



Munich Personal RePEc Archive

Practical Methods for Predicting Customer Retention

, and , and , and ,

«2 »

22 October 2024

Online at <https://mpa.ub.uni-muenchen.de/122483/>
MPRA Paper No. 122483, posted 02 Nov 2024 01:38 UTC

УДК 338.47:621.395.1:519.246.8:658.8

Черкашин Александр Михайлович

ООО «Т2 Мобайл», Москва, Россия

E-mail: a.cherkashin@t2.ru

Сахаджи Владислав Евгеньевич

ООО «Т2 Мобайл», Москва, Россия

E-mail: vladislav.sakhadzhi@t2.ru

Гулиев Руслан Чингизович

ООО «Т2 Мобайл», Москва, Россия

E-mail: ruslan.guliev@t2.ru

Большунова Елена Васильевна

ООО «Т2 Мобайл», Москва, Россия

E-mail: elena.bolshunova@t2.ru

Практические методы прогнозирования сохранения клиентской базы (перевод на русский язык)

Аннотация. Этот документ является переводом статьи «Practical Methods for Predicting Customer Retention» опубликованной на MPRA (<https://mpra.ub.uni-muenchen.de/id/eprint/122400>). Оригинальная статья была опубликована 15.10.2024. В статье рассматриваются методы анализа и прогнозирования сохранения активной абонентской базы в телекоммуникационной отрасли с использованием различных критериев активности абонентов. Результаты исследования показывают, что динамика сохранения активной абонентской базы достаточно точно аппроксимируется убывающей степенной функцией. Это позволяет строить среднесрочные прогнозы по усеченной базе начальных данных об активности абонентов. Вместе с тем нельзя не отметить потенциальные ограничения эффективности применения предложенного подхода для долгосрочного прогнозирования, связанные с изменением динамики оттока абонентов во времени.

Ключевые слова: абонентская база, удержание абонентов, отток абонентов, степенная функция, телекоммуникации, LTV.

Введение

Известные исследования в области прогнозирования оттока абонентов в основном фокусируются на использовании методов машинного обучения и анализа больших данных (Verhelst et al., 2021; Verbeke et al., 2012). Эти подходы часто требуют значительных вычислительных ресурсов и большого объема данных. При этом они, как правило, направлены на выявление факторов влияющих на склонность абонентов к оттоку (Ribeiro et al., 2023; Jain et al., 2021) или на предсказание поведения конкретных абонентов с целью воздействия на них. Данное исследование сфокусировано на анализе поведения групп абонентов, а также на величине обратной оттоку: сохранении активной абонентской базы или, заимствуя термин из биологии, на выживаемости абонентов. Исследование демонстрирует, что сохранение активной абонентской базы соответствует степенному закону. Это позволяет прогнозировать выживаемость группы абонентов, что в том числе необходимо для задач экономического моделирования и оценки совокупного дохода от абонента – LTV (Lifetime value) (Gupta & Zeithaml, 2006). Перед тем как переходить к моделированию сохранения активной абонентской базы необходимо обсудить практические вопросы выбора критериев активности абонентов и построения кривой выживаемости.

1. Признаки активности абонента

Анализ и прогнозирование сохранения активной абонентской базы во многом зависит от выбора критериев, по которым определяется активность абонентов. В данной главе будут рассмотрены различные подходы к определению активности абонентов и их влияние на оценку выживаемости базы. Предлагается следующая классификация критериев активности:

1. юридические критерии;
2. финансовые критерии;
3. критерии фактического использования услуг.

Каждая из этих категорий имеет свои преимущества и ограничения, которые будут подробно рассмотрены в последующих разделах.

1.1. Юридические критерии активности

Юридические критерии активности основаны на формальном статусе договорных отношений между компанией и абонентом. Основным показателем в этой категории является наличие действующего договора об оказании услуг связи. В сфере массового оказания услуг активность клиента и наличие действующего договора с ним, как правило, мало связаны. Наблюдается значительная задержка между фактическим прекращением использования услуг абонентом и юридическим расторжением договора. Многие операторы мобильной связи, включая Т2, предусматривают автоматическое расторжение договора после длительного периода неактивности. Например, согласно условиям Т2, договор расторгается, если абонент не совершает платных действий в течение 180 дней при нулевом или отрицательном балансе. В некоторых сферах, таких как розничная торговля или разовые услуги, юридические отношения могут оформляться для каждой отдельной транзакции, что делает этот

критерий неприменимым для долгосрочного анализа выживаемости. Соответственно, использование юридических критериев для анализа выживаемости абонентской базы имеет существенные ограничения.

Несмотря на ограниченную применимость в телекоммуникационной отрасли, юридические критерии могут быть релевантны в других секторах экономики или правовых системах, где формальный статус договора более тесно связан с фактическим использованием услуг.

1.2. Финансовые критерии активности

Финансовые критерии активности основаны на различных аспектах финансового взаимодействия между абонентом и оператором связи, включая генерацию выручки, платежи абонента, состояние баланса и связанные с обслуживанием затраты. Использование финансовых показателей в качестве критериев активности имеет ряд очевидных преимуществ. Прежде всего, они напрямую отражают экономическую ценность абонента для компании, что особенно важно с точки зрения экономического моделирования.

Однако, несмотря на это преимущество, использование финансовых критериев для оценки выживаемости абонентской базы сопряжено с рядом существенных ограничений. Одним из ключевых факторов, влияющих на достоверность финансовых критериев, является зависимость от учетной политики компании. Моменты начисления выручки могут значительно варьироваться в зависимости от принятых в организации правил. Например, при авансовом платеже за годовое обслуживание выручка может быть начислена единовременно или распределена на весь оплаченный период, что покажет совершенно разную активность абонента. Такая оплата услуг на год вперед может создать, как ложную постоянную активность, даже если абонент фактически прекратил пользоваться услугами через несколько месяцев после оплаты, так и, наоборот, показать полное отсутствие активности за пределами месяца аванса. Ни то, ни другое не даст достоверной информации о реальном поведении абонента, что не позволит использовать эти данные для прогнозирования. Это актуально в том числе в контексте различных маркетинговых акций и специальных предложений, которые могут стимулировать абонентов к долгосрочным предоплатам.

Отдельного внимания заслуживает проблема остаточных средств на счетах неактивных абонентов. Если у абонента, переставшего пользоваться услугами, остались средства на счете, их постепенное списание в выручку может показывать ложную активность. Это явление способно привести к систематическому искажению данных о выживаемости абонентской базы.

Пополнение баланса, как потенциальный критерий активности, с одной стороны, не зависит от учетной политики компании, что является его несомненным преимуществом. Однако он тоже не обязательно синхронизирован с реальным использованием услуг: абонент может внести крупную сумму и не пользоваться услугами длительное время, что приведет к искажению оценки его активности.

Использование затрат в качестве критерия активности сталкивается с теми же проблемами, что и использование выручки, но усугубляется более сложной системой учета и распределения затрат. Более того, затраты обычно связаны с использованием услуг, которое эффективнее измерять напрямую, о чем будет подробнее

рассказано в следующей главе, посвященной критериям фактического использования услуг.

Учитывая все вышеизложенное, можно заключить, что несмотря на важность финансовых показателей для экономической оценки абонентской базы, их использование в качестве критериев активности абонента требует крайне осторожного подхода. Эти ограничения могут привести к значительным искажениям в анализе и прогнозировании динамики сохранения активной абонентской базы. Для преодоления описанных ограничений рекомендуется тщательно анализировать учетную политику компании перед использованием финансовых критериев. При интерпретации финансовых показателей активности абонентов необходимо учитывать специфику бизнес-модели и особенности предоставления услуг. Финансовые критерии, хотя и важны для общей оценки экономической эффективности абонентской базы, не могут считаться универсальным и достаточным инструментом для анализа выживаемости абонентов. Тем не менее, финансовые критерии могут быть эффективны для решения узконаправленных задач, например, анализ оплачиваемости абонентской платы в тарифах с фиксированной ежемесячной оплатой.

1.3. Критерии фактического использования услуг

Критерии фактического использования услуг представляют собой наиболее объективный и непосредственный способ оценки активности абонентов в телекоммуникационной отрасли. В отличие от юридических и финансовых критериев, рассмотренных ранее, эти показатели основаны на реальном потреблении услуг связи, что обеспечивает высокую точность и актуальность данных о поведении абонентов. Основное преимущество данных критериев заключается в их независимости от учетной политики компании и других управленческих факторов. Использование услуг связи создает непосредственную нагрузку на сетевую инфраструктуру, отражая реальные физические процессы, которые фиксируются в момент их возникновения. Это позволяет получать актуальные данные о поведении абонентов без временных задержек, характерных для финансовых показателей.

В контексте оператора мобильной связи критерии фактического использования охватывают широкий спектр услуг, включая голосовые вызовы, передачу данных и обмен текстовыми сообщениями. Однако не все виды сетевой активности одинаково релевантны для оценки выживаемости абонента. Например, технический трафик, генерируемый устройством для поддержания связи с сетью, не может рассматриваться как показатель активности абонента (хотя и свидетельствует об активности устройства). При разработке системы критериев фактического использования услуг необходимо учитывать специфику предоставляемых сервисов и особенности их потребления.

Несмотря на очевидные преимущества, критерии фактического использования услуг тоже имеют определенные ограничения. Основным недостатком является потенциальный разрыв между фактическим использованием услуг и их экономической ценностью для компании. Абонент может активно пользоваться услугами, включенными в пакет, не генерируя дополнительной выручки. С другой стороны, неактивный по этим критериям абонент может регулярно оплачивать услуги, что важно с финансовой точки зрения. Например, абонент может иметь положительный баланс и регулярно вносить плату, но при этом не пользоваться услугами по различ-

ным причинам. Такая ситуация может возникать, например, при наличии у абонента нескольких SIM-карт от разных операторов или при сохранении номера, например, для регистрации на онлайн-сервисах. Для преодоления этого ограничения можно, например, оценивать фактическую разницу между абонентами «платящими» и «использующими услуги», и учитывать ее в моделировании. Другой способ – использовать комбинированный подход, сочетающий критерии фактического использования услуг и финансовые критерии.

Выбор конкретного критерия или комбинации критериев должен осуществляться с учетом специфики бизнеса, целей анализа и особенностей исследуемой абонентской базы. Универсального решения, подходящего для всех ситуаций, не существует, и каждой компании необходимо разработать свой подход к определению и измерению выживаемости абонентов. В рамках этой статьи мы будем использовать два специально разработанных критерия активности абонента:

1. Критерий А – учитывает только некоторое фактическое потребление услуг абонентом.
2. Критерий В – комбинированный, включает как фактическое потребление услуг, так и финансовые признаки активности абонента.

2. Методы расчета и построения кривой выживаемости

Кривая выживаемости абонентов – термин и инструмент, заимствованный из биологии и адаптированный для нужд бизнеса, позволяет визуализировать и количественно оценить динамику сохранения активной абонентской базы во времени. Она представляет собой графическое отображение доли абонентов, остающихся активными в течение определенного периода времени после подключения. В контексте телекоммуникационного бизнеса «выживаемость» означает продолжение использования услуг компании абонентом. Построение и анализ кривых выживаемости служат двум основным целям: диагностика текущего состояния и создание основы для построения прогнозных моделей.

Построить кривую выживаемости можно двумя разными методами. Далее эти методы будут рассмотрены подробнее, определены их преимущества и недостатки, а также их связь между собой.

2.1. Кривая выживаемости фиксированной группы абонентов

По первому методу рассчитывается и строится кривая выживаемости фиксированной группы абонентов. Метод основан на наблюдении за фиксированной группой абонентов, подключившихся в определенный момент времени и позволяет отслеживать динамику активности одной и той же группы пользователей на протяжении всего периода наблюдения, представляя собой когортный анализ (Fader & Hardie, 2009; Zhang & Chang, 2021).

Процесс построения кривой выживаемости абонентов включает следующие этапы:

1. определение начальной группы абонентов, подключившихся в выбранный период времени (обычно это календарный месяц);
2. отслеживание активности этой группы в соответствии с выбранным критерием активности на протяжении исследуемого периода;

3. расчет доли активных абонентов от их начального количества для каждого временного интервала.

Математически выживаемость для каждого периода может быть выражена следующей формулой:

$$R_i = \frac{A_i(T, c)}{G_T} \quad (1)$$

где:

- R_i – выживаемость за i -й период;
- $A_i(T, c)$ – количество активных (в соответствии с критерием активности c) абонентов в периоде i из подключений периода T ;
- G_T – количество подключений в период T ;
- $i \in [1; +\infty]$ – период для расчета выживаемости;
- T – фиксированный начальный период (обычно период подключения);
- c – критерий активности абонента.

Для иллюстрации данного метода рассмотрим на Рисунке 1 пример расчета выживаемости абонентов по критерию А для группы абонентов, подключившейся к сети Т2 в апреле 2023 года:

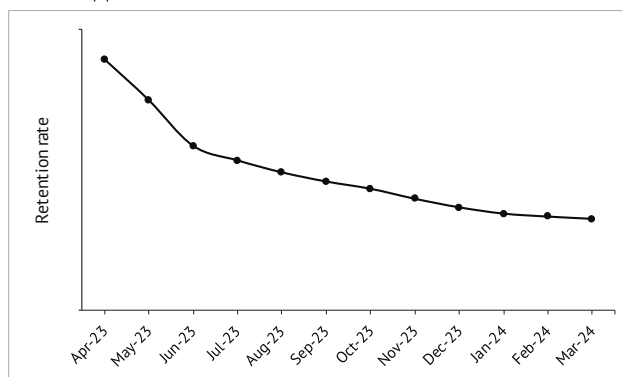


Рисунок 1. Кривая А-выживаемости абонентов, подключившихся в апреле 2023 года. Источник: разработано авторами на основе данных об активности абонентов Т2

Вертикальная ось должна иметь диапазон от 0 до 1 (от 0% до 100%) и, обычно, представляется в линейном масштабе. В рамках статьи для предотвращения раскрытия фактических данных Компании на этом и следующих рисунках намеренно изменен диапазон и убраны подписи значений, линейный масштаб оси сохраняется. С учетом вышеизложенного, Рисунок 1 можно читать так, что, например, в течение первого месяца (апрель 2023 года) активной по критерию А оставалась доля подключившихся абонентов близкая к 100%, в течение второго месяца – к 90%, в третьем месяце – допустим около 75% и так далее.

Важно отметить, что для построения кривой выживаемости за определенный период, например, 12 месяцев, необходимо иметь данные о группе абонентов, подключившихся как минимум за 12 месяцев до момента анализа. При этом выживаемость на первый месяц для этой кривой будет рассчитана по данным активности абонентов 12 месяцев назад.

2.2. Кривая выживаемости по когортам абонентов

Альтернативным методом является построение кривой выживаемости по когортам абонентов. Когорта – это группа абонентов, объединенных по времени подключения. Возраст когорты измеряется в полных календарных месяцах с момента подключения. Соответственно, этот подход основан на анализе текущей выживаемости групп абонентов разных возрастов.

Процесс построения кривой выживаемости по когортам включает следующие этапы:

1. разделение текущих активных абонентов на группы (когорты) по «абонентскому возрасту», например, выраженному в месяцах;
2. расчет выживаемости для каждой когорты, начиная с самой молодой;
3. построение кривой, где каждая точка представляет выживаемость соответствующей когорты на текущий момент.

Для иллюстрации рассмотрим пример:

- Когорта 1 (возраст 1 месяц): абоненты, подключившиеся месяц назад. Выживаемость рассчитывается как доля активных абонентов этой когорты в общем количестве подключений месяц назад.
- Когорта 2 (возраст 2 месяца): абоненты, подключившиеся два месяца назад. Выживаемость рассчитывается аналогично, но относительно подключений двухмесячной давности.

Математически выживаемость для каждой когорты может быть выражена похожей формулой:

$$R_j = \frac{A(j, c)}{G(P(j))} \quad (2)$$

где:

- R_j – выживаемость j -ой когорты абонентов;
- $A(j, c)$ – количество активных абонентов j -ой когорты, в соответствии с критерием активности c ;
- $G(P(j))$ – количество абонентов, подключившихся в период $P(j)$;
- $P(j)$ – период подключения, соответствующий когорте j ;
- $j \in [1; +\infty]$ – номер когорты;
- c – критерий активности абонента.

Обращаем внимание, что в отличие от кривой выживаемости фиксированной группы абонентов здесь для каждого периода рассчитывается текущая выживаемость разных групп абонентов, подключившихся в разные периоды. Значение знаменателя меняется для каждой когорты j .

2.3. Взаимосвязь двух кривых выживаемости

Кривая выживаемости по когортам может быть представлена как комбинация значений из нескольких кривых выживаемости. Схематично эта связь изображена на Рисунке 2:

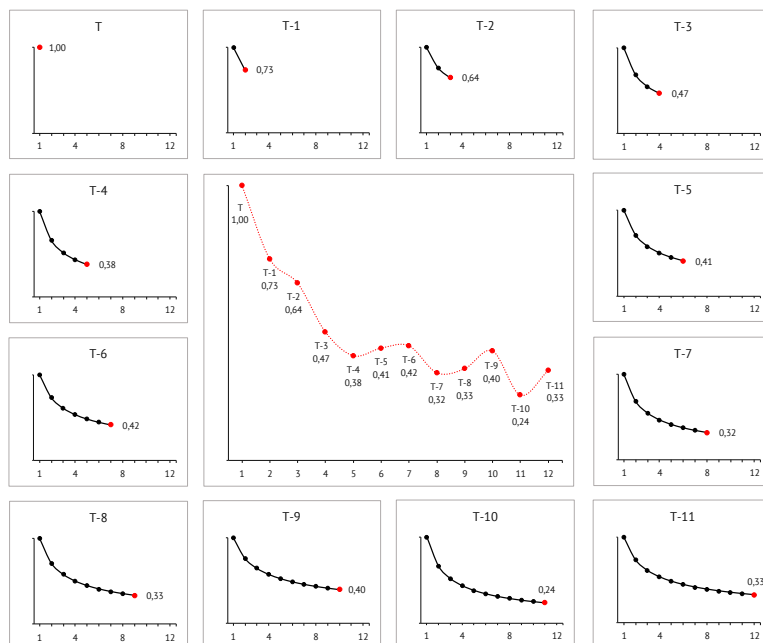


Рисунок 2. Связь между кривой выживаемости по когортам и кривыми выживаемости фиксированных групп. Источник: данные смоделированы случайным образом для иллюстрации.

На этой схеме Т представляет текущий (или последний доступный для анализа) месяц. Малые графики показывают кривые выживаемости для фиксированных групп абонентов, подключившихся в разные периоды (от Т до Т-11). Большой график посередине представляет кривую выживаемости по когортам, где каждая точка соответствует последней точке соответствующей кривой выживаемости: первая точка (Т – выживаемость на первый месяц) соответствует последней точке малого графика Т, вторая точка (Т-1 – выживаемость на второй месяц) соответствует последней точке малого графика Т-1 и так далее. Таким образом, кривая выживаемости по когортам состоит из последовательности значений выживаемости на первый, второй и последующие месяцы жизни разных групп абонентов, наблюдаемых в текущий момент времени.

Обратим внимание, что кривая выживаемости по когортам, в общем, повторяет динамику выживаемости кривых, из точек которых она состоит, при этом, динамика «зашумлена» различными факторами, связанными с тем, что это выживаемость разных групп абонентов.

2.4. Сравнительный анализ и выбор метода построения кривой выживаемости

Экономика – область знаний, вынужденная действовать в постоянно меняющихся внешних условиях. Эмпирические законы и зависимости из области экономики непостоянны. Данные и модели имеют тенденцию к быстрому устареванию, теряя актуальность. Это вынуждает постоянно искать баланс между количеством данных и их актуальностью (по мнению авторов данного исследования, это аналог известной

в статистике и машинном обучении дилеммы смещения-дисперсии – bias-variance tradeoff). С одной стороны, использование большего объема исторических данных может повысить статистическую устойчивость модели. С другой стороны, более старые данные могут не отражать текущие рыночные реалии, что приводит к увеличению смещения оценок. На наш взгляд, при выборе метода построения кривой выживаемости следует руководствоваться целью анализа (оперативный контроль, или долгосрочное прогнозирование, или исследовательские задачи) и спецификой бизнеса и рынка, на котором он работает.

Кривая выживаемости фиксированной группы абонентов отражает выживаемость одной и той же группы подключений на протяжении всего периода наблюдения. Кривая всегда строго невозрастающая, что соответствует логике процесса оттока абонентов. При этом она обладает низкой оперативностью: для построения долгосрочной кривой требуется длительный период наблюдения. Как следствие, к моменту получения полной кривой данные о выживаемости в первые периоды могут потерять актуальность.

Для задач оперативной оценки необходимо видеть, с одной стороны, максимально актуальные данные, а с другой стороны – общую картину в целом. Эту задачу решает кривая выживаемости по когортам. Она отражает текущее состояние абонентской базы на разных стадиях жизненного цикла. На ней представлены всегда самые последние доступные данные для каждого периода жизни. Но из-за того, что кривая выживаемости по когортам построена на основе разных групп абонентов, она подвержена высоким колебаниям, связанным с сезонностью и всеми другими факторами, влияющими на качество и состав подключений в разные периоды. Это создает, в том числе, возможность локального возрастания кривой, что контринтуитивно, так как противоречит логике процесса оттока.

В более теоретических исследованиях, в исследованиях влияния различных факторов на выживаемость базы или в исследованиях сравнения разных каналов подключений между собой, кривая выживаемости фиксированной группы абонентов является предпочтительной. Она обеспечивает более чистую картину динамики выживаемости конкретной группы абонентов, что может быть критически важно для выявления фундаментальных закономерностей и тенденций. Так как это исследование посвящено разработке методов прогнозирования сохранения абонентской базы, далее в статье будет использоваться кривая выживаемости фиксированной группы абонентов (в дальнейшем, просто – «кривая выживаемости»). В практической работе, при построении самих прогнозов или мониторинге качества подключений выбор может быть сделан в пользу кривой выживаемости по когортам.

Необходимо коротко упомянуть про построение кривых выживаемости по усредненным данным, что так же часто используется в практической работе. В условиях наличия сезонных колебаний и рыночных аномалий, усреднение по нескольким периодам позволяет сгладить динамику и исключить влияние разовых факторов. При этом, применение усреднения неизбежно приводит к использованию более старых данных, что может негативно сказаться на актуальности анализа.

3. Соответствие сохранения базы степенному закону

Для выявления устойчивых закономерностей в поведении абонентов желательно исследовать кривые выживаемости на значительных временных интервалах. В данном исследовании была проанализирована кривая выживаемости абонентов, подключившихся к сети Т2 в ноябре 2015 года, с последующим наблюдением их активности по критерию В на протяжении 105 месяцев (8,5 лет) до июля 2024 года (Рисунок 3).

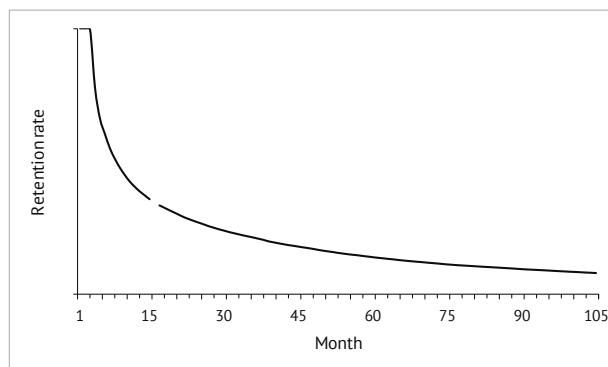


Рисунок 3. В-выживаемость абонентов, подключившихся к Т2 в ноябре 2015 года.

Источник: разработано авторами на основе данных об активности абонентов Т2

Примечание: Данные по активным абонентам за февраль 2017 года отсутствуют, что выглядит как пропуск на графике.

Визуальный анализ графика демонстрирует монотонный и гладкий характер кривой выживаемости, что позволяет выдвинуть гипотезу о возможности ее аппроксимации относительно простой математической функцией.

Для проверки выдвинутой гипотезы были рассмотрены четыре типа функций:

- степенная: ax^b ;
- экспоненциальная: ae^{bx} ;
- логарифмическая: $a\ln(x) + b$;
- линейная: $ax + b$.

Нелинейные функции были приведены к линейному виду, параметры функций определялись методом наименьших квадратов (МНК). Результаты аппроксимации представлены на Рисунке 4 и в Таблице 1:

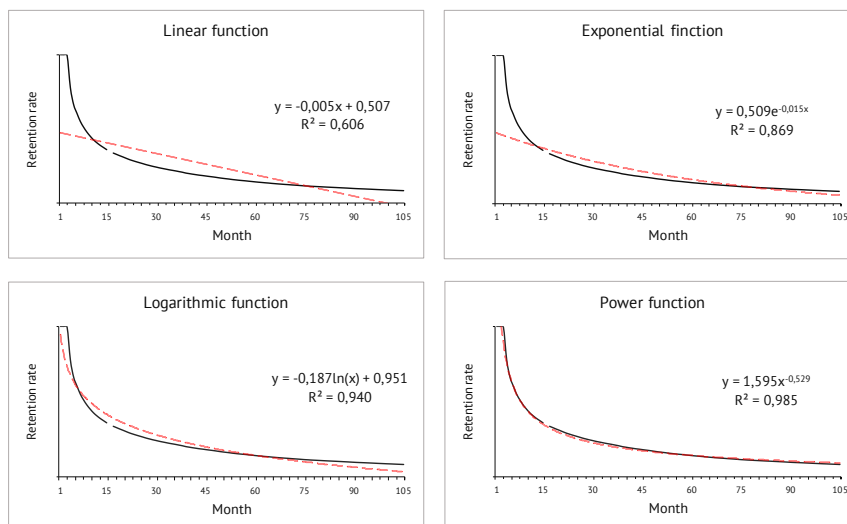


Рисунок 4. Аппроксимация кривой В-выживаемости различными функциями. Источник: разработано авторами на основе данных об активности абонентов Т2

Анализ результатов показывает, что степенная функция демонстрирует качественную и наилучшую аппроксимацию кривой выживаемости с коэффициентом детерминации (R^2) = 0,985.

Таблица 1

Характеристика точности аппроксимации выживаемости различными функциями

Function type	Function	R^2
Power	$y = 1,595x^{-0,529}$	0,985
Logarithmic	$y = -0,187\ln(x) + 0,951$	0,940
Exponential	$y = 0,509e^{-0,015x}$	0,869
Linear	$y = -0,005x + 0,507$	0,606

Составлено авторами на основе данных об активности абонентов Т2

Для верификации полученных результатов была проведена аппроксимация кривых выживаемости по критерию А для различных групп абонентов, подключившихся к сети компании в разные периоды и через разные каналы продаж. Результаты представлены на Рисунке 5 и в Таблице 2:

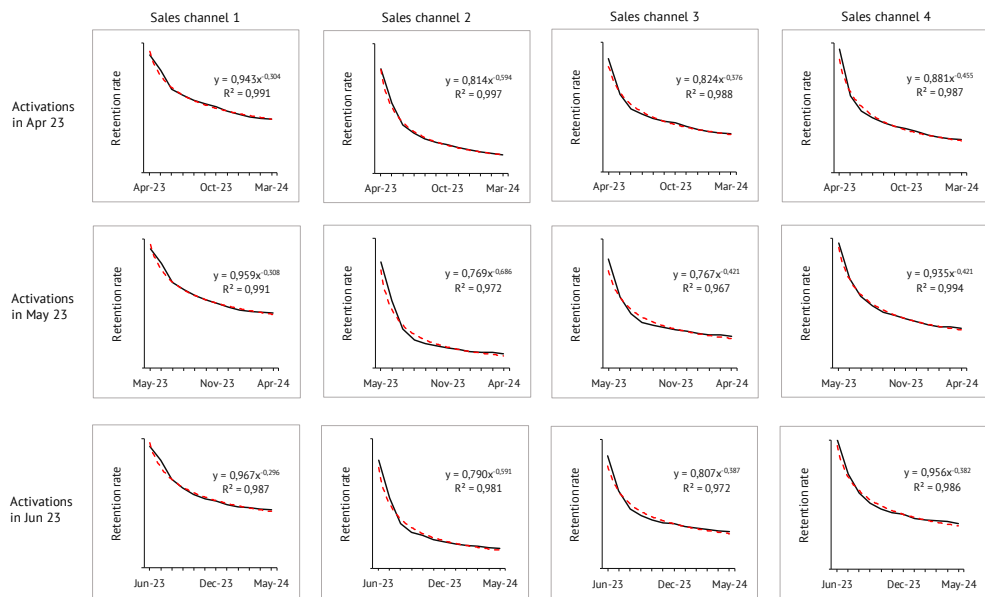


Рисунок 5. Аппроксимация кривых А-выживаемости степенной функцией для различных каналов продаж и периодов подключения. Источник: разработано авторами на основе данных об активности абонентов Т2

Анализ данных Таблицы 2 подтверждает высокую устойчивость результатов аппроксимации степенной функцией. Во всех рассмотренных случаях R^2 превышает 0,96, что свидетельствует о хорошем соответствии модели эмпирическим данным независимо от канала продаж и периода подключения абонентов.

Таблица 2

Характеристики точности аппроксимации степенной функцией для различных каналов и периодов продаж

Sales channel	Activation month	Function	R^2
1	Apr'23	$y = 0,943 x^{-0.304}$	0,991
	May'23	$y = 0,959 x^{-0.308}$	0,991
	Jun'23	$y = 0,967 x^{-0.296}$	0,987
2	Apr'23	$y = 0,814 x^{-0.594}$	0,997
	May'23	$y = 0,769 x^{-0.686}$	0,972
3	Jun'23	$y = 0,790 x^{-0.591}$	0,981
	Apr'23	$y = 0,824 x^{-0.376}$	0,988
	May'23	$y = 0,767 x^{-0.421}$	0,967
4	Jun'23	$y = 0,807 x^{-0.387}$	0,972
	Apr'23	$y = 0,881 x^{-0.455}$	0,987
	May'23	$y = 0,935 x^{-0.421}$	0,994
	Jun'23	$y = 0,956 x^{-0.382}$	0,986

Составлено авторами на основе данных об активности абонентов Т2

Полученные результаты позволяют сформулировать следующие наблюдения и гипотезы:

1. Универсальность степенной функции: высокая точность аппроксимации кривых выживаемости степенной функцией для различных критериев активности, сегментов абонентской базы и периодов времени может указывать на наличие фундаментальной закономерности в процессе сохранения и оттока абонентов.
2. Зависимость от начальных условий: вариации параметров аппроксимирующей степенной функции для разных групп абонентов могут отражать исключительно влияние начальных условий (канал продаж, период подключения) на долгосрочную динамику выживаемости.
3. Устойчивость к внешним факторам: сохранение формы кривых выживаемости, несмотря на влияние большого количества различных меняющихся факторов (качество услуг, активности по удержанию абонентов, активность конкурентов по привлечению базы). Возможно, это говорит о каком-то динамическом равновесии конкурентного рынка между усилиями компании по удержанию абонентов и факторами, способствующими их оттоку.
4. Степенные функции широко встречаются при описании закономерностей в самых разных областях окружающего мира: физике, астрономии, биологии, геологии, социологии, лингвистике, психологии, экономике и др. (Andriani & McKelvey, 2007). В исследовании М. Ньюмана (Newman, 2005) рассматриваются степенные законы в таких областях, как физика, биология и экономика, а также их проявления в виде распределения Парето и законе Ципфа. Степенные функции описывают следующие закономерности в экономике и финансах: распределение доходов и богатства, размеры компаний, доходность фондового рынка, объемы торгов, показатели международной торговли и др. (Gabaix, 2009). Преобладание степенной зависимости в описании выживаемости абонентов поднимает ряд фундаментальных вопросов о природе процессов, определяющих поведение абонентов. Например, в исследовании Дж. Стэддона (Staddon, 1978) сделана попытка обосновать возникновение степенных зависимостей в поведении живых организмов при реакции на различные стимулы.

Высокая эффективность аппроксимации кривых выживаемости абонентов степенными функциями открывает возможности для прогнозирования динамики абонентской базы. В следующей главе будут рассмотрены возможности применения полученных результатов для построения прогнозных моделей и обсуждены ограничения предложенного подхода.

4. Экстраполяция выживаемости и использование ее в качестве прогноза

Выявленная в предыдущей главе высокая точность аппроксимации кривых выживаемости степенными функциями открывает перспективы для их использования в прогнозировании. Экстраполяция этих функций может позволить оценить выживаемость абонентской базы на длительных временных интервалах без необходимости ожидания фактических данных. Это имеет существенное значение для экономического моделирования и оценки эффективности мероприятий по привлечению абонентов. Прогнозирование выживаемости абонентов играет ключевую роль в расчете

таких важных экономических показателей, как ожидаемая продолжительность жизни абонента (Customer Lifetime) и совокупный доход от абонента (LTV – Customer Lifetime Value) (Kumar, 2014). Эти метрики критически важны для сопоставления с затратами на привлечение абонентов (SAC – Sales/Subscriber Acquisition Cost) и, следовательно, для принятия обоснованных решений о запуске, продолжении или прекращении маркетинговых кампаний по наращиванию абонентской базы (Krstevski & Mancheski, 2016).

Для оценки эффективности экстраполяции степенных функций в прогнозировании выживаемости был проведен следующий анализ:

1. использованы данные о выживаемости абонентов по критерию А, подключившихся в апреле 2023 года, за период в 12 месяцев;
2. построены прогнозные модели на основе экстраполяции степенных функций с использованием разного количества исходных точек данных (от 2 до 11 месяцев);
3. рассчитана для каждой модели среднемесячная абсолютная ошибка прогноза как среднемесячная абсолютная разница между прогнозными и фактическими значениями (для расчета используются только периоды прогноза);
4. построен график зависимости среднемесячной абсолютной ошибки от количества используемых фактических данных при экстраполяции.

Результаты анализа представлены на Рисунках 6,7 и в Таблице 3.

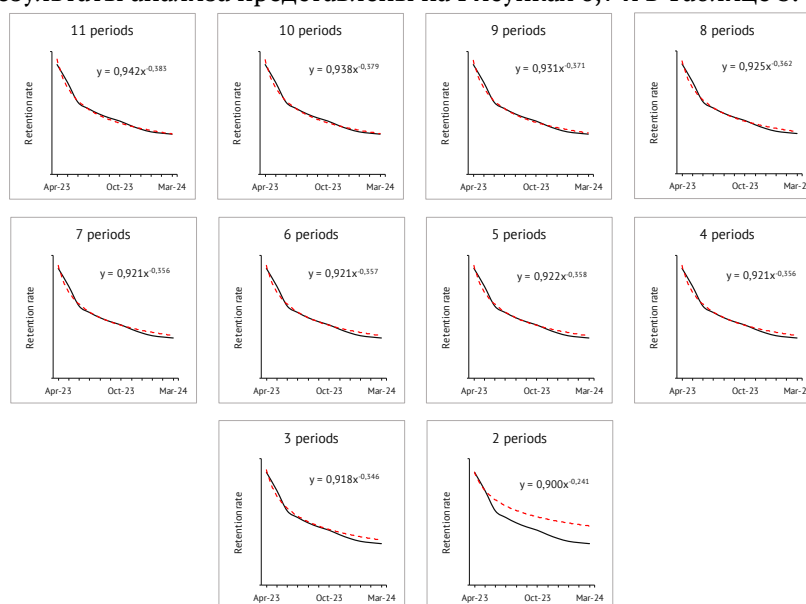


Рисунок 6. Сравнение фактической А-выживаемости с прогнозами на основе экстраполяции степенных функций. Источник: разработано авторами на основе данных об активности абонентов Т2

Наблюдается резкое повышение качества прогноза при использовании трех фактических точек данных, после чего качество прогноза показывает лишь незначительное улучшение. Небольшое anomalous увеличение ошибки прогноза, наблюдаемое в середине исследуемого диапазона (5-7 месяцев) связано с выбранной метрикой оценки качества прогноза и особенностями конкретного месяца подключений. Можно предположить, что в среднем ошибка будет монотонно снижаться при увеличении количества используемых фактических данных.

Таблица 3

Параметры аппроксимирующих функций и ошибки прогноза

Actual periods	Forecast periods	Function	Avg. abs. residuals
11	1	$y = 0,940x^{-0.383}$	0,004
10	2	$y = 0,938x^{-0.379}$	0,009
9	3	$y = 0,931x^{-0.371}$	0,015
8	4	$y = 0,925x^{-0.362}$	0,020
7	5	$y = 0,921x^{-0.356}$	0,021
6	6	$y = 0,921x^{-0.357}$	0,017
5	7	$y = 0,922x^{-0.358}$	0,015
4	8	$y = 0,921x^{-0.356}$	0,014
3	9	$y = 0,918x^{-0.346}$	0,020
2	10	$y = 0,900x^{-0.241}$	0,113

Составлено авторами на основе данных об активности абонентов Т2

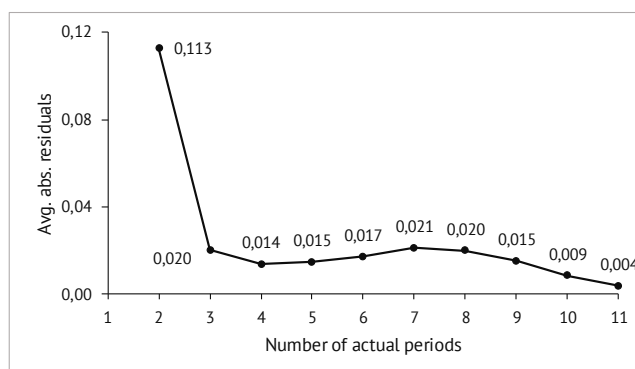


Рисунок 7. Зависимость среднемесячной ошибки прогноза от количества используемых фактических данных. Источник: разработано авторами на основе данных об активности абонентов Т2

Целью этого анализа было предложить общий подход к оценке качества прогноза, показать применимость прогноза на основе ограниченного набора фактических данных и предложить метод определения необходимого количества фактических данных для его построения. То, что при использовании нескольких фактических значений качество становится достаточно высоким, после чего уже лишь незначительно улучшается, имеет важное практическое значение. Например, можно использовать лишь первые три значения выживаемости для достаточно точной оценки выживаемости в будущем. Безусловно, для применения на практике этот вывод необходимо дополнительно валидировать на более широком наборе данных.

Следует отметить, что необязательно брать именно первые несколько значений. Этот метод дает возможность оценки «гипотетического прошлого» – то есть провести ретроспективный анализ и оценку альтернативных сценариев. Это может быть полезно в случаях, когда требуется проанализировать эффективность уже проведенных акций, направленных на удержание абонентов.

В практических задачах такой прогноз бывает целесообразно строить по кривой выживаемости по когортам, которая лучше улавливает актуальную дина-

мику выживаемости абонентской базы. Но, учитывая ее неоднородность, может потребоваться большее количество фактических данных для прогнозов требуемой точности.

Проведенное исследование демонстрирует высокий потенциал использования экстраполяции степенных функций для прогнозирования выживаемости абонентской базы. Однако, несмотря на выявленную эффективность, данный подход имеет определенные ограничения, которые требуют тщательного рассмотрения. В следующей главе будет проведен анализ этих ограничений, в частности, будет обсуждаться вопрос долгосрочной применимости степенных функций для моделирования выживаемости.

5. Ограничения применимости

Как было продемонстрировано в Главе 3, степенные функции вида ax^b демонстрируют высокую точность аппроксимации фактической выживаемости абонентской базы. В Главе 4 рассмотрена возможность использования этих функций для построения прогнозов. Однако, несмотря на их эффективность, очевидно, что высокое качество прогнозов, основанных на экстраполяции степенных функций, имеет определенные пределы применимости. С математической точки зрения, при $x \rightarrow +\infty$ функция вида ax^b , при $a > 0$ и $b < 0$, асимптотически стремится к нулю, не пересекая ось X:

$$\lim_{x \rightarrow +\infty} ax^b = 0, \quad \text{при } a > 0, b < 0 \quad (3)$$

Это означает, что в прогнозах, основанных на таких функциях, выживаемость никогда не достигнет нулевого значения. Однако, исходя из общих соображений, очевидно, что такое предположение не может соответствовать реальности. Выживаемость группы абонентов представляет собой агрегированный показатель индивидуальных состояний, каждое из которых в любой момент времени может принимать только два значения: 1 (абонент активен) или 0 (абонент неактивен). Логично предположить, что с течением времени количество активных абонентов будет неуклонно уменьшаться, пока не останется последний активный абонент. После его ухода выживаемость всей группы неизбежно станет равной нулю – событие, которое никогда не будет отражено функцией вида ax^b .

Более того, на выживаемость абонентской базы влияют не только факторы, связанные с желанием клиентов пользоваться услугами компании, но и объективные ограничения, такие как физическая возможность использования услуг связи. Абоненты – это живые люди (за исключением сегмента устройств с SIM-картами, которые также имеют ограниченный срок службы). Неизбежно наступит момент, когда каждый абонент физически не сможет продолжать пользоваться услугами.

Эти рассуждения указывают на потенциальное ограничение в применимости экстраполяции степенных функций для целей долгосрочного прогнозирования. Для более детального анализа этого ограничения обратимся к графику выживаемости по критерию В и его аппроксимации степенной функцией, представленными на Рисунках 3 и 4. При внимательном рассмотрении области графика, соответствующей более поздним периодам, можно заметить, что, несмотря на высокие показатели общего качества аппроксимации, с увеличением времени наблюдается тенденция к систематической переоценке фактической выживаемости аппроксимирующей кривой. Интересно отметить, что в первой половине графика

наблюдается обратная ситуация – функция несколько недооценивает фактическую выживаемость. Это можно объяснить тем, что коэффициенты аппроксимирующей функции подбираются таким образом, чтобы минимизировать общую ошибку по всей длине кривой.

Для более наглядной демонстрации эффекта переоценки выживаемости построим аппроксимацию, используя только 3 начальные фактических значения и экстраполируем выживаемость на весь период.

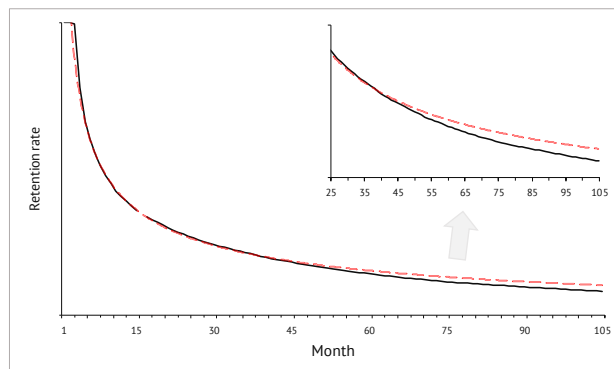


Рисунок 8. Экстраполяция В-выживаемости по начальным фактическим значениям. Источник: разработано авторами на основе данных об активности абонентов T2

Примечание: ввиду особенностей критерия В самые первые значения дают плохое представление о реальной активности абонента, поэтому для экстраполяции были выбраны фактические значения на 5, 6 и 7 месяцы. Отклонение в первые несколько месяцев так же объясняются спецификой выбранного критерия активности.

Результаты (Рисунок 8) демонстрируют, что кривая, построенная на основе ограниченного набора начальных данных, более точно описывает фактическую выживаемость в первой половине периода наблюдения. Однако теперь еще более отчетливо проявляется тенденция к расхождению между прогнозируемой и фактической выживаемостью по мере увеличения возраста группы абонентов. Такое наблюдение позволяет выдвинуть гипотезу о возникновении дополнительных факторов, влияющих на выживаемость абонентской базы с течением времени. Как уже говорилось выше, одним из таких факторов может быть физический возраст абонентов, влияние которого на использование услуг связи, предположительно, усиливается с увеличением срока пребывания в абонентской базе.

На основе кривой выживаемости крупных млекопитающих [Deevey, 1947; Pearl & Miner, 1935], в том числе людей (кривая типа I на Рисунке 9), можно сделать предположение об общем виде кривой выживаемости абонента с учетом фактора возраста.

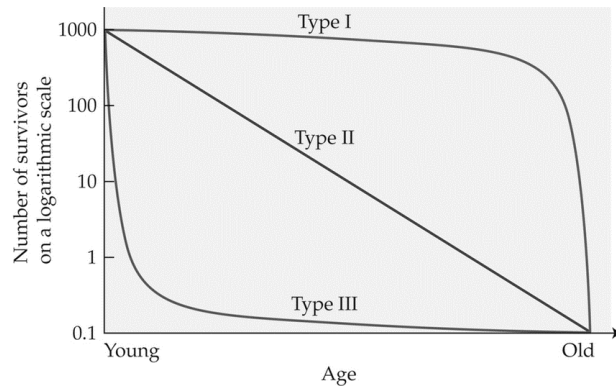


Рисунок 9. Три основных типа кривой выживаемости популяций животных в дикой природе по Э. Диви-мл. Источник: [Hill et al., 2021]

Если рассматривать кривую выживаемости абонента как вероятность сохранения его активности, то фактор физического возраста должен постепенно модифицировать кривую в соответствии с кривой типа I. Тогда, общий вид кривой выживаемости абонента должен иметь S-образную форму, состоящую из двух участков, разделенных промежутком неопределенной длительности (Рисунок 10). Первый участок отражает динамику выживаемости, обусловленную факторами, связанными с качеством услуг и лояльностью. Второй участок демонстрирует влияние иных объективных физических факторов:

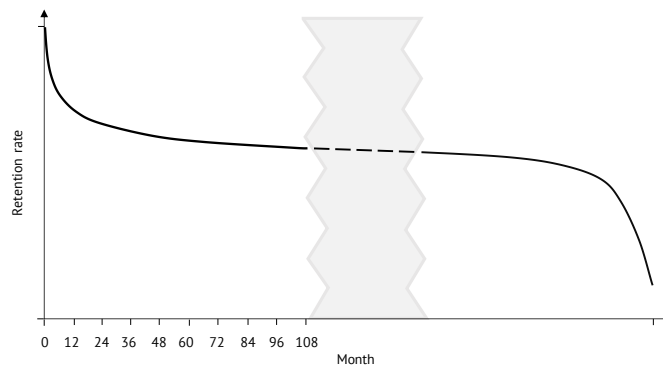


Рисунок 10. Общий вид кривой выживаемости абонента S-образной формы. Источник: разработано авторами

Не исключено, что этот вариант кривой с областями замедления и ускорения мог бы быть общим видом кривой физической выживаемости организмов. Возможно, если подобрать сопоставимые моменты начала наблюдения, то все три типа кривых являются вариантами такой общей кривой, а различия возникают в расположении областей замедления и ускорения, а также в периоде жизни, которому обычно уделяется внимание.

Изучение долгосрочного влияния фактора возраста и других демографических параметров на сохранение абонентской базы представляет собой направление для дальнейших исследований. Такой анализ мог бы способствовать разработке более точных моделей долгосрочного прогнозирования и помочь определить общий вид кривой выживаемости абонента на всем протяжении его жизненного цикла. Тем не менее, как было показано ранее, для задач краткосрочного и среднесрочного прогнозирования этим фактором можно пренебречь.

Заключение

В данном исследовании предложена классификация критериев активности абонентов и обсуждены вопросы их использования на практике. Комбинированные признаки фактического использования услуг, иногда в сочетании с финансовыми признаками, выглядят наиболее универсально для задач моделирования сохранения абонентской базы.

Рассмотрены два способа составления кривой выживаемости и определены их предпочтительные сферы применения. Для анализа последних фактических результатов бывает предпочтительно использовать кривую выживаемости по когортам, так как она содержит более актуальные данные. Для задач исследования формы кривой выживаемости и влияющих на нее факторов, для сравнения различных кривых между собой предпочтительно использовать кривую выживаемости фиксированной группы абонентов, так как она отражает чистую фактическую динамику активности.

Показано, что динамика сохранения абонентской базы соответствует степенному закону, а именно кривым вида ax^b ($a > 0$ и $b < 0$). Этот факт можно использовать для построения прогнозов сохранения абонентской базы на основе простой экстраполяции, используя, в нашем случае, всего три-четыре фактических значения. Возможность строить простые и точные прогнозы, зная лишь начальные значения выживаемости, находит большую применимость в задачах экономического моделирования и оценки полных совокупных ожидаемых эффектов от подключения абонентов, таких как средний срок жизни абонента и LTV.

Вместе с тем, у долгосрочных прогнозов, основанных на экстраполяции степенных функций есть ограничения, которые предположительно связаны с возрастающим со временем влиянием других факторов, таких как физический возраст абонента. Можно предположить, что общий вид функции выживаемости должен иметь какой-то комбинированный вид, в которой начальная часть кривой имеет скорее степенной вид, а конечная – отличается от него.

Данное исследование оставляет открытыми следующие вопросы, каждый из которых требует отдельного изучения:

1. Почему сохранение базы (выживаемость абонентов) так хорошо описывается именно степенными функциями?
2. Почему сохранение базы, (и, соответственно, отток) выглядит таким равномерным, и, как будто, зависящим только от начальных условий? Это проявление динамического равновесия усилий всех участников конкурентного рынка?
3. Какова могла бы быть общая форма долгосрочной кривой выживаемости абонентов?

Список литературы

1. Andriani, P. Beyond Gaussian averages: Redirecting international business and management research toward extreme events and power laws / P. Andriani, B. McKelvey // *Journal of International Business Studies*. – 2007. – №38(7). – P. 1212-1230.
2. Deevey, E. S., Jr. Life tables for natural populations of animals // *The Quarterly Review of Biology*. – 1947. – №22. – P. 283-314.
3. Fader, P. S. Probability models for customer-base analysis / P. S. Fader, B. G. Hardie // *Journal of interactive Marketing*. – 2009. – №23(1). – P. 61-69.
4. Gabaix, X. Power Laws in Economics and Finance // *Annual Review of Economics*. – 2009. – №1. – P. 255-293.
5. Gupta, S. Customer metrics and their impact on financial performance / S. Gupta, V. Zeithaml // *Marketing Science*. – 2006. – №25(6). – P. 718-739.
6. Hill, R. W. Youth in the study of comparative physiology: insights from demography in the wild / R. W. Hill, D. Sleboda, J. J. Millar // *Journal of Comparative Physiology B*. – 2021. – №191. – P. 1-16.
7. Jain, H. Telecom churn prediction and used techniques, datasets and performance measures: a review / H. Jain, A. Khunteta, S. Srivastava // *Telecommunication Systems*. – 2021. – №76(4). – P. 613-630.
8. Krstevski, D. Managerial Accounting: Modeling Customer Lifetime Value – An Application in the Telecommunication Industry / D. Krstevski, G. Manchenski // *European Journal of Business and Social Sciences*. – 2016. – №5(1). – P. 64-77.
9. Kumar, V. Customer lifetime value: The path to profitability. – Boston: Now Publishers, 2008. – 112 p.
10. Newman, M. E. J. Power laws, Pareto distributions and Zipf's law // *Contemporary Physics*. – 2005. – №46(5). – P. 323-351.
11. Pearl, R. Experimental studies on the duration of life. XIV. The comparative mortality of certain lower organisms / R. Pearl, J. R. Miner // *The Quarterly Review of Biology*. – 1935. – №10. – P. 60-79.
12. Ribeiro, H. Determinants of churn in telecommunication services: a systematic literature review / H. Ribeiro, B. Barbosa, A. C. Moreira & R. G. Rodrigues // *Management Review Quarterly*. – 2023. – №74(3). – P. 1327-1364.
13. Staddon, J. Theory of behavioral power functions // *Psychological Review*. – 1978. – №85(4). – P. 305-320.
14. Verhelst, T. Understanding Telecom Customer Churn with Machine Learning: From Prediction to Casual Inference / T. Verhelst, O. Caelen, J. C. Dewitte, B. Lebichot, G. Bontempi // *Artifical Intelligence and Machine Learning*. – 2021. – №1996. – P. 182-200.
15. Verbeke, W. Building comprehensible customer churn prediction models with advanced rule induction techniques / W. Verbeke, D. Martens, C. Mues, B. Baesens // *Expert Systems with Applications*. – 2011. – №38(3). – P. 2354-2364.
16. Zhang J. Z., Chang, C. W. Consumer dynamics: Theories, methods, and emerging directions / J. Z. Zhang, C. W. Chang // *Journal of the Academy of Marketing Science*. – 2021. – №49(1). – P. 166-196.