

Р. В. Амбарцумян

О пуассоновских наложениях скоплений

(Представлено академиком В. А. Амбарцумяном 31/III 1965)

§1. Введение. Точечные процессы в пространстве одного или многих измерений, которые мы называем пуассоновскими наложениями скоплений, вполне определяются заданием нестационарного точечного процесса Γ со средним числом происходящих в нем событий τ и некоторого числа $\lambda > 0$. Соответствующий процесс пуассоновского наложения скоплений строится как суперпозиция независимых реализаций процесса Γ , которые мы называем скоплениями и чьи начала составляют независимый пуассоновский процесс Λ интенсивности λ в пространстве соответствующего числа измерений. Процессы такой конструкции рассматривались в работах М. С. Бартлетта ⁽¹⁾ и Е. Неймана ⁽²⁾.

Обозначим построенный таким образом стационарный точечный процесс через Σ . Очевидно, что интенсивность процесса равна $\tau = \lambda\tau$.

Будем употреблять принятый термин „поток“ для обозначения точечного процесса в одном измерении (например, времени).

В этой заметке мы обсуждаем следующую проблему: в какой степени статистические свойства потока Γ могут быть восстановлены при помощи информации, содержащейся в достаточно длинной реализации потока Σ в условиях, когда заданы лишь моменты времени, в которых произошли события потока Σ и отсутствует дополнительная информация о принадлежности событий тем или иным скоплениям. Кроме того, мы дадим несколько простых рекомендаций для использования при практическом анализе пуассоновских наложений скоплений.

Мы ограничимся рассмотрением „одномерных свойств“ потока Σ , т. е. будем рассматривать только одномерные распределения, связанные с потоком Σ . Такой подход обоснован с точки зрения приложений.

Основные результаты будут касаться производящей функции

$$U(z, x) = \sum_{n=0}^{\infty} z^n u^{(n)}(x),$$

где $1-u^{<n>}(x)$ является функцией распределения общей длины $n+1$ последовательных интервалов между событиями в потоке Σ . Заметим, что вероятности $v_m(t)$ того, что за период времени длины t в потоке Σ произойдет более чем m событий, связаны с функциями $u^{<n>}(x)$ через формулы Пальма

$$v_0(t) = \sigma \int_0^t u^{<0>}(\xi) d\xi, \quad (1)$$

$$v_m(t) = \sigma \int_0^t |u^{<m>}(\xi) - u^{<m-1>}(\xi)| d\xi \quad (m = 1, 2, \dots),$$

и, таким образом, эти вероятности не могут дать добавочной информации о процессе Γ .

Целесообразное (одномерное) описание потока Γ получаем с помощью системы производящих функций

$$G_m(z, x) = \sum_{n=0}^{\infty} z^n g_m^{<n>}(x) \quad (m = 0, 1, 2, \dots),$$

которые описывают течение потока Γ , начиная с момента появления $m+1$ -го события этого процесса. Через $g_m^{<n>}(x)$ мы обозначили вероятность того, что промежуток времени между $m+1$ -м событием потока Γ (при условии, что таковое наступит) и $m+n+2$ -м событием потока Γ превзойдет величину x .

Мы будем рассматривать следующие два преобразования потоков.

Для каждого потока Π мы строим прореженный поток $\overset{p}{\Pi}$ путем выкидывания некоторых событий потока Π . Прореживание производится таким образом, что каждое событие потока Π выкидывается с вероятностью p или остается на месте с вероятностью $(1-p)$, причем выкидывание (или оставление) каждого из событий не зависит от характера расположения событий в данной реализации потока Π и от того, выкидывается или не выкидывается любое другое событие потока Π . Сохраним символы, уже введенные для потока Π для обозначения соответствующих понятий потока $\overset{p}{\Pi}$, снабжая лишь их "степенью" p . Например,

$$u^{<0>}(x) = (1-p) \sum_{n=0}^{\infty} p^n u^{<n>}(x) = (1-p) U(p, x). \quad (2)$$

Другое преобразование определяется только для потоков с конечным средним числом происходящих событий; оно представляет

собой некоторую процедуру усреднения $\sim$. Введем поток $\bar{\Gamma}$ путем непосредственного описания большого ансамбля его реализаций.

Обозначим через P_m вероятность того, что в скоплении произойдет больше чем m событий, а через Γ_m — течение потока Γ после момента появления его $m+1$ -го события (предполагая, что таковое в действительности произошло).

По определению, ансамбль, который строится путем смешения ансамблей реализаций потоков Γ_m , причем каждый из этих ансамблей берется с весом $\frac{1}{\gamma} P_m$, составляет большой ансамбль реализаций для потока $\bar{\Gamma}$.

Мы опять сохраним символы, введенные для потока Γ для обозначения соответствующих понятий, относящихся к $\bar{\Gamma}$, лишь снабжая их знаком тильда. Например,

$$\bar{G}_0(z, x) = \frac{1}{\gamma} \sum_{m=0}^{\infty} P_m G_m(z, x).$$

По аналогии с (2) получаем

$$\bar{g}_0^p(x) = (1-p) \bar{G}_0(p, x). \quad (3)$$

§ 2. Основные принципы. Возвращаясь к стационарному потоку пуассоновского наложения Σ , устанавливаем теорему, которую можно назвать „принципом расщепления“.

Теорема 1. *Течение потока Σ после его произвольного события (процесс Σ_1) может быть описано как наложение независимых потоков Σ и Γ*

$$\Sigma_1 = \bar{\Gamma} + \Sigma. \quad (4)$$

Под термином „произвольное событие“, как обычно, мы понимаем, что каждое событие потока Σ может быть выбрано в качестве произвольного события с одинаковым весом, независимо от конкретной картины потока Σ вблизи этого события.

Доказательство. Рассматривая скопление Γ' , которому принадлежит произвольное событие потока Σ , легко находим, что оно является событием номер $m+1$ в этом скоплении с вероятностью

$\frac{1}{\gamma} P_m$, и следовательно, хвост скопления Γ' ведет себя как поток

$\bar{\Gamma}$. Что касается событий, не принадлежащих к Γ' , то они составляют независимый поток Σ благодаря тому обстоятельству, что некоторая информация относительно местонахождения начала скопления Γ' , которой мы располагаем, не меняет пуассоновского характера начал других скоплений.

Теорема 2. Для каждого стационарного потока Π

$$(\Pi_1)^p = (\Pi)^p_1,$$

(Π_1 обозначает течение потока Π после его произвольного события).

Доказательство. Рассмотрим ансамбль реализаций отрезка потока Π с общим числом событий N достаточно большим.

Вероятность того, что заданное событие α будет избрано в качестве „произвольного события“ для потока Π , есть $1/N$. После описанной процедуры прореживания это событие α участвует в полученном таким образом ансамбле реализаций потока Π с вероятностью $(1-p)$. Средним числом событий после прореживания становится $N(1-p)$, и, таким образом, вероятность того, что данное событие будет избрано в качестве произвольного события для потока Π , равна $1/N$ и не зависит от p . Вместе с предположением о независимости каждого выбрасывания это позволяет выбирать произвольное событие в потоке Π перед проведением процедуры прореживания.

Применяя процедуру прореживания к потоку Σ_1 , из (4) получаем

$$(\Sigma_1)^p = \Sigma^p + \frac{p}{\Gamma}$$

я по теореме 2

$$(\Sigma_1)^p_1 = \Sigma^p_1 + \frac{p}{\Gamma}. \quad (5)$$

Знак $+$ опять означает наложение независимых потоков.

§3. Формальное решение и примеры. Используя (5), находим

$$u^{(p)}_0(x) = (1 - v_0(x)) \cdot \frac{p}{g_0} g_0(x). \quad (6)$$

С помощью (1), (2) и (3) сводим (6) к интегральному уравнению

$$(1-p)U(p, x) = \varepsilon(1-p)^2 \int U(p, \xi) d\xi \cdot (1-p)\bar{G}_0(p, x)$$

и, наконец,

$$U(p, x) = \bar{G}_0(p, x) e^{-(1-p)^2 \varepsilon \int_0^x \bar{G}_0(p, \xi) d\xi} \quad (7)$$

Это уравнение, конечно, остается верным, когда мы подставляем вместо p любое комплексное z , удовлетворяющее условию $|z| < 1$. Заметим, что производящая функция $U(z, x)$ определяет всего лишь функцию $\bar{G}_0(z, x)$

$$\bar{G}_0(z, x) = \frac{U(z, x)}{(1-z)^2 \int U(z, \xi) d\xi},$$

и поэтому всегда необходима дополнительная информация о природе потока Γ для того, чтобы полностью восстановить его одномерные свойства.

Практически некоторые сведения о характере скоплений можно получить, исследуя поведение кривых $u^{(0)}(x)$ и $M_1(t)$ — среднего числа событий в потоке Σ_1 за период $(0, t)$.

Располагая кривой $u^{(0)}(x)$, мы строим

$$\int dx \frac{u^{(0)}(x)}{\int u^{(0)}(\xi) d\xi} \quad (*) \quad (8)$$

Измеряя наклон асимптоты для (*), оцениваем отношение

$$\frac{P_0}{1} = \text{наклон асимптоты } (*). \quad (8)$$

Высота пересечения асимптоты для (*) с осью ординат может оказаться полезной для получения сведений о расстоянии между событиями внутри скоплений.

Например, если поток Γ строится путем вырезания куска однородного потока Π так, что этот кусок содержит больше чем m событий с вероятностью P_m и число событий на этом отрезке не зависит от их расположения, имеем

$$\bar{g}_0^{(0)}(x) = \frac{1}{1} P_0 + q^{(0)}(x) \left(1 - \frac{1}{1} P_0 \right),$$

где $q^{(0)}(x)$ — функция Пальма для потока Π . Отсюда для средней длины l интервала между событиями в потоке Π получаем оценку

$$\hat{l} = \frac{\text{высота пересечения асимптоты } (*) \text{ с осью ординат}}{1 - \text{наклон асимптоты } (*)}.$$

Что касается $M_1(t)$, то согласно (4)

$$M_1(t) = \lambda t + \bar{m}(t).$$

где $\bar{m}(t)$ есть среднее число событий, происходящих в потоке Γ за период $(0, t)$. Поэтому отношение $\frac{\sum n P_n}{\gamma} = \bar{m}(\infty)$ может быть оценено

$$\frac{\sum n P_n}{\gamma} = \text{высота пересечения асимптоты для } M_\gamma(t) \text{ с осью ординат} \quad (9)$$

§ 4. Обсуждение. Пуассоновское наложение скоплений может наблюдаться с различными степенями информации о пуассоновском потоке Λ , которые составляют начала скоплений. Таковыми могут быть:

- а) параметр потока Λ неизвестен;
- б) параметр потока Λ известен и равен λ ;
- в) наблюдаем реализацию Σ и одновременно порождающую ее реализацию потока Λ .

В случае а) практически невозможно найти величину γ в общем случае; с помощью (8) и (9) все же можно получить некоторую информацию о числе событий в скоплениях.

В случае б) полагаем

$$\gamma = \frac{\sigma}{\lambda}.$$

Если имеет место случай в), то „принцип расщепления“ можно сформулировать в следующем виде.

Обозначим через Σ' течение потока Σ после произвольного события потока Λ . Тогда

$$\Sigma' = \Sigma + \Gamma. \quad (10)$$

Таким образом, поток Γ может быть описан в терминах потоков Σ' и Σ .

Этот принцип остается неизменным, когда мы переходим к пуассоновским наложениям скоплений в пространстве двух или большего числа измерений.

Наиболее простым следствием (10) является (для плоскости)

$$M'(r) = \pi r^2 + m(r), \quad (11)$$

где $M'(r)$ (соответственно $m(r)$) есть ожидаемое число событий процесса Σ' (соответственно Γ) в круге радиуса r с центром в начале процесса.

Разумеется, как в (10), так и в (4) компонента Σ может расщепляться как некий шум, вносящий дополнительное слагаемое в дисперсию оценки для $m(r)$ (или для $m(t)$).

Желая применить (11) к исследованию ограниченной совместной реализации процессов Σ и Λ , покрываем данную карту системой $\{C_r\}$ непересекающихся кругов радиуса r с центрами в точках процесса Λ и полагаем, что в этих кругах реализации процесса Σ' независимы (практически). С ростом r число независимых подсчетов событий

процесса Σ' уменьшается. Отметим возможность использования смещенных оценок для $m(r)$ с тем, чтобы увеличить отношение сигнала к шуму при таком анализе. Их можно строить с помощью оценок величины $m_1(r)$ — математическое ожидание числа событий процесса Γ в круге S радиуса r , относительно которого начало процесса Γ занимает случайное положение; именно оно распределено равномерно в concentрическом кругу S радиуса ϵ .

Пусть η_n — число точек процесса Σ , лежащих в S_n , ζ_n — число точек процесса Λ , лежащих в concentрическом кругу S_n радиуса ϵ (включая центр), N_k — число кругов в системе $\{S_n\}$, для которых ζ_n принимает значение k , $N = \sum_{k=1}^{\infty} N_k$.

Смещенной оценкой для $m(r)$ служит

$$\hat{m}(r) = \frac{\sum_{n=1}^N \eta_n - N\sigma\pi r^2}{\sum_{k=1}^{\infty} kN_k}, \quad (12)$$

Имея большую свободу в построении системы $\{S_n\}$, выгодно выбирать ее таким образом, чтобы сумма $\sum_{n=1}^N \zeta_n$ достигала своего максимального значения.

При построении таких смещенных оценок следует учитывать действие „принципа неопределенности“: уменьшение дисперсии оценки (12) (которое достигается увеличением ϵ) вызывает увеличение смещения оценки, и наоборот. Заметим, что относительная эффективность применения (12) возрастает с ростом r .

Прием находит свое обоснование в следующем принципе, выражающем определенное свойство линейности пуассоновских наложений скоплений.

Обозначим через Σ^n течение процесса Σ после момента одновременного появления n событий процесса Λ (более строгое определение потока Σ^n дается с помощью очевидного предельного перехода).

Теорема 3.

$$\Sigma^n = \Sigma + \underbrace{\Gamma_1 + \Gamma_2 + \dots + \Gamma_n}_{n \text{ независимых процессов } \Gamma}.$$

В заключение рассмотрим однородное уравнение

$$\Pi_1 = \Pi, \quad (13)$$

которое приводит к новому определению пуассоновского потока.

Теорема 4. Пусть Π является стационарным потоком и

пусть Π_1 есть течение потока Π после его произвольного события.

Если Π удовлетворяет уравнению (13), то Π есть поток Пуассона.

Доказательно. Теорема немедленно следует из (7), когда мы полагаем

$$\bar{G}_0(p, x) = \frac{1}{1-p}.$$

В этом случае

$$U(p, x) = \frac{1}{1-p} e^{-x(1-p)},$$

что характеризует пуассоновский поток.

Институт математики и механики
Академии наук Армянской ССР

II. Գ. ՀԱՄԱՐԱՐՁՈՒՄԹԱՆ

Կույտերի պատահական վերադրման սուբիլ

Քաղաքում լինելու պատահական մի կետային հոսք Γ , որի համար իրականացվող պատահարների միջին թիվը վերջավոր է:

Քաղ $\lambda > 0$ լինել Λ Պուասոնի մի հոսքի ինտենսիվություն: Կատարվում են d Σ պրոցես, որը հանդիսանում է իրարից անկախ Γ հոսքերի (կույտերի) վերադրում, բնականորեն, որ այդ կույտերի սկիզբները համընկնում են Λ հոսքի պատահարների հետ:

Հաջորդում գուրս է բերվում $U(z, x) = \sum_{n=0}^{\infty} u^{(n)}(x)$ ֆունկցիայի տեսքով, ինչպես

$1 - u^{(n)}(x)$ — ի պատահարների միջև բնական $n+1$ իրար հաջորդող միջակայքների բնական երկարություն բաշխման ֆունկցիան է Σ հոսքում:

Ոչնչի հետևանքով է տրվում, որ հնարավոր է $U(z, x)$ ֆունկցիայի միջոցով նկարագրել

դրել միայն որոշակի միջինացման ձևափոխության միջոցով Γ հոսքից ստացվող $\bar{\Gamma}$ հոսքի համակարգչականները: Որոշ գործնական հանձնարարումներ են տրվում Σ հոսքերի վիճակագրության վերաբերյալ:

Վերջին թեորեմում բերվում է Պուասոնի հոսքի բնորոշման համար բավարարող պահանջների մի նոր միջինային սխեմա:

Л И Т Е Р А Т У Р А — Գ Ր Ա Կ Ա Ն ՈՒ Ք Յ ՈՒ Ն

¹ M. C. Bartlett, The spectral analysis of point processes J. R. Statist. Soc. B, 25, 264, 1963. ² E. Heijman, Զ. Պ. Երոտև և Ժ. Խան, Statistics of images of galaxies with particular reference to clustering. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 3, 1955.