Предсказание очередного события в редеющем потоке времени

Prediction of next event in the thin flow of events

Светлой памяти профессора Николая Михайловича Седякина посвящается

Ключевые слова: редеющий поток – thinning flow; случайное событие – accidental event; распределение величин – value distribution; временной интервал – time interval; случайный поток – random flow.

В статье рассматривается вопрос изучения редеющего потока случайных событий при произвольных распределениях величин временных интервалов между ними. На основании конечного числа наступивших событий случайного потока предсказывается наступление очередного, следующего события, которое ожидается в дальнейшем течении времени.

The article analyzes the problem of researching thin flow of casual events in the course of general distributions of time gap values between them. On the basis of the final number of occurred events of casual flow there is possibility to predict occurrence of the recurrent, next event which is expected in the following current of time.

СМАГИН / SMAGIN V.
Владимир Александрович

доктор технических наук,
профессор Военно-космической академии
им. А.Ф. Можайского,
Заслуженный деятель науки РФ,
действительный член
Международной академии
информатизации,
Санкт-Петербург

СОДЕРЖАТЕЛЬНАЯ ПОСТАНОВКА ЗАДАЧИ

Исследование свойств редеющего потока случайных событий является весьма важным как с теоретической, так и с практической точки зрения. С теоретической точки зрения получение расчетных алгоритмов и формул для такого потока при произвольных распределениях представляет достаточно трудную задачу.

Автору неизвестны какие-либо результаты в данной области. Н.М. Седякиным выполнено исследование свойств редеющего потока [1] в предположении экспоненциального распределения времени до наступления события в потоке. Им предложен широкий арсенал математических формул для достаточно полного изучения свойств и расчета показателей потока.

С практической точки зрения изучение потока случайных событий в настоящее время необходимо для исследования надежности функционирования программных средств (ПС) объектов военного и гражданского назначения. В исследованиях по надежности ПС хорошо известна модель Джелинского – Моранды [2]. Она позволяет рассчитывать показатели надежности ПС тогда, когда известны несколько интервалов времени между ошибками работы ПС, по которым требуется предсказать показатели надежности ПС в дальнейшем времени функционирования. Ограничением этой модели является предположение об экспоненциальном распределении времени между ошибками. Доцентом Военно-космической академии им. А.Ф. Можайского В.С. Солдатенко именно на основе исследований Н.М. Седякина и модели Джелинского – Моранды было выполнено практически важное для военных объектов исследование по надежности прикладного программного обеспечения.

Какова область приложения исследований редеющих потоков? Можно указать на то, что эта область достаточно широка и разнообразна. Во-первых, это многопроцессорные вычислительные системы, которые находят применение в различных сложных системах управления объектами, использующих принцип параллельности вычислений. Примером из военной области является применение подобных систем на вооружении ПВО страны. Во-вторых, это применение в самолетных, ракетных и космических системах нескольких двигателей, работающих одновременно. И, наконец, можно указать на современную многоспутниковую систему типа «ГЛОНАСС», структурно состоящую из множества одновременно действующих отечественных спутников. Можно привести и другие практически

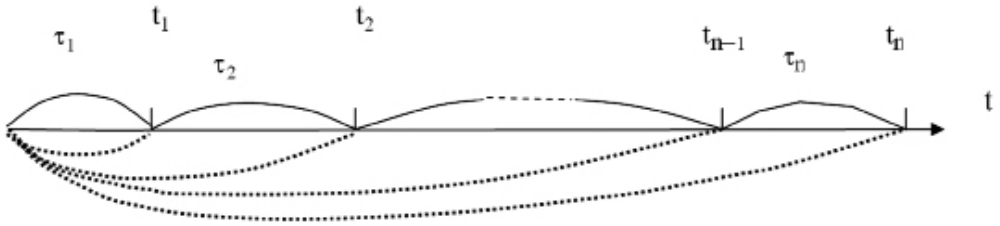


Рис. 1

важные примеры применения принципа параллелизма работы объектов, с течением времени деградирующих, теряющих сложность в силу различных причин.

Поэтому теоретическое решение задачи исследования свойств редеющего потока не только при экспоненциальном законе распределения времени, но и при произвольном, в большей степени, отвечающем потребностям практики, остается актуальным, прежде всего – с практической точки зрения. На наш взгляд, получение каких-либо новых результатов в данном направлении заслуживает внимания и поддержки.

Целью данной статьи является изучение редеющего потока случайных событий при произвольных распределениях величин временных интервалов между ними. При этом основная задача заключается в том, чтобы на основании конечного числа наступивших событий случайного потока предсказать наступление очередного, следующего события. Вопрос о том, какие меры при наличии этой информации могут приниматься в конкретной системе и ситуации, в данной статье обсуждаться не будет.

МАТЕМАТИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ И АЛГОРИТМ РЕШЕНИЯ

Необходимые сведения и формализации из теории редеющего потока

Интенсивность потока в момент появления каждого события уменьшается скачкообразно. Число событий является конечным, детерминированным или случайным. Временная диаграмма реализации редеющего потока показана на рис. 1.

Функция распределения длительности любого из n интервалов между моментами появления смежных событий потока определяется выражением [3]:

$$F_{\tau_i}(\tau) = A_n^i \int_0^\infty f(x_1) L$$

$$L \int_0^\infty f(\sum_{k=1}^{i-1} x_k) \int_0^\tau f(\sum_{k=1}^{i} x_k) P^{n-i}(\sum_{k=1}^{i} x_k) dx_i \; L\, dx_1, \qquad (1)$$

где A_n^i – число размещений из n по i; $f(x)$ – плотность распределения времени до наступления события; $P(x)$ – вероятность отсутствия события на интервале.

Плотность и функция распределения моментов появления событий потока определяются, соответственно, выражениями:

$$f_{t_i}(t) = A_n^i f(t) P^{n-i}(t) [\tfrac{1}{(i-1)!} + \sum_{j=1}^{i-1} (-1)^j \frac{P^j(t)}{j!}], \qquad (2)$$

$$F_{t_i}(t) = 1 - C_n^{i-1} P^{n-i+1}(t) - \sum_{j=1}^{i-1} (-1)^j \frac{A_n^i P^{n-i+j+1}(t)}{j!(n-i+j+1)}, \qquad (3)$$

где C_n^{i-1} – число сочетаний из n по $i-1$.

Вероятность появления в интервале [0,t) ровно i событий потока определяется через функции распределения событий потока, определяемые выражением (3):

$$P[n(t)=i] = \begin{cases} 1 - F_{t_1}(t), & i = 0, \\ F_{t_i}(t) - F_{t_{i+1}}(t), & 0 < i < n, \\ F_{t_n}(t), & i = n. \end{cases} \qquad (4)$$

С точки зрения практического применения представляет интерес вероятность появления ровно k событий потока на произвольно выбранном интервале [t,t+ Δt). Данная вероятность $P[n(t,t+\Delta t)=k]$ вычисляется через вероятности $P[n(t)=i]$, определяемые выражением (4), и условные вероятности $P[n(t,t+\Delta t)=k/i]$:

$$P[n(t,t+\Delta t)=k] = \sum_{i=0}^{n-k} P[n(t)=i] P[n(t,t+\Delta t)=k/i]. \qquad (5)$$

Условная вероятность $P[n(t,t+\Delta t)=k/i]$ появления k событий потока в интервале [t,t+ Δt) при условии, что до момента t произошло i событий потока, равна:

$$P[n(t,t+\Delta t)=k/i] = \\ = \int_0^{\Delta tt} [\frac{f_{t_{k+i}}(t+\xi)}{P_{t_{k+i}}(t)} - \frac{f_{t_{k+i+1}}(t+\xi)}{P_{t_{k+i+1}}(t)}] d\xi, \qquad (6)$$

где $f_{t_j}(t)$ – плотность вероятности момента появления j-го события потока, определяемая выражением (2), а $P_{t_j}(t) = 1 - F_{t_j}(t)$, определяемая выражением (3).

В дальнейшем для решения задачи прогноза нам потребуются только выражения (2) и (3), которые целесообразно представить в виде:

$$f_{t_k}(t) = kC_n^k \sum_{j=0}^{n-k} (-1)^j C_{n-k}^j F^{j+k-1}(t) f(t), \qquad (7)$$

$$F_{t_k}(t) = kC_n^k \sum_{j=0}^{n-k} (-1)^j C_{n-k}^j \frac{F^{j+k}(t)}{j+k}. \qquad (8)$$

Математическая постановка решения задачи прогнозирования ожидаемого события

Предположим, что нами зафиксированы длительности , $\tau_i, i = 1, 2, L\ k, k$ интервалов между смежными событиями редеющего потока. Предположим, что закон распределения длительности интервала одного источника нам известен и идентичны все n первоначальных источников. Требуется определить время и вероятность появления в будущем $k+1$-го события в потоке.

Для решения поставленной задачи воспользуемся выражениями (7) и (8) и методом максимального правдоподобия (ММП). Запишем функцию правдоподобия, соответствующую k временным интервалам фиксированных событий потока:

$$L(k) = \prod_{i=1}^{k} \{ \frac{N!}{(i-1)!} \sum_{j=0}^{N-i} [(-1)^j \frac{F^{i+j-1}(\sum_{j=1}^{i} \tau_j) f(\sum_{j=1}^{i} \tau_j)}{j!(N-k-j)!}] \}, \qquad (9)$$

в этой формуле принято $n = N -$ число первоначальных источников, образующих редеющий поток. Формула (9) состоит из k произведений выражений для плотностей вероятностей вида:

$$a(i,t) = \frac{N!}{(i-1)!} \sum_{j=0}^{N-i} [(-1)^j \frac{F^{i+j-1}(t) f(t)}{j!(N-i-j)!}. \qquad (10)$$

Плотности вероятности (10) соответствует функция распределения:

$$A(i,t) = \frac{N!}{(i-1)!} \sum_{j=0}^{N-i} [(-1)^j \frac{F^{i+j}(t)}{j!(N-i-j)!(i+j)}. \qquad (11)$$

В формуле (9) аргументы у функции распределения и плотности распределения представлены в виде суммы временных интервалов τ_j, соответствующей моменту наступления t_i i-го события потока.

Для решения задачи прогнозирования момента наступления $k+1$-го события потока введем новую независимую переменную z, которая означает величину временного интервала от момента наступления k-го события до момента наступления прогнозируемого $k+1$-го события потока. Переменной z соответствует случайная величина $\hat{Z}$, распределение которой неизвестно. Задача заключается в том, чтобы найти это распределение и воспользоваться им для получения прогнозной вероятностной оценки момента наступления события $k+1$.

Запишем новое значение функции правдоподобия, в которой содержится переменная z:

$$L(k,z) = L(k) \times$$
$$\times \frac{N!}{k!} \sum_{j=0}^{N-k-1i} [(-1)^j \frac{F^{k+j1}(z + \sum_{j=0}^{k} \tau_j) f(z + \sum_{j=0}^{ki} \tau_j)}{j!(N-k-1-j)!}] \}. \qquad (12)$$

Формуле (12) соответствует частная плотность вероятности:

$$a(k,z) = \frac{N!}{k!} \sum_{j=0}^{N-k} [(-1)^j \frac{F^{k+j}(z + \sum_{j=0}^{ki} \tau_j) f(z + \sum_{j=0}^{ki} \tau_j)}{j!(N-k-1-j)!}], \qquad (13)$$

в которой k может принимать переменные значения $1,2,\ldots,k$. Формула (12) может быть представлена как $L(k,z) = L(k) \cdot a(k,z)$, а ее логарифм $\ln[L(k,z)] = \ln[L(k)] + \ln[a(k,z)]$Из условия $\frac{\partial \ln[L(k,z)]}{\partial z} = 0$, учитывая, что $\frac{\partial \ln[L(k)]}{\partial z} = 0$, получаем $\frac{\partial \ln[a(k,z)]}{\partial z} = 0$. Это означает, что прогнозирование значения величины z непосредственно связано только с $a(k,z)$ и она может быть найдена из выражения (13).

ЧИСЛЕННЫЙ ПРИМЕР РЕАЛИЗАЦИИ АЛГОРИТМА

Рассмотрим простой пример решения задачи прогнозирования момента наступления будущего случайного события с использованием среды Mathcad 13.

На объекте предусматривается периодическая групповая замена четырех систем, работающих одновременно. Замена производится после окончания работы системы, вышедшей из строя последней. С точки зрения надежности в группе применено нагруженное резервирование. Требуется определить момент замены очередной группы. Предполагается, что отказ системы подчиняется закону Релея, параметр которого необходимо определять по моментам отказов систем в группе или исходя из физических соображений. Прогноз момента выхода из строя четвертой системы определяется после фиксации момента отказа третьей системы.

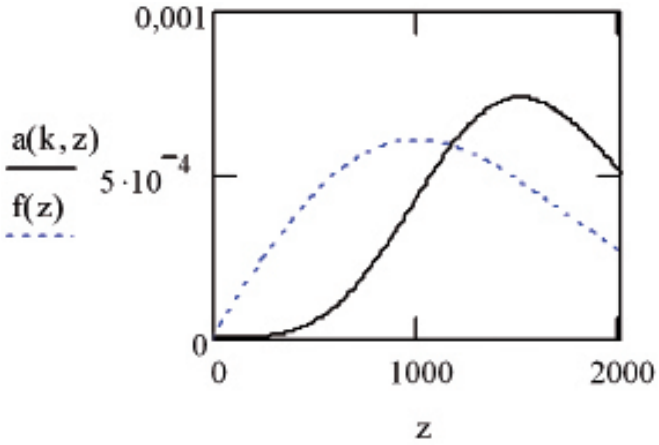


Рис. 2

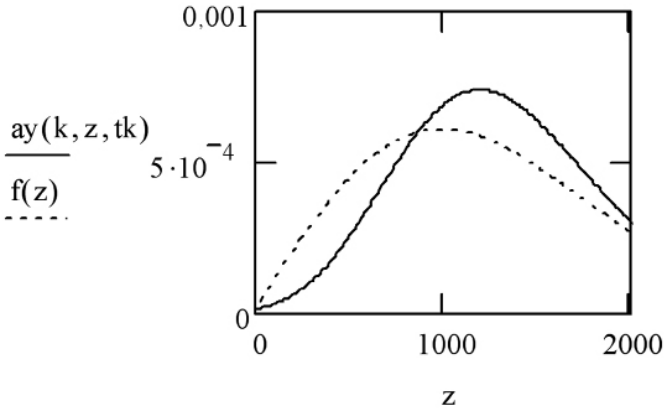


Рис. 3

Исходные данные для длительностей смежных интервалов между событиями потока равны: $\tau_1 = 21$ ч, $\tau_2 = 87$ ч, $\tau_3 = 203$ ч. На основе анализа физического процесса системы выбрано теоретическое распределение Релея времени до ее отказа со значением параметра $\sigma = 1000$ ч:

$$f(t) = \frac{t}{\sigma^2} e^{-\frac{t^2}{2\cdot\sigma^2}},\ F(t) =$$

$$= \int_0^t f(u)du,\ m = \int_0^\infty u \cdot f(u)du,\ m = 1,253\cdot 10^3,$$

где $f, F, m -$ плотность вероятности, функция распределения и среднее время до отказа системы. Требуется выполнить вероятностный прогноз момента наступления четвертого события (k=4) в потоке отказов.

С использованием выражения (13) получены графики функций $a(k,z)$, $f(z)$, показанные на рис. 2. Для сплошной кривой, соответствующей $a(k,z)$, рассчитаны значения математического ожидания, среднеквадратического отклонения и коэффициента вариации: $mo = 1,653\cdot 10^3$ ч, $cko = 557,47$ ч, $\eta o = 0,337$.

Затем была введена условная плотность вероятности момента наступления четвертого события z при условии, что время наступления третьего события в потоке равно $tk = \sum_{j=1}^{k} \tau_j$.

Она определяется как:

$$ay(k,z,tk) = \frac{a(k,z+tk)}{P(k,tk)},$$

где вероятность $P(k,tk) = \int_{tk}^{\infty} a(k,u)du$.

Рассчитаны значения условных математического ожидания, среднеквадратического отклонения и коэффициента вариации:

$$my = 1,343\cdot 10^3 \text{ ч},\ cko y = 556,008 \text{ ч},\ \eta y = 0,414.$$

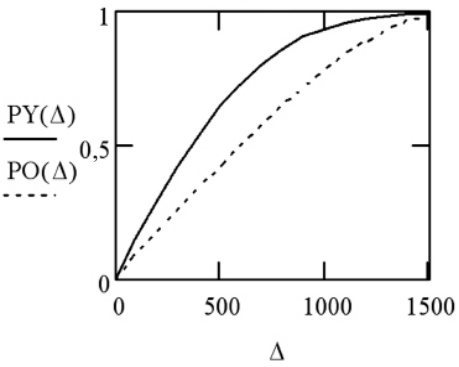
Графики плотностей $ay(k,z,tk)$, $f(z)$ показаны на рис. 3. Далее были рассчитаны вероятности попадания случайной величины $\hat{Z}$ – момента наступления четвертого события потока – в симметричную зону шириной 2Δ относительно математических ожиданий my и mo:

$$PY(\Delta) = \int_{my-\Delta}^{my+\Delta} a(k,u,tk)du, \quad PO(\Delta) = \int_{mo-\Delta}^{mo+\Delta} f(u)du.$$

Ширина зоны Δ изменялась в пределах $\Delta = 0,1500$ ч с шагом в 100 ч. Результаты расчетов представлены графиками на рис. 4 и числовыми значениями вероятностей в столбцах справа. На рисунке 5 представлена разность этих вероятностей $\Phi(\Delta) = PY(\Delta) - PO(\Delta)$. Из полученных расчетных данных следует, что наиболее вероятен прогноз замещения группы в интервале времени 743 – 1943 ч с вероятностью 0,725.

НЕСКОЛЬКО ЭВРИСТИЧЕСКИХ ЗАМЕЧАНИЙ

1. Прогнозирование, предсказание явлений и событий в самых различных областях человеческой деятельности всегда было и остается одной из самых злободневных задач. Этой задаче уделяли должное внимание такие выдающиеся ученые, как А.Н. Колмогоров [4] и Н. Винер [5]. Интерес к данному направлению чрезвычайно актуален и в наше время. Мы не можем претендовать на глубокие познания и полученные нами результаты. Наш интерес исключительно вызван достаточно узким вопросом предсказания ошибок в процессе функционирования программного обеспечения вычислительных средств. И хотя полученные результаты не являются значительными, сделана еще одна



PY(Δ) =
0
0.142
0.28
0.41
0.529
0.635
0.725
0.799
0.857
0.901
0.933
0.955
0.97
0.98
0.987
0.991

PO(Δ) =
0
0.084
0.168
0.252
0.335
0.416
0.496
0.572
0.646
0.715
0.779
0.836
0.886
0.927
0.959
0.981

Рис. 4

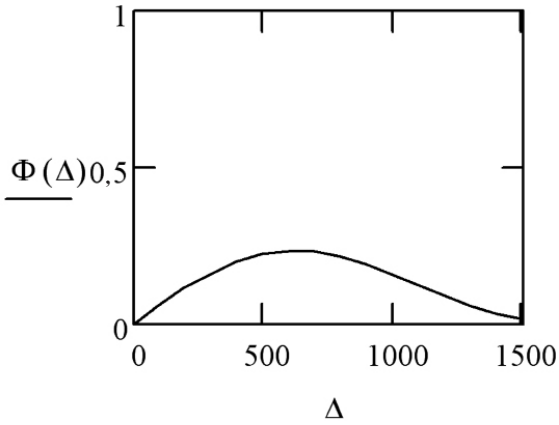


Рис. 5

попытка вероятностного предсказания случайных событий в классе редеющих нестационарных непуассоновских случайных потоков.

2. Представление случайного процесса возникновения ошибок в модели Джелинского – Моранды, как и детальное исследование редеющего потока Н.М. Седякина, для практики являются весьма ограничительными. Попытка автора осуществить прогноз будущего события в потоке случайных событий на основе их положений в [6], по сути, оказалась безрезультатной. Поэтому и был предпринят подход, излагаемый в данной статье. Использована модель описания случайного потока однородных событий при произвольном законе распределения времени наступления события.

3. В приведенном примере использован однопараметрический закон распределения, закон Релея. Для установления вида закона и значений его параметров в общем случае предполагается использование значений реализаций интервалов времени до наступления событий. При малом числе событий или интервалов данная задача становится достаточно трудной. При наличии другой возможности необходимо ею воспользоваться.

Второй, не менее трудной задачей является отыскание экстремальных значений функции правдоподобия или логарифмической функции от нее. Это обусловлено тем, что обе эти функции недифференцируемы из-за наличия в них факториалов. В статье для избавления от факториалов может быть применена формула Стирлинга, дающая достаточно высокую степень точности для их аппроксимации.

Принципиально устранение указанной второй трудности необходимо. Например, при оценивании надежности программного обеспечения по имеющимся статистическим данным необходимо определять число источников ошибок, вид закона распределения и значения его параметров, как это делается в модели Джелинского – Моранды. Если число начальных источников событий известно заранее, задача предсказания значительно упрощается. В этой статье вторую трудность нам устранить не удалось. Поэтому приводится достаточно простой пример реализации алгоритма. Можно это объяснить и нашим недостаточным знанием математических компьютерных систем.

4. При прогнозировании, конечно, следует всегда иметь в виду проблему малости выборок статистических данных. Чем больше объем выборки, тем точнее определяются вид закона распределения и его параметры, тем точнее прогноз момента ожидаемого события.

5. Использование условной плотности в процессе прогнозирования приводит к увеличению его достоверности, но незначительному.

6. Результаты выполненного, на наш взгляд, предварительного исследования прогнозирования момента наступления ожидаемого события в редеющем потоке нуждаются в дальнейшем изучении. Остается надеяться, что усилия в данном, важном практически направлении будут продолжены другими исследователями.

Литература

1. *Седякин Н.М.* Элементы теории случайных импульсных потоков. – М.: Сов. радио, 1965.

2. *Тейер Т., Липов М., Нельсон Э.* Надежность программного обеспечения. – М.: Мир, 1981.

3. *Тимофеев В.В.* Оценивание и обеспечение надежности функционально избыточных вычислительных систем // Вычислительные системы и сети. Методы оценивания и обеспечения качества / Под ред. В.А. Смагина. – СПб.: ВИККА им. А.Ф. Можайского, 1994.

4. *Колмогоров А.Н.* Интерполирование и экстраполирование случайных последовательностей. – Изв. АН СССР. – Сер. «Математика». – Т. 5. – 1941. – Стр. 3–14.

5. Extrapolation, interpolation and smoothing of stationary time series. – Technology Press of M. I. T. – New York: John Wiley & Son; London: Chapman & Hall, 1949.

6. *Смагин В.А.* Техническая синергетика. Вероятностные модели сложных систем. – СПб.: ВКА им. А.Ф. Можайского, 2004.