Лапина Мария Геннадьевна. Математическое моделирование процесса восстановления общего пассажирского спроса по цензурированной выборке продаж: диссертация ... кандидата Экономических наук: 08.00.13 / Лапина Мария Геннадьевна;[Место защиты: ФГБОУ ВО Санкт-Петербургский государственный экономический университет], 2017

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | |  |  |  |  | | --- | --- | --- | --- | | |  | | --- | | **Содержание к диссертации**  Введение  **ГЛАВА 1. Задача восстановления общего спроса как составляющая часть управления доходами14**  1.1 Управление доходами: структура и основные понятия 14  1.2 Оценка и прогнозирование спроса 18  1.3 Процесс цензурирования спроса: примеры и основные понятия 20  1.4 Обзор исследований, посвященных задаче восстановления общего спроса по цензурированной выборке  1.4.1 Исследования, посвященные работе с цензурированными данными 26  1.4.2 Исследования, посвященные работе с отдельными независимыми классами 29  1.4.3 Исследования, посвященные работе с несколькими связанными классами/ рейсами 31  1.5 Выводы к ГЛАВЕ 1 33  **ГЛАВА 2. Методы восстановления общего спроса по цензурированной выборке 35**  2.1 Математическая постановка задачи восстановления общего спроса по цензурированной выборке 35  2.2 Метод максимального правдоподобия при решении задачи восстановления общего спроса по цензурированной выборке 36  2.3 Описание существующих методов восстановления общего спроса, работающих с отдельными независимыми классами 45  2.3.1 «Наивные» методы 45  2.3.2 Методы, использующие информацию по нескольким временным срезам 46  2.3.3 Статистические итерационные методы 48  2.3.4 Непараметрический метод Каплана-Мейера 2.4 Относительная асимптотическая эффективность оценок, полученных с помощью метода максимального правдоподобия и метода Каплана-Мейера 54  2.5 Выводы к ГЛАВЕ 2. 65  **ГЛАВА 3. Сравнение точности оценок, полученных с помощью методов восстановления 67**  3.1 Сравнение точности оценок, полученных с помощью методов восстановления, использующих информацию одного временного среза 68  3.1.1 Формирование наблюдаемой цензурированной выборки 68  3.1.2 Результаты числовых расчетов и сравнительный анализ методов восстановления 76  3.2 Сравнение точности оценок метода максимального правдоподобия и метода, использующего информацию по нескольким временным срезам 86  3.2.1 Формирование наблюдаемой цензурированной выборки продаж86  3.2.2 Результаты числовых расчетов и сравнительный анализ методов восстановления  3.3 Влияние предположения о типе распределения на точность оценок, полученных с помощью методов восстановления. 90  3.4 Выводы к ГЛАВЕ 3. 97  **ГЛАВА 4. Анализ влияния методов восстановления общего спроса на доход авиакомпании 99**  4.1 Методика оценки влияния методов восстановления общего спроса на доход авиакомпании 99  4.2 Методы определения оптимальной стратегии продаж  4.3 Имитационное моделирование процесса продаж 104  4.4 Оценка влияния методов восстановления общего спроса на доход авиакомпании: сгенерированные исторические данные 105  4.5 Оценка влияния методов восстановления общего спроса на доход авиакомпании: реальные исторические данные 112  4.6 Выводы к ГЛАВЕ 4. 120  Заключение 122  Список литературы 124   * [Обзор исследований, посвященных задаче восстановления общего спроса по цензурированной выборке](http://www.dslib.net/mat-metody/matematicheskoe-modelirovanie-processa-vosstanovlenija-obwego-passazhirskogo-sprosa.html#7680901) * [Описание существующих методов восстановления общего спроса, работающих с отдельными независимыми классами](http://www.dslib.net/mat-metody/matematicheskoe-modelirovanie-processa-vosstanovlenija-obwego-passazhirskogo-sprosa.html#7680902) * [Результаты числовых расчетов и сравнительный анализ методов восстановления](http://www.dslib.net/mat-metody/matematicheskoe-modelirovanie-processa-vosstanovlenija-obwego-passazhirskogo-sprosa.html#7680903) * [Оценка влияния методов восстановления общего спроса на доход авиакомпании: сгенерированные исторические данные](http://www.dslib.net/mat-metody/matematicheskoe-modelirovanie-processa-vosstanovlenija-obwego-passazhirskogo-sprosa.html#7680904)   **Введение к работе**  **Актуальность темы исследования.** Управление доходами (Revenue management, далее RM) в последнее время находится в центре особого внимания как одна из наиболее успешных прикладных областей исследования операций. RM представляет собой набор особых стратегий и подходов, позволяющих так распределить имеющиеся ресурсы, чтобы в наиболее полной мере удовлетворить спрос на них и добиться максимизации прибыли. В связи с этим прогноз спроса является неотъемлемой и весьма важной составляющей многих моделей принятия решений, используемых в области управления доходами. В сферах услуг, таких как авиаиндустрия, гостиничный бизнес, индустрия проката автомобилей, железнодорожная индустрия и другие, оптимальное управление ресурсами может быть достигнуто путем контроля над наличием в продаже различных продуктов, а также установления ограничений на их количество (так называемых пределов бронирования и уровней защиты). На протяжении всего процесса контроля должен соблюдаться баланс между продажами дешевых продуктов (например, в случае пассажирской авиаиндустрии, билетов эконом-классов), пользующихся большим спросом, и «защитой» более дорогих продуктов (билетов бизнес - и премиум-классов), спрос на которые меньше и возникает, как правило, ближе к моменту окончания продаж. Любая стратегия продаж сильно зависит от прогноза спроса на все имеющиеся продукты. Очевидно, что неточный прогноз спроса может привести к негативным результатам, прежде всего к уменьшению дохода от реализации продуктов.  Именно поэтому одной из самых важных задач управления доходами является обеспечение надежных оценок спроса, не ограниченного пределами бронирования и вместимостью. Такой спрос носит название общего неограниченного (или просто общего). Процедура его прогнозирования усложняется тем, что исторические данные, собираемые компаниями, отражают уровень продаж, а не сам спрос на те или иные продукты. После того, как величина продаж какого-либо продукта достигнет заранее установленного для него предела бронирования, последующие запросы на приобретение будут, естественно, отклонены, при этом регистрация таких запросов практически невозможна. Таким образом, исторические данные о продажах будут соответствовать цензурированному спросу. Процедура оценивания параметров распределения общего спроса по исходной цензурированной выборке называется восстановлением общего  неограниченного спроса, а разработанные для этого методы – методами восстановления. Общий неограниченный спрос довольно сложно измерить, однако, несмотря на это, многие исследователи утверждают, что его использование делает прогноз более точным и тем самым увеличивает доход от реализации продуктов. Так, разными исследователями было выявлено, что влияние восстановления спроса на доход компаний может достигать 2-12%.  Диссертационная работа посвящена модельному анализу процесса восстановления общего спроса. В качестве ключевого примера рассмотрен процесс восстановления общего пассажирского спроса на авиабилеты по цензурированным данным о продажах.  **Степень разработанности научной проблемы.**  Вопросу восстановления общего неограниченного спроса, его оцениванию и прогнозированию посвящено достаточно много работ. Существуют труды, охватывающие различные аспекты управления доходами с отдельными главами, посвященными оцениванию и прогнозированию спроса. Среди них наиболее широко цитируемая книга K. Talluri и G.J. van Ryzin «The Theory and Practice of Revenue Management». В своих статьях к вопросу восстановления спроса неоднократно обращались такие зарубежные исследователи и ученые, как P.P. Belobaba, L.R. Weatherford, R.M. Ratliff, J.I. McGill, S. Polt, G.Vulcano и другие. Их работы посвящены новым методам восстановления, сравнению точности полученных оценок, анализу влияния выбранных подходов на доход компаний.  Однако вопрос о методе восстановления, дающего более точные оценки параметров распределения спроса, чем уже существующие подходы, менее затратного с точки зрения времени расчетов до настоящего момента остается открытым. Кроме того в опубликованных работах практически не уделяется внимание сравнению статистических свойств получаемых оценок, чувствительности методов ко всякого рода предположениям, в том числе к предположению о типе распределения спроса. Существующие подходы управления доходами не охватывают вопрос неполноты информации о цензурировании, когда для некоторых или всех наблюдений не известно, соответствуют они истинному спросу или подверглись цензурированию.  Таким образом, указанные вопросы являются актуальными и недостаточно разработанными, что и обусловило выбор темы, цель и задачи диссертационного исследования.  **Целью диссертационной работы** является разработка подхода к построению оценок параметров распределения общего спроса по цензурированной выборке продаж, обладающих высокой точностью и эффективностью.  Для достижения поставленной цели были сформулированы и решены следующие **основные задачи:**   1. проведен сравнительный анализ преимуществ, недостатков и особенностей реализации существующих методов восстановления общего спроса по цензурированным данным; 2. предложен и обоснован подход к решению задачи восстановления общего спроса, в основу которого лег метод максимального правдоподобия. В рамках подхода рассмотрены различные варианты, когда информация о цензурировании не является полной; 3. изучены асимптотические свойства оценок некоторых методов восстановления спроса (метода максимального правдоподобия, Каплана-Мейера); 4. проведены массовые числовые расчеты по моделированию спроса и расчету пределов бронирования, на базе которых выполнено сравнение точности оценок, полученных с использованием различных методов восстановления; 5. оценено влияние неполноты информации о цензурировании на точность оценок предложенного подхода; 6. для методов Expectation Maximization и максимального правдоподобия проведен сравнительный анализ чувствительности оценок к предположению о типе распределения общего спроса; 7. на основе результатов имитационного моделирования процесса продажи авиабилетов оценено влияние методов восстановления общего спроса на доход авиакомпании.   **Предметом исследования** является процесс восстановления общего пассажирского спроса по цензурированным данным о продажах в рамках управления доходами.  **Объектом исследования** являются пассажирские авиакомпании.  **Теоретической** **основой** **диссертационного** **исследования**  послужили научные работы отечественных и зарубежных авторов в области математической и прикладной статистики, исследования операций, в том числе управления доходами (A.P. Dempster, D.B. Rubin, A. Wald, M.D. Perlman, E.L. Kaplan, P. Meier, R. Fisher, Л.В. Канторович, А.Н. Колмогоров, K. Talluri, G.J. van Ryzin, P. Belobaba, D. C. Montgomery, A. C. Harvey и других).  **Методологическая основа исследования.** В работе применяются методы теории вероятностей, математической статистики, имитационное моделирование, методы исследования операций, компьютерное  моделирование. В качестве компьютерного инструмента для программной реализации разработанных моделей и проведения расчетов выбрана среда Mathematica 11.0.  **Информационную базу исследования** составили статистические данные о продажах авиабилетов одного из крупных авиаперевозчиков, а также модельные данные, полученные с помощью компьютерной среды Mathematica 11.0.  **Обоснованность и достоверность результатов исследования**  обеспечивается теоретическими принципами методов математической статистики и исследования операций, лежащих в основе диссертационной работы. В частных случаях полученные результаты соотносятся с более ранними экспериментальными данными этой же тематики.  **Соответствие диссертации Паспорту научной специальности.**  Область исследования, соответствующая паспорту специальности 08.00.13 – Математические и инструментальные методы экономики:  1. Математические методы.  п. 1.1. «Разработка и развитие математического аппарата анализа экономических систем: математической экономики, эконометрики, прикладной статистики, теории игр, оптимизации, теории принятия решений, дискретной математики и других методов, используемых в экономико-математическом моделировании»;  п. 1.2. «Теория и методология экономико-математического моделирования, исследование его возможностей и диапазонов применения: теоретические и методологические вопросы отображения социально-экономических процессов и систем в виде математических, информационных и компьютерных моделей»;  п. 1.4. «Разработка и исследование моделей и математических методов анализа микроэкономических процессов и систем: отраслей народного хозяйства, фирм и предприятий, домашних хозяйств, рынков, механизмов формирования спроса и потребления, способов количественной оценки предпринимательских рисков и обоснования инвестиционных решений».  2. Инструментальные средства.  п. 2.2. «Конструирование имитационных моделей как основы экспериментальных машинных комплексов и разработка моделей экспериментальной экономики для анализа деятельности сложных социально-экономических систем и определения эффективных направлений развития социально-экономической и финансовой сфер».  **Научная новизна результатов исследования** состоит в разработке подхода к решению задачи восстановления общего спроса, основанного на методе максимального правдоподобия. Подход позволяет получать оценки параметров распределения спроса, обладающие высокой точностью, и мало затратные, с точки зрения времени вычислений. Он также адаптирован к случаю, когда информация о цензурировании частично отсутствует, в то время как существующие подходы к восстановлению спроса имеют дело только с полной информацией. В работе впервые проведен анализ влияния неполноты информации о цензурировании на точность получаемых оценок параметров распределения спроса, а также на величину дохода при реализации продуктов авиакомпаний.  **Наиболее существенные результаты исследования, обладающие научной новизной и полученные лично соискателем:**  обоснован и разработан подход к построению оценок параметров распределения общего спроса по цензурированной выборке продаж, основанный на методе максимального правдоподобия. Данный подход имеет возможность работы с различными вариантами неполноты информации о цензурировании, в то время как существующие подходы к восстановлению спроса имеют дело только с полной информацией;  проведен сравнительный анализ асимптотической эффективности оценок, полученных в рамках методов Каплана-Мейера и максимального правдоподобия. Установлено, что преимущество параметрического метода  максимального правдоподобия, в случае корректного предположения о распределении случайной величины, особенно ярко проявляется для высокой доли цензурированных элементов наблюдаемой выборки и для значений функции надежности, близких к 0 или 1;  на основе результатов имитационного моделирования проведен сравнительный анализ точности оценок параметров распределения, полученных с помощью различных методов восстановления, а именно: методов Каплан-Мейера, Expectation Maximization, Projection Detruncation, Nave 3, двойного экспоненциального сглаживания, максимального правдоподобия. Моделирование проведено с учетом возможности изменения предположения о типе распределения спроса и работы с неполной информацией о цензурировании;  выявлено, что предлагаемый подход, основанный на методе максимального правдоподобия, для большого процента цензурирования дает более точные оценки параметров распределения, чем метод Expectation Maximization, который в настоящий момент является лучшим из используемых при решении задачи восстановления спроса;  выявлено, что методы Expectation Maximization и максимального правдоподобия малочувствительны к предположению о типе распределения общего спроса, однако с увеличением среднего квадратического отклонения устойчивость результатов снижается;  проведено имитационное моделирование процесса продажи авиабилетов и проанализировано влияние методов восстановления общего спроса на доход авиакомпании. Расчеты продемонстрировали, что с увеличением процента цензурирования величина дохода в большей мере зависит от выбранного метода восстановления общего спроса. На реальных данных показано, что использование подхода, основанного на методе максимального правдоподобия, позволяет получить статистически значимо больший выигрыш в доходе от реализации продуктов (в среднем до 1,4%) по сравнению с методом Expectation Maximization;  установлено, что предлагаемый подход при использовании исходных данных с высокой долей неполноты информации позволяет получать оценки параметров распределения общего спроса по точности сопоставимые с оценками методов, использующих полную информацию о цензурировании, как следствие даже значительное увеличение степени неполноты информации практически не приводит к потерям дохода.  **Теоретическая значимость результатов исследования** заключается в разработке подхода к решению задачи восстановления общего спроса по цензурированным данным, позволяющего эффективно работать с неполной информацией о цензурировании. Получаемые с помощью предлагаемого подхода оценки параметров распределения обладают высокой точностью и эффективностью.  **Практическая значимость исследования.**  Рассмотренный в диссертационной работе подход, основанный на методе максимального правдоподобия, может быть использован при восстановлении общего неограниченного спроса на авиабилеты, железнодорожные билеты, гостиничные номера, арендуемые автомобили. Его применение позволяет повысить точность получаемых оценок и, как следствие, уменьшить потери дохода от реализации продуктов. Кроме того, предложенный подход может быть использован при работе со случайно цензурированными данными любой природы, например, данными о выживаемости пациентов, продолжительности безработицы, сроке функционирования проектов. Преимуществом предлагаемого в работе подхода является не только высокая точность и эффективность получаемых с его помощью оценок, но и тот факт, что его использование возможно для случая неполноты информации о цензурировании. Ситуация, когда информация о цензурировании частично или полностью отсутствует, характерна для небольших авиакомпаний, в которых сбор информации осуществляется недостаточно тщательно.  **Апробация результатов исследования.** Основные результаты диссертации отмечены поощрительной премией на конкурсе имени профессора Б.Л. Овсиевича. Премией награждаются молодые ученые за фундаментальные экономико-математические исследования, выполненные в России, конкурс организован Российской академией наук, Санкт-Петербургским научным центром, Санкт-Петербургским экономико-математическим институтом РАН, Международным центром социально-экономических исследований «Леонтьевский центр», Фондом Егора Гайдара. Результаты докладывались и обсуждались на следующих конференциях и семинарах:  Научная конференция «Инновационная экономика: реалии и перспективы», НИУ ВШЭ в СПб, 27 сентября 2011 г.;  Международная научно-практическая конференция «Государственная статистика как общественное благо: стоимость, качество, использование». Федеральное государственное бюджетное учреждение науки «Социологический институт РАН», СПбГЭУ, СПб, 30 января - 1 февраля  2012 г.;  31th International Conference «Mathematical Methods in Economics 2013», Brno, 11-13 сентября 2013 г.;  Всероссийская конференция «Первые чтения памяти профессора Б.Л. Овсиевича «Экономико-математические исследования: математические модели и информационные технологии», СПб ЭМИ РАН, 21-23 октября  2013 г.;  Всероссийская конференция «Вторые чтения памяти профессора Б.Л. Овсиевича «Экономико-математические исследования: математические модели и информационные технологии», Санкт-Петербург, 26-28 октября 2015 г.;  городской семинар по экономико-математическим методам и моделям, проводимый Санкт-Петербургским экономико-математическим институтом РАН, 25 ноября 2015 г.;  семинар департамента экономики СПб ШЭМ НИУ ВШЭ, 11 апреля 2016 г.;  городской семинар по экономико-математическим методам и моделям, проводимый Санкт-Петербургским экономико-математическим институтом РАН, 13 марта 2017 г.  **Публикации результатов исследования**. Основные результаты исследования изложены в 10 научных работах, объемом 5,55 п.л. (авторский объем - 3,7 п.л.), 3 из которых опубликованы в русскоязычных изданиях из рекомендованного списка ВАК, 1 в англоязычном журнале, входящем в международную реферативную базу данных Scopus.  **Структура диссертации.** Диссертация состоит из введения, четырех глав, заключения, списка литературы. Общий объём диссертационной работы - 137 страниц. Обзор исследований, посвященных задаче восстановления общего спроса по цензурированной выборке Классические области применения RMS, такие как авиаиндустрия, гостиничный бизнес, индустрия проката автомобилей, объединяют следующие характеристики: продаже подлежат продукты, использование которых имеет ограничения по времени; спрос на эти продукты с момента начала до момента окончания продаж меняется; сферам присущи относительно низкие переменные затраты; существует возможность сегментации рынка и дифференциация цен.  Рассмотрим простейшую схему управления доходами на примере работы продавца, который имеет запас скоропортящихся фруктов. Ему следует установить цены на плоды, при которых доход будет максимальным, а плоды не успеют испортиться. Если установленные цены будут слишком низкими, все фрукты будут проданы в ближайшее время, а пришедшим позже покупателям придется отказать. Если же продавец установит слишком высокие цены, не все плоды будут проданы и их придется выбросить, понеся определенные потери.  Продавец принимает свои ценовые решения, основываясь на своем прогнозе спроса на фрукты. Он должен предсказать количество востребованных фруктов на разных ценовых уровнях. С течением времени плоды приближаются к своей конечной дате продажи, а продавец будет переоценивать спрос и корректировать устанавливаемые цены. Например, он может снизить цену, если видит, что плоды не продаются так быстро, как он ожидал, но постепенно начинают портиться. Прогнозирование спроса на продукты становится более сложным в случае, когда фрукты продаются полностью. Продавцу гораздо сложнее оценить, сколько фруктов он продал бы, если бы его запас был больше, и он смог удовлетворить весь спрос. В этом смысле его данные являются цензурированными.  Пример продавца фруктов иллюстрирует простейшую схему управления доходами и проблему оценивания общего спроса по цензурированным данным. Более сложное применение управления доходами практикуется авиакомпаниями. Они собирают большие массивы данных о продажах своих продуктов (в авиаиндустрии продукт — сочетание маршрута и ценового класса), которые используют для прогнозирования общего спроса. Затем используются сложные оптимизационные модели для расчета оптимального количества доступных мест, соответствующих различным продуктам. Процедуру прогнозирования и оптимизации повторяют до вылета самолета, то есть до момента, когда продукты теряют свою ценность.  Очевидно, что пассажиры в соответствии с их поведением и чувствительностью к ценам могут быть разделены на категории. Так, например, можно выделить бизнес - пассажиров и туристов. Бизнес - пассажиры, как правило, приобретают билеты близко к дате вылета, пребывание в месте назначения продолжается в течение короткого промежутка времени. Этот тип пассажиров, как правило, достаточно строг в своих планах, но готов платить более высокие цены за билеты.  Туристы, с другой стороны, бронируют билеты задолго до даты их путешествия. Их пребывание в месте назначения гораздо дольше, и они более гибкие в своих планах. Туристы часто решают вовсе не путешествовать, нежели покупать билеты с высокими тарифами. Задача в рамках систем управления доходами состоит в разработке плана, согласно которому билеты дешевых тарифов в определенный момент перестают быть доступными. Это совершается для того, чтобы не допустить покупку этих дешевых билетов потребителями, готовыми заплатить более высокие суммы. Так как пассажиры, приобретающие билеты дешевых ценовых классов, как правило, производят бронирования до пассажиров более дорогих классов, система управления доходами должна спрогнозировать количество бизнес - пассажиров, желающих приобрести билеты на данный рейс. Затем это количество должно быть отложено или «защищено», так чтобы они были доступны, когда бизнес -пассажиры придут за ними.  Если прогноз спроса для пассажиров более дорогих ценовых классов необоснованно высок, для них будет защищено слишком много мест, как следствие рейс будет совершаться с пустыми местами, которые могли быть проданы пассажирам более дешевых ценовых классов. Если прогноз наоборот слишком низок, то слишком мало мест будет защищено, места, которые могли быть проданы по более высоким тарифам, будут проданы по низким, часть дохода будет потеряна.  Сложность построения прогнозов состоит в том, что исторические данные, как уже упоминалось, часто не отражают общий спрос. После того, как авиакомпания из-за ограничений, установленных системой управления доходами, прекращает продажу билетов по определенному тарифу, она также прекращает сбор данных. Авиакомпания может получить гораздо больше запросов на продажу билетов определенного тарифа, но эти запросы не регистрируются -данные подвергнуты цензурированию и не представляют общий спрос.  Таким образом, необходимо воспользоваться определенными подходами, для того чтобы преобразовать цензурированные данные в более точные оценки общего неограниченного спроса. Существуют различные методы, задача которых состоит в «восстановлении» общего спроса по имеющимся цензурированным историческим данным. Эти подходы варьируются от простых к сложным, основываются на различных предположениях и подразумевают полную оценку вероятностного распределения или, по крайней мере, оценку параметров предполагаемого распределения спроса.  Определение цензурированного спроса  Рассмотрим процесс сбора исторических данных о продажах в рамках определенного ценового класса какого-либо рейса. В случае если запросы потребителей на бронирование подтверждаются, собранные данные о продажах будут соответствовать общему спросу (Рисунок 2 (а)). В противном случае, если в какой-то момент времени достигается установленный предел бронирования и, как следствие, какие-то запросы на бронирование были отклонены, исторические данные для данной временной точки отражают лишь цензурированный спрос (Рисунок 2 (б)). Описание существующих методов восстановления общего спроса, работающих с отдельными независимыми классами Если параметры \i2 и 2 заданы, тогда задача сводится к нахождению глобального максимума функции двух переменных [i± и а±. В противном случае решается задача при четырех неизвестных переменных. В обоих случаях информацию о том, является ли каждый отдельно взятый элемент zk цензурированным или нет, отсутствует. Поиск глобального максимума может быть произведен с использованием одного из хорошо известных алгоритмов, например метода градиента или любого другого.  Существует информация о том, какой из элементов выборки z цензурирован, а какой нет.  Как было сказано ранее, наблюдаемая выборка z частично состоит из элементов выборки х, частично из элементов выборки у. В рассматриваемом случае предполагается, что существует информация о том, что zk = xk или zk — Ук для каждого к = 1,..., п, то есть известно какой из элементов выборки х, соответствующей общему спросу, был цензурирован пределами бронирования. Так же подобного рода информацию можно интерпретировать, как наличие помимо выборки z = (z1,...,zn), соответствующей выборки значений индикаторов цензурированности 8 = (S1,..., 8п). Далее будут приведены формулы только для непрерывной случайной величина Z = min (X, Y), для дискретного случая формулы могут быть получены с помощью аналогичных рассуждений.  Очевидно, что zk = хк наблюдается в том случае, если хк ук. Для малого значения Лх вероятность д(хк)Ах осуществления этого события равна с точностью порядка о(Лх) вероятности осуществления следующих событий: попадание случайной величины X в интервал (хк,хк+Ах); вероятность осуществления этого события равна fx(xk; а±,..., аг)Ах; осуществление события Y хк; вероятность P(Y хк) осуществления этого события равна 1 - FY(xk; blt..., bs). Тогда g(xk)Ax fx(xk; ах,..., аг)Лх(і - FY(xk; blt..., bs)), и в точке zk = xk функция плотности вероятности fz(xk;a1,...,ar;b1,...,bs)=g(xk) принимает следующую форму: fz(xk;all...lar;bll...,bs)=fx(xk;all...lar)(l-FY(xk;bll...lbs)). С другой стороны Zj = у; наблюдается в том случае, если у,- xj. Поэтому, по аналогии, в точке z;- = у;- функция плотности вероятности fz(y}; «і, -, аг; Ьъ ..., bs) = g{yj) принимает следующую форму: fz{yf,a1,...,ar]b1,...,bs)=fY{yf,b1,...,bs)[l-Fx{yf,a1,...,ar)). Если воспользоваться введением индикаторной функции, будет получено выражение плотности распределения цензурированной случайной величины с информацией о цензурированиии конкретных элементов: fz{z)S;a1,...,ar)b1,...,bs) = [fx(z; %,..., а,.) X (1 - FY(z; Ьъ ..., bs))]s (10) X [(1 - Fx(z; at ar)) X fY(z; b± b,)]1"5 Как результат функция правдоподобия L выглядит следующим образом: n L(z; alt..., ar; b±,..., bs) = П fz(zk; %,..., ar; blt..., bs) = k=l m = Y\fx{xk;a1,...,ar){l-FY{xk;b1,...,bs))x (11) fe=i і x П/г(у;; b±,..., bs) (l - Fx(yj; %,..., ar)) где m - количество нецензурированных элементов, / - количество цензурированных элементов выборки z, то есть т + I = п. В обозначениях с индикаторной функцией: п L(z; аг,..., аг; Ьг,..., bs) = П fz{zk; аг,..., аг; Ьг,..., bs) = (12) к=1 п = Y][fx(zk; %,..., Or) X (1 - FY(zk; b±l..., bs))]s X fe=i x [(1 - Fx(zk; % ar)) x fY(zk; b± b,)]1- Как уже говорилось ранее, для удобства максимизируется логарифм функции правдоподобия: L(z;all...,ar;bll...lbs) max Если оценке подлежат только параметры распределения случайной величины X, а именно, а1,...,аг, в этом случае функция максимального правдоподобия запишется в следующем виде: m I yiogfx{xk)all...lar) + y(l-Fx(yj)all...lar)) max (13) fe=i j=i или в других обозначениях: п У log[[fx(zk; а± ar)Yk [1-Fx(zk;ai a,)]1 ] - max (14) fe=i Проведенное упрощение имеет дополнительное важное следствие: информация о распределении случайной величины Гне требуется. Результаты числовых расчетов и сравнительный анализ методов восстановления Сущность метода может быть описана только на примере с данными по нескольким временным срезам. На первом временном срезе все цензурированные наблюдения заменяются математическим ожиданием всех нецензурированных наблюдений. На втором временном срезе все цензурированные наблюдения заменяются пересмотренными наблюдениями первого временного среза, умноженные на коэффициент, равный отношению суммы математического ожидания всех нецензурированных наблюдений первого и второго временных срезов к математическому ожиданию всех нецензурированных наблюдений первого среза минус один. На третьем временном срезе все цензурированные наблюдения заменяются суммой пересмотренных наблюдений первого и второго срезов, умноженной на коэффициент, равный отношению суммы математического ожидания всех нецензурированных наблюдений первого, второго и третьего временных срезов к сумме математических ожиданий всех нецензурированных наблюдений первого и второго срезов минус один.  Метод двойного экспоненциального сглаживания  Метод двойного экспоненциального сглаживания (далее DES) или метод Хольта для решения вопроса цензурирования при прогнозировании спроса был предложен Ferguson и Queenan [98]. Для работы с данными о продажах продукта определенного рейса DES использует две сглаживающие константы: одну для сглаживания базовой составляющей спроса, другую для сглаживания трендовой составляющей. Пусть t представляет период времени между /, началом продаж, и В, моментом достижения спросом предела бронирования. Время рассчитывается в обратном порядке, то есть / и В - количество дней от отправления, тогда t Є [І,І-1,...,В],І В. После момента времени В запросы на бронирование продолжают появляться, но они уже не регистрируются. Если спрос не достигает предела бронирования, В = 0 (окончание продаж). Наблюдаемым спросом являются запросы, зарегистрированные между моментами времени I и В. Используются следующие обозначения: At истинный кумулятивный спрос за период t; Ft - экспоненциально сглаженная базовая компонента за период t; Tt - экспоненциально сглаженная трендовая компонента за период t; FITt - прогнозное значение кумулятивного спроса за период t; а сглаживающая константа для базовой составляющей; Р сглаживающая константа для трендовой составляющей. Прогнозное значение для предстоящего периода V. FITt = Ft + Tt, (19) где Ft = FITt+1 + a(At+1 - FITt+1) Tt = Tt+1+p(Ft-FITt+1). Сглаживающие константы рассчитываются за счет минимизации ошибки прогноза: в тт\ (At-FITt)2. (21) t=I  В качестве начальных значений для F7 и 7} выбираются истинное значение спроса в период / и средняя трендовая для доступных значений кумулятивного спроса, соответственно. Данная модель затем используется для получения прогнозных значений кумулятивного спроса для всех периодов, в рамках которых спрос подвергся цензурированию.  Такая процедура восстановления спроса проводится для каждой из имеющихся кривых бронирования. Далее можно воспользоваться восстановленными кривыми бронирования для получения оценок параметров спроса любого из временных срезов. Если Xt - оценка спроса выбранного временного спроса для і - ой кривой бронирования, тогда оценки математического ожидания и среднего квадратического отклонения распределения спроса этого временного среза определяются как а = -L — и п ЕГ=І( І-Ю2 о — ——. п Пример реализации данного метода представлен в вышеупомянутой статье.  Методы Projection Detruncation (PD) и Expectation Maximization (EM) — являются статистическими итерационными методами, состоящими на каждой итерации их двух шагов: оценочного шага (Е-шага), на котором производится замена всех цензурированных элементов выборки их верхними оценками и максимизационного (М-шага), на котором производится оценка параметров распределения за счет максимизации логарифмической функции правдоподобия, построенной по восстановленным на первом шаге данным. Оба шага повторяются до сходимости. Отличие методов состоит в выборе значений, которые будут взяты в качестве верхних оценок. Формулы для расчета верхних оценок в рамках каждого метода будут отличаться в зависимости от предположения о типе распределения общего спроса. Метод Expectation maximization  Термин «ЕМ» был введен в (Dempster, Laird and Rubin [39]). Как уже было упомянуто, каждая итерация метода ЕМ состоит из оценочного шага (Е-шага), на котором производится замена всех цензурированных элементов выборки их верхними оценками и максимизационного (М-шага), на котором ищется максимум логарифмической функции правдоподобия, построенной по восстановленным на E-шаге данным. Оценки параметров распределения общего спроса на каждой итерации определяются указанной точкой максимума. Оценка влияния методов восстановления общего спроса на доход авиакомпании: сгенерированные исторические данные Моделирование множества покупателей.  В данном разделе будет описана процедура моделирования множества покупателей, желающих приобрести имеющиеся продукты. Запросы на приобретение представляют собой совокупность «реальных покупателей», которые приходят в авиакомпанию за билетами. Каждый элемент представляет собой тройку вида: {N,t,vl где iV - уникальный идентификатор запроса, t - время появления запроса в минутах, р - номер продукта, на который поступил данный запрос. При необходимости можно учитывать покупателей, желающих вернуть ранее купленный билет. В этом случае элемент будет представлять собой четверку {N,t,p,q}, где q - идентификатор события, принимающий значение, равное 1, если покупатель желает купить билет, и -1 - если вернуть ранее купленный. В рамках же проведенного расчета вероятности возвратов не брались в расчет. Запросы генерируются на основе параметров «реального» спроса, которые необходимо задать исходя из каких-либо соображений. Так предполагается, что приход покупателей подчинен дискретному нормальному закону распределения, который задан формулой (35) или гамма-распределению. Тогда для создания «реального» спроса используются параметры др и ар, где (др, ар) - параметры прогнозируемого спроса, заданные по продуктам. Для моделирования множества покупателей можно использовать и другие распределения.  Моделирование процесса продаж.  Имея множество запросов на приобретение продуктов, можно запускать процесс их продаж, для чего была создана специальная функция. Рассмотрим подробнее ее работу.  В качестве входных данных функция имеет информацию о вместимости рейса и рассчитанные с ее использованием вложенные пределы бронирования для каждого продукта (использовалась стандартная иерархия вложения). Как упоминалось ранее, множество запросов представляет собой тройки: {iV, t, р). При поступлении запроса функция проверяет, положительна ли вместимость самолета и предел бронирования для продукта с номером, указанном на третьем месте в тройке запроса. Если обе эти величины больше нуля, то происходит продажа: к доходу прибавляется стоимость билета, а из вместимости вложенных пределов бронирования продуктов вычитается единица, а сама тройка добавляется в список удовлетворенных запросов. Если же хотя бы одна из величин неположительная, то продажа не происходит.  По окончанию обработки всех поступивших запросов известен доход авиакомпании и список всех удовлетворенных запросов на покупку.  В данном разделе проведен анализ влияния методов восстановления общего спроса (метод N3, ЕМ, ML) на доход авиакомпании в рамках продаж авиабилетов на один рейс с четырьмя ценовыми классами. Результаты, полученные с помощью различных подходов, сравнивались как между собой, так и с ситуацией, когда факт цензурирования исходного спроса не берется во внимание.  Первым этапом моделирования являлось создание статистики продаж. Рассматривался случай на момент окончания продаж. Были сгенерированы продажи для четырех ценовых классов из расчета 100 вылетевших рейсов, при этом обеспечены 25, 50, 75, 90, 97% цензурирования спроса (процедура создания статистики продаж аналогична процедуре, описанной в разделе подчинен дискретному нормальному закону, который задан формулой (35).  На следующем этапе моделирования к сгенерированным данным были применены методы восстановления общего спроса, которые позволили получить оценки параметров распределения общего неограниченного спроса. В рамках методов Expectation maximization и максимального правдоподобия использовалось предположение о том, что спрос на продукты подчинен нормальному закону. Полученные оценки параметров распределения общего спроса представлены в Таблице 10. Использовались следующие обозначения: наивный метод Nave 3 – N3, метод Expectation Maximization - EM, метод максимального правдоподобия с полной информацией о цензурировании – ML и при ее отсутствии – ML-100, оценки, полученные по невосстановленной выборке, соответствующей историческим данным о продажах - Const. | |  | |  | | |  | | |