, iso ~ r, main="L function", ylab="L(r)")
```

3) Для получения оценки  $\hat{g}(r)$  используется функция `pcf()`, причем ее аргументом может быть или сам объект точечного образа (`ppp.object`) – тогда оценка будет сделана методом ядерной аппроксимации, или полученная для точечного образа оценка  $\hat{K}(r)$  – тогда оценка  $\hat{g}(r)$  вычисляется через производную от сглаженной  $\hat{K}(r)$ :

```
my_gest1 <- pcf(ppp_object1, correction = "Ripley")
plot(my_gest1, iso ~ r, main="g function", ylim=c(0,15),
ylab="g(r)")
```

или

```
my_gest1 <- pcf(my_Kest1, method="a", spar=0.55)
```

```
plot(my_gest1,pcf ~ r, main="g function", ylim=c(0,15),
ylab="g(r) ")
```

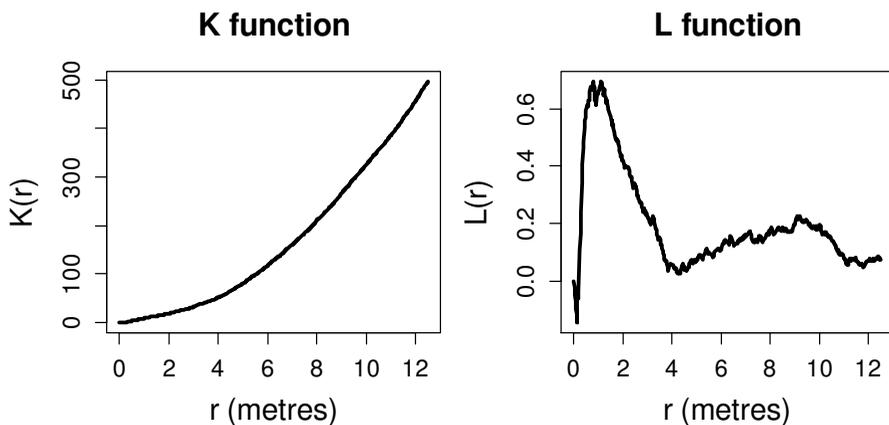


Рис. 40.  $\hat{K}(r)$  и  $\hat{L}(r)$  для точечного образа стволов *Tilia cordata* L.

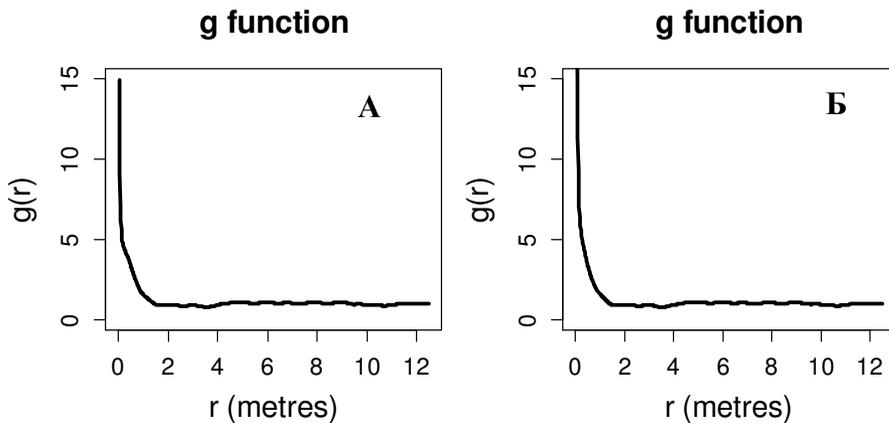


Рис. 41.  $\hat{g}(r)$  для точечного образа стволов *Tilia cordata* L.

А: получена ядерной аппроксимацией;

Б: получена как производная от сглаженной  $\hat{K}(r)$ .

### Пример 11

Оценим функции  $K(r)$ ,  $L(r)$ ,  $g(r)$  для точечного образа размещения особей *Adonis vernalis* L., построим графики изменения этих оценок в зависимости от  $r$ . При этом будем учитывать неоднородность соответствующего процесса: его интенсивность меняется в изучаемой области (см. пример 9).

1) Для получения оценки  $\hat{K}(r)$  в этом случае используется функция `Kinhom()`:

```
my_Kest2 <- Kinhom(ppp_object2, lambda= ppp_density2,  
                  correction = "Ripley")  
plot(my_Kest2, iso ~ r, main="K function", ylab="K(r)")
```

2) Оценку  $\hat{L}(r)$  получим с помощью функции `eval.fv()`:

```
r= my_Kest2$r  
my_Lest2 <- eval.fv(sqrt(my_Kest2 /pi)-r)  
plot(my_Lest2, iso ~ r, main="L function", ylab="L(r)")
```

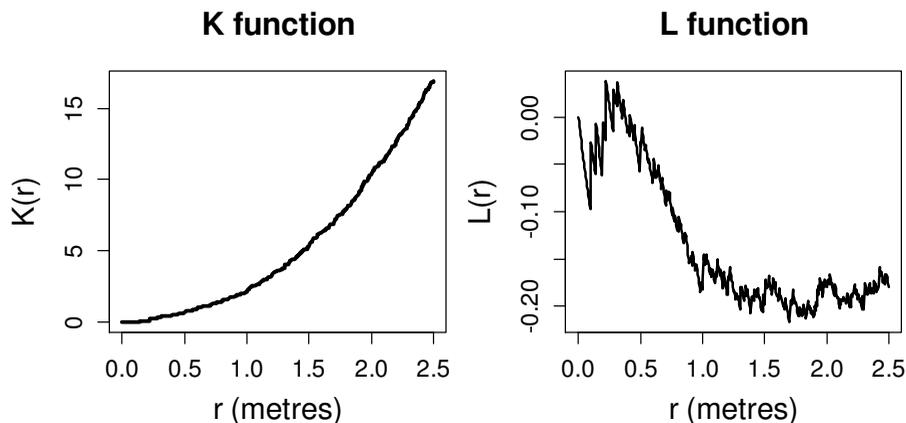


Рис. 42.  $\hat{K}(r)$  и  $\hat{L}(r)$  для точечного образа особей *Adonis vernalis* L.

3) Оценку  $\hat{g}(r)$  получим с помощью функции `pcf()`,

аргументом которой в этом случае будет  $\hat{K}(r)$  (оценка  $\hat{g}(r)$  вычисляется через производную от сглаженной  $\hat{K}(r)$ ):

```
my_gest2 <- pcf(my_Kest2, method="a", spar=0.55)
plot(my_gest2, pcf ~ r, main = "g function")
```

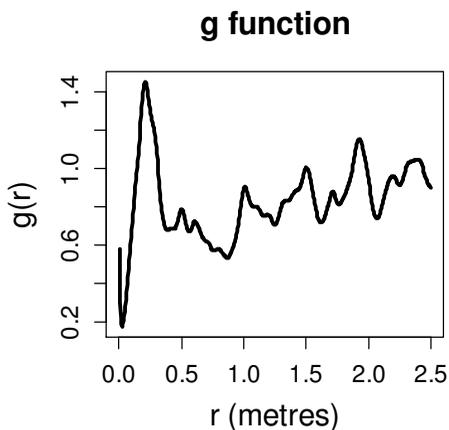


Рис. 43.  $\hat{g}(r)$  для точечного образа особей *Adonis vernalis* L.

### Оценки $\hat{G}(r)$ , $\hat{F}(r)$ , $\hat{J}(r)$

В предположении стационарности точечного процесса в изучаемой области для описания пространственных закономерностей в локальном масштабе используют функции  $G(r)$  и  $F(r)$ , основанные на распределении расстояний до ближайшего события, а также получаемую из них функцию  $J(r)$ . Их эмпирические оценки на основе наблюдаемого точечного образа обозначим  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$ .

Оценку  $\hat{G}(r)$  с адаптивной граничной коррекцией можно получить как:

$$\hat{G}(r) = \frac{\sum_i I(W_i < r, d_i > r)}{\sum_i I(d_i > r)},$$

где суммирование выполняется по всем  $n$  событиям точечного образа,  $W_i$  – расстояние от  $i$ -ого события до ближайшего события,  $d_i$  – расстояние от  $i$ -ого события до границы изучаемой области,  $I(\cdot)$  – индикаторная функция. То есть, для фиксированного  $r$  рассматриваются только события, удаленные от края области на расстояние, не меньшее  $r$ , и среди них подсчитывается доля таких событий, для которых расстояние до ближайшего соседа меньше  $r$ .

Подобным образом, оценка  $\hat{F}(r)$  с граничной коррекцией рассчитывается как:

$$\hat{F}(r) = \frac{\sum_i I(U_i < r, d_i > r)}{\sum_i I(d_i > r)},$$

здесь суммирование выполняется по  $m$  точкам, случайно выбранным в изучаемой области;  $U_i$  – расстояние от  $i$ -ой точки до ближайшего события.

Так же могут быть использованы более эффективные оценки функций  $F$  и  $G$ , получаемые методом Kaplan-Meier (Baddeley, Gill, 1997) (здесь не приводятся).

Используя  $\hat{G}(r)$  и  $\hat{F}(r)$  (если  $\hat{F}(r) < 1$ ), можно получить оценку функции  $J(r)$ :

$$\hat{J}(r) = \frac{1 - \hat{G}(r)}{1 - \hat{F}(r)}.$$

### Пример 12

Оценки  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$  реализованы в пакете **spatstat** функциями **Gest()**, **Fest()**, **Jest()**, соответственно. Все они имеют одинаковые параметры. Оценим для точечного образа стволов *Tilia cordata* L. функцию  $G(r)$  и построим график этой оценки.

```
my_Gest1<-Gest(ppp_object1, correction="han")  
plot(my_Gest1, han ~ r, main = "G function")
```

Параметры функций **Gest()**, **Fest()**, **Jest()** :

ppp_object	Объект типа ppp.object, содержащий точечный образ.
correction	Вид граничной коррекции, может принимать значения “none” – нет коррекции, “rs” – буферная зона, “km” – коррекция Kaplan-Meier, “han” – коррекция Hanish.

#### G function

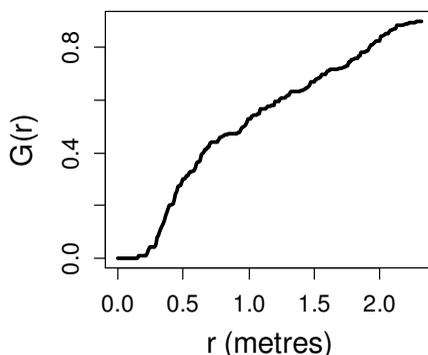


Рис. 44.  $\hat{G}(r)$  для точечного образа стволов *Tilia cordata* L.

Попробуйте повторить приведенный код, заменив функцию **Gest()** на функции **Fest()** и **Jest()**, и получить  $\hat{F}(r)$  и  $\hat{J}(r)$ .

## Проверка гипотез

Рассмотрим критерии проверки гипотез о виде пространственного процесса, порождающего наблюдаемый точечный образ, основанные на анализе свойств второго порядка. Эти методы исследуют пространственное взаимодействие изучаемых событий, позволяя ответить на вопрос: существует ли тенденция встречать события в группах, рядом друг с другом (агрегации), или события избегают друг друга, встречаются на удалении один от другого (регулярность), или же события размещаются независимо друг от друга. Простейшие из тестов используют статистики (или *индексы*), расчет которых базируется либо на методе квадратов, либо на расстояниях до ближайшего события. Более надежные критерии, позволяющие провести статистическую проверку гипотез, используют полученные оценки  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$ ,  $\hat{G}(r)$ ,  $\hat{F}(r)$  и  $\hat{J}(r)$ .

### *Индексы на основе «квадратов»*

Простейший индекс  $I = s^2 / \bar{n}$ , называемый индексом рассеивания (где  $\bar{n}$  - выборочное среднее числа событий в «квадратах»,  $s^2$  - выборочная дисперсия) основан на факте, что для распределения Пуассона среднее и дисперсия равны. Значения  $I$ , большие 1, служат признаком агрегации событий, тогда как значения менее 1 указывают на регулярность в масштабе размера «квадрата». Похожий индекс  $ICS = I - 1$  приблизительно равен среднему числу событий в кластере, если размер «квадрата» достаточно велик по отношению к размеру кластера. Индекс «среднего скопления» («mean crowding» index)  $n^* = \bar{n} + I - 1$  приблизительно равен числу событий, попадающих в

один «квадрат» с неким произвольным событием; этот индекс не зависит от пустых «квадратов». Другой индекс  $IP = n^* / \bar{x}$  - это относительная мера «пятнистости» (patchiness) точечного образа. Он показывает долю, на которую «среднее скоплений» превышает среднюю плотность событий. Эти статистики, получаемые на основе «квадратов», не задействуют в анализе пространственный аспект данных, а также сводят весь анализ к одному числу, что приводит к потере информации. К тому же, если сместить «квадраты» или изменить их размер, то можно получить уже иной результат.

Для преодоления этих недостатков используют регулярную сетку квадратов (сетку ячеек), покрывающую всю территорию исследования. При этом применяется метод *вложенных* квадратов, использующий *агломеративный* подход: последовательно рассматриваются сетки с увеличивающимся размером ячеек (рис. 45), на каждой из которых рассчитывается тот или иной индекс.

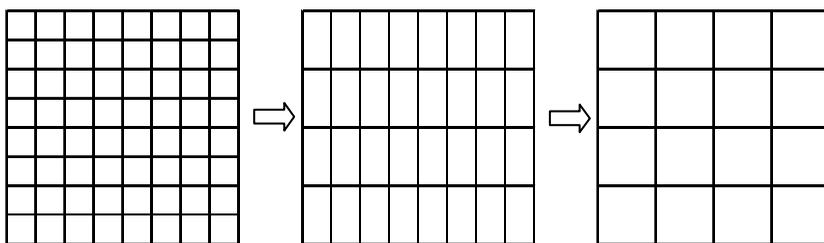


Рис. 45. Иллюстрация метода вложенных квадратов. Для каждой сетки рассчитывается индекс - статистика пространственной структуры.

Для анализа пространственной структуры строится график зависимости значений индекса от линейного размера квадратов (длины стороны или диагонали ячейки), на котором можно

попытаться различить масштабы пространственных зависимостей в точечном образе.

На вложенных сетках часто рассчитывается индекс Мориситы (Morishita):

$$I_{\delta} = \frac{m}{n(n-1)} \sum_{i=1}^m n_i(n_i - 1),$$

где  $m$  – число квадратов в сетке,  $n_i$  – число событий в  $i$ -ом квадрате,  $n$  – общее число событий точечного образа. В случае CSR  $I_{\delta}$  принимает значение, близкое к 1 при любом размере квадратов. Отклонение от 1 свидетельствует об отклонении от CSR.  $I_{\delta} > 1$  указывает на агрегированность событий в масштабе квадрата,  $I_{\delta} < 1$  выполняется для регулярных, разреженных образов (рис. 46).

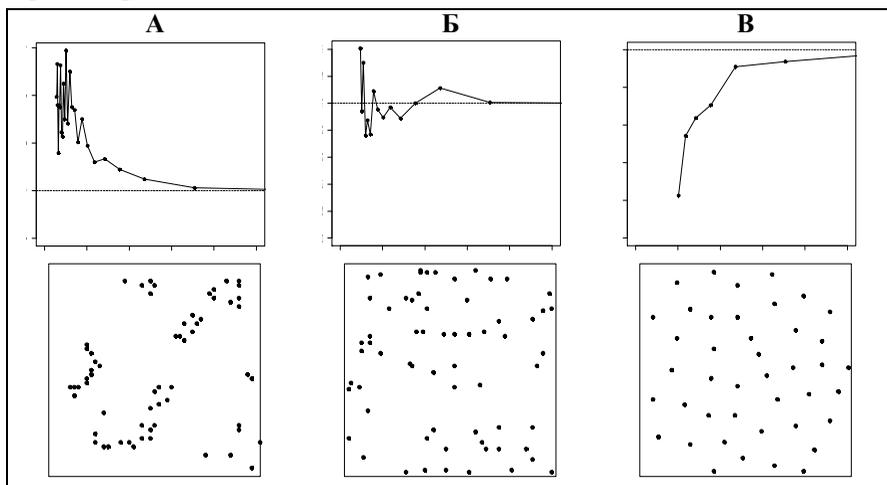


Рис. 46. Первый ряд: поведение индекса Мориситы в зависимости от размера ячеек сетки для агрегированного (А), близкого к полной пространственной случайности (Б) и регулярного точечного образа (В). Второй ряд: точечные образы, для которых рассчитывался индекс Мориситы.

### Пример 13

Для двух изучаемых точечных образов исследуем поведение индекса Мориситы с использованием метода вложенных квадратов. Этот тест реализован в функции `miplot()` пакета `spatstat`, а также может быть выполнен в R с помощью пользовательской функции `my_miplot()`. Данная функция, текст которой приведен в приложении 1, покрывает окно наблюдений сеткой 2×2 квадратов, затем сеткой 3×3 квадратов, и т.д. На каждой сетке вычисляется индекс Мориситы. Результат выдается в виде графика (рис. 47).

```
my_miplot(ppp_object1)
```

```
my_miplot(ppp_object2)
```

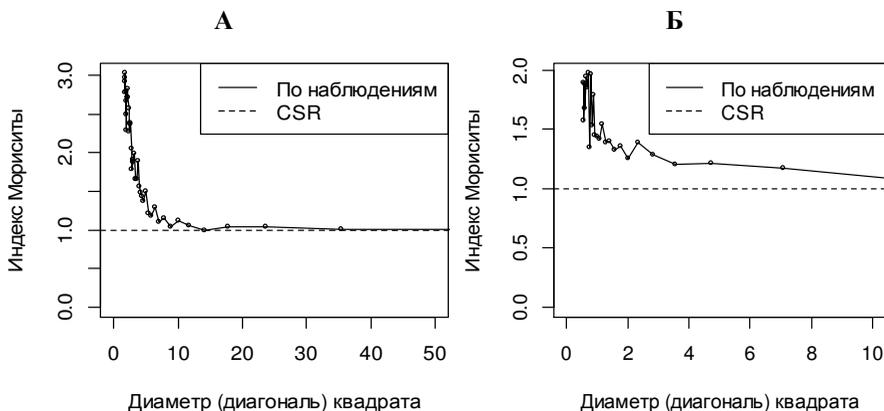


Рис. 47. Индекс Мориситы в зависимости от размера ячеек сетки (по горизонтали - длина диагонали ячейки в метрах).

А: для точечного образа размещения стволов *Tilia cordata* L.

Б: для точечного образа размещения особей *Adonis vernalis* L.

На графике рис. 47 А наблюдается сильно выраженный пик значений индекса Мориситы для квадратов небольших размеров, что говорит об отсутствии полной пространственной случайности

(CSR) в размещении стволов *Tilia cordata* L. Такое поведение индекса может отражать позитивное взаимодействие событий и характерный радиус их агрегаций.

График рис. 47 Б показывает, что значения индекса Мориситы для всех размеров квадратов превышают 1. Это указывает на отклонение от CSR, а именно на наличие агрегаций – подобластей с высокой плотностью особей *Adonis vernalis* L., что связано, скорее всего, с неоднородностью точечного процесса (интенсивность изменяется на исследуемой площадке).

### ***Индексы расстояний до ближайшего события***

Кроме статистик, рассчитываемых на основе «квадратов», при исследовании точечных образов могут применяться индексы, основанные на расстояниях до ближайшего события. Некоторые из этих индексов даны в таблице 1 (см. полный список в работе Cressie, 1993). Напомним обозначения, введенные ранее (рис. 6):

$W$  – расстояние от события до ближайшего события;

$U$  – расстояние от случайно выбранной точки пространства, не содержащей событие, до ближайшего события.

В случае однородного пуассоновского процесса с интенсивностью  $\lambda$  для этих статистик известно точное или асимптотическое теоретическое распределение. Поэтому они могут использоваться в качестве критериев для проверки гипотезы о CSR. Критические значения индексов определяются через квантили известного закона распределения (см. таблицу 1) при выбранном уровне значимости.

Таблица 1. Статистики расстояний до ближайшего события (с асимптотическим законом распределения для случая полной пространственной случайности)\*

Расстояние	Статистика	Распределение	Автор
$W$	$2\sqrt{\lambda} \sum_{i=1}^n W_i / n$	$N\left(1, \frac{4-\pi}{\pi n}\right)$	Clark, Evans, 1954
$W$	$2\pi\lambda \sum_{i=1}^n W_i^2$	$\chi_{2n}^2$	Skellam, 1952
$U$	$\frac{12m(m \log(\sum_{i=1}^m U_i^2/m) - \sum_{i=1}^m \log U_i^2)}{7m+1}$	$\chi_{m-1}^2$	Pollard, 1971
$U$	$\pi\lambda \sum_{i=1}^m U_i^2 / m$	$N(1, 1/m)$	Pielou, 1959
$W, U$	$\sum_{i=1}^m U_i^2 / \sum_{j=1}^n W_j^2$	$F_{2m, 2n}$	Hopkins, 1954

\* в таблице используются следующие обозначения:  $n$  – число событий в образе;  $m$  – число точек, не содержащих событий;  $N(a, \sigma^2)$  – нормальное распределение со средним  $a$  и дисперсией  $\sigma^2$ ;  $F_{k1, k2}$  – распределение Фишера с  $k1, k2$  степенями свободы.

Заметим, что согласно теории выборка должна быть независимой (только тогда по ней можно делать выводы о свойствах всей генеральной совокупности). Однако расстояния до ближайшего события, вычисленные по изучаемому точечному образу, часто основаны на одних и тех же событиях, поэтому набор таких расстояний не является независимой выборкой. Для того чтобы соблюсти требование независимости при вычислении статистик, основанных на расстояниях, рекомендуют использовать только 1/10 всех событий. Понятно, что это требование сложно выполнить для небольших образов точек. К

примеру, если наблюдаемый образ состоит из 100 событий, для вычисления индексов будет использовано только 10 расстояний до ближайшего события. То есть, выборка становится слишком малой для того чтобы по ней делать надежные статистические выводы.

Кроме этого, если площадь изучаемой территории мала (в масштабе работ), то при подсчете индексов нужно обязательно учитывать граничный эффект, иначе их значения будут завышены. Здесь могут быть применены те же способы граничной коррекции, что и в случае оценки интенсивности: буферная зона вдоль границы изучаемой области или тороидальная репликация. Однако в этом случае теоретические распределения перестают работать. Чтобы преодолеть эти трудности, для определения критических значений индексов рекомендуется использовать симуляции Монте-Карло.

В целом, все перечисленные здесь индексы расстояний до ближайшего события имеют преимущество над статистиками, получаемыми с помощью «квадратов», так как используют точные координаты событий и не зависят от формы или размера «квадратов». Их недостатком может считаться то, что в результате они также сводят всю информацию к одному числу, что не позволяет анализировать сложные пространственные структуры. Например, пусть в точечном образе наряду с группами близко расположенных событий (кластерами) есть и большое число изолированных событий, которые находятся достаточно далеко друг от друга. Среднее расстояние до ближайшего события для такого образа может быть близко к среднему для точечного образа однородного пуассоновского процесса, поэтому тесты, основанные на данных индексах, могут

не показать значимого отклонения от CSR. В общем случае рекомендуется применять методы, использующие эмпирические распределения (гистограммы) расстояний до ближайшего события, что позволяет преодолеть указанные недостатки и проводить более детальный анализ. Такие методы будут рассмотрены ниже.

### ***Пример 14***

Проверим гипотезу о CSR размещения особей *Tilia cordata* L., используя индекс Кларка-Эванса (Clark-Evans). В случае CSR значение этого индекса равно 1, для регулярных процессов значение индекса больше 1, для агрегационных процессов - меньше 1.

Тест с использованием индекса Кларка-Эванса может быть проведен с помощью функции `clarkevans.test()`. Его можно выполнить в двух вариантах: (а) без граничной коррекции, тогда критические значения критерия (или достигаемый уровень значимости - p-value) определяются по нормальному закону распределения, и (б) с граничной коррекцией, тогда тест проводится с использованием симуляций Монте-Карло.

Сначала применим тест Кларка-Эванса без граничной коррекции. При этом в качестве альтернативной выдвинем гипотезу о том, что имеет место процесс с агрегациями событий (значение индекса Кларка-Эванса меньше 1):

```
clarkevans.test(ppp_object1, correction="none",  
                alternative="less")
```

```
Clark-Evans test  
No edge correction  
Z-test
```

```
data: ppp_object1
```

```
R = 0.8519, p-value = 3.619e-07
```

```
alternative hypothesis: clustered (R < 1)
```

Значение индекса  $\approx 0.85$ , достигаемый уровень значимости  $p\text{-value} \approx 0$ . Нулевая гипотеза о CSR отвергается с надежностью не менее 99%. Значение индекса меньше единицы, что может свидетельствовать об агрегированности точечного процесса (нужно помнить, что значение индекса, рассчитанное без граничной коррекции, несколько завышено из-за граничного эффекта).

Применение теста Кларка-Эванса с граничной коррекцией дает такой же результат:

```
clarkevans.test(ppp_object1, correction="Donnelly",
                alternative="less")

Clark-Evans test
Donnelly correction
Monte Carlo test based on 999 simulations of CSR
with fixed n
data: ppp_object1
R = 0.8315, p-value = 0.001
alternative hypothesis: clustered (R < 1)
```

Гипотеза о CSR отвергается с надежностью не менее 99% (достигаемый уровень значимости меньше 0.01).

### ***Критерии на основе $\hat{G}(r)$ , $\hat{F}(r)$ , $\hat{J}(r)$ . Метод «конверта»***

Для исследования пространственных закономерностей в локальном масштабе (в предположении стационарности точечного процесса) вместо индексов, описывающих точечный рисунок одним числом, предпочтительнее использовать функции распределения расстояний до ближайшего события  $G(r)$  и  $F(r)$ , а также основанную на них функцию  $J(r)$ . Проверка гипотезы о CSR проводится с помощью сравнения их эмпирических оценок  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$ , получаемых по наблюдаемому образу точек, с

теоретическими функциями, соответствующими однородному пуассоновскому процессу:

$$G(r) = F(r) = 1 - e^{-\lambda \cdot 2\pi r^2}, \quad J(r) = 1.$$

Это сравнение распределений из-за несоблюдения требования о независимости выборки (расстояний до ближайшего события) не может быть выполнено с помощью классических тестов, например, Колмогорова-Смирнова. Поэтому для проверки гипотез рекомендуется использовать **метод «конверта»** (simulation-envelopes), реализуемый с помощью симуляций Монте-Карло.

Согласно этому методу создается  $k$  симуляций точечных образов (обычно  $k$  - большое число, например, 9999), соответствующих гипотетическому процессу. В данном случае - однородному пуассоновскому процессу с интенсивностью равной  $\hat{\lambda} = n / A_s$ . Для каждой реализации (симулированного образа) рассчитываются функции-оценки, например  $\hat{G}(r)$ . Затем при фиксированном расстоянии  $r$  значения всех  $k$  функций-оценок ранжируются, и выбираются два значения: с заданным рангом  $rank$  и с рангом  $k - rank + 1$ . Например, если ранг равен 1, тогда выбираются самое большое и самое маленькое значения; если ранг равен 2, то выбираются значения второе сверху и второе снизу; и т.д. В результате получим две функции:  $G_{lo}(r)$  и  $G_{hi}(r)$ , которые и задают границы «конверта» для принятия нулевой гипотезы о соответствии CSR (рис. 48). Выход из этих границ функции-оценки  $\hat{G}(r)$ , рассчитанной по выборочным данным, трактуется как статистически значимое отклонение от CSR.

Уровень значимости отклонения при фиксированном  $r$  будет равен:

$$\text{p-value} = \frac{2 \cdot \text{rank}}{k + 1},$$

где  $\text{rank}$  – ранг граничных значений. Аналогичным образом строятся «конверты» для функций  $\hat{F}(r)$  и  $\hat{J}(r)$ .

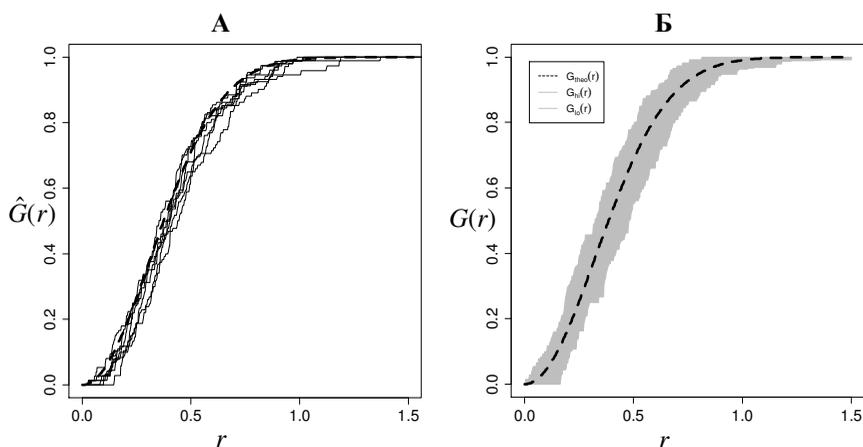


Рис. 48. Иллюстрация построения границ принятия гипотезы о CSR (метод «конверта»).

А: функции  $\hat{G}(r)$ , вычисленные по симулированным данным (взято 5 симуляций).

Б: «конверт» принятия нулевой гипотезы – между функциями  $G_{lo}(r)$  и  $G_{hi}(r)$  (показан серой заливкой). Пунктирная линия – поведение функции  $G(r)$  в случае CSR.

Если гипотеза о CSR отвергается, то значения функции  $\hat{J}(r)$  при небольших  $r$  могут быть использованы для определения типа отклонения от CSR. А именно, если на некотором расстоянии  $r$ , где имеет место выход функции  $\hat{J}(r)$  из границ

«конверта»,  $\hat{J}(r) < 1$ , то это указывает на агрегацию событий точечного образа, тогда как  $\hat{J}(r) > 1$  указывает на некоторую регулярность распределения событий на данном расстоянии.

На рис. 49 представлены примеры точечных образов, рассчитанные по ним функции-оценки  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$  и «конверты» принятия гипотезы о CSR, построенные с помощью симуляций Монте-Карло. Можно видеть, что в случаях А и Б гипотеза о CSR отвергается, в случае В – принимается. При этом в случае А в размещении событий имеет место регулярность ( $\hat{J}(r) > 1$  при небольших  $r$ ), а в случае Б отклонение от CSR проявляется наличием агрегаций событий ( $\hat{J}(r) < 1$  при больших  $r$ ).

Следует отметить, что метод Монте-Карло может использоваться при проверке и других гипотез, не только о соответствии CSR. Симулирован может быть любой другой пространственный точечный процесс, к примеру, кластерный пуассоновский процесс, процесс Штрауса, процесс твердого ядра. Отклонение эмпирической функции, рассчитанной по наблюдаемым данным, от «конверта», полученного на основе симулированных реализаций, будет свидетельствовать о несоответствии наблюдаемого точечного образа процессу, модель которого использовалась для симуляции.

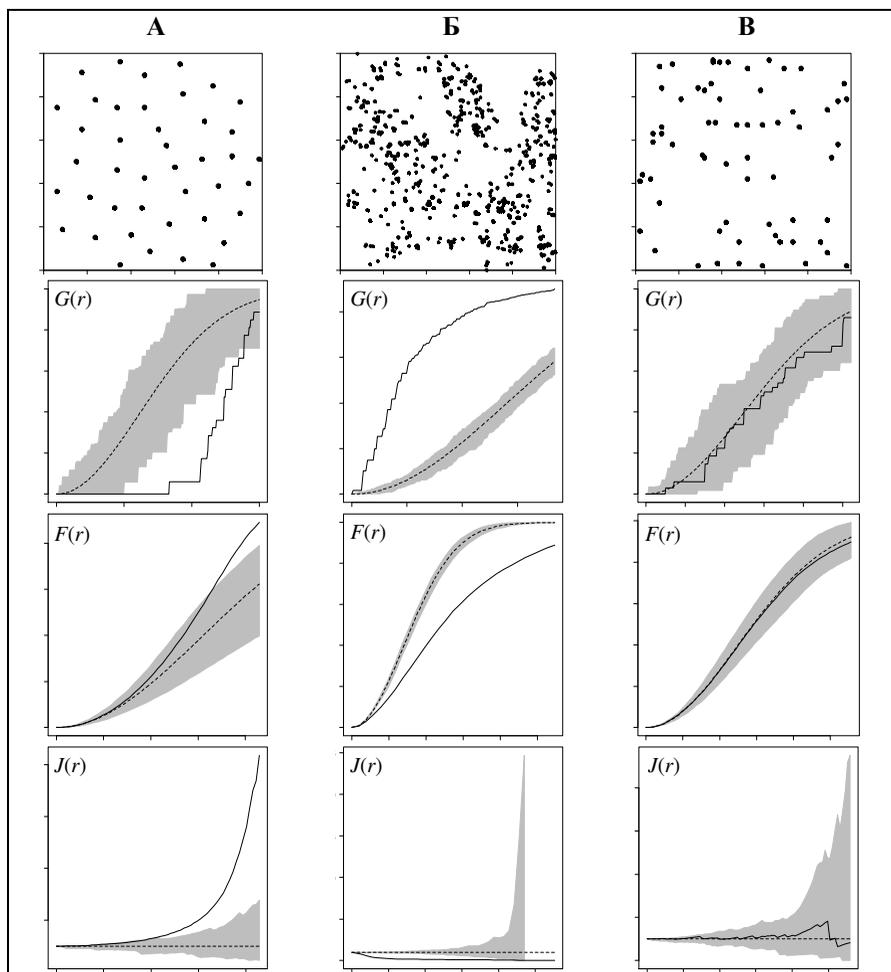


Рис. 49. А: центры клеток среза ткани, полученные под увеличением (Crick, Lawrence, 1975; Ripley, 1977); Б: проростки ежевики (Hutchings, 1979; Diggle 1983); В: проростки японской сосны *Pinus thunbergii* (Numata, 1961; Diggle 2003).

Первый ряд: точечные образы; второй ряд:  $\hat{G}(r)$ ; третий ряд:  $\hat{F}(r)$ ; четвертый ряд:  $\hat{J}(r)$ . Пунктирные линии – поведение функций в случае CSR. Серая заливка – 2% «конверты» принятия нулевой гипотезы о CSR, полученные с помощью 99 симуляций однородного пуассоновского процесса.

### **Пример 15**

Проверим гипотезу о CSR размещения стволов *Tilia cordata* L. на основе анализа  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$  - оценок функций распределения расстояний до ближайшего события. Для этого точечного образа анализ интенсивности не обнаружил противоречия гипотезе об однородности процесса, порождающего его (пример 7). Поэтому применение критерия на основе  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$  будет корректным.

1) Оценим функции  $G(r)$ ,  $F(r)$ ,  $J(r)$  и построим графики оценок, на которых также отобразим поведение этих функций в случае CSR (рис. 50 А):

```
my_Gest1<-Gest(ppp_object1, correction="han")
plot(my_Gest1, main = "G function")
my_Fest1<-Fest(ppp_object1, correction="cs")
plot(my_Fest1, main = "F function")
my_Jest1<-Jest(ppp_object1, correction="best")
plot(my_Jest1, main = "J function",ylim=c(0,2))
```

2) Для того чтобы проверить, является ли наблюдаемое отклонение от теоретического поведения статистически значимым, построим «конверты» на основе симуляций Монте-Карло с помощью функции `envelope()` (рис. 50 Б):

```
Gest_envelope <- envelope(unmark(ppp_object1),
                          fun=Gest,correction="han",nsim=199)
plot(Gest_envelope)
Fest_envelope <- envelope(unmark(ppp_object1),
                          fun = Fest,correction="cs",nsim=199)
plot(Fest_envelope)
Jest_envelope <- envelope(unmark(ppp_object1),
                          fun=Jest,correction="best",nsim=199)
plot(Jest_envelope,ylim=c(0,2))
```

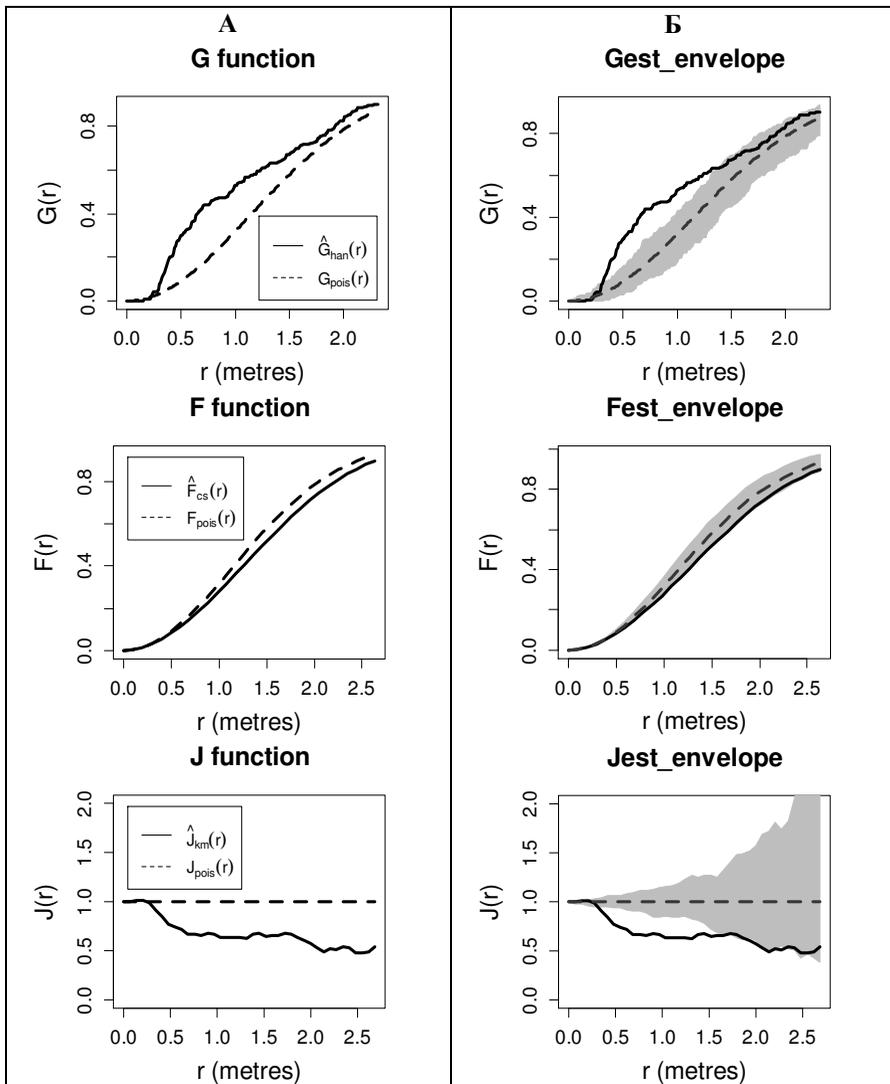


Рис. 50. Сплошной линией показаны оценки  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$ , рассчитанные для точечного образа стволов *Tilia cordata* L. Пунктирная линия – поведение функций  $G(r)$ ,  $F(r)$ ,  $J(r)$  в случае CSR. Серой заливкой на графиках справа показаны «конверты» принятия гипотезы о CSR, полученные с помощью 199 реализаций однородного пуассоновского процесса.

Параметры функции **envelope ()** :

pre_ppp	Объект типа ppp.object, содержащий точечный образ.
fun	Функция, для которой нужно вычислить «конверт». Это могут быть Gest, Fest, Jest, Kest, Lest, pcf.
correction	Вид граничной коррекции.
nsim	Число симуляций Монте-Карло.
nrank	ранг граничных значений (по умолчанию равен 1)

В этом примере было использовано 199 симуляций, что при фиксированном  $r$  соответствует уровню значимости  $p\text{-value} = 2/(199+1) = 0.01$ . Рекомендуется увеличивать число симуляций, чтобы уменьшить вероятность совокупной ошибки первого рода (Grabarnik et al., 2011).

На графиках рис. 50 можно видеть, что поведение функций  $\hat{G}(r)$  и  $\hat{J}(r)$  на расстояниях примерно от 0.4 до 1.5 м показывает значимое отклонение от CSR. При этом  $\hat{J}(r) < 1$ , что говорит о том, что в точечном образе присутствуют скопления (агрегации) особей, находящихся на таких расстояниях друг от друга. Вероятно, данные скопления образованы прегенеративными побегам *Tilia cordata* L. порослевого происхождения.

Так как функции  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$  предоставляют информацию о соседстве первого порядка, характерном для малых, локальных расстояний, их анализ не позволяет сказать, имеются ли другие пространственные особенности размещения липы сердцелистной, характеризующиеся более протяженными расстояниями.

Заметим, что для исследования точечного образа *Adonis vernalis* L. применение критерия на основе статистик  $\hat{G}(r)$ ,  $\hat{F}(r)$ ,  $\hat{J}(r)$  не является корректным, так как процесс, порождающий

этот точечный образ, не может считаться стационарным в изучаемой области: анализ интенсивности (примеры 8, 9) показал, что свойства первого порядка этого процесса меняются по территории исследования.

### *Критерии на основе $\hat{K}(r)$ , $\hat{L}(r)$ , $\hat{g}(r)$*

Статистики расстояния до ближайшего соседа, приведенные выше, характеризуют локальное взаимодействие событий и не показывают закономерности точечного рисунка в более протяженном пространственном масштабе. Так, например, для пространственного рисунка размещения растений в силу особенностей размножения на малых расстояниях могут быть характерны агрегации особей. А в более глобальном масштабе подобные первичные агрегации могут быть организованы в новые, более протяженные скопления из-за неоднородности ресурсов (пятна влажности в понижениях рельефа, световые пятна в пологе леса). При этом, если рассмотреть рисунок размещения особей в еще большем пространстве, то эти скопления в свою очередь могут быть размещены регулярно по отношению друг к другу, подчиняясь пространственному рисунку ресурсных пятен. Проанализировать подобные закономерности помогут функции  $K(r)$ ,  $L(r)$ ,  $g(r)$ .

Для проверки гипотезы о полной пространственной случайности (CSR) процесса, генерирующего наблюдаемый точечный образ, эмпирические оценки  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$  сравниваются с теоретическими функциями, соответствующими однородному пуассоновскому процессу:

$$K(r) = \pi r^2, \quad L(r) = 0, \quad g(r) = 1.$$

Для оценки значимости отклонений одних от других применяется описанный выше метод «конверта» (simulation-envelopes) на основе симуляций Монте-Карло.

Значимое отклонение функций  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$ , полученных по наблюдаемому точечному образу, от поведения, соответствующего однородному пуассоновскому процессу, свидетельствует о наличии взаимодействия между событиями либо позитивного, приводящего к их агрегированию, либо негативного, приводящего к регулярности их размещения.

Если при некотором  $r$  имеет место выход функции  $\hat{L}(r)$  (или функции  $\hat{K}(r)$ ) из границ «конверта», и значение  $\hat{L}(r) > 0$  (или  $\hat{K}(r) > \pi r^2$ ), то это указывает на агрегацию событий на расстояниях меньших  $r$ . Если же при выходе из границ «конверта»  $\hat{L}(r) < 0$  (или  $\hat{K}(r) < \pi r^2$ ), то точечный процесс проявляет свойство регулярности на расстояниях меньших  $r$ .

Как уже говорилось  $K$ - и  $L$ -функции являются кумулятивными и отражают агрегацию или регулярность, наблюдаемую вплоть до данного расстояния  $r$ , а парная корреляционная функция  $g$  - на данном расстоянии  $r$ . Если при некотором  $r$  функция  $\hat{g}(r)$  выходит из границ «конверта», и  $\hat{g}(r) > 1$ , то это свидетельствуют об агрегации (скоплениях) событий, расположенных на таком расстоянии друг от друга (весьма вероятно, что в точечном образе встретятся точки, разделенные расстоянием  $r$ ). Если же  $\hat{g}(r)$  выходит из границ «конверта», и  $\hat{g}(r) < 1$ , то это говорит о регулярности (отталкивании) событий на расстоянии  $r$  (маловероятно, что в точечном образе встретятся точки, разделенные таким  $r$ ).

На рис. 51 представлены примеры точечных образов (тех же, что и на рис. 49), рассчитанные по ним функции-оценки  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$  и «конверты» принятия гипотезы о CSR, полученные путем симуляций однородного пуассоновского процесса. Здесь также можно видеть, что в случаях А и Б гипотеза о CSR отвергается, в случае В – принимается.

В случае А все функции говорят о наличии регулярности в размещении событий (центров клеток):  $\hat{L}(r) < 0$ ,  $\hat{K}(r) < \pi r^2$ ,  $\hat{g}(r) < 1$  при  $r$  меньших  $\approx 0.1$ . То есть центр каждой клетки окружен пространством («радиуса»  $\approx 0.1$ ), где мала вероятность встречи центра другой клетки. Можно предположить, что это значение ( $\approx 0.1$ ) определяется размером клетки.

В случае Б (проростки ежевики) четко выделяется пик первичных агрегаций с радиусом около 0.02:  $\hat{L}(r) > 0$ ,  $\hat{K}(r) > \pi r^2$ ,  $\hat{g}(r) > 1$  при  $r$  меньших  $\approx 0.02$ . Кривые  $\hat{L}(r)$  и  $\hat{K}(r)$  и на больших расстояниях  $r$  не входят в границы «конверта». Такое поведение свидетельствует о наличии более масштабных неоднородностей (крупные «пятна» зарослей проростков, состоящие из первичных агрегаций), связанных либо с особенностями размножения данного вида, либо с неоднородностью среды. Функция  $\hat{g}(r)$  значимо больше 1 лишь при  $r$  меньших  $\approx 0.12$ , что дает оценку характерного размера таких «пятен».

В случае В все функции-оценки находятся в пределах «конвертов», следовательно, можно сделать вывод, что особи сосны размещаются в пространстве полностью случайно.

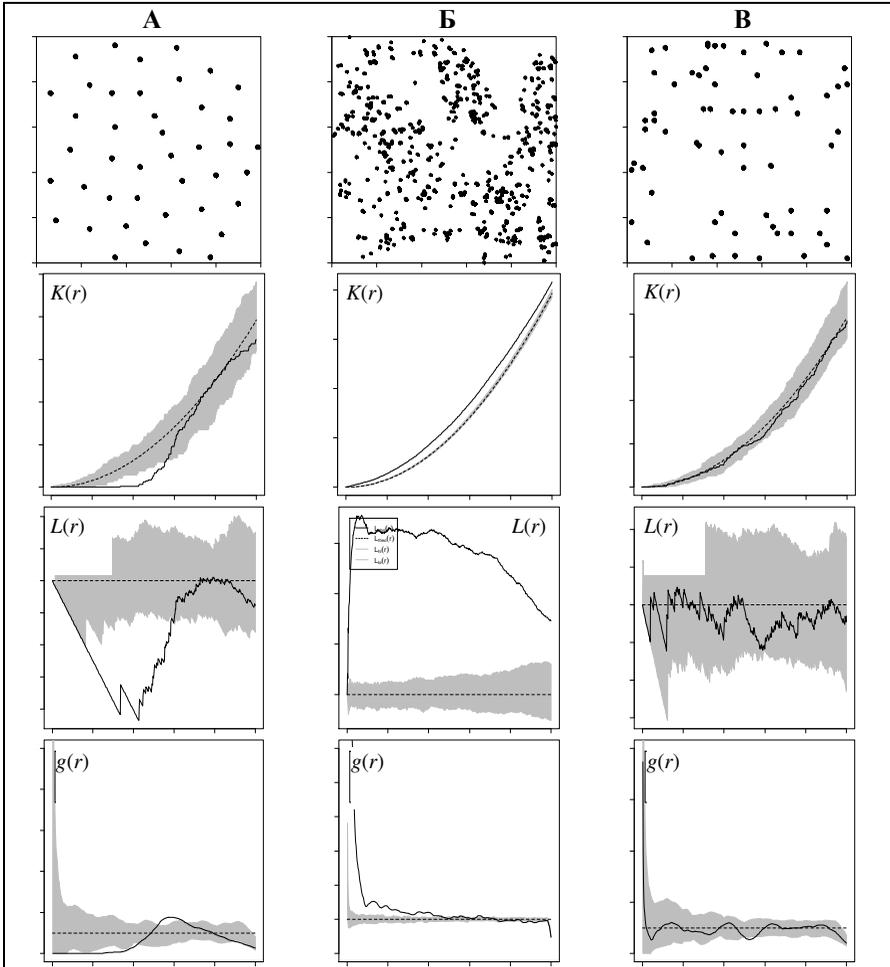


Рис. 51. А: центры клеток среза ткани, полученные под увеличением (Crick, Lawrence, 1975; Ripley, 1977); Б: проростки ежевики (Hutchings, 1979; Diggle 1983); В: проростки японской сосны *Pinus thunbergii* (Numata, 1961; Diggle 2003).

Первый ряд: точечные образы, второй ряд:  $\hat{K}(r)$ , третий ряд:  $\hat{L}(r)$ , четвертый ряд:  $\hat{g}(r)$ . Пунктирные линии – поведение функций в случае CSR. Серая заливка – 2% «конверты» принятия нулевой гипотезы о CSR, полученные с помощью 99 симуляций однородного пуассоновского процесса.

Для проверки других гипотез о виде процесса, порождающего изучаемый точечный образ, эмпирические оценки  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$  сравниваются с «конвертами», полученными на основе симулированных реализаций гипотетического процесса.

### **Пример 16**

Проверим гипотезу о CSR размещения стволов *Tilia cordata* L. с помощью критериев на основе  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$ .

Оценим для наблюдаемого точечного образа функции  $K(r)$ ,  $L(r)$ ,  $g(r)$ . Построим графики оценок, на которых также отобразим поведение этих функций в случае CSR. Чтобы определить, является ли наблюдаемое отклонение от теоретического поведения статистически значимым, построим и отобразим на этих же графиках «конверты» принятия гипотезы о CSR, полученные на основе симуляций Монте-Карло с помощью функции `envelope()` (рис. 52):

```
Kest_env1 <- envelope(ppp_object1, fun = Kest,
                     correction = "Ripley", nsim = 199)
plot(Kest_env1, lwd=2)
Lest_env1 <- envelope(ppp_object1, fun = Lest,
                     correction = "Ripley", nsim = 199)
Lest_env1$obs <- Lest_env1$obs - Lest_env1$r
Lest_env1$lo <- Lest_env1$lo - Lest_env1$r
Lest_env1$hi <- Lest_env1$hi - Lest_env1$r
Lest_env1$theo <- Lest_env1$theo - Lest_env1$r
plot(Lest_env1, lwd=2)
pcf_env1 <- envelope(ppp_object1, fun = pcf,
                    correction = "Ripley", nsim = 199)
plot(pcf_env1, lwd=2, ylim=c(0,10))
```

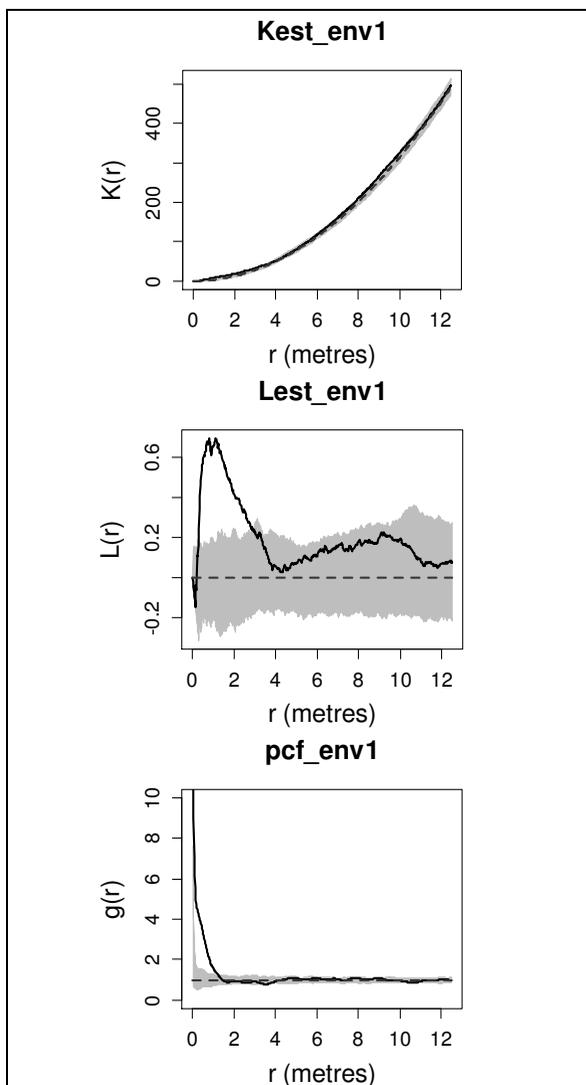


Рис. 52. Сплошной линией показаны оценки  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$ , рассчитанные для точечного образа стволов *Tilia cordata* L. Пунктирная линия – поведение функций в случае CSR. Серой заливкой показаны «конверты» принятия гипотезы о CSR, полученные с помощью 199 реализаций однородного пуассоновского процесса.

Было использовано 199 симуляций, что при фиксированном  $r$  обеспечивает уровень значимости  $p\text{-value} = 0.01$ .

Все функции-оценки  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$  имеют превышение над «конвертом», что говорит о значимом отклонении от CSR, а именно, о наличии не соответствующих пуассоновскому процессу агрегаций событий.

$\hat{L}(r)$ -функция (более наглядно, чем  $\hat{K}(r)$ ) показывает, что на изучаемой площадке имеют место скопления стволов липы сердцелистной, находящихся на расстоянии примерно от 0.5 до 1.5 м друг от друга. Превышение  $\hat{L}(r)$  над границами «конверта» на промежутке от 1.5 до 3 м, где наблюдается спад значений  $\hat{L}(r)$ , обусловлен, по-видимому, ее кумулятивным характером. Оценка парной корреляционной функции  $\hat{g}(r)$  позволяет более точно увидеть характерные размеры агрегаций, так как не обладает кумулятивным эффектом. Ее поведение свидетельствует, что размер агрегаций не превышает 1.5 м, что согласуется с результатом предыдущего примера. На более протяженных расстояниях анализ не показывает пространственных зависимостей в размещении изучаемых объектов.

### ***Пример 17***

Так как предыдущий пример выявил отклонение от полной пространственной случайности, проявляющееся наличием агрегаций, проверим гипотезу о том, что размещение стволов *Tilia cordata* L. является реализацией процесса Матерна, принадлежащего классу процессов Неймана-Скотта, моделирующих агрегационные точечные процессы.

Эта гипотеза формулируется с указанием значений параметров гипотетического процесса. Процесс Матерна имеет три параметра:  $\kappa$  – интенсивность родительского процесса (однородного пуассоновского),  $R$  – радиус круга вокруг «родителя», где размещаются «потомки»,  $\mu$  – параметр распределения Пуассона, определяющего случайное число «потомков» каждого «родителя». В качестве значений параметров берутся их оценки, получаемые методом минимального контраста, представленным в пакете **spatstat** функцией **kppm()**. Данная функция оценивает параметры кластерной модели путем «подгонки» теоретической  $K$ -функции этой модели к эмпирической  $\hat{K}(r)$ , полученной по выборочным точкам. Вместо  $K$ -функции может использоваться парная корреляционная функция  $g$ .

Параметры функции **kppm()**:

ppp_object	Объект типа rpp.object, содержащий точечный образ.
clusters	Модель гипотетического кластерного процесса: «Thomas» - процесс Томаса, «MatClust» – процесс Матерна, и др.
trend	$R$ -формула, задающая модель интенсивности (см. выше функцию <b>rppm()</b> ), если гипотетический кластерный процесс является неоднородным. В случае однородного процесса - trend = ~1.
statistic	«Подгоняемая» статистика: «K» – $K$ -функция, «pcf» – парная корреляционная функция.

1) Оценим параметры модели Матерна:

```
clust_model <- kppm(unmark(ppp_object1), trend = ~1,
                  "MatClust", statistic="pcf")
```

```
clust_model
```

```
Stationary cluster point process model
Fitted to point pattern dataset 'unmark(ppp_object1)'
Fitted by minimum contrast
```

```

Summary statistic: pair correlation function
Uniform intensity:      0.1224
Cluster model: Matern Cluster process
Fitted cluster parameters:
      kappa      R
0.1591189 0.6132708
Mean cluster size: 0.769236 points

```

Здесь  $\kappa$  - интенсивность родительского процесса (~0.16 особей на единицу площади);  $R$  - радиус круга вокруг «родителя», где размещаются «потомки» (~0.61). Среднее число «потомков» в каждом кластере - Mean cluster size = intensity/ $\kappa$  (~0.77).

Заметим, что в данном случае интенсивность процесса Матерна предполагалась постоянной на изучаемой площадке. Если моделируется неоднородный процесс Матерна, то среднее число потомков в кластере также будет меняться по территории.

2) Построим «конверты» принятия нулевой гипотезы на основе оценок функций  $K(r)$ ,  $L(r)$ ,  $g(r)$ , выведем соответствующие графики (рис. 53):

```

kest_model <-envelope(clust_model, fun= Kest,
                    correction = "Ripley", nsim = 999)
plot(kest_model,main="MatClust",xlim=c(0,2),ylim=c(0,8))
lest_model <- envelope(clust_model, fun="Lest",
                    correction = "Ripley", nsim = 999)
lest_model$obs<-lest_model$obs-lest_model$r
lest_model$lo<-lest_model$lo-lest_model$r
lest_model$hi<-lest_model$hi-lest_model$r
lest_model$mmean<-lest_model$mmean-lest_model$r
plot(lest_model,main="MatClust")
pcf_model <-envelope(clust_model, fun=pcf,
                    correction = "Ripley", nsim = 999)
plot(pcf_model, main="MatClust", ylim=c(0, 8))

```

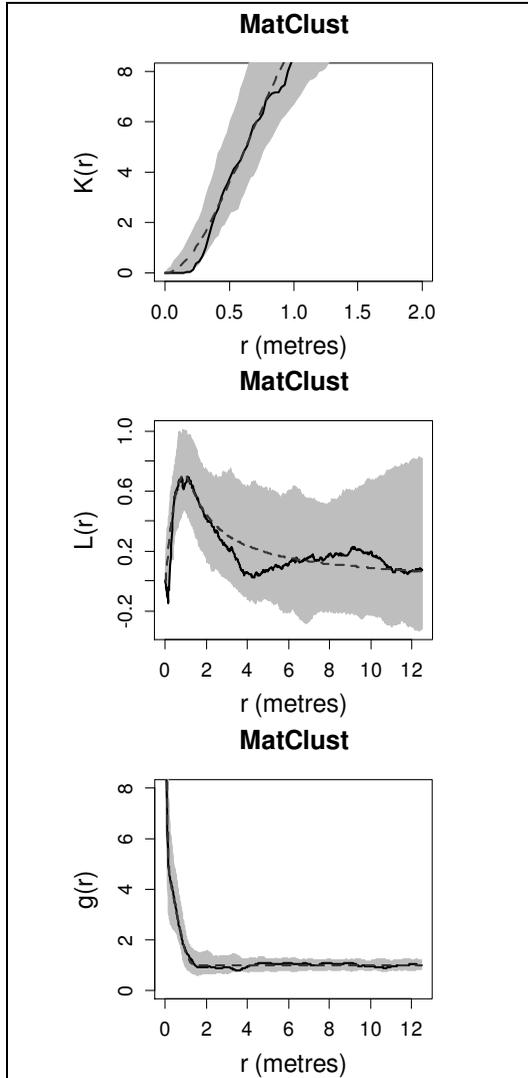


Рис. 53. Сплошная линия - оценки  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$  для точечного образа стволов *Tilia cordata* L. Пунктирная линия – поведение функций в случае процесса Матерна. Серая заливка - «конверты» принятия нулевой гипотезы, полученные с помощью 999 реализаций процесса Матерна.

Анализ полученных графиков показывает (рис. 53), что все функции-оценки находятся в пределах «конвертов», то есть изучаемые данные не дают оснований отвергнуть выдвинутую гипотезу. При фиксированном  $r$  уровень значимости  $p\text{-value} = 0.002$ .

Таким образом, точечный образ стволов *Tilia cordata* L. в исследуемой области может быть описан моделью однородного кластерного процесса Матерна.

### Пример 18

Согласно критерию хи-квадрат на основе «квадратов» (пример 9), данные о размещении особей *Adonis vernalis* L. не противоречат гипотезе о том, что этот точечный образ является реализацией неоднородного пуассоновского процесса с интенсивностью  $\lambda(x)$ , задаваемой моделью (4). Проверим эту же гипотезу с помощью критериев на основе  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$ , исследующих пространственные взаимодействия событий.

Вычислим оценки с учетом неоднородности процесса, построим «конверты» принятия нулевой гипотезы, отобразим их на графиках (рис. 54):

```
Kest_env2 <- envelope(intensity_model, fun = Kinhom,
                      correction = "Ripley", nsim= 199)
plot(Kest_env2, main="PoissonInhom")
Lest_env2 <- envelope(intensity_model, fun = Linhom,
                      correction = "Ripley", nsim =199)
Lest_env2$obs <- Lest_env2$obs - Lest_env2$r
Lest_env2$lo <- Lest_env2$lo - Lest_env2$r
Lest_env2$hi <- Lest_env2$hi - Lest_env2$r
Lest_env2$mmean <- Lest_env2$mmean - Lest_env2$r
plot(Lest_env2, main="PoissonInhom")
pcf_env2 <- envelope(intensity_model, fun = pcfinhom,
                     correction = "Ripley", nsim = 199)
plot(pcf_env2, main="PoissonInhom", ylim=c(0,10))
```

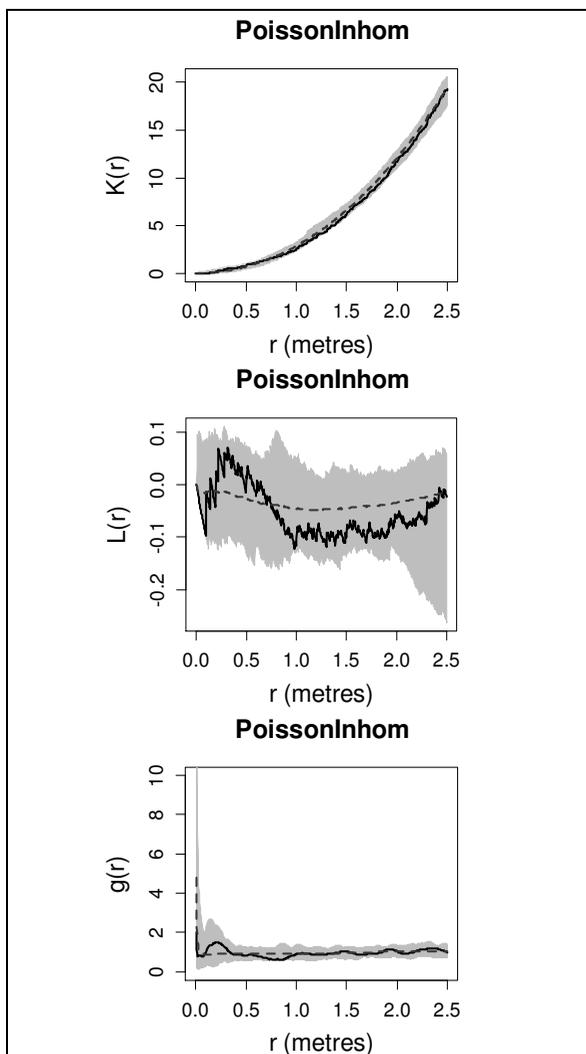


Рис. 54. Сплошной линией показаны оценки  $\hat{K}(r)$ ,  $\hat{L}(r)$ ,  $\hat{g}(r)$  для точечного образа особей *Adonis vernalis* L. Серая заливка - «конверты» принятия нулевой гипотезы, полученные по 199 реализациям неоднородного пуассоновского процесса с интенсивностью  $\lambda(x)$ , даваемой формулой (4). Пунктирная линия - среднее поведение функций для этих реализаций.

Число использованных симуляций (199) обеспечивает при фиксированном  $r$  уровень значимости  $p\text{-value} = 0.01$ .

Можно видеть (рис. 54), что все функции-оценки находятся в пределах «конвертов», следовательно, данные о размещении особей адониса не противоречат выдвинутой гипотезе. Особи адониса не показывают статистически значимого позитивного или негативного пространственного взаимодействия (агрегаций или регулярностей в размещении).

*Таким образом, моделью пространственного процесса, порождающего точечный образ особей *Adonis vernalis* L. на исследуемой площадке, может считаться неоднородный пуассоновский процесс с интенсивностью*

$\lambda(x) = \lambda(Y) = e^{-0.4736+0.1527 \cdot Y}$ . Полученный вывод согласуется с априорными знаниями об изучаемой территории, рельеф которой понижается по направлению к южной границе площадки. Северная часть площадки находится на пологой вершине склона, где скот поедает растения конкурирующих видов, не трогая адонис в силу его ядовитости. Это создает более благоприятные экологические условия для данного вида, увеличивая интенсивность процесса в северном направлении.

Более детальный анализ точечного рисунка размещения этого вида может быть проведен с учетом тематической информации – «меток» каждой особи (ее онтогенетическое состояние, число цветков). Для этого должны привлекаться методы анализа маркированных точечных образов. Теория маркированных точечных процессов и методы анализа маркированных точечных образов в этом пособии не рассматриваются.

## ЛИТЕРАТУРА

1. *Василевич В. И.* Статистические методы в геоботанике / Л.: Наука, 1969. – 273 с.
2. *Гмурман В.Е.* Теория вероятностей и математическая статистика / М.: Высшая школа, 2000. - 479 с.
3. *Грабарник П.Я.* Моделирование и методы статистического анализа пространственной структуры древостоев на основе случайных точечных полей. Автореферат дисс. на соиск. учен. степ. доктора физ.-мат. наук / Петрозаводск, 2013. – 43 с.
4. *Грабарник П.Я.* Статистические методы в пространственной экологии: новый взгляд на типичные задачи анализа экологических данных // Принципы и способы сохранения биоразнообразия: материалы Второй Национальной конференции с международным участием. - Пушино, ИФХиБПП РАН, 2011. - С. 80–81.
5. *Грабарник П.Я.* Моделирование пространственной структуры древостоев // Моделирование динамики органического вещества в лесных экосистемах. - М.: Наука, 2007. - С. 132–146.
6. *Грабарник П.Я.* Марковские модели пространственных точечных процессов с несколькими типами точек // Математическое моделирование популяций растений и фитоценозов. - М.: Наука, 1992. - С. 85–90.
7. *Грабарник П.Я., Комаров А.С.* Статистический анализ горизонтальной структуры древостоя // Моделирование биогеоэкологических процессов. - М.: Наука, 1981. - С. 119–135.
8. *Грейг–Смит П.* Количественная экология растений / М.: Наука. 1967. - 359 с.
9. *Епанечников В.А.* Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и ее применение. – 1969. - Т. 14 (1). - С. 156–161.
10. *Савельев А.А., Мухарамова С.С., Пилюгин А.Г., Чижикова Н.А.* Геостатистический анализ данных в экологии и природопользовании (с применением пакета R). Учебное пособие / Казань: Издательство Казанского университета, 2012. - 120 с.

11. *Савельев А.А., Пилюгин А.Г., Мухарамова С.С.* Основные понятия языка R. Учебно-методическая разработка / Казань: Изд-во Казан. ун-та, 2007. – 30 с.
12. *Савельев А.А., Пилюгин А.Г., Мухарамова С.С.* Использование языка R для статистической обработки данных. Учебно-методическая разработка / Казань: Изд-во Казан. ун-та, 2007. – 30 с.
13. *Фардеева М.Б., Саидова Н. В.* Жизненность и виталитетная структура ценопопуляций *Adonis vernalis* (Ranunculaceae) в различных зональных условиях Республики Татарстан // Растительные ресурсы. – 2010. - Т. 46 (1). - С. 17–27.
14. *Чижикова Н.А.* Некоторые закономерности формирования дискретных структур в континууме растительного покрова. Автореферат дисс. на соиск. учен. степ. кандидата биологических наук / Казань, 2008. - 24 с.
15. *Baddeley A.* Analysing spatial point patterns in R / CSIRO, 2011. – 232 p.
16. *Baddeley A., Turner R.* Spatstat: an R package for analyzing spatial point patterns // Journal of Statistical Software. – 2005. - V. 12 (6). - P. 1–42. URL <http://www.jstatsoft.org/v12/i06/>.
17. *Baddeley A., Møller J., Waagepetersen. R.* Non- and semi-parametric estimation of interaction in inhomogeneous point patterns // Statistica Neerlandica. -2000. – V.54. – P.329 – 350.
18. *Baddeley A., Gill R.D.* Kaplan–Meier estimators of distance distributions for spatial point processes // Annals of Statistics. – 1997. - V. 25. - P. 263–292.
19. *Bailey T.C., Gatrell A.C.* Interactive Spatial Data Analysis / Harlow: Longman, 1995. - 413 p.
20. *Berman M., Diggle P.* Estimating weighted integrals of the second-order intensity of a spatial point process // Journal of the Royal Statistical Society, Series B. – 1989. - V. 51. - P. 81–92.
21. *Besag J., Diggle P.J.* Simple Monte Carlo tests for spatial pattern // Applied Statistics. – 1977. - V. 26. - P. 327–333.
22. *Besag J.E.* Comments on Ripley's paper // Journal of the Royal Statistical Society, Series B. – 1977. - V. 39. - P. 193–195.
23. *Byth K., Ripley, B. D.* On sampling spatial patterns by distance methods // Biometrics. – 1980. - V. 36. - P. 279–284.

24. *Clark P.J., Evans F.C.* Distance to nearest neighbour as a measure of spatial relationships in populations // *Ecology*. – 1954. – V. 35. – P. 445–453.
25. *Condit R., P. S. Ashton, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S.P. Hubbell, R.B. Foster, L. Hua Seng, A. Itoh, J.V. LaFrankie, E. Losos, N. Manokaran, R. Sukumar, T. Yamakura.* Spatial patterns in the distribution of common and rare tropical tree species: a test from large plots in six different forests // *Science*. – 2000. – V. 288. – P. 1414–1418.
26. *Cox D. R.* Some Statistical Methods Connected with Series of Events // *Journal of the Royal Statistical Society*. – 1955. – V. 17 (2). – P. 129–164.
27. *Crick F. H., Lawrence P. A.* Compartments and polyclones in insect development // *Science*. – 1975. – V.189. – P.340–347.
28. *Cressie N.A.C.* *Statistics for Spatial Data* / New York: Wiley, 1993. – 928 p.
29. *Dale M.R.T.* *Spatial pattern analysis in plant ecology* / Cambridge: Cambridge University Press, 1999. – 340 p.
30. *Daley D.J., Vere-Jones D.* *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods* / Springer, 2003. – 471 p.
31. *Diggle P.J.* *Statistical Analysis of Spatial Point Patterns* (second edition) / London: Edward Arnold Publishers, 2003. – 159 p.
32. *Diggle P. J.* A kernel method for smoothing point process data // *Journal of the Royal Statistical Society, Series C*. – 1985. – V. 34. – P. 138–147.
33. *Diggle P.J.* *Statistical analysis of spatial point patterns* / London: Academic Press, 1983. – 148 p.
34. *Dixon P.M.* Nearest neighbor methods // *Encyclopedia of Environmetrics*. – 2002. – V.3. – P. 1370–1383.
35. *Donnelly K.* Simulations to determine the variance and edge-effect of total nearest neighbour distance // *Simulation methods in archaeology*. – Cambridge University Press, 1978. – P. 91–95.
36. *Fisher R. A., Thornton H. G., Mackenzie W. A.* The accuracy of the plating method of estimating the density of bacterial populations // *Annals of Applied Biology*. – 1922. – V.9. – P. 325–359.

37. *Gatrell A.C., Bailey T.C., Diggle P.J., Rowlingson B.S.* Spatial point pattern analysis and its application in epidemiology // Transactions of the Institute of British Geographers. – 1996. - V. 21 (1). - P. 256–274.
38. *Getis A., Franklin J.* Second-order neighbourhood analysis of mapped point patterns // Ecology. – 1987. - V. 68. - P. 473–477.
39. *Grabarnik P., Myllymäki M., Stoyan, D.* Correct testing of mark independence for marked point patterns // Ecological Modelling. – 2011. - V. 222. - P. 3888–3894.
40. *Grandell J.* Doubly stochastic Poisson processes // Lecture Notes in Mathematics, V. 529. - Berlin–New York: Springer–Verlag, 1976. - 234 p.
41. *Greig-Smith P.* The use of random and contiguous quadrats in the study of the structure of plant communities // Ann. Bot. London, N.S. – 1952. – V.16. – P. 293-316.
42. *Greig-Smith P.* Quantitative Plant Ecology / London: Butterworth, 1964. – 256 p.
43. *Guttorp P.* Statistical inference for branching processes / New York: Wiley, 1991. - 224 p.
44. *Hopkins B.* A new method for determining the type of distribution of plant individuals // Annals of Botany. – 1954. - V. 18. - P. 193–127.
45. *Hutchings M.J.* Weight–density relationships in ramet populations of clonal perennial herbs, with special reference to the negative three–halves power law // Journal of Ecology. – 1979. - V. 67. - P. 21–34.
46. *Kershaw K. A.* The use of cover and frequency in the detection of pattern in plant communities // Ecology. – 1957. - V. 38. - P. 291–299.
47. *van Lieshout M.N.M., Baddeley A.J.* A nonparametric measure of spatial interaction in point patterns // Statistica Neerlandica. – 1996. - V. 50. - P. 344–361.
48. *Lloyd M.* Mean crowding // Journal of Animal Ecology. – 1967. – V.36. - P. 1–30.
49. *Matern B.* Spatial Variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations // Meddelanden fran Statens Skogsforskningsinstitu. – 1960. - V. 49 (5). - 144 p. (second edition: 1986, Lecture Notes in Statistics, no. 36, Springer, New York, 151 p.)
50. *Møller J.* Spatial Statistics and Computational Methods // Springer Lecture Notes in Statistics, V. 173. - New York: Springer, 2003. – 205 p.

51. *Møller J., Waagepetersen R.* Statistical Inference and Simulation for Spatial Point Processes / Chapman and Hall/CRC, 2003. - 320 p.
52. *Morisita M.* Measuring the dispersion and the analysis of distribution patterns // Memoires of the Faculty of Science, Kyushu University, Series E. Biology. – 1959. - V.2. - P. 215–235.
53. *Neyman J.* On a new class of “contagious” distribution, applicable in entomology and bacteriology // Annals of Mathematical Statistics. – 1939. - V. 10. - P. 35–55.
54. *Numata M.* Forest vegetation in the vicinity of Choshi // Coastal flora and vegetation at Choshi, Chiba Prefecture. IV. Bulletin of Choshi Marine Laboratory, Chiba University. – 1961. - V.3. - P. 28–48.
55. *Ohser J.* On estimators for the reduced second moment measure of point processes // Mathematische Operationsforschung und Statistik, series Statistics. – 1983. - V. 14. - P.63–71.
56. *Parzen E.* On estimation of a probability density function and mode // Annals of Mathematical Statistics. – 1962. - V. 33(3). - P. 1065–1076.
57. *Penttinen A., Stoyan D., Henttonen H.* Marked point processes in forest statistics // Forest Science. – 1992. - V. 38. - P. 806–824
58. *Pielou E.C.* The use of point-to-plant distances in the study of the pattern of plant populations // Journal of Ecology. – 1959. - V. 47. - P.607–613.
59. *Pollard J.H.* On distance estimators of density in randomly distributed forests // Biometrics. – 1971. - V. 27. - P. 991–1002.
60. *R Core Team.* R: A language and environment for statistical computing / R Foundation for Statistical Computing, Vienna, Austria, 2014. URL <http://www.R-project.org/>.
61. *Ripley B.D.* Statistical Inference for Spatial Processes / Cambridge University Press, 1988. - 148 p.
62. *Ripley B. D.* Spatial Statistics / New York: Wiley, 1981. – 252 p.
63. *Ripley B. D.* Modelling spatial patterns (with discussion) // Journal of the Royal Statistical Society, Series B. – 1977. - V. 39. - P. 172–212.
64. *Ripley B. D.* The second-order analysis of stationary point processes // Journal of Applied Probability. – 1976. - V. 13. - P. 255–266.
65. *Rosenblatt M.* Remarks on some nonparametric estimates of a density function // Annals of Mathematical Statistics. – 1956. - V. 27. - P.832–837.
66. *Silverman B.W.* Density estimation for statistics and data analysis / London: Chapman and Hall, 1986. - 176 p.

67. *Skellam J.D.* Studies in statistical ecology. 1. Spatial pattern // *Biometrika*. – 1952. - V. 39. - P. 346–362.
68. *Stoyan D., Penttinen A.* Recent applications of point process methods in forestry statistics // *Statistical Science*. – 2000. - V. 15 (1). - P. 61–78.
69. *Stoyan D., Stoyan H.* *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics* / Wiley, 1994. – 406 p.
70. *Thompson H.R.* Spatial point processes, with application to ecology // *Biometrika*. – 1955. - V. 42 (1/2). - P. 102–115.
71. *Thompson C.J.* *Mathematical statistical mechanics* / New York: The Macmillian Company, 1972. - 278 p.
72. *Turner M.G., Gardner R.H., O'Neill V.R.* *Landscape Ecology in Theory and Practice: Pattern and Process* / Springer, 2001. - 414 p.
73. *Watt A. S.* Pattern and Process in the Plant Community // *The Journal of Ecology*. – 1947. - V. 35 (1/2). - P. 1–22.
74. *Wiegand T., Moloney K.A.* Rings, circles and null-models for point pattern analysis in ecology // *Oikos*. – 2004. - V. 104. - P. 209–229.
75. *Zuur A.F., Ieno E.N., Smith G.M.* *Analysing Ecological Data (second edition)* / Springer, 2011. - 672 p.

## ПРИЛОЖЕНИЕ

```
my_miplot<-function (X, ...)
{
  Xname <- deparse(substitute(X))
  X <- as.ppp(X)
  W <- X$window
  N <- X$n
  if (W$type != "rectangle")
    stop("Площадка не прямоугольна - индекс не
определен")
  a <- min(diff(W$xrange), diff(W$yrange))
  maxnquad <- floor(a/mean(nndist(X)))
  if (maxnquad <= 1)
    stop("Недостаточно точек для теста")
  mindex <- numeric(maxnquad)
  for (nquad in 1:maxnquad) {
    qq <- quadratcount(X, nquad, nquad)
    tt <- as.vector(as.table(qq))
    mindex[nquad] <- length(tt) * sum(tt * (tt -
1))/(N *
      (N - 1))
  }
  quadsizesize <- diameter(W)/(1:maxnquad)
  unitinfo <- summary(unitname(W))$axis
  plot(quadsizesize, mindex, xlim = c(0, a), ylim = c(0,
max(mindex)),
    xlab = "Диаметр (диагональ) квадрата", ylab =
"Индекс Мориситы",
    main = "", cex = 0.5)
  lines(quadsizesize, mindex)
  abline(h = 1, lty = 2, col = "red")
  legend("topright", lty = c(1, 2), col = c(1, 2),
    legend = c("По наблюдениям", "CSR"))
}
```

*Учебное издание*

**Теория пространственных точечных процессов  
в задачах экологии и природопользования  
(с применением пакета R):**

Учебное пособие

Дизайн обложки

***М.А. Ахметов***

Подписано в печать 20.11.2014.

Бумага офсетная. Печать ризографическая.

Формат 60x84 1/16. Гарнитура «Times New Roman».

Усл. печ. л. 9,25. Уч.-изд. л. 4,1. Тираж 100 экз. Заказ 238.

Отпечатано с готового оригинал-макета  
в лаборатории оперативной полиграфии Издательства КФУ

420012, г. Казань, ул. Бутлерова, 4

тел. (843) 291-13-88, 291-13-47