УДК 519.7

ВЫЯВЛЕНИЕ «РАЗЛАДКИ» ПУАССОНОВСКОГО ПРОЦЕССА

Дзанагова И.Т., Хугаева Л.Т., Дигурова А.М., Мамсурова Ф.Х.

ФГБОУ ВПО «Северо-Осетинский государственный университет имени Коста Левановича Хетагурова», Владикавказ, Россия (362025, РСО-Алания, г. Владикавказ, ул. Ватутина, 44-46), e-mail: nosu@nosu.ru

Приведены тесты проверки «разладки» пуассоновского процесса. Для выявления «разладки» случайной последовательности событий используют обычно как непоследовательные, так и последовательные алгоритмы. Для ускорения выявления «разладки» пуассоновского процесса предпочтительным является применение последовательных методов, ориентированных на выявление момента «разладки» по первым наблюдениям и ограниченной информации. При этом момент «разладки» эквивалентен переходу от пуассоновской модели редких событий к одной из возможных её модификаций. Одной из таких модификаций, описывающей статистическую структуру, возникающую из комбинации простейших пуассоновских потоков, является распределение Пуассона степени k, формирование которого тоже рассматривается в этой статье. Рассматриваемые в работе элементы теории построения экстремальных распределений экстремальных случайных величин в предложенной постановке возникли из сущности и содержания практических задач.

Ключевые слова: «разладка» пуассоновского процесса, распределение, производящая функция.

IDENTIFICATION OF POISSON PROCESS "DISCORD"

Dzanagova I.T., Khugaeva L.T., Digurova A.M., Mamsurova F.H.

The North Ossetian state university n.a. Costa Levanovich Khetagurov, Vladikavkaz, Russia (362025, RSO-Alania, Vladikavkaz, Vatutin St., 44-46), e-mail: nosu@nosu.ru

We provideherecheck tests of "discord" of Poisson process. For identification of "discord" of casual sequence of events we usually use both inconsistent and consecutive algorithms. For acceleration of identification of Poisson process "discord" application of the consecutive methods focused on identification of the moment of "discord" on the first supervision and limited information is preferable. Thus the moment of "discord" is equivalent to transition from Poisson model of rare events to one of its possible modifications. One of such modifications describing the statistical structure arising from a combination of the elementary Poisson streams is Poisson's distribution degrees k, which formation too is considered in this article. The elements of the theory of creation of extreme distributions of extreme random variables considered in work in the offered statement arose from essence and the maintenance of practical tasks.

Keywords: "discord" of Poisson process, distribution, the making function.

Задача выявления изменения свойств пуассоновского процесса состоит в следующем. На промежутке действительной оси времени (пространства) наблюдается некоторое количество точек, характеризующее появление однородных событий.

Число точек n и их расположение на выбранной оси является случайным (координаты $t_1, t_2, ..., t_n$).

Последовательность независимых одинаково распределённых случайных величин $\Delta t(\Delta t_1 = t_1,\ \Delta t_2 = t_2 - t_1,\ \Delta t_3 = t_3 - t_2, ...\ \Delta t_n = t_n - t_{n-1})$, для пуассоновского потока имеет функцию распределения $F_1(\Delta t) = 1 - exp(-\lambda \Delta t)$.[1]

Пусть $\Delta t_{n+1} = t_{n+1} - t_n$, имеют функции распределения $F_1(\Delta t) \neq F_0(\Delta t)$. Требуется определить момент «разладки» пуассоновского потока n.

После наблюдения на интервале $/0,t_{n+1}/\ n+1$ моментов появления случайного события ($n=1,2,...$) будем решать задачу о различении следующих двух гипотез H_0: «разладки» нет, H_1: «разладка» есть [5].

Зафиксируем уровень значимости α и определим тест для различения этих гипотез неравенством

$$\frac{2}{2-\alpha} \geq \left(\frac{t_{n+1}}{t_n}\right)^n \leq \frac{2}{\alpha}. \qquad (1)$$

Выбор теста (1) можно объяснить следующим образом. Введем в рассмотрение тестовую статистику h_1 (разность статистических оценок энтропии).

$$h_1 = H_\varepsilon^Э(n+1) - H_\varepsilon^Э(n),$$

где $H_\varepsilon^Э(n+1), H_\varepsilon^Э(n)$ - оценка дифференциальной энтропии, определенная по $n+1$ (соответственно по n) наблюдениям случайной последовательности $t_1, t_2, ..., t_n, t_{n+1}$ с экспоненциальным распределением времени между двумя смежными событиями. Можно показать, что тестовая статистика имеет вид

$$h_1 = \ln\frac{n}{n+1} + \ln(1+\aleph_1),$$

где $\aleph_1$ – случайная величина, определяемая следующим соотношением $\aleph_1 = \frac{\Delta t_{n+1}}{t_n}$ [6].

Действительно, так как эмпирическое значение дифференциальной энтропии для выборок из экспоненциальных генеральных совокупностей имеет вид

$$H_\varepsilon^Э = 1 + \ln m_{\Delta t},$$

где $m_{\Delta t}$ – оценка математического ожидания промежутков времени между двумя смежными событиями, то

$$h_1 = H_\varepsilon^Э(n+1) - H_\varepsilon^Э(n) + 1 + \ln\frac{\sum_{i=1}^{n+1}\Delta t_i}{n+1} - 1 - \ln\frac{\sum_{i=1}^{n}\Delta t_i}{n} = \ln\frac{n}{n+1} + \ln(1+\aleph_1).$$

Функция распределения случайной величины $\aleph_1$, является инвариантной к параметру λ экспоненциального распределения $F_0(\Delta t) = 1 - \exp(-\lambda\Delta t)$ и имеет плотность [6;7]

$$f_1(\aleph_1) = \frac{n}{(1+\aleph_1)^{n+1}}.$$

Для доказательства этого утверждения прежде всего заметим, что если случайная величина t_n представляет собой сумму экспоненциально распределённых случайных величин $\Delta t_i (i = 1,2,...,n)$, то она в свою очередь будет распределена по закону гамма-распределения с параметром формы n [2]. Отсюда следует, что плотность отношения двух независимых случайных величин Δt_{n+1} и $t_n (\aleph_1 = \frac{\Delta t_{n+1}}{t_n})$ определяется следующим образом

$$f_1(\aleph_1) = \int_0^\infty \lambda e^{-\lambda \aleph_1 t_n} t_n \frac{\lambda(\lambda t_n)^{n-1} e^{-\lambda t_n}}{\Gamma(n)} dt_n. \quad (2)$$

Учитывая, что

$$\int_0^\infty \frac{\lambda(\aleph_1 + 1)^{n+1} t_n^n e^{-\lambda(\aleph_1+1)t_n}}{\Gamma(n+1)} dt_n = 1,$$

после очевидного преобразования интеграла (2) находим

$$f_1(\aleph_1) = \int_0^\infty \frac{\lambda^{n+1} t_n^n e^{-\lambda(\aleph_1+1)t_n}}{\Gamma(n)} dt_n =$$

$$= \frac{\lambda^{n+1}}{\Gamma(n)} \int_0^\infty \frac{\lambda(\aleph_1 + 1)^{n+1} t_n^n e^{-\lambda(\aleph_1+1)t_n} \Gamma(n+1)}{[\lambda(\aleph_1 + 1)]^{n+1} \Gamma(n+1)} dt_n = n(\aleph_1 + 1)^{-n-1}.$$

Используя плотность распределения $f_1(\aleph_1)$, можно определить плотность распределения статистики h_1 [7]

$$f_2(h_1) = \frac{n^{n+1}}{(n+1)^n} e^{-nh}.$$

Для этого достаточно преобразовать плотность распределения с учётом соотношения между случайными величинами h_1 и $\aleph_1$

$$\aleph_1 = (n+1)n^{-1} \exp h_1 - 1.$$

Функция распределения тестовой статистики h_i непосредственно определяется из её плотности [3]

$$F_2(h_2) = \int_{ln\frac{n}{n+1}}^{h_1} \frac{n^{n+1}}{(n+1)^n} e^{-nh_1} dh_1 = 1 - \left(\frac{n}{n+1}\right)^n e^{-nh_1}.$$

Заметим, что область задания статистики критерия ограничена снизу значением $ln\frac{n}{n+1}$. Доверительные границы для теста проверки «разладки» пуассоновского процесса, соответствующие доверительному уровню $1-\alpha$, могут быть определены с помощью функции распределения $F_2(h_1)$ следующим образом:

$$\frac{a}{2} \le 1-\left(\frac{n}{n+1}\right)^n e^{-nh_1} \le 1-\frac{a}{2}.$$

Несложные преобразования этого неравенства дают удобную формулу (1) для теста (решающее правило).

Введённый в рассмотрение тест (1) представляется целесообразным дополнить ещё одним тестом, повышающим эффективность процедуры выявления момента «разладки».

Из рассмотрения теста (1) вытекает, что он становится малоэффективным («некритичным»), если пуассоновский поток имеет тенденцию к вырождению в регулярный поток. Указанный недостаток можно устранить, если ввести в рассмотрение тестовую статистику h_2 [3].

$$h_2 = H_\varepsilon^H(n+1) - H_\varepsilon^H(n),$$

где $H_\varepsilon^H(n+1), H_\varepsilon^H(n)$ – оценка дифференциальной энтропии, определённая по $n+1$ (соответственно по n) стандартным нормализованным случайным величинам $Z_i(i=1,...,n,n+1)$, определяемым по зависимости

$$Z_n = \sum_{\nu=0}^{\infty} \frac{\left[\frac{1}{2} - e^{\frac{\Delta t_n(n-1)}{t_n}}\right]^\nu}{\nu!} (2\pi)^{\frac{\nu}{2}} \psi_\nu(0)$$

по наблюдениям последовательности $t_1, t_2, ..., t_n, t_{n+1.}$.

Как и в предыдущем случае можно показать, что

$$h_2 = \frac{1}{2}\ln\frac{n}{n+1} + \frac{1}{2}\ln(1+\aleph_2),$$

$$\aleph_2 = \frac{Z_{n+1}^2}{\sum_{n-1}^{n} Z_n^2}$$

а плотность распределения случайной величины имеет вид:

$$f_3(\aleph_2) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)\sqrt{\aleph_2}(1+\aleph_2)^{\frac{n+1}{n}}}. \qquad (3)$$

Соответственно и статистика h_2 будет иметь плотность распределения вида

$$f_4(h_2)=\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{n}{n+1}\right)^{\frac{n-1}{2}}e^{-nh_2}\left(\frac{n+1}{n}e^{-nh_2}-1\right)^{-\frac{1}{2}}. \tag{4}$$

Доверительные границы для второго теста, соответствующие уровню $1-\beta$, могут быть определены на основе полученных выражений для плотностей распределения тестовых статистик h_2 (или $\aleph_2$) [7; 9].

Однако получить решающее правило в удобной форме, как это удалось получить для первого теста, в данном случае не представляется возможным. С вычислительной точки зрения процедура проверки гипотезы с использованием плотностей распределений (3) или (4) принципиальных трудностей не вызывает.

Действительно, в силу того, что интегральная функция распределения $F_3(\aleph)$ не может быть получена в аналитическом виде, границы критической области определяются численными методами приближённых вычислений из уравнений

$$\frac{\beta}{2}=\int_0^{\aleph_{кр.п}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)\aleph^{1/2}(1+\aleph)^{\frac{n+1}{2}}}d\aleph, \tag{5}$$

$$1-\frac{\beta}{2}=\int_0^{\aleph_{кр.п}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)\aleph^{1/2}(1+\aleph)^{n+1}}d\aleph. \tag{6}$$

Следовательно, окончательно решающее соотношение запишется в виде следующего неравенства

$$\aleph_{кр.п}\left(\frac{\beta}{2},n\right)\le\frac{Z_{n+1}^2}{\sum_{i=1}^{n}Z_i^2}\le\aleph_{кр.п}\left(1-\frac{\beta}{2},n\right).$$

По зависимостям (5) и (6) определяются правая $\aleph_{кр.п}$ и левая $\aleph_{кр.л}$ границы критической области теста 2.

Результаты вычислений для различных уровней значимости β представлены в табл. 1, из которой нетрудно заметить, что с увеличением объёма последовательности n и уровня значимости β область доверия сужается [3;8;10].

Таблица 1

Границы критической области для теста 2

n+1	$\beta = 0{,}2$		$\beta = 0{,}3$		$\beta = 0{,}4$	
	$\aleph_{кр.л}$	$\aleph_{кр.п}$	$\aleph_{кр.л}$	$\aleph_{кр.п}$	$\aleph_{кр.л}$	$\aleph_{кр.п}$
2	0,0273	51,49	0,0625	20,57	0,1153	10,79
3	0,0108	5,39	0,0251	3,07	0,0446	2,03
4	0,0068	2,36	0,0153	1,47	0,0278	1,03
5	0,0048	1,47	0,0109	0,95	0,0197	0,69
6	0,0038	1,07	0,0085	0,71	0,0155	0,52
7	0,0031	0,85	0,007	0,56	0,0128	0,42
8	0,0026	0,7	0,0058	0,47	0,0106	0,35
9	0,0023	0,6	0,0051	0,41	0,0092	0,3
10	0,002	0,53	0,0045	0,36	0,0082	0,26
11	0,0018	0,47	0,004	0,32	0,0073	0,24
12	0,0016	0,43	0,0037	0,29	0,0067	0,21
13	0,0015	0,39	0,0034	0,27	0,0059	0,18
14	0,0013	0,57	0,0031	0,24	0,0055	0,17
15	0,0012	0,34	0,0028	0,23	0,0051	0,16
16	0,0011	0,32	0,0026	0,21	0,0048	0,15

Заметим также, что введённый в рассмотрение второй тест в силу его чувствительности к дисперсии нормализованной случайной величины является критичным к тенденции вырождения пуассоновского потока в регулярный.

Таким образом, введённые в рассмотрение тесты выявления пуассоновского потока позволяют выявить момент изменения свойств случайной последовательности событий и перейти от пуассоновской модели редких событий к одной из её модификаций.

Одной из таких модификаций, описывающей статистическую структуру, возникающую из комбинации простейших пуассоновских потоков, является распределение Пуассона степени k, производящая функция которого имеет вид: [7;8]

$$f(t) = \boldsymbol{exp}\left\{\frac{\nu}{k}\sum_{n=1}^{k}(t^n - 1)\right\}.$$

Это распределение можно охарактеризовать следующими параметрами:

математическим ожиданием [4;5]

$$E_n = \left.\frac{df(t)}{dt}\right|_{t=1} = \boldsymbol{exp}\left\{\frac{\nu}{k}\sum_{n=1}^{k}(t^n - 1)\right\}\frac{\nu}{k}\sum_{n=1}^{k}n(t^{n-1})\Bigg|_{t=1} = \frac{\nu}{k}\sum_{n=1}^{k}(n-1) = \frac{\nu(k-1)}{2};$$

дисперсией (вторым центральным моментом)

$$\sigma^2 = En^2 - E[n^2] = \ddot{f}(t)|_{t=1} + \dot{f}(t)|_{t=1} - (En)^2 = \frac{\nu(k-1)}{2} + \frac{\nu}{k}\sum_{n=2}^{k}[n(n-1)-1].$$

Из полученных соотношений следует, что отношение дисперсии к математическому ожиданию для этого распределения зависит только от параметра k.

Следовательно, при помощи этой зависимости можно определить параметр kпо выборочным значениям моментов [5]

Обобщённое распределение Пуассона порядка k формируется следующим образом.

Пусть $g_k(t)$ – производящая функция вероятностного распределения, сосредоточенного на множестве целых чисел $1,2,...,k$, а $\phi(t)$– производящая функция распределения Пуассона.

Тогда производящая функция $f(t)$ обобщённого пуассоновского распределения порядка k определяется формулой [7]

$$f(t) = \phi[g_k(t)].$$

Так, например, если $g_k(t)$ – производящая функция биномиального распределения

$$g_k(t) = [1 + p(t-1)]^k,$$

то

$$f(t) = e^{\nu\{[1+p(t-1)]^k - 1\}}$$

$$p_0 = e^{\nu[(1-p)^k - 1]};$$

$$p_1 = p_0 \nu k (1-p)^{k-1} p \quad \text{и т.д.}$$

Однопараметрическое распределение $p_n = a\xi(n+1, a+1)$ может рассматриваться как составное пуассоновского (усечённое в конце) распределения, параметры которого подчиняются показательному распределению со средним значением, равным a^{-1}.

Список литературы

1. Гнеденко Б.В. Курс теории вероятностей. – М.: Наука, 1969.
2. Дзанагова И.Т., Кокоева З.Т. Экстремальные модели редких событий // Известия ГГАУ. – 2007. – № 44. – Ч. 2.
3. Дзанагова И.Т. Теоретико-информационные методы построения экстремальных и непредельных распределений сумм случайного числа случайных величин. – Владикавказ: ГУП Изд. «Олимп», 2009.

4. Дзанагова И.Т., Галаванова З.Е., Высшая математика. Ч. I. – ГУП Изд. «Олимп» г. Владикавказ, 2012.

5. Дзанагова И.Т., Хугаева Л.Т. Высшая математика. Ч. II. – ГУП Изд «Олимп» г. Владикавказ, 2012.

6. Колмогоров А.Н. Основные понятия теории вероятностей. – М.: ОНТИ, 1936.

7. Крамер Г. Математические методы статистики. – М.: Мир, 1975.

8. Мартыщенко Л.А. Введение в статистическое моделирование технических систем. – МО СССР, 1982.

9. Мартыщенко Л.А. Экстремальное распределение экстремальных случайных величин. – Л.: МО СССР, 1989.

10. Фишборн П. Принятие решений по неполной информации о вероятностях исходов // Экспресс-информация. Техническая кибернетика. – 1965. – № 34.

Рецензенты:

Хасцаев Б.Д., д.т.н., профессор, декан факультета электронной техники, Северо-Кавказский Ордена Дружбы Народов горно-металлургический институт (Государственный технологический университет), г. Владикавказ.

Шелехов П.Ю., д.т.н., профессор кафедры «Детали машин», Северо-Кавказский Ордена Дружбы Народов горно-металлургический институт (Государственный технологический университет), г. Владикавказ.