

ЕМКОСТНЫЙ МЕТОД АНАЛИЗА РЕДКИХ СОБЫТИЙ В СФЕРЕ УСЛУГ¹

*Ю.А. Кораблев, П.С. Голованова,
Т.А. Кострица*

DOI: 10.33293/1609-1442-2020-3(90)-132-142

Представьте, что вы являетесь владельцем некоторого сервиса. Вам необходимо составить план работы своих мастеров на определенный будущий период, определить количество необходимых расходных материалов. Для этого вам требуется прогноз будущих необходимых вам услуг. Классические математические методы работы с временными рядами не подходят для решения этой задачи. Агрегирование данных об услугах по месяцам и составление временного ряда способны только запутать. Прогнозирование услуг необходимо выполнять средствами методов работы с редкими событиями. Редким событиям посвящено сравнительно мало работ. Методов исследования редких событий значительно меньше, чем методов для анализа частых событий (временных рядов). Наиболее популярным методом исследования редких событий на данный момент является теория случайных процессов, в которой используется поток событий Пуассона или Эрланга. Однако с помощью случайных потоков нельзя предсказать сам момент

возникновения события. В работе описан подход к анализу редких событий, который базируется на разделении событий по идентификаторам источников, в которых они образованы; на регрессии параметров процесса, происходящего внутри источников, в результате которого образуются эти события; на поиске любым известным методом закономерностей изменения параметров; на экстраполяции параметров на будущее; и, наконец, – на запуске процесса для получения прогноза моментов возникновения следующих событий. Для процессов потребления и нарастания возмущения, которые являются наиболее распространенными процессами образования событий в экономике, предложен метод восстановления скорости потребления или накопления возмущения из истории редких событий. Услуги можно моделировать как процесс накопления возмущения до определенного уровня. Статья посвящена применению емкостного метода анализа редких событий на реальных данных в сфере услуг (стрижка в парикмахерской, маникюр в салоне красоты, услуги сотовой связи). Задача заключается в том, чтобы восстановить функцию, которая приводит к событиям приобретения услуг, после чего спрогнозировать следующие события.

Ключевые слова: емкостный метод, редкие события, анализ, прогнозирование, сфера услуг, стрижка, маникюр, связь.

JEL: C1, C15, C4, C5, C53.

© Кораблев Ю.А., Голованова П.С., Кострица Т.А., 2020 г.

Кораблев Юрий Александрович, к.э.н., доцент, доцент кафедры, Финансовый университет при Правительстве РФ, Москва, Россия; yura-korablyov@yandex.ru

Голованова Полина Сергеевна, студент, Финансовый университет при Правительстве РФ, Москва, Россия; polin-mart@ya.ru

Кострица Татьяна Андреевна, студент, Финансовый университет при Правительстве РФ, Москва, Россия; 1tanyalilt@gmail.com

¹ Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект № 19-010-00154).

1. ВВЕДЕНИЕ

Анализ и прогнозирование играют большую роль в планировании развития организации, занимающейся любым видом деятельности, и сфера услуг здесь не исключение. Прогноз приобретения клиентами услуг может позволить организовать распорядок работы мастеров, определить необходимое число самих мастеров и объем их занятости, необходимое количество расходных материалов и загрузку устройств.

Процесс приобретения услуг, как правило, имеет характер редких событий. Иногда частые события отличаются от редких в зависимости от того, насколько число одних со-

бытий за наблюдаемые интервалы времени превосходит число других событий (либо как сильно различаются вероятности их возникновения). Однако все это относительно: для одних систем даже тысяча событий в секунду может быть не достаточной, чтобы такие события назывались частыми, а для других – события, которые возникают раз в несколько месяцев, по-прежнему могут являться частыми событиями. Есть и другие толкования понятия редкости событий. Нассим Талеб называет события, которых никто не ожидал (аномальные) и которые обладают огромным воздействием и заставляют человека искать объяснение их возникновению, черными лебедями (Талеб, 2010). В теории информации утверждается, что редкие события уже из-за того, что они редкие, несут большую информацию (Шеннон, 1963). Между частыми и редкими событиями сложно провести четкую грань, которая бы определяла, где заканчиваются частые события и начинаются редкие. Существует известная философская проблема – второй закон диалектики Гегеля о переходе количества в качество. Возможно, четкой грани в принципе невозможно провести, и исследователь вынужден проводить такую грань каждый раз, ориентируясь на свой субъективный опыт и здравый смысл. Именно на исследователя ложится груз ответственности за первичное формирование набора данных и выбор метода анализа этих данных, которые он же представил и структурировал определенным образом. Следует обратить внимание на следующее обстоятельство. Выбор дальнейшего метода анализа данных зависит от того, как изначально представлены эти данные, и при этом начальное представление данных зависит от того, какие методы анализа данных известны нашему исследователю. Вполне возможно, что именно знание метода является основополагающей отправной точкой ко всему исследованию. Если бы все исследователи заранее знали наилучший метод анализа данных, то и исходный набор данных был бы сформирован и представлен соответствующим образом. На практике мы обычно наблюдаем си-

туацию, когда один исследователь формирует набор данных (записывает исходные данные определенным образом), не зная, как другой исследователь будет их анализировать. В то же время этот второй исследователь становится ограниченным в выборе метода анализа из-за исходного представления данных.

2. ОБЗОР СОВРЕМЕННОГО СОСТОЯНИЯ ИЗУЧАЕМОЙ ПРОБЛЕМЫ

Иногда редкие события представляют в виде обычных временных рядов, содержащих множество нулевых значений. Обычные математические и статистические методы к анализу редких событий не подходят. Существуют специальные методы, которые позволяют анализировать редкие события, представленные временным рядом (содержащим нулевые значения). Метод Кростона (Croston, 1972) предлагает в торговле временной ряд с редкими продажами разбивать на два: ненулевых значений и интервалов времени между ненулевыми значениями, после чего экспоненциально сглаживать эти ряды и давать прогноз в виде последнего сглаженного значения. Метод Виллемейна (или бут-стреппинга) (Willemain, Park et al., 2001; Кисляков, 2019) предлагает (в логистике) получать подвыборки заданной длины из исходного временного ряда, суммировать элементы подвыборки и строить эмпирические функции распределения для суммы этих элементов, после чего определять квантиль для заданной доверительной вероятности (т.е. размер страхового запаса определяется из уровня удовлетворения спроса 95 или 99%). Метод поиска «ближайшего соседа» (Cover, Hart, 1967) в предыстории позволяет находить наиболее похожую последовательность наблюдений (на основе некоторой нормы), после чего выдает прогноз – как значение, следующее за этим «соседом». В селективных методах на основе ошибки прогноза на предшествующем шаге выбирают либо правило

расчета значений параметров метода, либо сам метод прогнозирования.

Задачу анализа редких событий иногда сводят к задаче классификации (*распознавания*). В этом случае формируют обучающий набор данных (обучающую выборку), который содержит наблюдаемые события вместе с некоторым множеством наблюдаемых признаков каждого события. В методах классификации определяют закономерность между наблюдаемыми признаками и классом событий. Задача классификации состоит в том, чтобы по наблюдаемым признакам предсказать, какой тип событий ожидается в соответствующий этим признакам период времени.

Помимо этого, редкие события в теории случайных процессов также представляют в виде Пуассоновских потоков событий или потоков Пальма, где события возникают в дискретные моменты времени через случайные интервалы времени с экспоненциальным или любым другим законом распределения. Для моделирования сверхредких событий используют модифицированные Пуассоновские процессы (Дзанагова, Хугаева, 2015). На практике чаще всего применяют классические Пуассоновские процессы. Так, в логистике из имеющихся данных редких продаж определяют интенсивность Пуассоновского потока, после чего, исходя из доверительной вероятности, рассчитывают размер запасов (Лукинский, Замалетдинова, 2015). Причем в теории случайных процессов разговора о частоте или редкости событий не ведется, интенсивность потока может быть задана как одно событие за несколько лет или как миллион событий за секунду, и все равно такое возникновение событий называют законом наступления редких событий. Нас будет интересовать именно такое представление, когда у каждого события есть свой момент времени возникновения, помимо этого, у каждого события может быть некоторое значение, отмечающее влияние этого события на мир.

Недостаток методов, основанных на представлении редких событий в виде временных рядов (содержащих подавляющее

число нулей), состоит в том, что их точность не всегда даже достигает даже 50%. При анализе данных редких событий – как задачи классификации – наблюдаемые признаки часто не предшествуют наблюдаемому событию, а наблюдаются уже после возникновения события, в результате чего невозможно предсказать появление самого такого события. Недостаток анализа редких событий с помощью Пуассоновских потоков состоит в том, что метод может дать только оценку вероятности числа событий за выбранный период времени, но не способен дать прогноз момента наступления события. Потоки Пальма всегда стационарны, прогнозирование следующих событий возможно лишь путем откладывания от предыдущего события математического ожидания интервалов времени между событиями. Использование непурассоновских потоков событий с нестационарной интенсивностью для прогнозирования будущих событий на практике и в теории не встречается. Необходимо разрабатывать новые математические и инструментальные средства анализа редких событий.

3. ОСНОВНАЯ ИДЕЯ

Попробуем задаться вопросом, почему, если мы представляем данные в виде потока событий (как в теории случайных процессов), такой поток событий мы начинаем считать потоком случайных событий? Неужели у нас совсем нет никакой информации о том, как возникают эти события? Что представляет собой такая случайность? Ведь случайность – это мера неопределенности или отсутствия знания.

Предлагаемый нами подход заключается в том, чтобы использовать всю возможную информацию о процессе образования событий. Схема предлагаемого подхода исследования редких событий представлена на рис. 1. На первом этапе (а) предполагается разделение всего набора данных событий по источникам,

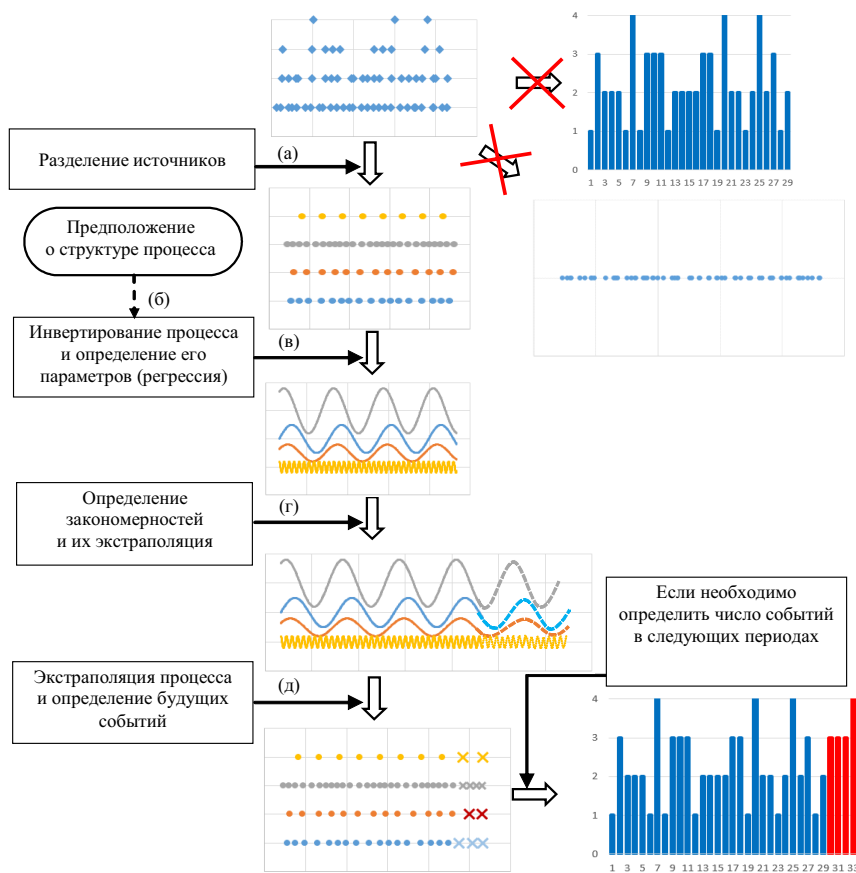


Рис. 1. Схема анализа и прогнозирования редких событий

в которых образовались эти события. Иными словами, если события образуются в разных источниках, то их надо распределить по разным выборкам, чтобы в каждой выборке оказались события, которые образовались в одном источнике. Второй этап (б), наверное, самый важный: необходимо сформулировать предположение о способе образования событий в этих источниках. Как раз на этом этапе мы используем всю доступную нам информацию, чтобы описать процесс образования событий в источниках. На следующем этапе (в) нужно из имеющихся данных восстановить параметры процесса образования событий. Получается нечто похожее на регрессионный анализ, только на этом этапе определяются не параметры математического выражения (модели),

а параметры процессов (алгоритмов). На этом этапе нужно инвертировать процесс образования событий, чтобы по имеющимся событиям получить параметры процесса. Затем (этап (г)), когда параметры процесса уже определены, необходимо произвести экстраполяцию динамики поведения этих параметров на будущее. Для этого предварительно необходимо найти закономерности в динамике изменения этих параметров, если такие изменения были. Целью этой экстраполяции параметров процесса является запуск самого процесса образования событий на следующем этапе (д) с установленными значениями параметров для получения прогноза будущих событий.

В экономике можно выделить процесс потребления или процесс накопления возму-

щения как самые распространенные процессы образования редких событий. Эти процессы можно моделировать как процессы наполнения или опустошения емкости. Такой метод анализа мы называем *емкостным методом* (Кораблев, 2017–2019). Основным параметром таких процессов является скорость накопления возмущения или расхода запаса $f(t)$. Примерами могут являться: функция ежедневного спроса у неподконтрольного нам клиента, которую мы напрямую не наблюдаем; интенсивность потребления; скорость нарастания ощущения потребности в приобретении услуги (причем не у нас, а у неподконтрольных нам клиентов, т.е. в источниках событий).

Для таких процессов (процессов потребления) можно построить обратную задачу, когда по имеющимся данным (t_i, y_i) о моментах времени и величинах воздействия события (объемах закупок) восстанавливается функция $f(t)$ (скорость потребления). Основное предположение – размер совершенного события y_i есть интеграл от функции $f(t)$ скорости изменения запаса (потребления) за время от момента этого события t_i до момента времени следующего события t_{i+1} . Однако на практике наблюдения происходят с погрешностью, $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$. В результате для восстановления неизвестной функции $f(t)$ строим оптимизационную задачу, в которой минимизируется сумма квадратов отклонений наблюдаемых значений интегралов и интегралов от искомой функции плюс штраф на нелинейность (C – произвольная большая константа):

$$\sum_{i=1}^{n-1} \left(y_i - \int_{t_i}^{t_{i+1}} f(t)dt \right)^2 + C \int_{t_1}^{t_n} (f''(t))^2 dt \rightarrow \min.$$

Данная задача решена путем нахождения особого интегрального кубического базисного сплайна, минимизирующего сумму

разницы между интегралом от функции и наблюдением. Привести решение этой задачи из-за ограниченного объема статьи не представляется возможным, решение можно найти в (Кораблев, 2020).

После восстановления функции скорости расхода запаса $f(t)$ начинается следующий этап, на котором происходят поиск закономерностей и экстраполяция их на будущее. Для этого можно использовать любые известные математические методы, любую внешнюю информацию, чтобы определить зависимость от внешних факторов. Последним этапом будет прогнозирование будущих дискретных событий, которые получаются в результате процесса потребления с заданной на предыдущем этапе функцией скорости $f(t)$.

4. РЕЗУЛЬТАТЫ

Целью данной статьи является демонстрация описанного метода для анализа и прогнозирования редких событий в сфере услуг. В отличие от торговли, где величина y_i является известным объемом покупки, нужно определить, чему будет равна величина y_i при приобретении услуги. Если услуга может быть выражена некоторым объемом, то тогда величина y_i представляет собой объем каждой приобретенной услуги. Однако услуги часто не измеряются, в этом случае в качестве величины y_i можно брать единицу, обозначающую сам факт приобретения услуги; прогнозирование упрощается, так как не нужно определять объем покупок.

Следует дополнительно прояснить, что мы анализируем события по каждому клиенту отдельно. Причем клиент приобретает одну и ту же услугу. Тем самым в анализируемой выборке у нас имеются данные только об одном клиенте и только об одном виде услуг (ни в коем случае не путать с обычными потоками событий из теории случайных процессов, когда анализируется поток всех событий

от всех возможных источников и все данные перемешаны). Рассматриваемые примеры условны и имеют в основном демонстрационный характер, мы не являемся специалистами в соответствующих прикладных областях, а лишь анализируем данные о редких событиях.

4.1. Стрижка

Имеются данные о датах стрижки одного постоянного клиента, который непрерывно стригся на протяжении 4 лет в одной и той же парикмахерской, г. Красногорска Московской области (табл. 1).

Шаг «а» (разделение источников событий) уже выполнен, так как мы анализируем данные одного клиента.

Шаг «б» (вводим предположение о процессе). Предполагаем, что процесс образования событий соответствует процессу накопления воздействия до определенного уровня, и моделируем его как емкость.

Шаг «в» (регрессия параметров процесса). При условии что объем приобретенной услуги всегда равен единице $y_i = 1$ (площадь под ступенькой), восстанавливаем скорость накопления воздействия из последовательности интегралов (рис. 2).

Функцию, которая показывает скорость возрастания потребности в стрижке, можно физически проинтерпретировать как относительную скорость роста волос. Относительной она является оттого, что выражена не

в единицах длины за день, а в долях ко всей услуге. Значение точек этой функции показывает, насколько в долях от всего объема услуги мы каждый день приближаемся к тому, чтобы вновь отправиться приобретать эту услугу. Можно также дать еще одну интерпретацию этой функции: как скорость роста желания человека подстричься.

Шаг «г» (поиск закономерностей и экстраполяция). Для запуска процесса и определения будущих событий потребуется знать значение функции в будущем, для этого производим поиск закономерностей, строим модель и производим экстраполяцию на будущее по полученной модели. На этом графике (рис. 2) можно заметить некоторую закономерность, так, в зимние месяцы и в середине лета заметен провал функции («потребность стрижься»). Для того чтобы перейти к прогнозированию, надо понять закономерность, которую можно было бы обобщить на буду-

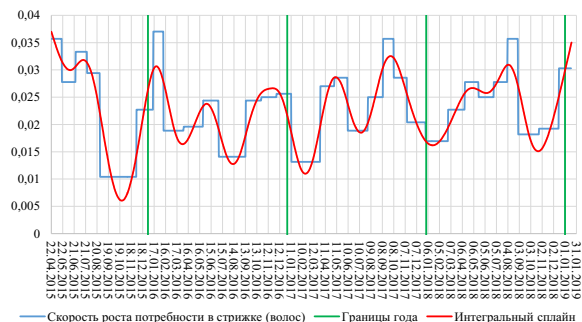


Рис. 2. Функция скорости роста волос

Таблица 1
Данные о стрижке в парикмахерской г. Красногорска

i	Дата	i	Дата	i	Дата	i	Дата
1	22.04.2015	9	04.04.2016	17	05.05.2017	25	13.04.2018
2	20.05.2015	10	25.05.2016	18	09.06.2017	26	19.05.2018
3	25.06.2015	11	05.07.2016	19	01.08.2017	27	28.06.2018
4	25.07.2015	12	14.09.2016	20	10.09.2017	28	03.08.2018
5	28.08.2015	13	25.10.2016	21	08.10.2017	29	31.08.2018
6	02.12.2015	14	04.12.2016	22	12.11.2017	30	25.10.2018
7	15.01.2016	15	12.01.2017	23	31.12.2017	31	16.12.2018
8	11.02.2016	16	29.03.2017	24	28.02.2018	32	18.01.2019

щее. В последних двух годах функция имеет два пика: в конце весны и в начале осени. Изначально было желание предположить, что скорость роста волос должна жестко зависеть от времени года, однако из рис. 2 этого не следует (вершины пиков сдвинуты по времени, расстояние между двумя этими пиками различное). В открытых источниках, посвященных скорости роста волос, говорится о том, что рост волос зависит от температуры: чем выше температура, тем больше активность организма, и, как следствие, выше скорость роста волос. Однако после того как удалось найти данные о средней дневной температуре² и наложить график средней температуры в заданном районе, это утверждение не подтверждается. Наоборот, в жаркие месяцы клиент реже ходил в парикмахерскую: похоже, весной и перед осенними похолоданиями скорость роста волос возрастала. Конечно, вполне возможно, что в жаркие летние месяцы клиент находился в отпуске и отдыхал по туристической путевке в другой стране, и, возможно, он очередной раз стригся в это время в местной парикмахерской, в результате чего в исходном наборе данных были пропущены некоторые наблюдения. Если бы такая информация была, то можно было бы скорректировать данные, назначив соответствующий объем услуги $y_i = 2$ (а не 1), т.е. за время от t_i до t_{i+1} произведено две стрижки. Однако такой информации у нас нет. К тому же в последний год в летние месяцы интервалы времени между стрижками отличаются не так сильно, чтобы можно было предположить, что пропущена еще одна услуга, но небольшой спад летом все равно заметен. Из-за недостатка информации остается сделать вывод, что скорость роста волос больше зависит от внутреннего состояния человека, например от изменения уровня гормонов и других факторов.

Определим среди имеющихся данных закономерность. Так как наблюдается колеба-

тельный процесс, можно попробовать представить функцию в виде суммы гармонических колебаний. Для экстраполяции хорошо подходит метод Куинна и Фернандеса (Quinn–Fernandes algorithm) (Quinn, Fernandes, 1991; Quinn, Hannan, 2001), который представляет исходные данные в виде суммы ограниченного числа гармонических функций, которые можно с легкостью экстраполировать. На рис. 3 пунктирной линией изображен график экстраполированных значений функции (скорости роста волос) на год вперед (предполагаем, что динамика будущего соответствует динамике в прошлом). Дополнительно отметим, что модели для экстраполяции можно строить различными способами, – необходимы специальные исследования в каждой прикладной области (не беремся утверждать, что построенный график экстраполированных значений функции будет единственно верным).

Шаг «д» (запуск процесса и прогноз будущих событий). Прогноз даты следующего обращения клиента за услугой происходит следующим образом. С самой последней даты приобретения клиентом услуги начинаем от нуля (суммарное воздействие) прибавлять значения функции с увеличением времени. Как только накопленное воздействие становится больше единицы, мы считаем, что клиент в эту дату вновь захочет обратиться за приобретением услуги, и в эту дату сбрасываем накопленное воздействие на ноль. Далее снова начинаем прибавлять значение функции, пока не полу-

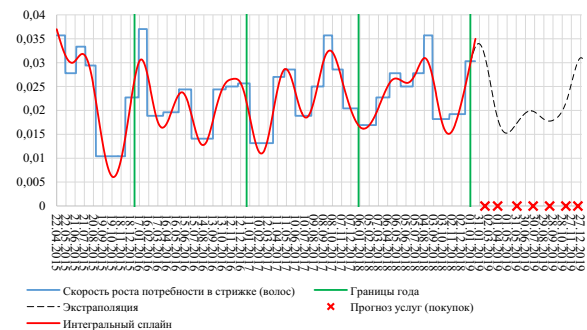


Рис. 3. Экстраполяция скорости роста волос и прогноз услуг

² Погода в Красногорске. Погодный сервис World Weather. URL: <https://world-weather.ru/pogoda/russia/krasnogorsk/> (дата обращения: 14.04.2019).

чим дату очередной покупки (можно действовать и наоборот: от единицы производить вычитание, как если бы для покупателя, а не для нас происходил обычный процесс потребления (из раздела систем управления запасами) услуги). Прогноз будущих стрижек нашего клиента изобразим крестиками на оси времени (рис. 3).

4.2. Маникюр

Имеется набор данных о маникюре постоянного клиента, который на протяжении 2 лет ходил в один и тот же салон красоты (табл. 2). Используя описанный выше подход, также осуществляя поиск закономерности в виде гармонических функций, мы получим следующий результат (рис. 4).

Большая потребность в маникюре наблюдается у клиента перед праздниками, она падает после окончания праздников. Стоит учесть, что у разных клиентов может быть разная динамика потребности в услуге. Если получится определить закономерности всех клиентов, можно предсказывать будущие мо-

менты обращения каждого, определять число обращений за месяц и т.д.

4.3. Услуги связи

Имеются данные за четыре года о пополнении баланса сотовой связи абонента, представленные в табл. 3. Пользователь услуги связи подключен к тарифу, в котором отсутствует ежемесячная или любая другая абонентская плата. Расходование средств происходит только за счет звонков и СМС-сообщений.

Восстанавливаем скорость расходования средств на связь (рис. 5). Если определить закономерность, можно спрогнозировать следующую дату пополнения баланса абонентом, также как это было сделано выше. Однако не обязательно строить модель расхода средств полностью только на прошлых данных. Можно предположить тенденцию поведения искомой функции на данных из других источников. Например, можно предположить, что стоимость всех услуг со временем всегда рас-

Таблица 2
Данные о маникюре постоянного клиента

<i>i</i>	Дата	<i>i</i>	Дата	<i>i</i>	Дата
1	19.07.2017	9	07.02.2018	17	03.10.2018
2	11.08.2017	10	04.03.2018	18	05.12.2018
3	27.09.2017	11	15.04.2018	19	29.12.2018
4	18.10.2017	12	08.05.2018	20	06.02.2019
5	07.11.2017	13	29.05.2018	21	26.02.2019
6	29.11.2017	14	17.06.2018	22	13.03.2019
7	24.12.2017	15	12.07.2018	23	02.04.2019
8	16.01.2018	16	29.08.2018	—	—

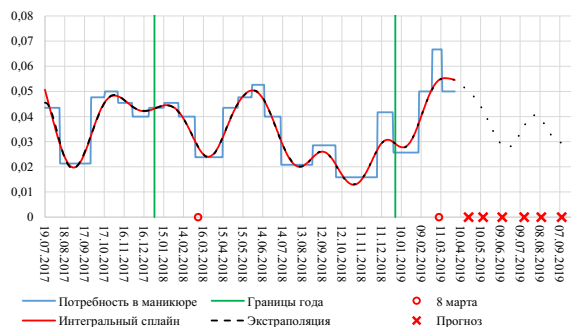


Рис. 4. Потребность в маникюре постоянного клиента

Таблица 3
Данные пополнения баланса

<i>i</i>	Дата	Пополнение	<i>i</i>	Дата	Пополнение	<i>i</i>	Дата	Пополнение
1	17.09.2016	300	4	29.05.2017	300	7	28.08.2018	300
2	19.11.2016	300	5	07.01.2018	300	8	26.11.2018	400
3	02.04.2017	300	6	03.06.2018	300	9	16.04.2019	300



Рис. 5. Предположение о росте расходов на связь клиента и прогноз услуг

тет, или абонент с развитием цифровых технологий и различных сервисов будет более активно использовать средства связи. Эти дополнительные предположения можно комбинировать с закономерностями, найденными из данных.

Стоит отметить, что этот вид обработки данных для определения скорости расходов на связь может быть не интересен, так как в личном кабинете любого абонента есть способ получить выписку о расходах на связь за определенный период. Например, в личном кабинете МТС можно получить выписку только за последние три месяца, однако при обращении в салон связи можно получить выписку за более длительный период. Получается, что восстанавливать скорость потребления из редких событий для областей, где уже развита цифровизация, не всегда актуально (клиент может обратиться и получить нужную информацию, и сам поставщик услуги может извлечь ее из своего хранилища). Однако для хранения информации о том, как меняется баланс каждый день, необходимо значительно больше объема памяти, чем для хранения данных о пополнении баланса, из которых можно восстановить динамику изменения баланса.

5. ЗАКЛЮЧЕНИЕ

Анализ редких событий – сложная задача, для решения которой обычные математические и статистические методы не подходят. В то же время многие математические методы не используют всей доступной информации, относящейся к этим редким событиям, а рассматривают их как сухие данные и абстрактные числа. Однако у нас все же есть понимание того, как образуются эти редкие события, и это понимание надо также использовать. Необходимо рассматривать редкие события с точки зрения процессов, которые формируют эти редкие события. Тогда анализ редких событий с точки зрения этих процессов превращается в восстановление параметров этих процессов. В экономике России, да и в мире в целом процессами образования редких событий могут являться процессы потребления или накопления воздействия, которые можно моделировать как процессы опустошения или наполнения емкости. Из данных редких событий можно восстановить нестационарную скорость, с которой приближался момент возникновения каждого события, для этого используется метод восстановления функции по последовательности интегралов. Благодаря рассмотренной в работе методике можно получить прогноз будущих событий, вновь запустив процесс образования событий с установленными значениями параметров. В работе показано, что в сфере услуг можно также использовать этот метод для анализа и прогнозирования событий. Для этого предполагается, что величина воздействия от приобретения услуги равна единице, после чего можно восстановить функцию потребности в услуге. Восстанавливаемая этим методом функция имеет интересную интерпретацию, например, для услуги стрижки волос функция может интерпретироваться как скорость роста волос, для услуг маникюра – как скорость роста ногтей, для услуг связи – как скорость изменения баланса. На примерах показано, как получить прогноз будущих услуг.

Данный метод находится в начале развития, но уже сейчас можно заметить его перспективы. Однако для успешного применения этого метода необходимо, чтобы на начальном этапе сбора данных информация о событиях записывалась раздельно. Не должно быть агрегирования, когда все события сваливаются в одну кучу (например, формируя временной ряд или суммарный поток событий). При соблюдении этих условий представленный метод будет крайне эффективным средством исследования.

Список литературы / References

- Дзанагова И.Т., Хугаева Л.Т. (2015). Информационно-статистические методы построения экстремальных моделей редких событий // *Фундаментальные исследования*. Академия Естествознания (Пенза). № 11. Ч. 6. С. 1081–1084. [Dzanagova I. T., Khugaeva L. T. (2015). Information-statistical methods for constructing extremal models of rare events. *Fundamental Research. Academy of Natural History* (Penza), part 6, no. 11, pp. 1081–1084 (in Russian).]
- Кисляков А.Н. (2019). Метод виртуального увеличения выборки при прогнозировании редких продаж в условиях информационной асимметрии // *Вестник Алтайской академии экономики и права*. Т. 2. № 1. С. 47–54. [Kislyakov A. N. (2019). Method of virtual increase in the sample when predicting rare sales in conditions of information asymmetry. *Vestnik of Altai Academy of Economics and Law*, vol. 2, no. 1, pp. 47–54 (in Russian).]
- Кораблев Ю.А. (2017). Емкостный метод анализа редких продаж в Excel. Экономика и управление: проблемы, решения. Т. 3 (66). № 6. С. 224–230. [Korablev Yu. A. (2017). Capacity method for analyzing rare sales in Excel. *Ekonomika i Upravlenie: Problemy I Resheniya*, vol. 3 (66), no. 6, pp. 224–230 (in Russian).]
- Кораблев Ю.А. (2019). Погрешность емкостного метода анализа редких событий, удаленность от конечного потребителя // *Известия Кабардино-Балкарского научного центра (КБНЦ) РАН*. № 3 (89). С. 48–77. [Korablev Yu. A. (2019). Error of the capacity method of rare events analysis, remoteness from the end user. *The News of K.-B. Sc. Center of RAS*, no. 3 (89), pp. 48–77 (in Russian).] DOI: 10.35330/1991-6639-2019-3-89-48-77.
- Кораблев Ю.А. (2020). Метод восстановления функции по интегралам для анализа и прогнозирования редких событий в экономике // *Экономика и математические методы*. Т. 56. № 3. С. 114–125. [Korablev Yu. A. (2020). The function restoration method by integrals for analysis and forecasting of rare events in the economy. *Economics and Mathematical Methods*, vol. 56, no. 3, pp. 114–125 (in Russian).]
- Лукинский В., Замалетдинова Д. (2015). Методы управления запасами: расчет показателей запаса для товарных групп, относящихся к редким событиям (часть I) // *Логистика*. № 1 (98). С. 28–33. [Lukinsky V., Zamaletdinova D. (2015). Methods of inventory management: the calculation of inventory indicators for product groups related to rare events (Part I). *Logistics*, no. 1 (98), pp. 28–33 (in Russian).]
- Лукинский В., Замалетдинова Д. (2015). Методы управления запасами: расчет показателей запаса для товарных групп, относящихся к редким событиям (часть II) // *Логистика*. № 2 (99). С. 24–27. [Lukinsky V., Zamaletdinova D. (2015). Methods of inventory management: the calculation of inventory indicators for product groups related to rare events (Part II). *Logistics*, no. 2 (99), pp. 24–27 (in Russian).]
- Тaleb Н. (2010). Черный лебедь. Под знаком непредсказуемости: сборник. М.: Азбука-Аттикус. [Taleb N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. N.Y., The New York Times.]
- Шеннон К. (1963). Работы по теории информации и кибернетике: пер. с англ. М.: Иностранная литература. 832 с. [Shannon C. (1963). *Works on information theory and cybernetics*. Moscow (IL): Science and Education. Moscow, Inostrannaya literatura (in Russian).] URL: <http://www.sciepub.com/reference/5832>

- Claude E. (1949). Shannon and Warren Weaver: The Mathematical Theory of Communication. Urbana: The University of Illinois Press.
- Cover T., Hart P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, vol. 13 (1), pp. 21–27.
- Croston J.D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly* (1970–1977), vol. 23 (3), pp. 289–303.
- Quinn B.G., Fernandes J.M. (1991). A Fast Efficient Technique for the Estimation of Frequency. *Биометрика*, vol. 78, no. 3, September, pp. 489–497.
- Quinn B.G., Hannan E.J. (2001). The Estimation and Tracking of Frequency. Cambridge: Cambridge University Press. 278 p.
- Willemain T.R., Park D.S., Kim Y.B., Shin K.I. (2001). Simulation output analysis using the threshold bootstrap. *European Journal of Operational Research*, vol. 134 (1), pp. 17–28.

Рукопись поступила в редакцию 10.03.2020 г.

CAPACITY METHOD OF RARE EVENTS ANALYSIS IN THE AREA OF SERVICES

*Yu.A. Korablev, P.S. Golovanova,
T.A. Kostritsa*

DOI: 10.33293/1609-1442-2020-3(90)-132-142

Yuri A. Korablev, PhD, Financial University under the Government of the Russian Federation, Moscow, Russia; yura-korablyov@yandex.ru;

Polina S. Golovanova, Financial University under the Government of the Russian Federation, Moscow, Russia; polinmart@ya.ru

Tatyana A. Kostritsa, Financial University under the Government of the Russian Federation, Moscow, Russia; 1tan-yalilt@gmail.com

Acknowledgment. The study was funded by Russian Foundation for Basic Research according to the research, project № 19-010-00154.

Imagine that you are owner of some service. You need to determine for a certain future period the work plan of your craftsmen, the number of consumables needed. To do this, you need to make a forecast of future services number. Classical mathematical methods of working with time series are not suitable for this task. Aggregation of data on services by months and the compilation of a time series can only confuse. Forecasting services should be performed using methods designed to work with rare events. Rare events are devoted to relatively few works. Methods for the study of rare events are significantly less than methods for analyzing frequent events (time series). The most popular method of studying rare events at the moment is the use of the theory of random processes, which uses a stream of Poisson or Erlang events. However, using random streams, one cannot predict the very moment of the occurrence of an event. The paper describes an approach to the rare events analysis, which is based on: dividing events by identifiers of the sources in which they are formed; regression process parameters occurring within the sources, resulting in these events formation; search by any known method of parameters change patterns; the process start itself to obtain a forecast of the following events time occurrence. For the consumption processes and the disturbances growth process, which are the most common processes of the events formation in the economy, a method is proposed for restoring the consumption or accumulating disturbances rate from the rare events history. Services as can be modeled as the process of accumulating disturbances to a certain level. The article is devoted to the application of the capacity method of rare events analysis on real data in the service sector (haircut in a hairdresser, a manicure in a beauty salon, cellular communication services). The task is to restore the function that leads to the acquisition of services, and then predict the following events.

Keywords: capacity method, rare events, analysis, forecasting, services, haircut, manicure, communication.

JEL: C1, C15, C4, C5, C53.

Manuscript received 10.03.2020