

Анализ случайных факторов процесса самообразования

Целью исследования является статистическое описание случайных факторов процесса самообразования – такого этапа процесса непрерывного образования, при котором отсутствует целенаправленное воздействие на обучающегося образовательной организацией и разработка алгоритмов оценки этих факторов.

Предполагается, что мотивациями самообразования являются внутренние факторы, характеризующие личность обучающегося и внешние, связанные с изменяющейся средой и возникающими новыми задачами. Явлениями, доступными для анализа процесса самообразования (наблюдаемыми данными), считаются события, имеющие отношение к этому процессу, которые моделируются точками на оси времени, число и положение которых предполагается случайными. Каждой точке процесса может быть поставлен в соответствие неизвестный и ненаблюдаемый случайный или неслучайный фактор (параметр), который влияет на интенсивность образования точек. Цель состоит в описании наблюдаемых и ненаблюдаемых данных и разработке алгоритмов их оптимальной оценки. Далее такие оценки могут быть использованы для индивидуальной характеристики процесса самообразования или для сравнения различных обучающихся.

Для анализа статистических характеристик процесса самообразования применен математический аппарат теории точечных случайных процессов, который позволяет определить ключевые статистические характеристики неизвестных случайных факторов процесса самообразования.

Работа состоит из логически завершенной модели, включающей следующие составные части.

- Обоснование базовой статистической модели появления точек в процессе самообразования в виде пуассоновского процесса, единственной характеристикой которого является интенсивность возникновения событий.
- Методика проверки гипотезы о пуассоновском распределении наблюдаемых событий.
- Обобщение базовой модели на случай, когда функция интенсивности зависит и от времени и от неизвестного фактора (параметра),

который может быть как случайным, так и не случайным. Такие факторы интерпретируются как факторы мотивационного типа, поскольку непосредственно влияют на интенсивность образования точек.

- Обобщение базовой статистической модели другого типа, когда каждому случайному событию приписывается случайное или неслучайное число. Эти числа интерпретируются как ресурс (цена), который расходуется при появлении каждого события и сопоставляются внешним факторам процесса самообразования.

Для каждой частной модели указаны оптимальные алгоритмы оценок соответствующих факторов по выбранным критериям, в простейших случаях получены аналитические выражения.

Показано, что для случайного параметра, не зависящего от времени достаточной статистикой является число точек на интервале наблюдения, а для изменяющегося во времени случайного параметра применим алгоритм оптимальной линейной фильтрации.

Для внешних факторов самообразования, получены выражения для математического ожидания и дисперсии этих факторов.

Рассмотрен сквозной числовой пример применения теории, включающий вычислительный эксперимент.

Применение математического аппарата случайных точечных процессов позволяет сформулировать модель случайных факторов процесса самообразования в виде случайной последовательности точек, которые отождествляются с некоторыми событиями, сопровождающими процесс самообразования.

Плодотворность упомянутого подхода подтверждается тем, что указаны алгоритмы для определения всех основных статистических характеристик всех рассмотренных типов случайных процессов возникновения событий, а для простейших случаев получены аналитические выражения.

Ключевые слова: самообразование, мотивация, событие, точечный случайный процесс.

Aleksandr A. Solodov

ООО «Технопрогресс 2000», Moscow, Russia

Analysis of random factors of the self-education process

The aim of the study is the statistical description of the random factors of the self-education process, namely that stage of the process of continuous education, in which there is no meaningful impact on the student's educational organization and the development of algorithms for estimating these factors. It is assumed that motivations of self-education are intrinsic factors that characterize the individual learner and external, associated with the changing environment and emerging challenges. Phenomena available for analysis a self-learning process (observed data) are events relevant to this process, which are modeled by points on the time axis, the number and position of which is assumed to be random. Each point can be mapped with the unknown and unobserved random or nonrandom factor (parameter) which affects the intensity of formation of dots. The purpose is to describe observable and unobservable data and developing algorithms for optimal evaluation. Further, such evaluations can be used for the individual characteristics of the process of self-study or for comparison of different students.

For the analysis of statistical characteristics of the process of self-education applied mathematical apparatus of the theory of point random

processes, which allows to determine the key statistical characteristics of unknown random factors of the process of self-education.

The work consists of a logically complete model including the following components.

- Study the basic statistical model of the appearance of points in the process of self-education in the form of a Poisson process, the only characteristic is the intensity of occurrence of events
- Methods of testing the hypothesis about Poisson distribution of observed events.
- Generalization of the basic model to the case where the intensity function depends on the time and unknown factor (variable) can be both random and not random. Such factors are interpreted as motivational factors, as directly affect the intensity of formation of dots.
- Generalization of the basic model of a different type, when each random event is attributed to random or non-random number. These numbers are interpreted as a resource (price), which is consumed with the appearance of each event and are mapped to external factors self-learning process.

For each private model provided optimal algorithms for estimating the relevant factors according to selected criteria, in the simplest cases, the analytical expressions are indicated.

It is shown that for a random parameter that is not time-dependent sufficient statistics is the number of points on the observation interval, and for time-varying random parameter; we apply the algorithm of optimal linear filtering. For external factors of self-education process, expressions for mathematical expectation and dispersion are obtained.

Considered a numerical example of application of the theory, including computational experiment.

The use of mathematical apparatus of random point processes allows us to formulate the model of the random factors of the process of self-education in the form of random sequence of points which are identified with some of the events that accompany the process of self-education. The fruitfulness of the approach is confirmed by the fact that algorithms to determine all the basic statistical characteristics of all the considered types of random processes of the occurrence of events are indicated and for simple cases analytical expressions are obtained.

Keywords: self-motivation, event, point casual process

Введение

Разработке математических моделей процесса обучения под руководством (преподавателя, тьютора и т.п.) посвящена обширная литература. Вместе с тем, концепция непрерывного образования предполагает продолжение образовательного процесса на протяжении всей жизни личности. При этом, очевидно, наиболее продолжительный период обучения приходится на процесс самообразования, особенностью которого является отсутствие учебных планов и программ, контроля и целенаправленной деятельности преподавателя. На этом этапе мотивацию личности определяют внутренние факторы (параметры, характеристики) обучения, которые характеризуют собственно личность обучающегося, и внешние, характеризующие независящие от обучающегося обстоятельства -изменение внешних условий, появление новых задач и т.п. [1].

Анализ воздействия и учета случайных факторов в процессе самообразования является малоизученной задачей, решение которой позволило бы определить закономерности этого процесса и выявить его измеримые характеристики. В связи с этим в работе предпринят анализ влияния случайных факторов на процесс самообразования на основе математической теории случайных точечных процессов.

В работе приняты следующие основные допущения.

Считается, что самообразование, являясь непрерывным во времени, проявляется в качестве некоторых событий, которые могут характеризовать процесс самообразования. На появление таких событий могут влиять как внутренние факторы процесса самообразования (например, осознание того, что появилась новая задача, требующая дополнительного самообразования), так и внешние (например, завершение решения новой задачи, выразившееся в получении патента).

События, требующие реакции самообразования, появляются случайным образом. Это объясняется непредсказуемостью появления внутренних и внешних факторов, а также тем обстоятельством, что новое знание генерируется тогда, когда имеется возможность случайного выбора из множества [2]. В связи с этим события моделируются точками на оси времени, появление и число которых является случайным. Основной характеристикой точечного процесса является интенсивность появления событий размерностью [1/время], имеющая смысл среднего числа событий, появившихся в течение выбранной единицы времени.

Получены соотношения, позволяющие на основа-

нии изучения последовательности точек сделать оценки внутренних и внешних факторов процесса самообразования и создать предпосылки для применения электронных систем обучения [3]. Приведен сквозной гипотетический пример, включающий вычислительный эксперимент.

1. Точечный процесс появления случайных событий

Введем в рассмотрение пространство состояний обучающегося, представляющее собой множество точек, каждая из которых характеризует некоторым образом состояние обучающегося и назовем его пространством состояний обучающегося. В процессе обучения, развивающемся в непрерывном времени, обучающийся скачкообразно меняет свое состояние и совершает тем самым переход между точками, описывая в пространстве состояний траекторию, которую назовем индивидуальной траекторией обучающегося

Точечная структура пространства состояний объясняется в рассматриваемой задаче его большей общностью по сравнению с непрерывным пространством, поскольку, плотно упаковывая точки, можно сколь угодно близко приблизиться к непрерывной траектории. Кроме того, диалектический закон перехода количества в качество, сопровождаемый скачкообразными изменениями, по-видимому, лучше описывает непознанные закономерности психологии и педагогики.

Пусть в начальный момент времени t_0 обучающийся находился за пределами пространства состояний, а через некоторое время t_1 текущий момент времени W_1 в результате обучения перешел в состояние, которому припишем номер 1. Через некоторое другое время t_2 в текущий момент времени W_2 обучающийся перешел в другое состояние, которому припишем номер 2 и т.д.

Введем в рассмотрение (рис. 1) процесс $N(t)$ счета точек, в которых побывал обучающийся и назовем его процессом счета точек или счетным точечным процессом. Таким образом, процесс $N(t)$ является кусочно-постоянным, имеет единичные приращения в моменты появления точек W_i и показывает, сколько точек появилось на интервале времени $[t_0, t)$.

Сделаем ключевое предположение о том, что времена появления точек W_i , а поэтому и межточечные интервалы t_i и число точек $N(t)$ являются случайными величинами. Это означает, что обучающемуся необходимо случайное время для изменения своего состояния в пространстве состояний.

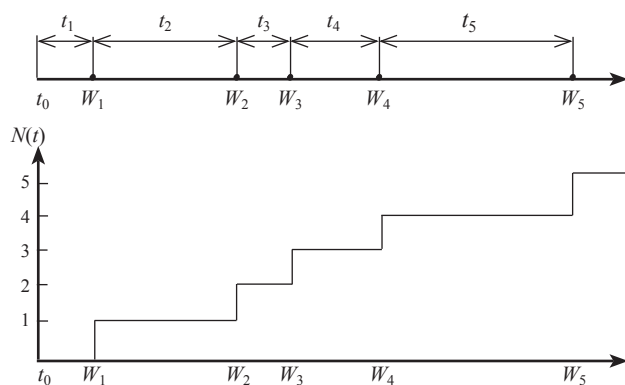


Рис. 1. Точечный случайный процесс и процесс счета

Действительно, поскольку обучающийся в рассматриваемой постановке задачи уже перешел на уровень самообразования, то естественно предположить, что процесс его обучения, в основном, обусловлен его реакциями на появляющиеся перед ним внешние и внутренние факторы. Такие факторы появляются в случайные моменты времени, обладают различной сложностью и требуют от обучающегося целенаправленных усилий различного объема, направления и интенсивности с целью адекватной реакции.

Конечно, реакция обучающегося на эти факторы, т.е. собственно обучение должно осуществляться по осмысленной программе самостоятельно, за межточечное время в сформулированной модели. В этих терминах осознание возникновения новой задачи, неполноты уже имеющихся знаний, подготовка публикации, защита диссертации и т.п. являются точками в пространстве состояний. Понятно, кроме того, что точки могут размножаться, т.е. после достижения обучающимся некоторой точки в пространстве состояний число точек может случайным образом измениться, что отражает идею непрерывного в течение активной жизни образования.

Будем далее полагать, что наблюдаемыми, т.е. доступными для изучения величинами являются точки на временной оси, отвечающие появлению упомянутых событий.

2. Статистическое описание наблюдаемых величин

Для создания стохастической модели случайных факторов необходимо задать базовые статистические характеристики доступных наблюдению случайных величин. Очевидно, что единственными наблюдаемыми величинами являются времена и число появившихся к некоторому моменту времени точек. Обозначим непрерывную случайную величину произвольного межточечного интервала через T и рассмотрим плотность вероятности экспоненциального распределения

$$P_T(T) = \lambda e^{-\lambda T}, \quad (2.1)$$

где λ – положительная величина.

Экспоненциальное распределение (2.1) широко применяется в науке и технике и описывает, в част-

ности, процессы переработки информации, т.е. интеллектуальную деятельность, например, длительность телефонного разговора или сеанса в сети Интернет. С другой стороны, это распределение применяется для описания случайного времени на осуществление некоторые работы – обслуживание посетителей в магазине, прием самолетов в аэропорте и т.п. Таким образом, экспоненциальное распределение времени, по-видимому, учитывает характерные особенности процесса обучения – интеллектуальную и физическую работу.

В связи с этим предположим, что распределение (2.1) описывает также и случайное время между соседними точками в процессе самообразования.

Рассмотрим произвольный интервал времени $[s, t)$, такой, что $t - s = T$ и предположим, что число точек, появившихся к моментам времени t и s равно соответственно $N(t)$ и $N(s)$. Обозначим через $N(t, s) = N(t) - N(s)$ число точек, появившихся на этом интервале, а через $P(N(t, s) = n)$ вероятность того, что это число точек окажется равным n .

Известно [4], что экспоненциальное распределение времени (2.1) между точками точечного процесса соответствует пуассоновскому закону распределения числа точек

$$P(N(t, s) = n) = \frac{1}{n!} (\lambda T)^n e^{-\lambda T} \quad (2.2)$$

и обратно, пуассоновское распределение числа точек (2.2) соответствует экспоненциальному распределению времени между точками (2.1). Таким образом, будем теперь полагать, что точечный процесс, изображенный на рис. 1. является пуассоновским случайным точечным процессом или просто пуассоновским точечным процессом, в котором времена появления точек W_1, W_2, \dots, W_i и их число $N(t)$ к моменту времени t являются случайными величинами.

Важнейшим свойством пуассоновского процесса является то, что он имеет независимые приращения. Это означает, что число точек, появившихся на любом интервале $[s, t)$, не зависит от того, какие события произошли до этого.

Из соотношения (2.2) следует, что математическое ожидание числа точек, появившихся на интервале времени T равно λT , поэтому параметр λ характеризует интенсивность появления точек пуассоновского процесса, т.е. указывает среднее число точек, появляющихся в единицу времени и имеет размерность [1/время]. При увеличении λ точки будут в среднем появляться чаще и наоборот. При увеличении безразмерного произведения λT дискретность процесса самообразования становится все менее выраженной и траектория приближается к непрерывной, а распределение (2.2) – к нормальному. Изучение подобных моделей является темой отдельного исследования, поэтому здесь мы будем полагать, что величина параметра λ в модели обучения согласована с общим масштабом рассматриваемого в модели времени таким образом, что безразмерное произведение λT составляет единицы. Так, если рассматривается время обучения протяженностью в годы, то параметр λ будет составлять несколько единиц в год.

Таким образом, параметр λ характеризует интенсивность появления существенных событий в процессе самообразования, поэтому он характеризует и интенсивность процесса обучения. Наличие единственного параметра в распределениях является удобной особенностью, позволяющей получать математические выражения, допускающие простые интерпретации.

При постановке и решении задач статистического анализа случайных факторов процесса самообразования в рамках сформулированной модели будем полагать, что наблюдаемыми величинами являются как случайные времена появления точек W_1, W_2, \dots, W_i , так и их число $N(t)$.

3. Проверка гипотезы о виде распределения событий

Одним из основных и типичных вопросов разработки статистической модели любого процесса является проверка соответствия модели реальным данным, получаемым на основании опыта. Для пуассоновской модели (2.2) этот вопрос сводится к оценке совместимы ли теоретические вероятности появления определенного числа точек на определенных интервалах времени с числом реально подсчитанных точек на этих же интервалах. Для проверки этой гипотезы воспользуемся широко известным критерием согласия хи-квадрат, в соответствии с которым весь интервал времени, в котором проводились наблюдения, разбивается, например, на M одинаковых подинтервалов, в которых подсчитывается число появившихся событий. Пусть число временных интервалов с определенным количеством событий i равно n_i и будем полагать, что функция интенсивности процесса пуассона неизвестна. Определим статистику хи-квадрат следующим образом:

$$\chi^2 = \sum_{i=0}^{i=m} (n_i - Mp_i)^2 / Mp_i, \quad (3.1)$$

где m – наибольшее число точек, появившихся на любом из подинтервалов,

p_i – теоретические вероятности появления m точек на подинтервале, рассчитанные по формуле (2.1).

Случайная величина (3.1) подчиняется хи-квадрат распределению с единственным параметром $k = m - 1 - 1$ (последняя единица появляется потому, что приходится делать оценку единственного неизвестного параметра – функции интенсивности).

Теперь, зная плотность распределения величины хи-квадрат, можно задать вероятность $q = \varepsilon$, с которой конкретное, полученное в результате расчетов значение хи-квадрат может находиться вне соответствующего этой вероятности интервала

$$\chi_1^2(q) \leq \chi^2 \leq \chi_2^2(q) \quad (3.2)$$

Имеются подробные таблицы вероятностей $P(\chi^2 \geq \chi_q^2) = q$, по которым находим критическое значение $\chi_1^2(q)$ и $\chi_2^2(q)$ и сравниваем их с уже имеющимся числом χ^2 . Если окажется, что условие (3.2) удовлетворяется, то это означает, что с вероятностью $1 - q$

наблюдаемые данные соответствуют гипотезе о законе распределения случайной величины. Если окажется обратное, то делается вывод о сомнительности гипотезы.

По определению случайной величины хи-квадрат ясно, что она неотрицательна, поэтому нижняя граница в неравенстве (3.2) принципиально ограничивается нулем. Поэтому нет смысла для критерия хи-квадрат устанавливать нижнюю границу, которая должна быть вблизи нуля. Поэтому часто рассматривают и используют односторонний критерий хи-квадрат, когда нижняя граница по умолчанию принимается равной нулю. а по уровню значимости q определяются только верхняя граница $\chi_2^2(q)$.

Пример 3.1. Рассмотрим следующий пример применения критерия хи-квадрат. Пусть на протяжении $M = 56$ подинтервалов были обнаружены 113 событий, причем не обнаружено ни одного события на 13 подинтервалах, одно событие на 12 подинтервалах и т.д. Распределение числа подинтервалов с фиксированным числом событий дается таблицей 1.

Таблица 1

Число событий	0	1	2	3	4	5	6
Число подинтервалов	13	12	10	10	6	3	2

Поставим задачу подбора закона распределения случайных чисел, представляющих собой число событий на одном подинтервале ($T = 1$) и проверки непротиворечивости такого выбора с применением критерия хи-квадрат.

Выдвинем гипотезу о том, что пороки распределены по закону Пуассона, т.е. применим простую пуассоновскую модель вида (2.2). При этом неизвестным является интенсивность закона распределения. Поскольку интенсивность является математическим ожиданием числа событий на единице времени, то в качестве ее оценки λ^* примем среднее число событий на одном подинтервале, которое, очевидно, равно $113 / 56 = 2,02$ [событий / подинтервал]. Ниже будет показано, что это оценка максимального правдоподобия. Поскольку нами выбрана единица измерения времени [подинтервал], то все дальнейшие расчеты будут проводиться для этой выбранной единицы времени. Зная теперь интенсивность закона Пуассона, легко определить теоретические вероятности появления одного, двух, и т. д. событий в одном подинтервале, а также вероятность непоявления событий, которая рассчитывается путем подстановки в формулу (2.2) $n = 0$. Расчет произведен по формуле (2.2) путем подстановки в нее значения оценки интенсивности, единичного временного подинтервала и искомым чисел событий. Рассчитанные величины, фигурирующие в формуле (3.1) сведены в табл. 2.

По данным этой таблицы и по формуле (3.1) рассчитываем конкретное значение случайной величины хи-квадрат, которое оказывается равным 9,49. Поскольку при формулировке гипотезы о пуассоновском законе распределения числа событий сделана статистическая оценка интенсивности пуассоновского распределения (число параметров, подлежащих оценке равно едини-

це), то случайная величина хи-квадрат распределена по закону хи-квадрат с $k = 7 - 1 - 1 = 5$ степенями свободы.

Таблица 2

Экспериментальные		Теоретические	
Число событий	Число подинтервалов n_i	Вероятность p_i	Число событий $m_i = Mp_i$
0	$n_0 = 13$	$p_0 = 0,132$	$m_0 = 0,132 \times 56 = 7,43$
1	$n_1 = 12$	$p_1 = 0,268$	$m_1 = 0,268 \times 56 = 15,00$
2	$n_2 = 10$	$p_2 = 0,271$	$m_2 = 0,271 \times 56 = 15,15$
3	$n_3 = 10$	$p_3 = 0,182$	$m_3 = 0,182 \times 56 = 10,21$
4	$n_4 = 6$	$p_4 = 0,092$	$m_4 = 0,092 \times 56 = 5,15$
5	$n_5 = 3$	$p_5 = 0,037$	$m_5 = 0,037 \times 56 = 2,08$
6	$n_6 = 2$	$p_6 = 0,013$	$m_6 = 0,013 \times 56 = 0,7$

Зададимся уровнем значимости критерия q , например, 0,05. Тогда по таблицам квантилей распределения хи-квадрат определяем верхний порог $\chi^2_2(q)$, равный 11,07. Поскольку рассчитанное значение случайной величины оказалось меньше, то сформулированная гипотеза совместима с наблюдаемыми данными. В свою очередь, это означает, что результатам наблюдений соответствует распределение (2.2) и применимы все возможности для дальнейшего определения статистических характеристик траектории.

4. Модель внутренних неизвестных факторов точечного случайного процесса

Для разработки содержательной модели предположим, что величина λ является теперь известной функцией времени, а также известной функцией многомерного неслучайного вектора X . Обозначим эту функцию через $\lambda(t, X)$. Появление в описании функции интенсивности фактора X приводит к тому, что характер наблюдаемых случайных величин существенно меняется, т.е. он оказывает влияние на интенсивность появления точек и, следовательно, на интенсивность процесса самообучения. Поскольку, кроме того, фактор X определяет поведение функции интенсивности, которая, в свою очередь, характеризует усилия обучающегося, то будем называть его внутренним фактором интенсивности самообучения. Для функции интенсивности $\lambda(t, X)$ формула (2.1) принимает вид [5].

$$P(N(t, s) = n | X) = \frac{1}{n!} \left[\int_t^s \lambda(\tau, X) d\tau \right]^n \exp \left(- \int_t^s \lambda(\tau, X) d\tau \right), \quad (4.1)$$

а процесс, описываемый этой формулой, становится неоднородным пуассоновским процессом. Очевидно, что при постоянной функции интенсивности формула (4.1) принимает вид (2.1). Математическое ожидание числа точек, появившихся на интервале $[s, t]$ равно, очевидно

$$\Lambda(t, X) = \int_t^s \lambda(\tau, X) d\tau \quad (4.2)$$

Обозначим через $\omega = \{W_1, W_2, \dots, W_n, N(t) = n\}$ совокупность всех данных, которые могут содержаться в реализации пуассоновского процесса на интервале

времени $[t_0, t]$, т.е. времена появления первых n точек и факт того, что число точек к моменту времени равно n . Сформируем вероятностную характеристику $p(\omega)$ этой совокупности следующим образом:

$$p(\omega) = P(N(t) = n | W_1, W_2, \dots, W_n) p(W_1, W_2, \dots, W_n), \quad (4.3)$$

т.е. произведение вероятности того, что число точек равно n при условии, что точки появились во времена W_1, W_2, \dots, W_n на совместную плотность вероятности этих времен появления. По математическому смыслу это произведение вероятности на плотность вероятности и называется плотностью вероятности реализации точечного процесса.

Можно показать [5], что для функции интенсивности $\lambda(t, X)$ плотность вероятности реализации (4.3) принимает вид

$$p(\omega) = \exp \left\{ - \int_t^s \lambda(\tau, X) d\tau + \sum_{i=1}^{N(t)} \ln[\lambda(W_i, X)] \right\}. \quad (4.4)$$

Характеристика (4.4) содержит всю статистическую информацию, получаемую при известных данных $\omega = \{W_1, W_2, \dots, W_n, N(t) = n\}$ и может быть использована для формирования статистических оценок.

5. Оптимальная оценка внутренних неизвестных факторов процесса самообразования

Основной целью разработки модели случайных факторов самообучения является возможность постановки и решения задач оценки этих факторов путем обработки тех данных, которые доступны наблюдению. Оценке параметров стохастических процессов посвящена обширная литература, причем до аналитических выражений доводятся, в основном, задачи, ограниченные непрерывными гауссовыми распределениями фигурирующих в них величин. Значительно меньше результатов известно для точечных пуассоновских случайных процессов. Далее, в основном, будем следовать математическим результатам работы [5].

Общая постановка задачи оценки факторов самообучения состоит в следующем. Пусть на интервале времени $[t_0, T]$ пуассоновского процесса с функцией интенсивности $\lambda(t, X)$ наблюдаются данные $\omega = \{W_1, W_2, \dots, W_n, N(T) = n\}$. Необходимо на основании анализа этих данных определить наилучшим образом функцию $\lambda(t, X)$ и вектора X . Поскольку имеется плотность вероятности реализации (4.4), которая в терминах оценок является функцией правдоподобия, то будем применять оценку максимального правдоподобия, которая максимизирует функцию (4.4). Поскольку логарифм является монотонной функцией, то отыскивать оценку максимального правдоподобия можно искать, максимизируя логарифм (4.4) при $t = T$:

$$L(\omega) = \left(- \int_{t_0}^T \lambda(\tau, X) d\tau \right) + \sum_{i=1}^{N(T)} \ln[\lambda(W_i, X)] \quad (5.1)$$

Функцию (5.1) будем в дальнейшем называть функцией правдоподобия. Максимизация функции прав-

доподобия (5.1) в аналитическом (конечном) виде возможна только для простейших ситуаций, одна из которых рассмотрена ниже. Тем не менее, сформулированный результат позволяет применять известные численные методы отыскания экстремумов функционалов этого вида.

Рассмотрим простейший случай масштабирования известной функции интенсивности $\mu(t)$ неизвестным скалярным положительным неслучайным фактором X , т.е. положим, что на интервале времени $[t_0, T)$ известен набор данных $\omega = \{W_1, W_2, \dots, W_n, N(T) = n\}$ реализации пуассоновского процесса с функцией интенсивности

$$\lambda(t, X) = X\mu(t). \quad (5.2)$$

Еще раз подчеркнем, что в наборе ω содержится вся доступная при данных наблюдениях информация, в том числе и количество событий на временных подинтервалах.

Функция правдоподобия (5.1) принимает, очевидно, вид

$$\begin{aligned} L(\omega) &= \left(-X \int_{t_0}^T \mu(\tau) d\tau \right) + \sum_{i=1}^{N(T)} (\ln X + \ln \mu(W_i)) = \\ &= \left(-X \int_{t_0}^T \mu(\tau) d\tau \right) + N(T) \ln X + \sum_{i=1}^{N(T)} \ln \mu(W_i). \end{aligned} \quad (5.3)$$

Очевидно, последний член в правой части не зависит от X , поэтому оценка максимального правдоподобия не зависит ни от времен появления точек, несмотря на то, что они, несомненно, несут информацию, ни от индивидуальных чисел появления событий в подинтервалах всего интервала наблюдения. Это объясняется простейшей моделью функции интенсивности (5.2), которая является известной с точностью до масштабирующего множителя.

Максимум (5.3) легко отыскивается путем приравнивания производной правой части к нулю. В результате получим следующую точную оценку максимального правдоподобия

$$X_{МП} = \frac{N(T)}{\int_{t_0}^T \mu(\tau) d\tau}. \quad (5.4)$$

С точки зрения процесса самообучения оценка (5.4) может рассматриваться как уточнение по результатам наблюдения за характерными событиями (точками) уже известной функции интенсивности $\mu(t)$, причем результат вполне понятен: чем больше точек появилось, тем интенсивнее идет процесс самообучения. Поскольку число точек $N(T)$ является случайным, то и оценка (9) тоже является случайной. Очевидно, что математическое ожидание оценки совпадает с X , поэтому оценка является несмещенной. Можно показать, что эта оценка имеет минимальное среднее квадратическое отклонение $\sigma_{МП}$, равное

$$\sigma_{МП} = \frac{X_{МП}}{\sqrt{N(T)}}. \quad (5.2)$$

Таким образом, дисперсия медленно уменьшается с ростом числа наблюдений.

Если положить в формуле (5.2) $\mu(t) = 1$, то получим модель неизвестной постоянной функции интенсивности $\lambda(t, X) = X$ и оценка (5.4) принимает вид

$$X_{МП} = \frac{N(T)}{(T - t_0)}, \quad (5.2)$$

сохраняя перечисленные выше свойства.

Пример 5.1. Рассмотрим табл. 1. В соответствии с формулой (5.6) оценка максимального правдоподобия функции интенсивности $\lambda(t, X)_{МП} = X_{МП}$ равна 113[событий] / 56[подинтервалов] = 2,02[событий/подинтервал]. Эта оценка уже использовалась в п. 3. Среднеквадратическое отклонение $\sigma_{МП}$ этой оценки минимально и равно 0,19[событий / подинтервал] в соответствии с формулой (5.6).

Если до проведения наблюдений имелись основания полагать, что функция интенсивности $\mu(t)$ в этом примере равнялась, например, 2 [событий/подинтервал], то, в соответствии с (5.4), данные таблицы 1 могут быть использованы для уточнения значения интенсивности:

$$X_{МП} = \frac{N(T)}{\int_{t_0}^T \mu(\tau) d\tau} = \frac{113}{\int_0^{56} 2 d\tau} = 1,009,$$

а уточненная оценка в соответствии с (5.2) $\lambda(t, X)_{МП} = X_{МП}\mu(t) = 1,009 \times 2 = 2,018$.

6. Модель случайных внутренних факторов. Оптимальная оценка случайных мотивационных характеристик

Естественным развитием предыдущего рассмотрения является предположение о том, что функция интенсивности $\lambda(t, X(t))$ зависит теперь от случайного процесса $X(t)$. Случайный точечный процесс, в котором функция интенсивности появления точек является случайной, называется процессом с двойной случайностью

Случайную функцию $X(t)$, являющуюся обобщением внутренней характеристики X из раздела 5, будем интерпретировать как изменяющуюся во времени мотивационную характеристику обучающегося. В самом деле, влияние на интенсивность появления событий в процессе обучения, носит принципиально случайный характер и определяется множеством факторов, зависящих от среды, в которой существует обучающийся, его психологического состояния, его представлений о перспективах научного и карьерного роста и т.п. В совокупности эти обстоятельства и составляют фактор текущей мотивации, который будем характеризовать случайным процессом $X(t)$.

Подчеркнем, что характеристика $X(t)$ не наблюдается (не измеряется) непосредственно.

Проблематичным представляется вопрос о способе описания процесса $X(t)$. С одной стороны, он должен описывать характерные черты изменения мотиваций, а с другой — позволять развить содержательную математическую теорию. В качестве компромисса будем полагать, что процесс $X(t)$ является одномерным

и определяется следующей парой линейных соотношений:

$$\lambda(t, X(t)) = \lambda_0[1 + X(t)], \quad (6.1)$$

$$\frac{dX(t)}{dt} = -\frac{1}{\tau} X(t) + \frac{1}{\tau} n(t), X(0) = 0, \quad (6.2)$$

где λ_0 – математическое ожидание функции интенсивности, $n(t)$ – стандартный внешний процесс, влияющий на обучающегося, τ – постоянная времени реакции обучающегося на стандартный внешний процесс, $X(0)$ – значение процесса в начальный момент времени.

Целесообразность модели (6.1) и (6.2) подтверждается следующими аргументами. В правой части уравнения (6.2) фигурирует стандартный, т.е. одинаковый для всех обучающихся процесс. Вслед за теорией динамических систем будем полагать, что представляет собой белый гауссовский шум со спектральной плотностью R . Это означает, что математическое ожидание процесса $X(t)$ равно нулю, его средний размах определяется величиной R (и может быть выбран одинаковым для всех обучающихся), а его динамика, т.е. поведение во времени, определяется постоянной времени τ . Действительно, из уравнения (6.2) следует, что при уменьшении τ правая часть, т.е. производная процесса $X(t)$ увеличивается, а это и означает более динамичное поведение процесса.

Важно подчеркнуть, что модель (6.1) и (6.2) наиболее подходит для сравнения динамики мотивационных факторов у разных обучающихся, поскольку все параметры модели, кроме постоянной времени τ могут быть выбраны одинаковыми.

Таким образом, в процессе оценки наблюдается случайный точечный процесс $N(t)$ с функцией интенсивности $\lambda(t, X(t))$, удовлетворяющей соотношениям (6.1) и (6.2).

Задача оптимальной оценки мотивационной характеристики состоит в формировании на основании обработки процесса $N(t)$ оптимальной по критерию минимума среднеквадратической ошибки (с.к.о.) оценки $X^*(t)$ процесса $X(t)$.

Последовательное решение [6] сформулированной выше задачи приводит к системе связанных дифференциальных уравнений, которые называются фильтром Калмана:

$$\begin{aligned} \frac{dX^*(t)}{dt} = \\ = -(1/\tau)X^*(t) + D(t)[N(t) - (1 + \lambda_0 X^*(t))] X^*(0) - 0. \end{aligned} \quad (6.3)$$

$$\frac{dD(t)}{dt} = -2D(t)/\tau + (R/\tau^2) - D^2(t)\lambda_0, D(0) = D_0. \quad (6.4)$$

$$\lambda^*(t) = \lambda_0[1 + X^*(t)]. \quad (6.5)$$

В уравнениях (6.3) и (6.4) через $X^*(t)$ обозначена оптимальная по критерию минимума с.к.о. оценка процесса $X(t)$, а через $D(t)$ – текущая дисперсия ошибки.

Важнейшей особенностью уравнения (6.4) является то, что в него не входит наблюдаемый процесс $N(t)$, а

все коэффициенты известны. Это означает, что данное уравнение может быть решено до начала процедуры оценки, т.е. можно получить функцию $D(t)$ для любого аргумента t независимо от решения уравнения фильтрации (3.15). В свою очередь, это означает, что качество оценки не зависит от поступающих данных, т.е. от характера изменения процесса $N(t)$.

По математической классификации уравнение (6.4) является нелинейным дифференциальным уравнением первого порядка и называется уравнением Риккати. Аналитическое решение этого уравнения получить не удается.

Таким образом, решение задачи оптимальной линейной оценки в рассмотренной постановке задачи дается парой дифференциальных уравнений (6.3) и (6.4). Отметим, что приведенный алгоритм Калмана является оптимальным, т.е. качество оценки, достигаемое с помощью этого алгоритма, не может быть улучшено в классе линейных алгоритмов.

Практическое применение алгоритма Калмана сводится к решению в темпе поступления наблюдаемого процесса $N(t)$ указанных уравнений.

Пример 6.1. Рассмотрим данные табл. 1 из примера 3.1. Поскольку для применения алгоритма Калмана нужно сформировать наблюдаемый процесс $N(t)$, то распределим случайным образом 113 событий по всем 56 подинтервалам так, чтобы сохранилось указанное в Табл. 1 распределение числа событий в одном подинтервале. Результат этой процедуры представлен на рис. 2, где отмечено количество событий на данном подинтервале, позволивший получить соотношения, пригодные для алгоритмизации.

Конкретные времена появления событий, т.е. их текущие времена t в каждом подинтервале распределим также случайно. Таким образом, последовательный просмотр всех 56 подинтервалов сформирует одну из возможных реализаций случайного точечного процесса, который в соответствии с изложенной моделью будем считать пуассоновским процессом с двойной случайностью. При этом, если на текущем шаге появляется событие, то $N = 1$, а если нет, то $N = 0$.

Вычислительный эксперимент, осуществим для следующих значений переменных, нормированных к

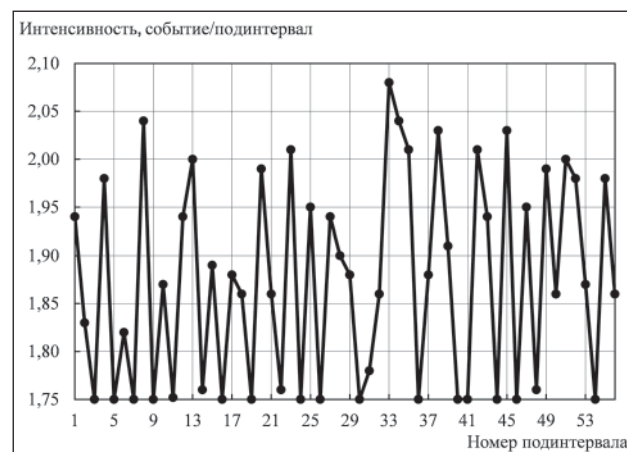


Рис. 2. Случайное распределение интенсивности по подинтервалам

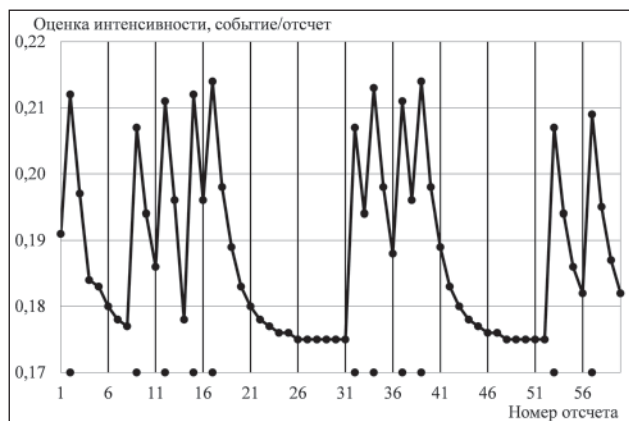


Рис. 3. Поведение оценки интенсивности

длительности подинтервала: $X^*(0) = 0, \lambda_0 = 2; \tau = 0,28$. Для обеспечения выполнения для любого t с вероятностью 0,999 физического очевидного неравенства $X(t) \geq 1$, положим в формуле (6.4) $R = 4/7\tau$. Моделирование уравнений проведено с шагом $\Delta t = 0,1$ длины подинтервала, поэтому всего в эксперименте сделано 560 отсчетов.

На рис. 3. представлено детальное поведение оценки интенсивности (фрагмент общей картины), рассчитанное в соответствии с приведенными алгоритмами и формулой (6.5). В соответствии с выбранным шагом всего представлено 60 отсчетов на 6 подинтервалах, число событий, которые отмечены точками на оси аргумента, при этом составило 11. Рис. 3 иллюстрирует тот важнейший факт, что алгоритм позволяет оценивать интенсивность возникновения событий даже в промежутках между событиями, что можно использовать для оценки интенсивности появления неравномерных, т.е. либо частых, либо редких событий.

Для проверки согласия теоретических данных с экспериментальными, рассчитаем текущие (для каждого подинтервала) значения критерия хи квадрат. Для этого рассчитаем по формуле (2.8) с учетом соотношений (3.28) и (3.29) вероятности появления событий для каждого подинтервала (не представлены). Результаты этих расчетов представлены на рис. 4. В соответствии с изложенной в разделе 3 методикой, можно сделать вы-

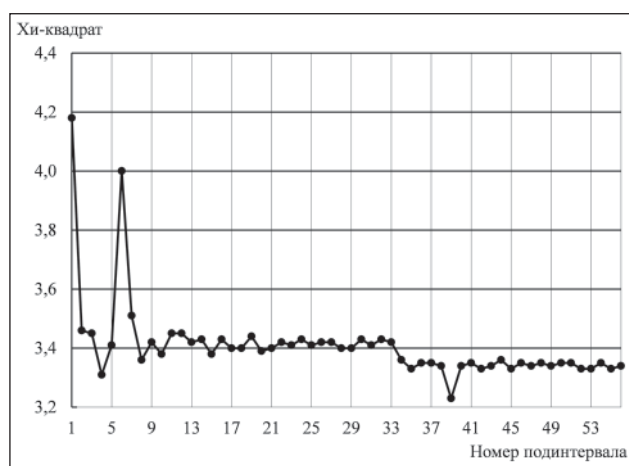


Рис. 4. Поведение хи-квадрат

вод о том, что все рассчитанные вероятности согласуются с наблюдаемыми данными.

Смысл применения алгоритмов оптимальной фильтрации состоит в том, что он позволяет сделать оценки по весьма ограниченному набору наблюдений. Это подтверждается тем, что значение критерия хи квадрат, как видно на рис. 4, не меняется существенно при обработке дополнительных наблюдаемых данных после 9 подинтервала.

Общий вывод, который можно сделать по результатам эксперимента состоит в том, что все рассчитываемые величины – оценка интенсивности, величина критерия хи-квадрат – существенно не меняются на протяжении всего эксперимента (за исключением нескольких начальных подинтервалов до установившегося режима). Это свидетельствует о том, что в эксперименте процесс самообразования проходил без выбросов в том смысле, что на всем времени эксперимента не было необходимости менять модель.

7. Анализ статистических характеристик внешних факторов

Рассмотрим случайный точечный процесс $\{N(t), t \geq t_0\}$ появления точек на индивидуальной траектории обучающегося. Припишем каждой появившейся i -той точке некоторую случайную величину u_i , которую назовем меткой и которая в общем случае может быть векторной. Эти величины по определению являются внешними по отношению к точечному процессу в том смысле, что не могут влиять на интенсивность появления точек, поэтому будем считать, что они являются внешними по отношению к траектории обучающегося характеристиками.

Типичными внешними характеристиками могут быть некоторые ресурсы, выделяемые (потребляемые) внешними источниками при достижении точечным процессом очередной точки. В связи с этим представляет интерес изучение процесса накопления меток $\{u(t), t \geq t_0\}$ на интервале времени $[t, s)$:

$$u(t) = \sum_{i=1}^{N(t)} u_i, \tag{7.1}$$

В формуле (7.1) метки u_i могут быть либо случайными независимыми одинаково распределенными величинами, которые также не зависят от $N(t)$, либо неслучайными величинами.

Если u_i в (7.1) случайны, то, очевидно, $u(t)$ является суммой случайного числа случайных слагаемых. Практический интерес в этой ситуации представляет математическое ожидание $Mu(t)$, процесса $u(t)$, которое легко вычисляется с использованием свойств условных математических ожиданий:

$$Mu(t) = M_N \{M_u [u(t)/N(t)]\} = M_N \left\{ \sum_{i=1}^{N(t)} m_i \right\}.$$

Здесь применены очевидные обозначения для математических ожиданий с индексом, указывающим случайную величину, по которой происходит усреднение.

Таблица 3

Если все математические ожидания случайных меток равны m , то

$$Mu(t) = mM_N\{N(t)\} = m \int_0^t \lambda(\tau) d\tau. \quad (7.2)$$

Рассмотрим ситуацию, в которой метки U_i , приписываемые точкам, неслучайны. Тогда процесс накопления меток принимает вид

$$U(t) = \sum_{i=1}^{N(t)} U_i, \quad (7.3)$$

Поскольку число точек $N(t)$ в (7.3) случайно, то и сумма $U(t)$ является случайной. Для определения математического ожидания

$$M_N\{U(t)\} = M_N\left\{\sum_{i=1}^{N(t)} U_i\right\}$$

применим декомпозицию пуассоновского процесса, т.е. представление его в виде суммы.

Пусть точки пуассоновского процесса $N(t)$, которым приписываются одинаковые неслучайные метки первого типа $U^{(1)}$, появляются с интенсивностью $\lambda_1(t)$ и т.д. точки, которым приписываются одинаковые метки типа $U^{(k)}$ появляются с интенсивностью, $k = 1, \dots, R$. Можно показать, что тогда пуассоновский процесс $N(t)$ представляется в виде суммы пуассоновских процессов $N_k(t)$, причем

$$\lambda(t) = \sum_{k=1}^R \lambda_k(t) \quad (7.4)$$

$$N(t) = \sum_{k=1}^R N_k(t) \quad (7.5)$$

Очевидно, что число точек $U_{t,s}^{(k)}$, появившихся на интервале времени $[t, s]$ и снабженных метками k -го типа, $U^{(k)}$ равно $U^{(k)}N_k(t, s)$, а их математическое ожидание с учетом формулы (7.5) определяется как

$$U_M^{(k)} = U^{(k)} \int_t^s \lambda_k(\tau) d\tau,$$

а математическое ожидание U_M меток всех типов, с учетом формулы (7.5) очевидно,

$$U_M = \sum_{k=1}^R U^{(k)} \int_t^s \lambda_k(\tau) d\tau \quad (7.6)$$

Пример 7.1. Рассмотрим появление событий на случайной траектории в соответствии с примером 3.1 и положим, что события могут быть четырех типов, которые перечислены в таблице 2. В этой же таблице указаны возможные функции интенсивностей появления событий, удовлетворяющие условию (7.4), а также средние числа появления событий на интервале времени $[0, 56)$, равное $\int_0^{56} \lambda_k d\tau = 56\lambda_k$ средняя цена $U_M^{(k)}$ каждого события на всей траектории, равная $U^{(k)} \int_0^{56} \lambda_k d\tau = 56U^{(k)}$ и итоговые характеристики событий.

Указанные в табл. 3 характеристики процесса самообразования отличаются от простого подсчета цен событий тем, что они относятся к случайному процессу появления событий и представляют собой математи-

Тип события k	Описание события	Цена события $U^{(k)}$	Интенсивность появления событий λ_k	Среднее число событий $56\lambda_k$	Средняя цена событий $U_M^{(k)}$
1	Изучение научной статьи	1	1,4	77	77
2	Участие в конференции	4	0,4	22	88
3	Публикация научной статьи	10	0,2	12	120
4	Публикация научной монографии	50	0,04	2	100
Всего			2,02	113	385

ческие ожидания этих событий. Таким образом, модель позволяет прогнозировать средние ресурсы (цену), которые сопровождают процесс самообразования.

Заключение

Применение математического аппарата случайных точечных процессов позволяет сформулировать модель случайных факторов процесса самообразования в виде случайной последовательности точек в пространстве состояний обучаемого, которые отождествляются с некоторыми событиями, сопровождающими процесс самообразования. Основной характеристикой случайного точечного процесса является функция интенсивности появления точек.

Введение в рассмотрение неизвестного и ненаблюдаемого фактора (параметра), от которого зависит функция интенсивности, и который интерпретируется как внутренний фактор процесса самообразования, позволило получить оптимальные статистические оценки этого фактора. Более того, теория случайных точечных процессов позволяет определить оптимальные оценки параметра в том случае, когда он является ненаблюдаемым случайным непрерывным процессом. Это, в свою очередь, означает, что анализ случайных событий в процессе самообразования позволяет сформировать оценки того процесса, который оказывает непрерывное случайное влияние на процесс самообразования и который может рассматриваться как процесс формирования мотивации.

Модель позволяет также ввести в рассмотрение внешние факторы процесса самообразования, которые не влияют на интенсивность появления событий и могут рассматриваться как ресурс или цена, сопровождающие появление событий.

Плодотворность упомянутого подхода подтверждается тем, что указаны алгоритмы для определения всех основных статистических характеристик всех рассмотренных типов случайных процессов возникновения событий, а для простейших случаев получены аналитические выражения.

Дальнейшее развитие вопросов анализа случайных факторов процесса самообразования может быть связано с уточнением моделей непрерывных процессов мотивационного типа, а также применением многомерных точечных процессов для учета факторов разных категорий.

Литература

1. Солодова Е.А. Новые модели в системе образования: Синергетический подход / Предисл. Г.Г. Малинецкого. – М.: Книжный дом «ЛИБРОКОМ», 2012. – 344 с.
2. Кастлер Г. Возникновение биологической организации. – М.: Мир, 1967. 90 с.
3. Трембач В.М. Система управления базами эволюционирующих знаний для решения задач непрерывного образования. – М.: МЭСИ, 2013. – 255 с.
4. Тихонов В.И., Миронов М.А. Марковские процессы. – М.: Советское радио, 1977. – 488 с.
5. Donald L. Snyder, Michael I. Miller. Random Point Processes in Time and Space. Second Edition Springer-Verlag New York Inc, 1991. 488 с.
6. Солодов А.В., Солодов А.А. Статистическая динамика систем с точечными процессами. – М.: Наука, 1988. 284 с.

Сведения об авторе

*Александр Александрович Солодов, д.т.н, проф,
Генеральный директор ООО «Технопрогресс 2000»
Тел.: (903) 726 10 09, E mail: aasol@rambler.ru
ООО «Технопрогресс 2000», Москва, Россия*

References

1. Solodova E.A. Novye modeli v sisteme obrazovaniya: Sinergeticheskij podhod / Predisl. G.G. Malineckogo. – M.: Knizhnyj dom «LIBROKOM», 2012. – 344 s.
2. Kastler G. Vozniknovenie biologicheskoy organizacii. – M.: Mir, 1967. 90 s.
3. Trembach V.M. Sistema upravleniya bazami ehvolucioniruyushchih znaniy dlya resheniya zadach nepreryvnogo obrazovaniya. – M.: MEHSI, 2013. – 255 s.
4. Tihonov V.I., Mironov M.A. Markovskie processy. – M.: Sovetskoe radio, 1977. – 488 s.
5. Donald L. Snyder, Michael I. Miller. Random Point Processes in Time and Space. Second Edition Springer-Verlag New York Inc, 1991. 488 s.
6. Solodov A.V., Solodov A.A. Statisticheskaya dinamika sistem s tochechnymi processami. – M.: Nauka, 1988. 284 s.

Information about the author

*Aleksandr A. Solodov, doctor of technical Sciences, Professor,
General Director OOO "Tekhnoprogress 2000"
Tel.: (903) 726 10 09, E mail: aasol@rambler.ru
OOO "Tekhnoprogress 2000", Moscow, Russia*