

Санкт-Петербургский государственный университет

Направление: «Математика» (010101)

Профиль: «Теория вероятностей и математическая статистика» (010105)

Данилова Елена Владимировна

Методы оценки периода циклического пуассоновского процесса

Дипломная работа

Научный руководитель:

профессор, д. физ.-мат. наук Грибкова Н.В.

Рецензент:

профессор, д. физ.-мат. наук Боровских Ю.В.

Санкт-Петербург

2016

SAINT-PETERSBURG STATE UNIVERSITY

Main Field of Study: "Mathematics" (010101)

Area of Specialisation: "Probability theory and mathematical statistics" (010105)

Elena Danilova

Methods of estimating the period of a cyclic Poisson process

Graduation Thesis

Scientific Supervisor:

Professor, Doctor of Physics and Mathematics N. V. Gribkova

Reviewer:

Professor, Doctor of Physics and Mathematics Yu.V. Borovskikh

Saint-Petersburg

2016

### **Аннотация.**

Данилова, Е.В. Методы оценки периода циклического пуассоновского процесса. Научный руководитель: Грибкова Н.В. Направление: математика, кафедра теории вероятностей и математической статистики.

В работе изучаются и систематизируются основные методы оценки периода циклического процесса Пуассона. Библ. – 9 назв.

Данилова, Е.В. Методы оценки периода циклического пуассоновского процесса: дипломная работа: защищена 09.06.16. / Данилова Елена Владимировна. – СПб., 2016 – 21 с. – Библиогр.: с. 21.

## Содержание.

|  |    |
|--|----|
| Глава 1.   |    |
| 1.1 Определения и основные обозначения.....        | 5  |
| 1.2 Введение.....                                  | 6  |
| 1.3 Постановка задачи и обзор литературы.....      | 8  |
| Глава 2. Обзор основных результатов                |    |
| 2.1 Метод Хелмерса – Мангку [1].....               | 10 |
| 2.2 Метод Беббингтона – Зитикиса [2].....          | 13 |
| 2.3 Метод Белитсера – Серра – ван Зантена [3]..... | 18 |
| Заключение.....                                    | 20 |
| Список литературы.....                             | 21 |

## Глава 1.

### 1.1 Определения и основные обозначения.

В этой работе мы будем рассматривать специальный случай неоднородного процесса Пуассона – циклический процесс Пуассона, т.е. процесс с периодической интенсивностью. Дадим определение такого процесса.

#### Определение 1.

*Неоднородный процесс Пуассона с периодической интенсивностью.*

Пусть  $X(s)$ ,  $s \geq 0$  – процесс с независимыми приращениями. Тогда

$$P(X(s) - X(t) = k) = \frac{[\Lambda(s) - \Lambda(t)]^k * e^{-[\Lambda(s) - \Lambda(t)]}}{k!}, 0 \leq t \leq s,$$

где  $\Lambda$  называется ведущей функцией процесса. Производная этой функции называется интенсивностью  $\Lambda'(s) = \lambda(s)$ . Причем в данном случае  $\lambda(s)$  – периодическая функция. У *однородного* процесса  $\lambda$  постоянна:  $\Lambda(s) = \lambda s$ , а у *неоднородного*  $\Lambda(s) = \int_0^s \lambda(t) dt$ .

Из определения процесса Пуассона видно, что приращения этого процесса имеют распределение Пуассона с параметрами  $\Lambda(s) - \Lambda(t)$ . Отсюда следует, что *математическое ожидание приращения* для любых двух точек  $t \leq s$  равно разности этих функций.

$E(X(s) - X(t)) = \Lambda(s) - \Lambda(t) = \int_t^s \lambda(t) dt$  И, соответственно, если есть интенсивность, то это интеграл. *Дисперсия* совпадает с математическим ожиданием:  $D(X(s) - X(t)) = \Lambda(s) - \Lambda(t)$ .

## 1.2 Введение.

Понятие потока однородных событий возникло в математике как отражение различных реальных явлений, например, потока вызовов на АТС, потока клиентов, несчастных случаев, прибытия пациентов в отделении интенсивной терапии больницы, при анализе транспортных потоков на магистралях. Он также применяется в страховании и анализе финансовых механизмов, таких как поток платежей и других реальных потоков. Теорию потока однородных событий разработал советский математик А. Я. Хинчин, впоследствии она легла в основу теории массового обслуживания. Целью исследований теории массового обслуживания является рациональный выбор структуры системы обслуживания и процесса обслуживания на основе изучения потоков, поступающих в систему и выходящие из неё, а также длительности ожидания.

Методы для анализа натуральной периодической или циклической информации хорошо понятны в случае, когда данные формируют временные ряды. Тем не менее, часто данные состоят из событий, случайным образом расположенных в промежутке времени. Точечные процессы образуют естественный класс стохастических моделей для таких данных. Мы называем данные циклическими, если функция интенсивности является периодической. Такие процессы стали рассматривать в связи с необходимостью решать прикладные задачи, в которых определенные события происходят в случайные моменты времени и с некоторой периодичностью. Хорошую интерпретацию циклического процесса дает процесс моментов отказа технического оборудования, зависящего от сезонных факторов: мороз, жара, влажность и т.д. Интенсивность таких процессов часто можно считать периодической, она зависит от времени года. Для моделирования таких потоков используют процесс Пуассона.

Во многих задачах обычно используется то, что период известен. Но на практике это часто бывает не так. Оценка периода является интересной

проблемой сама по себе, но также и является важным компонентом для непараметрического оценивания функции интенсивности. Если мы не знаем периода, то мы не можем оценить и функцию интенсивности, так как существующие методы оценки функции интенсивности основаны на знании периода и нуждаются в оценке периода до построения непараметрической оценки. Проблема оценки может быть формально сформулирована так: мы хотим оценить одномерный параметр (период) в присутствии бесконечномерного дополнительного параметра (функция интенсивности). Поэтому для нас основной интерес составляет оценка или аппроксимирование этой функции интенсивности  $\lambda(s)$  и существующие методы для этого.

Рассмотрим  $X$  неоднородный точечный процесс Пуассона на действительной прямой  $\mathbb{R}$  с неизвестной локально интегрируемой функцией интенсивности  $\lambda(s)$  для некоторого периода  $\tau > 0$ . Мы рассмотрим ситуацию, когда интенсивность  $\lambda$  – это периодическая функция, т.е. выполняется такое равенство:  $\lambda(s) = \lambda(s + k\tau)$ ,  $k = \pm 1, \pm 2, \dots$ , где  $\tau$  – это наименьшая величина, для которой это выполняется – период. Предположим, что мы имеем только одну реализацию  $X$  в ограниченном временном интервале или «окне»,  $W \subset \mathbb{R}$ .

Если лямбда произвольная параметрическая функция, допустим:  $\lambda(s) = \sin(ax + b)$ , то для оценки параметров можно воспользоваться методом максимального правдоподобия.

Часто бывает так, что параметрическая модель не верна, возникает ситуация с многомодальностью, когда функция интенсивности внутри периода имеет несколько максимумов, т.е. может не описываться никакой параметрической моделью. Поэтому очень важно иметь непараметрические методы оценки периода.

### 1.3 Постановка задачи и обзор литературы.

Основной целью работы является изучение и систематизация материалов, сравнение методов оценки периода. На данный момент существует не очень много работ по этой теме. Приведем ниже наиболее примечательные из них.

Впервые метод оценки периода циклического процесса Пуассона был представлен в 1982 году в работе Вере-Джонса [7]. С помощью спектрального анализа он рассчитал оценку на основе максимума периодограммы Бартлетта для частоты  $\omega$  для циклической Пуассоновской интенсивности  $\lambda(s) = A \exp(\rho \cos(\omega s + \phi))$ , а также исследовал ее асимптотическое поведение. Вере-Джонс также отметил, что эта оценка эффективна только для специальных моделей точечных процессов, когда функция интенсивности имеет форму, близкую к предполагаемой.

От стандартных параметрических методов для оценки интенсивности циклического процесса Пуассона, которыми являются метод периодограмм и подходы спектрального анализа, Беббингтон и Зитикис [2] в своей работе в 2004 году перешли к полу-параметрическому методу, который чувствителен к форме интенсивности, в частности к числу пиков в цикле. Они дали устойчивую оценку для периода интенсивности функции Пуассона, построив семейство непараметрических оценок для периода, и протестировали ее на искусственных данных. Эти непараметрические оценки оказались достаточно точны, в то время как метод периодограмм вообще не мог справиться с мультимодальными циклами.

Хелмерс, Мангку и Зитикис [5], [8] исследуют состоятельность простой непараметрической оценки ядерного типа  $\hat{\tau}_n$  периода  $\tau$  для функции интенсивности циклического процесса Пуассона. Они рассматривают ситуацию при фиксированной  $\lambda$  и наблюдении только одной реализации  $X(\omega)$  процесса Пуассона в ограниченном, расширяющемся окне  $W_n$ . Предлагаемая оценка состоятельна при бесконечном расширении окна.



Получена полная сходимость. Одновременно с ними, Хелмерс и Мангку [1] продолжают исследовать свойства оценок для периода в предложенных выше условиях. Авторы также используют идею предложенную Беббингтоном и Зитикисом [2] для построения своей оценки. Они улучшили результаты и доказали состоятельность оценки.

Белитсер, Серра и ван Зантен [3] продолжают исследовать свойства полу-параметрических оценок периода неоднородного процесса Пуассона, предлагая свою M-оценку для периода. Авторами было получено новое семейство оценок для периода функции интенсивности и доказана состоятельность таких оценок. В работе также проводится сравнение качества полученной оценки  $\hat{t}_n$  на тестах на основе искусственных данных, предложенных Беббингтоном и Зитикисом в [2]. Авторы сопоставляют свою оценку с параметрической оценкой Вере-Джонса на основе метода периодограмм [7], оценкой  $\hat{t}_{n,max}$  Хелмерса и Мангку [6] и «сглаженными» оценками  $\hat{t}_{n,M}$  из работы [2]. Метод, предложенный Хелмерсом, Мангку и Зитикисом демонстрируют хорошие показатели на искусственном наборе данных, а также на реальном наборе из колл-центра.

## Глава 2. Обзор основных результатов.

### 2.1 Метод Хелмерса – Мангку [1].

Рассмотрим идею метода и основные результаты. На вероятностном пространстве  $(\Omega, \mathcal{A}, P)$  задан циклический точечный процесс Пуассона  $X$ ,  $\sigma$ -конечная мера  $\mu$ , абсолютно непрерывная относительно меры Лебега  $\nu$ , и локально интегрируемая функцией интенсивности  $\lambda: \mathbb{R} \rightarrow \mathbb{R}^+ \cup 0$ .  $\lambda$  задана таким образом, что  $\forall$  ограниченного борелевского множества  $B$ , имеем  $\mu(B) = \int_B \lambda(s) ds < \infty$ .

Также  $\lambda$  циклическая с периодом  $\tau$ , т.е. для некоторого  $\tau \in \mathbb{R}^+$  выполнено  $\lambda(s + k\tau) = \lambda(s)$  для всех  $s \in \mathbb{R}$  и  $k \in \mathbb{Z}$ .

Период  $\tau$  и функцию интенсивности  $\lambda$  будем считать неизвестными и предположим, что для некоторого  $\omega \in \Omega$ , мы наблюдаем лишь одну реализацию  $X(\omega)$  пуассоновского процесса  $X$  через ограниченный интервал, называемый окном  $W \subset \mathbb{R}$ .

Так как  $\lambda$  локально интегрируема, точечный процесс Пуассона всегда содержит только конечное число точек в любом ограниченном подмножестве  $\mathbb{R}$ . Для того чтобы исследовать состоятельность оценки  $\tau$ , предположим, что окно  $W$  зависит от "времени"  $n = 1, 2, \dots$ , таким образом, что  $|W_n| \rightarrow \infty$  при  $n \rightarrow \infty$ , где  $|W_n|$  - это размер окна  $W_n$ , или мера Лебега.

Необходимым условием для существования состоятельной оценки  $\tau$  будет следующее равенство:  $\int_{\mathbb{R}} \lambda(s) ds = EX(\mathbb{R}) = \infty$ , откуда следует, что с вероятностью почти наверно точечная реализация  $X(\omega)$  содержит бесконечно много точек.

Для циклических функций интенсивности это требование выполняется автоматически при условии, что глобальная интенсивность  $\theta = \tau^{-1} \int_0^\tau \lambda(s) ds$  является положительной. Будем считать, что  $\theta > 0$ .

*Идея методики.*

Следующая конструкция была предложена Беббингтоном и Зитикисом [2].

Обозначим интервал, на котором мы рассматриваем реализацию периодического процесса Пуассона через  $W_n := (a, b_n]$ . Этот интервал конечен,  $a \in \mathbb{R}$  фиксирована и правая граница будет стремиться к бесконечности  $b_n \rightarrow \infty$ , когда время  $n \rightarrow \infty$ . При данной реализации мы хотим оценить период интенсивности  $\tau$  от  $\lambda$ . В это окно  $W_n$  мы вписываем непересекающиеся интервалы  $I_{1, \dots, I_K}$  эквивалентной длины  $\ell$ . Мы можем менять  $\ell$  и двигать эту конструкцию вправо и влево с тем, чтобы в интервалы попало примерно одинаковое количество точек. По-видимому, период будет найден верно, если дисперсия относительно среднего числа точек (попаданий) будет как можно меньше.

Основываясь на этом, Хелмерс и Мангку предложили оценку  $\hat{\tau}_n$  периода  $\tau$ :

$$Q_n(\delta) = \frac{1}{|W_n|} \sum_{i=1}^{N_{n,\delta}} (X(U_{\delta,i}) - \frac{1}{N_{n,\delta}} \sum_{j=1}^{N_{n,\delta}} X(U_{\delta,j}))^2, \forall \delta \in \Theta$$

$\Theta$  - параметрическое пространство,  $\tau \in \Theta$  и пусть  $\Theta$  ограниченный открытый интервал в  $\mathbb{R}^+$ , такой, что никаких чисел кратных  $\tau$  не содержится в  $\Theta$ . Где  $N_{n,\delta} = \lfloor \frac{|W_n|}{\delta} \rfloor$  - это (максимальное) число смежных непересекающихся интервалов  $U_{\delta,i}$  длины  $\delta$  в окне  $W_n$ . Обозначим через  $a_n, b_n$  левую и правую конечные точки, т.е.  $W_n = [a_n, b_n]$ .

И пусть  $U_{\delta,i}$  - интервалы такого вида  $[a_n + r + (i - 1)\delta, a_n + r + i\delta]$  для некоторого  $r \in [0, (|W_n| - \delta N_{n,\delta})]$ . Иначе выбор  $r$  свободен и в принципе не имеет значения (см. Условие E).

Определим оценку периода  $\tau$ :

$$\hat{\tau}_n = \arg \min_{\delta \in \Theta} Q_n(\delta)$$

Параметр  $\tau$  может быть оценен следующим образом:

сперва оценим  $k\tau$ , для некоторого положительного  $k$ , удовлетворяющего  $k = o(|W_n|)$ ,  $k\hat{t}_{n,k}$  задается

$$k\hat{t}_{n,k} = \arg \min_{\delta \in \Theta_k} Q_n(\delta)$$

Пусть  $\hat{t}_{n,k}$  - полученная оценка  $\tau$ .

Главным результатом работы Мангку и Хелмерса является следующая теорема.

### Теорема 1.

Предположим, что  $\lambda$  циклическая (с периодом  $\tau$ ) и Липшицева. Пусть условие (E) выполнено и для некоторого  $0 \leq c < \frac{1}{3}$  верно  $k = k_n |W_n|^c = 1$ .

Тогда для любого  $\gamma < \frac{1}{2}$  выполняется

$$|W_n|^\gamma |\hat{t}_{n,k} - \tau| \xrightarrow{p} 0$$

при  $n \rightarrow \infty$ .

Иными словами, разность имеет порядок «о»-малое по сравнению с гаммовой степенью величины стремящейся к нулю.

Требование  $c < \frac{1}{3}$  задает границу, которая определяет максимальное количество, которое мы можем оценить без потери состоятельности.

### Условие E.

Если существует  $t \in (0, \tau)$ : для каждого  $n \geq 1$ ,

$$\int_{U_{t,i}} \lambda(s) ds = t\theta$$

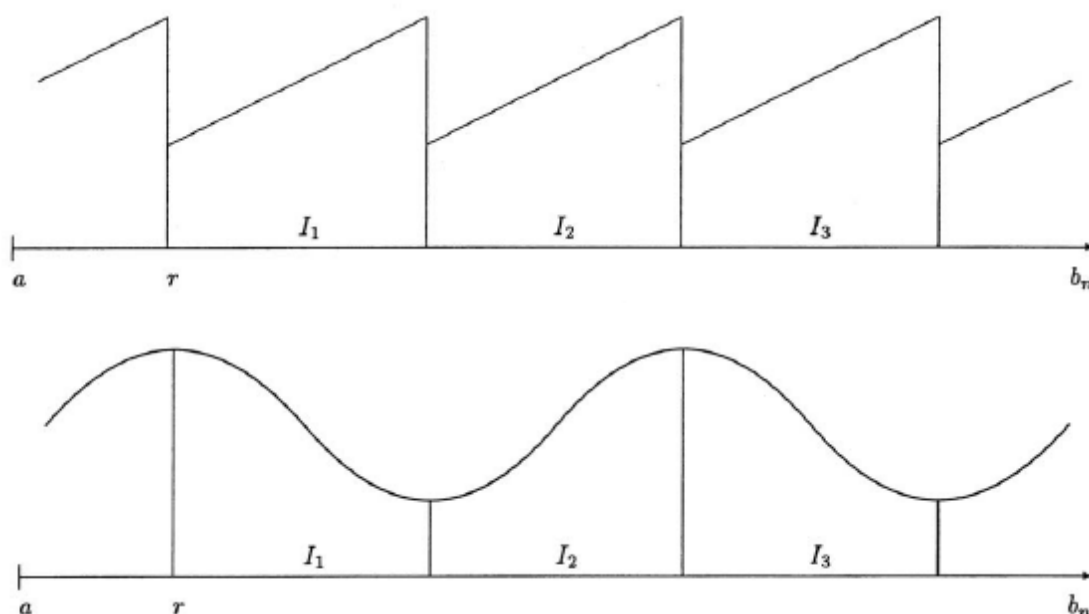
с  $U_{t,i} = [a_n + r + (i - 1)t, a_n + r + i t]$  для  $i = 1, \dots, N_{n,t}$ , тогда мера множества, для которого верно выражение в скобках, равна 0:

$$v \left\{ r: \int_{U_{t,i}} \lambda(s) ds = t\theta, i = 1, \dots, N_{n,t} \right\} = 0.$$

## 2.2 Метод Беббингтона – Зитикиса [2].

В данной статье Беббингтон и Зитикис предложили устойчивую оценку для периода интенсивности функции Пуассона, построив семейство непараметрических оценок, а затем провели ее испытания на искусственных данных. В работе отмечены ограничения различных оценок, проведено сравнение их производительности и надежности.

*Непараметрическая оценка.*



*Рис.1.*

Для нахождения оценки построим конструкцию, которую мы описали в параграфе 2.1 и рассмотрим две периодические функции интенсивности и соответствующий выбор интервалов для них  $I_1, \dots, I_K$ . Для первой функции интенсивности длины интервалов – это период  $\tau$ , для нижней – длина половины периода  $\frac{\tau}{2}$ .

Если длина  $\ell$  была получена вышеописанным способом, то  $m\ell$ ,  $m \in \mathbb{N}$  будут также удовлетворять критерию, описанному в предыдущем параграфе. При этом мы стремимся найти интервалы  $I_1, \dots, I_K$  с минимальной длиной  $\ell$ .

$X(I_k)$  – число точек в любом  $I_k$  из полученных интервалов. Согласно нашей конструкции, все различия  $X(I_k) - X(I_{k+1}), k = 1, \dots, K - 1$  должны быть настолько малы, насколько возможно. Мангку [6] требует, чтобы разница была малой в среднем и дает следующую оценку минимальной длины  $\ell$ :

$$\hat{\tau}_{n,min} := \arg \min_{\ell} \min_{(1)} \min_{(2)} \sum_{k=1}^{K-1} (X(I_k) - X(I_{k+1}))^2 \quad (2),$$

$\min_{(2)}$  - это минимум, взятый по всем смежным интервалам  $I_1, \dots, I_K \subset W_n$  длины  $\ell$ ,  $\min_{(1)}$  - берется по всем длинам  $\ell > 0$ , а  $\arg$  – это длина  $\ell$  на которой  $\min_{(1)}$  достигается.

Из рис. 1 мы видим, что оценка  $\hat{\tau}_{n,min}$  может иногда оценить период  $\tau$ , а иногда только половину его. Вышеописанная оценка  $\hat{\tau}_{n,min}$  имеет свои недостатки. При поиске смежных интервалов  $I_1, \dots, I_K \subset W_n$  минимальной длины, мы получаем очень большое их количество. К тому же разности  $X(I_k) - X(I_{k+1})$  на практике, обычно не равняются нулю. Таким образом, из-за большого количества слагаемых, суммы  $\sum_{k=1}^{K-1} (X(I_k) - X(I_{k+1}))^2$  обычно принимают большое значение. В результате этого при помощи оценки  $\hat{\tau}_{n,min}$ , мы можем оценить только от одной трети до половины длины окна  $W_n$ . Беббингтон и Зитикис пришли к выводу, что оценка

$$\tau_{n,min} := \arg \min_{\ell} \min_{(1)} \min_{(2)} \sum_{k=1}^{K-1} \left( \int_{I_k} \lambda(s) ds - \int_{I_{k+1}} \lambda(s) ds \right)^2,$$

на которой основана оценка (2), работает хорошо, пока мы можем найти интервалы  $I_1, \dots, I_K$ : разности  $\int_{I_k} \lambda(s) ds - \int_{I_{k+1}} \lambda(s) ds$  будут в точности равны 0 для любого  $k = 1, \dots, K - 1$ .

Беббингтон и Зитикис предложили следующую модификацию оценки  $\hat{\tau}_{n,min}$ :

$$\hat{t}_{n,max} := 2 \arg \max_{\ell} \max_{(1)} \max_{(2)} \sum_{k=1}^{K-1} (X(J_k) - X(J_{k+1}))^2 \quad (3),$$

где  $\max_{(2)}$  берется по всем смежным интервалам  $J_1, \dots, J_K \subset (a, b_n]$  длины  $\ell$ ,  $\max_{(1)}$  берется по всем длинам  $\ell$ , а  $\arg$  - это длина  $\ell$ , на которой минимум  $\max_{(1)}$  достигается. Максимум должен быть получен на половине этого периода, хотя можно построить интенсивности таким образом, что максимум находится на меньших его частях. Это видно из рис. 2 и рис. 3.

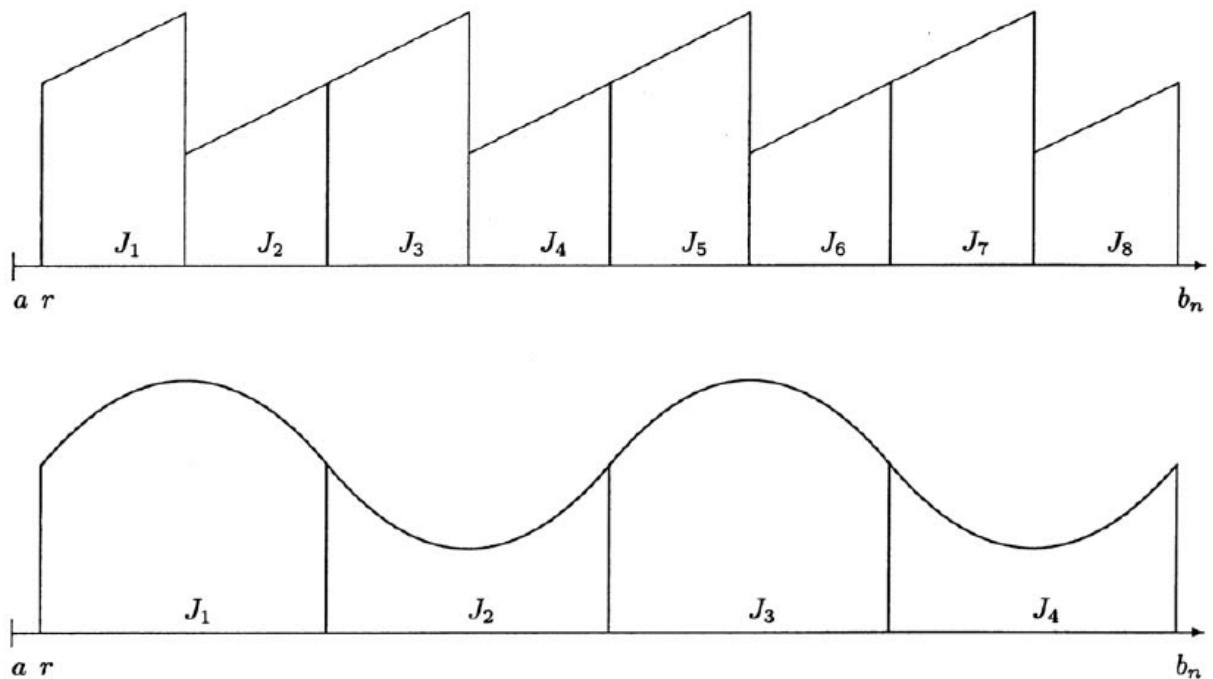


Рис.2.

На рисунке показаны две периодические функции интенсивности и соответствующие выборы интервалов  $J_1, \dots, J_K$ . Для обеих функций интенсивности длина интервалов  $\frac{\tau}{2}$ .

$\hat{t}_{n,max}$  отдает предпочтение наиболее возможному числу  $K$  интервалов  $J_k$ , т.е. интервалам минимальной длины  $\ell$ . Тем не менее, главное различие между  $\hat{t}_{n,max}$  и предыдущей оценкой,  $\hat{t}_{n,min}$  в том, что мы ищем смежные интервалы, такие что в среднем, разница между числом точек в любых двух смежных интервалах максимальна. Для предыдущей оценки мы по возможности искали интервалы с равным числом точек. Мы используем

смежные интервалы, чтобы использовать тот факт, что высокие и низкие интенсивности чередуются в циклическом процессе.

Бebbингтон и Зитикис используют выражение:

$$\hat{t}_{n,max} = 2 \arg \sup_{\varepsilon} \sup_{r \geq a} \sum_{i=0}^{\infty} I\{r + (i, i + 2] \varepsilon \in W_n\} Y_i^2(\varepsilon, r),$$

где  $I\{\cdot\}$  обозначает индикаторную функцию, которая равна 1, если выражение внутри скобок истинно, и 0 иначе, и где

$$Y_i(\varepsilon, r) := X(r + (i, i + 1] \varepsilon) - X(r + (i + 1, i + 2] \varepsilon)$$

Авторы вместо квадрата разности  $(X(J_k) - X(J_{k+1}))^2$  предлагают использовать  $H(X(J_k) - X(J_{k+1}))$  для некоторой неотрицательной функции  $H$ . Отмечая при этом, что они не нашли какую-либо функцию, по своим свойствам лучше, чем  $H(x) = x^2$ . Последнее замечание позволяет убрать неправильные периоды и при этом не исключить верный период из-за случайного шума в наблюдаемой реализации. Сумма квадратов отклонений чувствительна к пикам, и поэтому предлагаемая оценка стремится выявить интервалы высокой и низкой интенсивности.

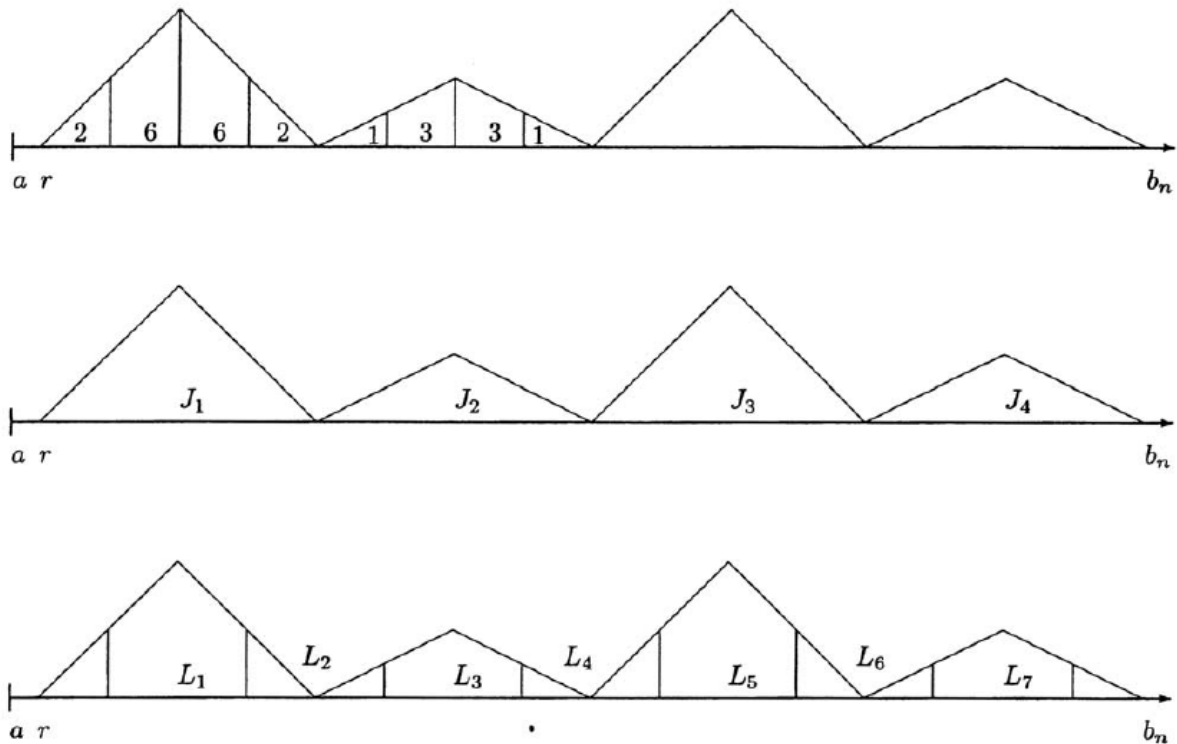


Рис.3.



На рисунке приведен пример, в котором оценка  $\tau_{n,max}$  находит  $\frac{\tau}{2}$  вместо истинного периода  $\tau$ . Верхняя фигура показывает площади под кривой интенсивности. Средняя фигура разделяет  $W_n$  на интервалы  $J_1, \dots, J_4$  длины  $\frac{\tau}{2}$ , а нижняя на интервалы  $L_1, \dots, L_7$  длины  $\frac{\tau}{4}$ . Сумма  $\sum_{k=1}^6 (X(L_k) - X(L_{k+1}))^2$  больше, чем  $\sum_{k=1}^3 (X(J_k) - X(J_{k+1}))^2$ .

*Обобщение оценки.*

Вместо того, что брать разницу только между двумя последовательными интервалами в определении  $Y_i(\varepsilon, r)$ , можно взять несколько различных разностей. Беббингтон и Зитикис определяют обобщенную оценку при фиксированном  $M \in \mathbb{N}$ :

$$\hat{t}_{n,M} := 2 \arg \sup_{\varepsilon} \sup_{r \geq a} \sum_{i=0}^{\infty} I\{r + (i, i + 2M] \varepsilon \in W_n\} Y_{i,M}^2(\varepsilon, r),$$

где

$$Y_{i,M}(\varepsilon, r) := \sum_{m=1}^M (X(r + (i + 2m - 2)\varepsilon, r + (i + 2m - 1)\varepsilon) - X(r + (i + 2m - 1)\varepsilon, r + (i + 2m)\varepsilon)).$$

Когда  $M = 1$ ,  $Y_{i,M}(\varepsilon, r)$  равно  $Y_{i,1}(\varepsilon, r)$ , откуда  $\hat{t}_{n,1} \equiv \hat{t}_{n,max}$ . Таким образом  $\hat{t}_{n,M}$  действительно является обобщением  $\hat{t}_{n,max}$ . Беббингтон и Зитикис отмечают, что  $\hat{t}_{n,2}$  обычно превосходит по качеству  $\hat{t}_{n,1}$ , по крайней мере, в среднем. Но при дальнейшем исследовании  $\hat{t}_{n,M}$  для  $M \geq 3$  им не удалось выявить каких-либо значительных улучшений оценки  $\hat{t}_{n,2}$ .

В заключение Беббингтон и Зитикис дают следующую оценку для использования с некоторыми функциями  $H$ , отличными от  $H(x) = x^2$ :

$$\hat{t}_{n,M}(H) := 2 \arg \sup_{\varepsilon} \sup_{r \geq a} \sum_{i=0}^{\infty} I\{r + (i, i + 2M] \varepsilon \in W_n\} H(Y_{i,M}(\varepsilon, r)).$$

$\hat{t}_{n,M}$  это «сглаженная» версия  $\hat{t}_{n,max}$ , где  $M$  – степень «сглаживания».

### 2.3 Метод Белитсера – Серра – ван Зантена [3].

В статье представлена М-оценка для периода. Для ее построения авторы сначала рассматривают траекторию неоднородного процесса Пуассона с периодической функцией интенсивности  $\lambda$  с периодом  $\tau > 0$  и вводят вспомогательный параметр  $T > 0$ , разделяя интервал времени наблюдения  $[0, n)$  на  $[n/T]$  частей эквивалентной длины  $[(i-1)T, iT)$ ,  $i = 1, \dots, [n/T]$ . Затем для каждого из этих интервалов сравнивают число событий произошедших в первые  $\theta \leq T/2$  и последние  $\theta$  временных отрезков.

Это приводит к построению критерия сходимости целевой функции к предельной функции, нули которой связаны с неизвестным периодом в явном виде. Рассмотрим это подробнее.

Из основных свойств процесса Пуассона видно, что ожидаемая разница между двумя этими числами событий равна

$$\int_{(i-1)T}^{(i-1)T+\theta} \lambda(s) ds - \int_{iT-\theta}^{iT} \lambda(s) ds \quad (1)$$

Из периодичности функции  $\lambda$  следует, что интегралы от  $\lambda$  по двум различным интервалам совпадают, если длины интервалов одновременно равны одному и тому же периоду  $\tau$  с учетом кратности или, если интервалы имеют одинаковую длину и они кратны периоду друг друга.

Из этого следует, что если  $T$  такое, что  $(\ell-1)\tau < T < \ell\tau$  для некоторого  $\ell \in \mathbb{N}$ , тогда равенство (1) обращается в 0 для  $\theta$  равному любому из значений

$$0 < T - (\ell-1)\tau < \tau < T - (\ell-2)\tau < 2\tau < \dots < T - (\ell-k)\tau < k\tau < \dots \quad (2)$$

Авторы доказали (см. Теорема 2), что при соответствующих ограничениях на вспомогательный параметр  $T$ , это единственные точки, в которых функция  $\psi_n(\theta, T)$  стремится к 0 для любого  $n$ .

$$\psi_n(\theta, T) = \frac{1}{[n/T]} \sum_{i=1}^{[n/T]} \left( \int_{(i-1)T}^{(i-1)T+\theta} \lambda(s) ds - \int_{iT-\theta}^{iT} \lambda(s) ds \right)^2 \quad (3)$$

Белитсер, Серра и ван Зантен основывают свой метод на знании целевой функции, которая оценивает функцию  $\psi_n$ . Они определяют случайную функцию  $\Psi_n$  на  $[0, T/2]$  следующим образом:

$$\Psi_n(\theta, T) = \frac{1}{[n/T]} \sum_{i=1}^{[n/T]} (N_i^-(\theta, T) - N_i^+(\theta, T))^2 - N_i^-(\theta, T) - N_i^+(\theta, T),$$

где  $N_i^-(\theta, T) = N_{(i-1)T+\theta} - N_{(i-1)T}$  – число точек в первых  $\theta$  временных отрезках интервала  $[(i-1)T, iT)$  и, аналогично,  $N_i^+(\theta, T) = N_{iT} - N_{iT-\theta}$  – число точек в последних  $\theta$  временных отрезках этого интервала.

Из доказательства к теореме 2, представленного в статье, видно, что  $\Psi_n$  – действительно последовательная оценка для  $\psi_n$ , т.е. для фиксированного  $\theta$  и  $T$ ,  $\Psi_n(\theta, T) - \psi_n(\theta, T)$  сходится по вероятности к 0 при  $n \rightarrow \infty$ .

*Процедура оценки при известном периоде.*

Пусть период известен:  $\tau \in [a, b]$ , где  $0 < a < b < 2a$ . Тогда, если мы возьмем значение вспомогательного параметра  $T \in (b, 2a)$ , получим, что  $\tau < T < 2\tau$  и  $T - \tau > T - b > 0$ .

Белитсером, Серра и ван Зантенном было получено (см. Теорема 2), что, при техническом условии, что  $T/\tau$  – иррационально, целевая функция  $\theta \mapsto \Psi_n(\theta, T)$  сходится по вероятности, равномерно на  $[T - b, T - a]$  к гладкой, неотрицательной функции  $\psi$ , которая имеет 0 в точке  $[T - \tau]$ . Основываясь на этом, авторы получили оценку:

$$\hat{t}_n = \hat{t}_n(T) = T - \hat{\theta}_n, \quad (4)$$

где  $\hat{\theta}_n$  минимизирует  $\theta \mapsto |\Psi_n(\theta, T)|$  на  $[T - b, T - a]$ , т.е. это точка, такая что

$$|\Psi_n(\hat{\theta}_n, T)| \leq \inf_{\theta \in [T-b, T-a]} |\Psi_n(\theta, T)| + o_p(1). \quad (5)$$

Такая точка всегда существует, так как функция  $\theta \mapsto \Psi_n(\theta, T)$  почти наверно кусочно-постоянная функция, которая меняет значения только на конечном числе точек.

### **Теорема 1.**

Для всех  $T > 0$ , таких что  $T/\tau$  – иррациональное число,

$$\sup_{\theta \in [0, T/2]} |\Psi_n(\theta, T) - \psi(\theta, T)| \xrightarrow{P} 0$$

При  $n \rightarrow \infty$ .

### **Теорема 2.**

Пусть  $T/\tau$  – иррациональное число. Тогда все нули функции  $\theta \mapsto \psi(\theta, T)$ ,  $\theta \in (0, T)$  являются точками  $\theta$  такими, что каждая  $\theta$  – кратна  $\tau$ , или  $T - \theta$  множитель  $\tau$ .

Используя в дальнейшей работе теоремы 1 и 2 и стандартную теорию М-оценок, авторы доказали состоятельность оценки (4).

### **Заключение.**

В работе были изучены и систематизированы наиболее известные на данный момент методы оценки периода циклического процесса Пуассона.

Большой интерес представляет дальнейшее исследование статистических свойств имеющихся оценок, а также их сравнительный анализ, например, методом компьютерного моделирования.

## Список литературы.

- [1] *R. Helmers, I. W. Mangku*, On estimating the period of a cyclic Poisson process / Lecture Notes-Monograph Series Vol. 42, Mathematical statistics and applications: Festschrift for Constance van Eeden, 345-356, Institute of Mathematical Statistics, Beachwood, OH, 2003.
- [2] *M. Bebbington, R. Zitikis*, A Robust Heuristic Estimator for the Period of a Poisson Intensity Function / Methodology And Computing In Applied Probability 6(4), 441-462, January 2004.
- [3] *E. Belitser, P. Serra and H. van Zanten*, Estimating the Period of a Cyclic Non-Homogeneous Poisson Process / Online: 18 July 2012, Scandinavian Journal of Statistics Vol. 40, Issue 2, 204–218, June 2013.
- [4] *R. Helmers, I. W. Mangku*, Statistical Estimation of Poisson Intensity Functions / Probability, Networks and Algorithms – R9913, December 31, 1999.
- [5] *R. Helmers, I. W. Mangku and R. Zitikis*, Consistent estimation of the intensity function of a cyclic Poisson process / Journal of Multivariate Analysis, Vol. 84, #1, 19-39, Academic Press, 31 January 2003a.
- [6] *I.W. Mangku*, Estimating the intensity of a cyclic Poisson process / Ph.D. Thesis, University of Amsterdam, 2001.
- [7] *D. Vere-Jones*, On the Estimation of Frequency in Point-Process Data / J. of Appl. Probab. Vol. 19A, Essays in Statistical Science, 383-394, 1982.
- [8] *R. Helmers, I. W. Mangku, R. Zitikis*, Statistical properties of a kernel type estimator of the intensity function of a cyclic Poisson process / J. Multivariate Anal, 2003b.
- [9] *R. Helmers, I. W. Mangku*, Predicting a cyclic Poisson process / Online: 17 March 2012, The Institute of Statistical Mathematics, Tokyo, 2012