

Федеральное государственное автономное образовательное учреждение
высшего профессионального образования «Национальный
исследовательский университет «Высшая школа экономики»

Факультет экономических наук
Образовательная программа «Экономика»

БАКАЛАВРСКАЯ ВЫПУСКНАЯ РАБОТА

*«Выявление индивидуальных предпочтений с помощью
Байесовской иерархической мультиномиальной логит
модели»*

Выполнил: студент группы БЭК

139

Федотова Мирослава Андреевна

Научный руководитель: стар-
ший преподаватель департа-
мента прикладной экономики
Демешев Борис Борисович

Москва 2017

Содержание

1	Введение	4
2	Обзор литературы	8
3	Теоретические основы	14
3.1	<i>Функция полезности и вероятность выбора</i>	14
3.2	<i>Байесовский подход к оценке коэффициентов и предска- ниям</i>	19
3.3	<i>Математическое построение Байесовской агрегированной мультиномиальной логит модели</i>	20
3.4	<i>Математическое построение Байесовской иерархической мультиномиальной логит модели</i>	22
4	Практическая часть	26
4.1	<i>Описание данных</i>	26
4.2	<i>Байесовская иерархическая мультиномиальная логит мо- дель на примере данных по мобильным тарифам</i>	29
4.3	<i>Чувствительность Байесовской иерархической мульти- номиальной логит модели к априорным распределениям параметров</i>	32
4.4	<i>Различные спецификации Байесовской иерархической мульт- номиальной логит модели</i>	39
5	Заключение	46
6	Литература	48

1 Введение

Компьютерная революция прошлого столетия и прогресс в сборе и хранении данных породили новые тенденции во всех областях. В частности, это коснулось маркетинга, где одной из таких тенденций стало широкое применение совместного анализа (conjoint analysis). Он же, привлёкший к себе внимание академического и промышленного мира ещё в 1970-х годах, и по сей день является неотъемлемой частью маркетинговых исследований.

Своей популярностью совместный анализ обязан множеству задач, которые он способен решать. К их ряду можно отнести прогнозирование различных показателей фирмы, таких как: прибыльность и доля рынка для потенциального продукта, реакция потребителей на изменение ценовой политики, оптимальный набор характеристик для нового продукта, необходимость усовершенствований некоторых свойств текущих продуктов, влияние новых продуктов конкурентов на доход компании. На самом деле перечень вопросов, адресованных совместному анализу, намного шире. Однако все они обладают общей чертой: для ответа на них требуется найти вид функций полезностей потребителей, с помощью которых возможно симулирование поведения индивидов, а именно, выбор того или иного продукта.

Довольно распространённым методом для решения задачи поиска функций полезности является агрегированная мультиномиальная логистическая регрессия. С помощью неё можно получать адекватные оценки коэффициентов ¹, поскольку число респондентов в совместных исследованиях достаточно велико. Однако эти оценки будут характеризовать некоего среднего потребителя, что влечёт за собой высокую неточность в

¹Понятие "коэффициенты" вводится в [Разделе 3.1](#).

симуляциях выбора отдельных индивидов. Это, в свою очередь, делает ответы на обозначенные выше вопросы более размытыми. Кроме того, такое унифицированное представление потребителей упускает из виду возможность сегментации рынка, выбора ниши и целевой аудитории. В действительности же мы с вами знаем, насколько люди неодинаковы в своих вкусах, и для исследователя представляют особый интерес потребители во всём их разнообразии.

Совокупность всех вышеуказанных факторов свидетельствует об актуальности рассмотрения моделей, предполагающих гетерогенность индивидов. К таковым можно отнести Байесовский иерархический мультиномиальный логит, который занимает центральное место в данном исследовании.

До середины 1980-х Байесовский подход не использовался на практике ввиду невозможности подсчёта апостериорного распределения (за исключением сопряжённых априорных распределений к семейству функций правдоподобия). Появление методов сэмплирования, в частности, Монте Карло по схеме марковской цепи (МСМС), освободили от оков большое количество моделей, которые довольно быстро нашли своё применение в маркетинге. Однако в русскоязычной литературе они до сих пор не получили широкого распространения, несмотря на имеющийся на то спрос прикладного характера.

В рамках данного исследования была поставлена следующая научно-практическая цель — написание численной Байесовской иерархической мультиномиальной логит модели, её применение, а также сравнение её модификаций между собой и с агрегированным подходом.

Для достижения поставленной цели были выделены следующие задачи:

1.1 математическое построение Байесовской агрегированной мультиномиальной логит модели;

1.2 математическое построение Байесовской иерархической мультиномиальной логит модели;

2. изучение языка программирования Stan;

3.1 написание численной Байесовской агрегированной мультиномиальной логит модели на языке R в связке со Stan'ом;

3.2. написание численной Байесовской иерархической мультиномиальной логит модели на языке R в связке со Stan'ом;

4.1 применение написанных моделей в различных модификациях на данных о мобильных тарифах;

4.2 сравнение результатов различных моделей, применённых к данным о мобильных тарифах;

5. подведение итогов.

В качестве объекта исследования выступают Байесовские иерархические мультиномиальные логит модели.

Предметом исследования является построение численной Байесовской иерархической мультиномиальной логит модели на языках R и Stan, а также сравнение предсказательной силы её различных модификаций и агрегированной логистической регрессии.

Говоря о новизне данной работы, стоит отметить, что в русскоязычной литературе проблематика поиска индивидуальных коэффициентов в рамках совместного анализа практически не раскрыта. В англоязычных же источниках зачастую делается упор на теоретическую составляющую различных Байесовских моделей в маркетинге, но крайне мало информации о том, как провести численных анализ.

Для этих целей существует популярный во всём мире софт Sawtooth

Software, однако стоимость делает его недоступным для многих исследователей. Самостоятельное же написание иерархической модели, позволяющей считать индивидуальные коэффициенты, большинству социологов и маркетологов не под силу, в связи с недостаточной теоретической подготовленностью. Также имеется пакет в R, но он не обладает достаточной гибкостью (ChoiceModelR). Данное исследование стремится стать полезным для тех, кто занимается совместным анализом, но не имеет возможности воспользоваться популярным дорогостоящим софтом. В этом и заключается практическая значимость настоящей работы.

2 Обзор литературы

Основной целью анализа рынка является использование ценной информации для принятия важных стратегических решений, в частности, маркетинговых. В этом русле могут возникать совершенно различные вопросы. Например, фирма может быть заинтересована в предсказании рыночной ситуации в будущем. Ей может руководить желание узнать реакцию потребителей на рекламу или новый продукт. Компания может задаться целью выбрать оптимальную цену для своего товара или услуги. Она может быть заинтересована в том, какими людьми являются её потребители. Для ответа на эти и многие другие вопросы могут быть использованы различные методологии по исследованию рынка.

Глобально можно выделить два подхода: качественный и количественный. Первый помогает понять поведение людей с помощью интервьюирования отдельных индивидов, фокус-групп, проективных методик, ситуативного анализа, наблюдений за потребителями, этнографии (Carson, 2001; Goulding, 2005; Belk, 2007; Daymon, Holloway, 2010). Качественный анализ скорее направлен на общее формирование стратегии.

Настоящее исследование будет относиться к количественным методам, которые, основываясь на данных, преследуют получение устойчивых результатов. Они, в свою очередь, используются для принятия конкретных решений по определённым вопросам, спектр которых крайне широк. Однако нас будет интересовать именно исследование выбора потребителей, для чего используются модели дискретного выбора. Ряд источников охватывает сразу большое число различных моделей (Greene 2009; Franses, Паар, 2001). Другие же, наоборот, фокусируются на конкретных. Так, в статье Kuhfeld, 2001 автор подробно рассматривает применение мультиномиальной логистической регрессии на различных при-

мерах: выбор конфет, тканей, места проведения отпуска, бренда, продуктов питания, лекарств. Для каждого случая автор подробно описывает не только методологию и результаты, но и каждый шаг проведённого исследования (в частности, дизайн эксперимента, сбор и обработку данных). В другой статье (Dow, Endersby, 2004) авторы сравнивают мультиномиальную логит с пробит моделью на примере голосования на политических выборах. Они приходят к тому, что предпочтительнее оказывается логит модель ввиду своей компактной формы. А рассуждая о её главном недостатке - предположении о независимости альтернатив, авторы заключают, что на практике в большинстве случаев оно не является ограничительным.

В упомянутых выше статьях используется классический подход к оценке параметров. Однако многие исследователи применяют именно Байесовский подход, поскольку он даёт более ценную информацию - распределение параметров, а не их точечные оценки. Множество работ посвящено Байесовским мультиномиальным пробит и логит моделям (Albert, Chib, 1993; Allenby G. M., Lenk 1994, McCulloch, Polson, 2000; Imai, van Dyk, 2005). Объединяют все предложенные ранее в этом разделе статьи то, что авторы оценивают агрегированные предпочтения потребителей. Респонденты рассматриваются в своей общности, то есть предполагаются некие усреднённые предпочтения по всем индивидам. Такой подход является скудным описанием действительности, поскольку не предполагает различий между потребителями.

Некоторые исследователи обращаются к подходам, которые предполагают кластеризацию индивидов, например, алгоритм поиска латентных классов (DeSarbo 1992; Ramaswamy, Cohen, 2000; Magidson, Vermunt, 2004). Предпочтения потребителей, оценённые с помощью таких моде-

лей, разнятся от группы к группе, но одинаковы для принадлежащих к одной группе.

Однако существуют методы, которые позволяют получить полную гетерогенность индивидов, что подразумевает наличие индивидуального вектора полезностей для каждого человека. Речь идёт об иерархических Байесовских моделях. С их помощью в статье Solgaard, Hansen, 2003 авторы выявляли индивидуальные предпочтения потребителей относительно различных видов супермаркетов. Другие исследователи (Bradlow Е. Т., Рао, 2000) использовали тот же подход для ответа вопрос о том, какой ассортимент наиболее привлекателен для каждого отдельного клиента .

В действительности иерархический Байесовский подход особенно популярен в совместных исследованиях (conjoint analysis), которые являются одним из основных инструментов в маркетинге. В качестве примера применения иерархического Байесовского подхода в совместных исследованиях можно привести статью Wellman, Vidican, 2008, где авторы оценивали индивидуальные предпочтения потребителей относительно различных характеристик страховых планов. В работе Loock, 2012 анализировалось поведение инвесторов, вкладывающих средства в возобновляемые источники энергии. На Байесовских иерархических мультиномиальных логит моделях также основано множество других исследований (Sinha, Gazley, Ashill, 2008; Yu, Goos, Vandebroek, 2011; Byun H., Lee 2017).

В научной работе Orme, 2009 показано, что данный подход лучше других прогнозирует выбор потребителей, однако имеются некоторые ограничения по его применимости. Байесовские иерархические модели реализованы в программе Sawtooth Software, но её использование сопря-

жено с большими денежными вложениями. Также имеется пакет в R (ChoiceModelR), но он недостаточно гибок по своим характеристикам. В связи с упомянутыми причинами написание Байесовской иерархической мультиномиальной модели представляет практическую ценность для совместных исследований. Однако прежде, чем описывать эту модель и результаты, которые она даёт, ознакомим читателя поподробнее с тем, что из себя представляет совместный анализ.

Появление совместного анализа предвосхитила статья, опубликованная в журнале, не имеющем отношения к маркетингу (Luce and Tukey, 1964). Методы, использовавшиеся в ней, были подмечены другими исследователями, как возможный способ объяснения поведения покупателей, оценки их предпочтений (важности определённых свойств продуктов) и предсказания их выбора в будущем (Green, Rao, 1971). Эта работа положила начало бурному развитию нового направления в маркетинге. С тех пор появилось огромное количество литературы о совместном подходе.

Итак, обратимся к одной из книг (Orme, 2010), посвящённых совместному анализу, чтобы пояснить, что он из себя представляет. Эта методология основана на непосредственном сборе информации для каждого отдельного исследования, а не на исторических данных. В процессе анализа проводится тщательно разработанный эксперимент, который нацелен на выявление потребительских предпочтений. Во время него опрашиваемым предоставляется некоторое число продуктов, описанных определённым набором характеристик (другими словами: черт, атрибутов). Иногда продукты также сопровождаются картинкой. Задача респондентов может быть сформулирована несколькими способами: выбрать наиболее понравившийся продукт, оценить все продукты по некоторой шкале или отранжировать в порядке предпочтительности.

Автор (Orme, 2010) уделяет особое внимание преимуществам совместного анализа. Он отмечает, что черты, которыми наделяются продукты, могут как уже существовать на рынке, так и быть представленными лишь в будущем. В связи с этим совместный анализ позволяет изучать сценарии появления абсолютно новых продуктов. Ни одна методология, основанная на исторических данных, не может справиться с подобной задачей.

Само название «conjoint» подразумевает, что респонденты должны оценить продукт во всей его целостности, то есть их предпочтения относительно каждой характеристики рассматриваются совместно.

Существуют также методологии, нацеленные на выявление предпочтений относительно атрибутов самих по себе. Они могут базироваться на заданиях следующего характера: оцените важность наличия двух сим-карт в телефоне; выберите максимальную цену, которую Вы готовы заплатить за бутылку воды.

При таком сравнении подходов проявляется главное преимущество совместного анализа, на котором акцентируют внимание автор (Orme, 2006): эксперимент максимально походит на реальность. В действительности потребители выбирают именно между целостными продуктами, поэтому эксперимент формата совместного анализа им близок. А вопросы о важности отдельных характеристик в количественном выражении могут вводить респондентов в заблуждение, поскольку людям крайне трудно определить своё отношение к конкретной черте, не рассматривая продукт в целом. Опрашиваемые склонны утверждать, что многие атрибуты настолько необходимы для конкретного товара, что без них просто не стали бы приобретать его, но в реальности же всё обстоит иначе.

Как было упомянуто ранее, существует три типа подходов в совмест-

ных исследованиях. На практике наиболее популярным является совместный анализ, основанный на дискретном выборе (choice-based conjoint, СВС). Это объясняется тем, что данный тип задания наиболее близок к поведению потребителя в реальной среде, где приходится принимать решение о покупке товара, сравнивая его с субститутами (Johnson, Orme, 2007). Эксперимент обычно выглядит следующим образом: каждому респонденту предоставляются несколько карточек. Каждая карточка содержит набор альтернатив (профилей), число которых варьируется в от 2 до 5, поскольку в реальной жизни люди обычно выбирают из такого количества вариантов. Респондент должен отметить наиболее понравившийся профиль на каждой карточке. Именно такой тип исследований будет положен в основу данной работы для написания Байесовской иерархической мультиномиальной логит модели.

3 Теоретические основы

3.1 Функция полезности и вероятность выбора

Модели дискретного выбора представляют собой способ описания выбора рациональных агентов. Данный подход основан на идее о том, что поведение индивида зависит от его функции полезности. Совершая выбор, он сравнивает полезности от различных альтернатив и отдаёт предпочтение той, которая доставляет ему наибольшую полезность.

Будем предполагать, что имеется N агентов и J альтернатив. Обозначим истинную функцию полезности агента n от альтернативы j за U_{nj} , а его выбор за y_n . С этого момента и далее, используя индексы n , j , nj , будем подразумевать $\forall n \subset N, \forall j \subset J$, если не оговорено иное.

Нам неизвестна истинная функция полезности индивидов, однако наша промежуточная цель ² - понять, как потребитель совершает свой выбор. Поэтому создаётся смоделированная полезность, которую будем обозначать за V_{nj} для агента n , выбравшего альтернативу j .

Для построения V_{nj} используется вся имеющаяся информация об альтернативах, а именно, их характеристики. Так, например, если бы речь шла о телевизорах, каждая альтернатива могла бы быть описана диагональю экрана, брендом, стоимостью и другими атрибутами. Характеристики альтернативы j , предложенной индивиду n , будем описывать вектором x_{nj} . Предполагается также наличие вектора коэффициентов β_n , который связывает функцию полезности индивида n со свойствами альтернативы j , то есть $V_{nj} = V(x_{nj}, \beta_n)$.

²Конечная цель при проведении совместного анализа может быть иметь, как было сказано ранее, следующий вид: прогнозирование различных показателей фирмы, таких как: прибыльность и доля рынка для потенциального продукта, реакция потребителей на изменение ценовой политики, оптимальный набор характеристик для нового продукта, необходимость усовершенствований некоторых свойств текущих продуктов, влияние новых продуктов конкурентов на доход компании

Истинная полезность складывается из смоделированной и некой случайной ошибки: $U_{nj} = V_{nj} + \varepsilon_{nj}$, где ε_{nj} включает в себя факторы, которые влияют на истинную полезность U_{nj} , но не описываются V_{nj} .

Далее необходимо специфицировать аналитический вид V_{nj} и ε_{nj} . Будем описывать функции полезности с помощью линейной формы, то есть: $V_{nj} = \beta'_n x_{nj}$, где вектора β_n и x_{nj} имеют размерность $A+1$, A - количество атрибутов, описывающих каждую альтернативу. Ошибки ε_{nj} неизвестны, поэтому рассматриваются как случайные величины. Будем предполагать, что остатки независимо и одинаково распределены по Гумбелю, тогда их функция плотности имеет вид:

$$f_{\varepsilon_{nj}}(x) = \frac{\exp(-x)}{\sigma} \cdot \exp\left(\frac{-x - \mu}{\sigma}\right). \quad (3.1.1)$$

Функция распределения в таком случае выражается следующим образом:

$$F_{\varepsilon_{nj}}(x) = \exp\left(-\exp\left(-\frac{-x - \mu}{\sigma}\right)\right). \quad (3.1.2)$$

Мы будем рассматривать стандартное распределение Гумбеля, то есть $\mu = 0$, $\sigma = 1$, тогда функция распределения случайной величины ε_{nj} примет более простой вид:

$$F_{\varepsilon_{nj}}(x) = \exp(-\exp(-x)). \quad (3.1.3)$$

Напомним нашу задачу. Для предсказания выбора индивида нам нужно понять, как устроена его функция полезности. Поскольку вектор x_{nj} известен для всех индивидов и альтернатив, остаётся оценить вектор параметров β_n для $\forall n \subset N$ исходя из предпосылок, введённых выше.

Для оценки параметров (как при использовании классического под-

хода, так и Байесовского) требуется знать вид функции вероятности, с коей потребитель выберет конкретную альтернативу.

Итак, вероятность, что индивид n выберет альтернативу i может быть представлена следующим образом:

$$P_{ni} = P(y_n = i) = P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i) =$$

$$P(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj}, \forall j \neq i). \quad (3.1.4)$$

Как было сказано ранее, случайные величины ε_{ni} и ε_{nj} независимы и одинаково распределены в соответствии со стандартным распределением Гумбеля. Тогда:

$$F_{\varepsilon_{nj}}(\varepsilon_{ni} + V_{ni} - V_{nj}) = \exp(-\exp(-(\varepsilon_{ni} + V_{ni} - V_{nj}))). \quad (3.1.5)$$

По предположение о независимости случайных величин ε_{ni} и ε_{nj} имеем:

$$P(y_n = i | \varepsilon_{ni}) = \prod_{j \neq i} \exp(-\exp(-(\varepsilon_{ni} + V_{ni} - V_{nj}))). \quad (3.1.6)$$

Получаем:

$$\begin{aligned} P(y_n = i) &= \int P(y_n = i, \varepsilon_{ni}) d\varepsilon_{ni} = \int P(y_n = i | \varepsilon_{ni}) P(\varepsilon_{ni}) d\varepsilon_{ni} = \\ &= \int \prod_{j \neq i} \exp(-\exp(-(\varepsilon_{ni} + V_{ni} - V_{nj}))) \exp(-\varepsilon_{ni}) \exp(-\exp(-\varepsilon_{ni})) d\varepsilon_{ni}. \end{aligned} \quad (3.1.7)$$

Для удобства положим $\varepsilon_{ni} = a$, тогда

$$P_{ni} = \int_{-\infty}^{+\infty} \prod_{j \neq i} \exp(-\exp(-(a + V_{ni} - V_{nj}))) \exp(-a) \exp(-\exp(-a)) da. \quad (3.1.8)$$

Заметим, что при $i = j$:

$$\exp(-\exp(-a + V_{ni} - V_{nj})) = \exp(-\exp(-a + V_{ni} - V_{ni})) = \exp(-\exp(-a)). \quad (3.1.9)$$

Тогда выражение (3.1.8) можем переписать следующим образом:

$$P_{ni} = \int_{-\infty}^{+\infty} \prod_{\forall j} \exp(-\exp(-(a + V_{ni} - V_{nj}))) \exp(-a) da. \quad (3.1.10)$$

Перемножив все экспоненты, имеем:

$$\begin{aligned} P_{ni} &= \int_{-\infty}^{+\infty} \exp\left(-\sum_{\forall j} \exp(-(a + V_{ni} - V_{nj}))\right) \exp(-a) da = \\ &= \int_{-\infty}^{+\infty} \exp(-\exp(-a)) \sum_{\forall j} \exp(-(V_{ni} - V_{nj})) \exp(-a) da. \end{aligned} \quad (3.1.11)$$

Для вычисления интеграла воспользуемся заменой $\exp(-s) = t$, тогда получим:

$$- \int_{+\infty}^0 \exp(-t \sum_{\forall j} \exp(-(V_{ni} - V_{nj}))) dt = \int_0^{+\infty} \exp(-t \sum_{\forall j} \exp(-(V_{ni} - V_{nj}))) dt =$$

$$\begin{aligned}
&= \frac{\exp(-t \sum_{\forall j} \exp(-(V_{ni} - V_{nj}))}{-\sum_{\forall j} \exp(-(V_{ni} - V_{nj}))} \Big|_0^{+\infty} = \frac{1}{\sum_{\forall j} \exp(-(V_{ni} - V_{nj}))} = \\
&= \frac{1}{\exp(-V_{ni}) \sum_{\forall j} \exp(V_{nj})} = \frac{\exp(V_{ni})}{\sum_{\forall j} \exp(V_{nj})}. \tag{3.1.12}
\end{aligned}$$

Таким образом, вероятность выбора индивидом n альтернативы i есть логистическая функция:

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_{\forall j} \exp(V_{nj})}. \tag{3.1.13}$$

Тот факт, что вероятность выбора, выражается в краткой форме - преимущество логит модели перед другими моделями дискретного выбора. В первую очередь по причине того, что это упрощает генерирование выборки. Например, в случае пробит-модели, пришлось бы аппроксимировать n -мерный интеграл.

Однако исследователь должен крайне внимательно отнестись к предположению о независимости случайных величин ε_{ni} и ε_{nj} , которое мы использовали при выводе вероятности выбора альтернативы i индивидом n .

Остатки ε_{ni} и ε_{nj} - это ненаблюдаемая часть полезности, получаемой индивидом n при выборе альтернативы i или j , соответственно. Иными словами, это разница между реальной полезностью и смоделированной исследователем. Получается, необходимым условием независимости этих остатков является хорошо специфицированная модель, в которую включена вся информация, используемая потребителем в процессе выбора. Тогда ненаблюдаемая часть полезности будет лишь шумом.

3.2 Байесовский подход к оценке коэффициентов и предсказаниям

Для умения объяснять и предсказывать выбор потребителей, нужен способ для оценки вектора β_n для каждого $n \in N$. В данной работе для этой цели будет использоваться Байесовский подход.

Вкратце напомним, что он из себя представляет. Обозначим за β все параметры модели, а за D все имеющиеся данные (в нашем случае y и X ³). Наша доопытная вера в значения параметров будет описываться априорной функцией распределения $p(\beta)$. Информацию о наших данных будет отражать функция правдоподобия $p(D|\beta)$ ⁴. Тогда по теореме Байеса апостериорная функция распределения параметров может быть выражена следующим образом:

$$p(\beta|D) = \frac{p(D, \beta)}{p(D)} = \frac{p(D|\beta) \cdot p(\beta)}{p(D)}. \quad (3.2.1)$$

Знаменатель представим в виде: $p(D) = \int p(D, \beta) d\beta = \int p(D|\beta) p(\beta) d\beta$. Поскольку $p(D)$ - константа, выражение (3.2.1) можем переписать следующим образом:

$$p(\beta|D) \propto p(D|\beta) \cdot p(\beta). \quad (3.2.2)$$

Рассчитать отсюда функцию плотности апостериорного распределения можно лишь в случае сопряжённого априорного распределения к семейству функций правдоподобия ⁵. В остальных же случаях, то есть

³Под y понимается совокупность выборов, совершённых всеми респондентами, под X - характеристики всевозможных альтернатив, предложенных респондентам во время опроса. Стоит напомнить, что профили не повторяются, у каждого человека свои индивидуальные карточки.

⁴Также обозначается как $l(\beta)$.

⁵ Распределение $p(\theta)$ называется сопряжённым априорным распределением к семейству функций правдоподобия $p(x|\theta)$, если апостериорное распределение $p(\theta|x)$ принадлежит тому же семейству вероятностных распределений, что и априорное распределение $p(\theta)$ (т.е. имеет тот же вид, но с

почти во всех, приходится прибегать к численным методам ⁶. С их помощью можно получить выборку из апостериорного распределения, а далее рассчитать по ней искомую функцию от параметра (например, среднее, медиану, моду).

Предсказания, так же как и параметры, определяются утверждением об их распределении. Обозначим показатели, которые мы хотим предсказать, за D_f , тогда:

$$p(D_f|D) = \int p(D_f, \beta|D)d\beta = \int p(D_f|\beta, D)p(\beta|D)d\beta. \quad (3.2.3)$$

В большинстве случаев предполагается, что D и $D_f|\beta$ независимы. Тогда предиктивное распределение принимает вид:

$$p(D_f|D) = \int p(D_f|\beta)p(\beta|D)d\beta. \quad (3.2.4)$$

3.3 Математическое построение Байесовской агрегированной мультиномиальной логит модели

В данной работе, как было сказано ранее, будет рассмотрен совместный анализ, основанный на дискретном выборе (choice based conjoint, CBC). В действительности он же и является самым распространённым в силу своих преимуществ.

Итак, напомним эксперимент. Каждому респонденту предоставляются карточки. Каждая карточка содержит набор альтернатив. Респондент

другими параметрами).

⁶Именно поэтому Байесовский подход долгое время не мог применяться на практике.

должен выбрать наиболее понравившуюся альтернативу на каждой карточке.

Обозначим число респондентов за N , число карточек за C , число альтернатив за J .

Поскольку было опрошено N человек, каждому из которых предлагалось C карточек, содержащих по J альтернатив, описанных A характеристиками, функция правдоподобия принимает вид: ⁷

$$p(y, X|\beta) = \prod_{n=1}^N \prod_{c=1}^C \prod_{j=1}^J P(y_{nc} = j)^{I[y_{nc}=j]}. \quad (3.3.1)$$

Используя (3.1.1), выпишем вероятность выбора индивидом n альтернативы i на карточке c , представляющую собой логистическую функцию:

$$P(y_{nc} = i) = \frac{\exp(V_{nci})}{\sum_{\forall j}^J \exp(V_{ncj})}. \quad (3.3.2)$$

В агрегированном случае, когда мы не предполагаем гетерогенности индивидов и рассматриваем некие усреднённые предпочтения, смоделированная полезность имеет вид: $V_{ncj} = \beta' x_{ncj}$. Тогда выражение (3.3.2) преобразуется:

$$P(y_{nc} = i) = \frac{\exp(\beta' x_{nci})}{\sum_{\forall j}^J \exp(\beta' x_{ncj})}. \quad (3.3.3)$$

А функция правдоподобия, в свою очередь, принимает вид:

$$p(y, X|\beta) = \prod_{n=1}^N \prod_{c=1}^C \prod_{j=1}^J \left(\frac{\exp(\beta' x_{nci})}{\sum_{\forall j}^J \exp(\beta' x_{ncj})} \right)^{I[y_{nc}=j]}. \quad (3.3.4)$$

Байесовский подход также требует от нас введения априорных распределений параметров. Обратимся к руководству по языку Stan (Stan

⁷Величины C , J и A одинаковы для всех респондентов.

Development Team, 2016, версия 2.12.0, стр. 55). В случае логистической регрессии авторы рекомендуют использовать слабо-информативно распределение о параметрах, а именно:

$$\beta \sim Normal(0, 5), \quad (3.3.5)$$

где элементы вектора $\beta = (\beta_1, \beta_2, \dots, \beta_A)$ независимы.

3.4 *Математическое построение Байесовской иерархической мультиномиальной логит модели*

Иерархической модель - способ описать индивидов во всём разнообразии их предпочтений. Поскольку в таком случае каждый респондент имеет свой вектор β_n , его смоделированную полезность можно представить как $V_{ncj} = \beta_n' x_{ncj}$. Тогда функция правдоподобия принимает вид:

$$p(y, X|\beta) = \prod_{n=1}^N \prod_{c=1}^C \prod_{j=1}^J \left(\frac{\exp(\beta_n' x_{nci})}{\sum_{\forall j} \exp(\beta_n' x_{ncj})} \right)^{I[y_{nc}=j]}. \quad (3.4.1)$$

При выборе априорных распределений будем полагаться на кембриджское пособие по анализу данных с помощью иерархических моделей (Gelman and Hill, 2007, гл. 13, гл. 17). Авторы предполагают, что для каждого индивида вектор коэффициентов β_n имеет многомерное нормальное распределение с вектором средних μ и ковариационной матрицей Σ :

$$\beta_n \sim MultiNormal(\mu, \Sigma). \quad (3.4.2)$$

Авторы включают дополнительные знания об индивидах в распределение μ , но мы не располагаем таковыми, поэтому будем использо-

вать слабо-информативное априорное распределение для вектора средних, элементы которого независимы:

$$\mu_n \sim \text{Normal}(0, 5). \quad (3.4.3)$$

В качестве априорного распределения для ковариационной матрицы авторы (Gelman and Hill, 2007, гл. 13, гл. 17) предлагают использовать обратное распределение Вишарта, мотивируя это тем, что данное предположение упрощает сэмплирование по Гиббсу. Но мы не последуем за ними, поскольку для численной модели будем использовать язык Stan, в котором заложен алгоритм Гамильтона.

Итак, в руководстве по Stan (Stan Development Team, 2016, версия 2.12.0, стр. 73) рекомендуется определять ковариационную матрицу Σ через корреляционную матрицу Ω и вектор стандартных отклонений τ :

$$\Sigma = D(\tau) \cdot \Omega \cdot D(\tau), \quad (3.4.4)$$

где $D(\tau)$ - матрица, элементы которой на главной диагонали совпадают с элементами вектора τ , а остальные - нули.

Продолжая следовать рекомендациям из руководства по Stan, зададим следующие распределения для τ и Ω :

$$\tau_a \sim \text{Cauchy}(0, \gamma), \tau_a > 0, \quad (3.4.5)$$

$$\Omega \sim \text{LKJcorr}(\nu), \nu \geq 1. \quad (3.4.6)$$

Поскольку распределение LKJ не достаточно популярно, остановимся на нём поподробнее. Будем использовать Stan, чтобы сперва получить

выборку из этого распределения, а затем визуализировать его ⁸.

Для понимания того, что из себя представляет распределение корреляционной матрицы LKJ , будем рассматривать распределение одного элемента корреляционной матрицы. Другие же будут распределены аналогично (за исключением диагональных элементов по понятным причинам).

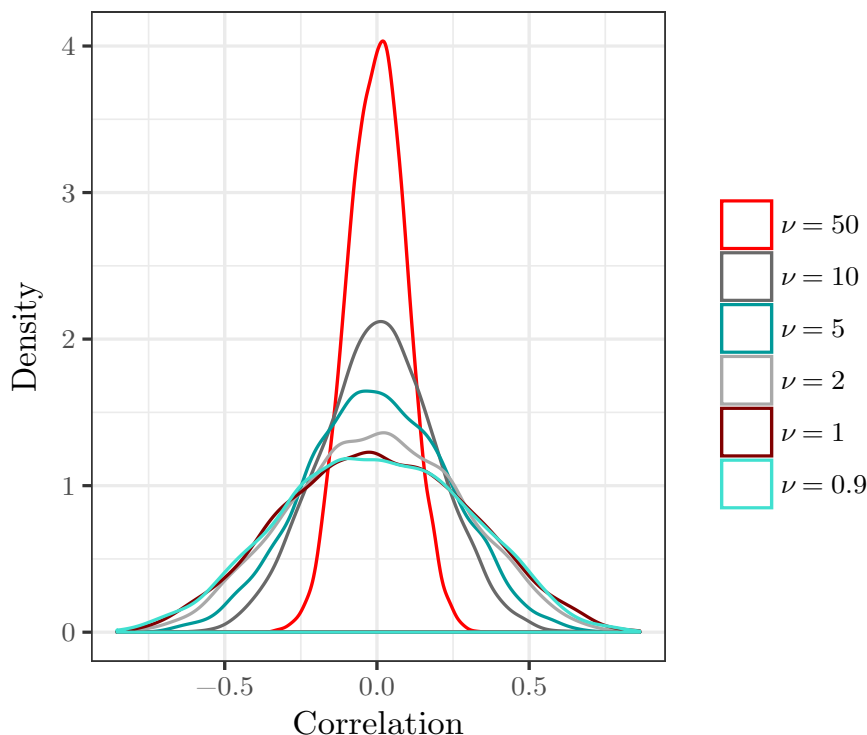


Рис. 1: Визуализация корреляции из LKJ распределения

График демонстрирует нам то, что подтверждается теорией. С ростом параметра ν корреляционная матрица начинает больше походить на единичную. С помощью LKJ распределения можно контролировать ожидаемую корреляцию между параметрами β_a .

Итак, для внесения полной ясности в априорные распределения нашей модели, заложим в базовую модификацию значения параметров γ и ν , рекомендуемые в руководстве по Stan:

⁸В Приложении 1 приведён код для сэмплирования из ЛКЖ распределения и визуализации.

$$\tau_a \sim \textit{Cauchy}(0, 2.5), \quad (3.4.7)$$

$$\Omega \sim \textit{LKJcorr}(2). \quad (3.4.8)$$

4 Практическая часть

4.1 Описание данных

Перейдём к нашему практическому примеру, на котором строится Байесовская мультиномиальная иерархическая логит модель.

Эксперимент для получения данных проводился следующим образом: каждому из 296 респондентов было предложено 7 карточек, каждая из которых содержала 5 альтернатив. Каждая альтернатива описывалась 7 характеристиками мобильного тарифа: количество доступных минут в месяц, количество доступных гигабайт в месяц, ежемесячный платёж, мобильный оператор, возможность персонализации, характеристика счёта: общий/раздельный, количество сим-карт. Каждый респондент должен был выбрать на каждой карточке одну альтернативу, которая казалась ему наиболее привлекательной.

Итак, в [Таблице 1](#) продемонстрированы уровни каждого атрибута, иными словами, значения, которые может принимать каждая характеристика.

Таблица 1: Дизайн эксперимента

Кол-во минут	Кол-во гигабайт	Плата	Оператор	Персонализация	Счёт	Кол-во сим-карт
100	1	150	1	Есть	Общий	1
200	2	300	2	Нет	Раздельный	2
300	3	450	3			3
400	4	600	4			
500	5	750				
600	6					
700	7					
	8					
	9					
	10					

Стоит отметить, что дизайн эксперимента был разработан не нами, а компанией, которая занимается маркетинговыми исследованиями. То же касается и сбора данных. Тем не менее мы обсудим несколько основных моментов.

Профили генерируются таким образом, чтобы не повторяться, поэтому каждый респондент получает свои индивидуальные карточки. Однако для анализа не используют всевозможные комбинации уровней, иначе число альтернатив было бы слишком велико ($\sum_{a=1}^A q(l_a)$, где a - номер атрибута, $q(l_a)$ - количество уровней для атрибута a). Чтобы уменьшить число сочетаний, составляется ортогональный план. Этот метод обладает рядом преимуществ, основным из которых является получение качественных предсказаний (Рао, 2013, гл. 2, стр. 47-48). Условие формирования ортогонального плана завязано на правиле пропорциональности: каждый уровень атрибута a_i должен встречаться в k раз чаще каждого уровня атрибута a_j , где $k = \frac{q(l_{a_j})}{q(l_{a_i})}$.

Для наглядности в [Таблице 2](#) приведён пример карточки, которая попала одному из респондентов.

Таблица 2: Карточка

Кол-во минут	Кол-во гигабайт	Плата	Оператор	Персонализация	Счёт	Кол-во сим-карт
400	1	600	4	Есть	Раздельный	2
300	5	750	3	Есть	Раздельный	1
700	2	300	1	Нет	Общий	3
200	8	150	1	Нет	Общий	3
100	3	450	2	Нет	Общий	2

Для анализа данных нам необходимо перекодировать значения уровней. Тогда из [Таблицы 2](#) получаем [Таблицу 3](#) и [Таблицу 4](#)

Исходные данные были получены в файле с расширением .sav. Когда же мы загружаем их в R и приводим в наглядный вид, у каждого челове-

Таблица 3: Преобразованные уровни атрибутов

Количество минут	Количество гигабайт	Ежемесячный платёж	Оператор 1	Оператор 2
1	1	1	1	1
2	2	2	0	0
3	3	3		
4	4	4		
5	5	5		
6	6			
7	7			
	8			
	9			
	10			

Таблица 4: Преобразованные уровни атрибутов

Оператор 3	Наличие персонализации	Наличие общего счёта	2 сим-карты	3 сим-карты
1	1	1	1	1
0	0	0	0	0

ка появляется своя матрица X и вектор y . Столбцы матрицы X отвечают за характеристики в перекодированном виде, аналогично [Таблице 4](#). Также матрица X содержит единичный столбец. Строки отражают альтернативы, при том карточки выставляются друг за другом. Вектор y содержит выбор человека: номера наиболее предпочитаемых альтернатив, представленных на предложенных ему карточках. Таким образом, каждому человеку соответствует собственная матрица X размерности 35 на 11 и вектор y длиной 7.

4.2 *Байесовская иерархическая мультиномиальная логит модель на примере данных по мобильным тарифам*

Напомним, что нас интересуют не только оценки индивидуальных коэффициентов сами по себе, но и возможность на их основе проводить симуляции в различных рыночных условиях. В связи с этим модель подразумевает наличие тренировочной и тестовой выборки, то есть для каждого человека имеется две матрицы X , \tilde{X} и один вектор y . Далее исследователь получает выборку⁹ из апостериорного распределения параметров β_n и \tilde{y}_n , благодаря которой может найти желаемую функцию от параметра, например, среднее или моду.

Мы не станем приводить код модели в основном тексте данной работы, однако обсудим параметры, используемые численной моделью, но не введённые при описании математической, а именно L и ll . Их добавлением мы хотели добиться большей гибкости. Созданная нами иерархическая модель позволяет вводить любое количество кластеров от 1 до N , за которое отвечает параметр L . Если мы объединим всех людей в одну группу, придём к агрегированной логит модели. Если же мы впадём в другую крайность, разделив всех людей на N групп, получим индивидуальные коэффициенты для каждого из N респондентов. Параметр ll отвечает за принадлежность индивида к некоторой группе. Например, если мы заинтересованы именно в индивидуальных коэффициентах, ll можно представить в виде вектора, элементами которого являются последовательно идущие числа от 1 до N .

Итак, сперва будем полагать, что $L = 296$, добиваясь максимальной

⁹В нашем случае выборка состоит из 2000 значений для каждого параметра

гетерогенности в предпочтениях индивидов.

Далее разделим выборку на тренировочную и тестовую. В нашей работе мы делаем это для оценки качества модели. В реальной же ситуации компания может обучить модель на данных, собранных с помощью проведённого ею эксперимента, а затем получить симуляции выбора индивидов на основе характеристик товаров, действительно присутствующих на рынке. Полученные результаты могут быть пересчитаны в так называемую долю рынка, исходя из которой уже принимаются стратегические маркетинговые решения.

Итак, вернёмся к нашим данным. Деление на тренировочную и тестовую выборку в нашем случае подразумевает следующее: из профиля каждого человека рандомно исключаем одну карточку. Таким образом, в тренировочную выборку входит матрица X размерности 30 на 11, а в тестовую - матрица \tilde{X} размерности 5 на 11. К сожалению, созданная нами модель требует одинакового количества карточек для всех индивидов как в тренировочной, так и в тестовой выборке. Поэтому мы не исключаем человеко-карточки случайным образом.

После приведения данных в требуемый моделью вид и запуска модели, было получено 2000 значений каждого параметра из его апостериорного распределения. Для наглядности посчитаем среднее для β и приведём в [Таблице 5](#) значения коэффициентов при нескольких атрибутах для первых четырёх респондентов.

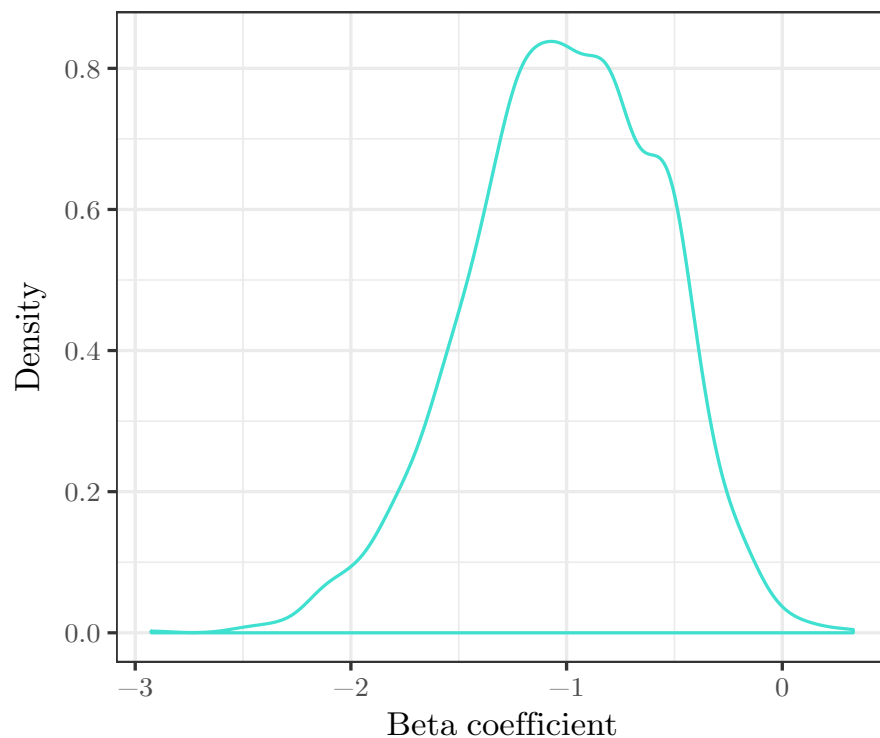
Однако мы можем иначе взглянуть на полученные оценки коэффициентов. Одним из преимуществ Байесовского подхода является возможность извлечения большей информации, нежели при использовании классического подхода. На основе полученной выборки построим оценённую функцию плотности коэффициента β для атрибута 'Плата' для

Таблица 5: Оценки индивидуальных коэффициентов

Номер инд.	Свободный коэфф.	Количество минут	Количество гигабайт	Плата	Оператор 1
1	-0.1065607	0.80699223	0.2621687	-0.9336447	0.7833177
2	0.2070330	0.56243958	0.2413598	-0.4733364	0.1433198
3	0.1742506	0.45178105	0.2142404	-0.3351748	-0.1039047
4	-0.2856412	0.04208365	0.2523574	-1.0340263	-0.7794623

респондента под номером 103 Рис.2.

Рис. 2: Оценённое апостериорное распределение коэффициента $\beta_{9,103}$



Также с помощью Байесовской иерархической мультиномиальной логит модели мы получили 2000 так называемых предсказаний для каждого из 296 респондентов. Сперва мы можем опять же найти некоторую функцию от полученной выборки из апостериорного распределения \tilde{y} . Поскольку данная случайная величина является дискретной, считать среднее мы не станем. Разумнее будет, например, вычислить оценку вероятности выбора той или иной альтернативы конкретным человеком.

Данный результат можно найти в [Таблице 6](#).

Таблица 6: Оценка вероятности выбора различных альтернатив

Номер инди- вида	Вер-ть выбора альт. 1	Вер-ть выбора альт. 2	Вер-ть выбора альт. 3	Вер-ть выбор альт. 4	Вер-ть выбора альт. 5
217	0.038	0.2595	0.0905	0.596	0.016
55	0.007	0.046	0.112	0.6115	0.2235
3	0.077	0.1475	0.1775	0.0405	0.5575
4	0.037	0.0395	0.1245	0.0895	0.7095

С помощью данных оценённых вероятностей можно высчитать, как уже не раз было сказано, "долю рынка".

4.3 *Чувствительность Байесовской иерархической мультиномиальной логит модели к априорным распределениям параметров*

Естественно, могут возникнуть сомнения в устойчивости результатов, полученных с помощью Байесовской иерархической мультиномиальной логит модели. Основной причиной тому большое число априорных распределений параметров, заданных, казалось бы, наугад.

Чтобы проверить, насколько сильно априорные распределения влияют на апостериорный вывод, изменим базовую спецификацию модели, описанную в [Разделе 3.4](#). Вспомним, что были введены следующие предположения:

$$\mu_n \sim \text{Normal}(0, 5)$$

$$\tau_a \sim \text{Cauchy}(0, 2.5),$$

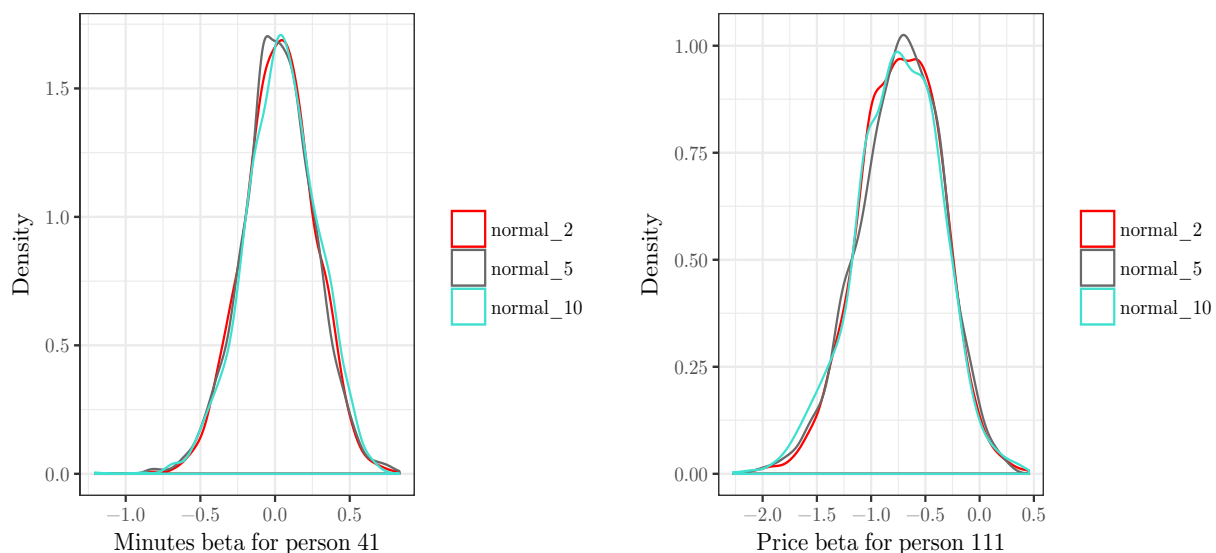
$$\Omega \sim \text{LKJcorr}(2).$$

Теперь выберем случайно несколько респондентов и номеров коэффициентов β . Генерируя два целых числа от 1 до 296 и два целых числа от 1 до 11¹⁰, получаем следующие номера индивидов: 41, 111, и коэффициентов: 6, 9 (отвечающие за количество доступных минут и сумму ежемесячного платежа, соответственно). Будем менять априорные распределения и смотреть, как это отражается на оценённом апостериорном распределении β .

Далее, под выражением "апостериорное распределение" будем подразумевать распределение, полученное на основе выборки из 2000 значений.

Начнём с параметра μ . Сравним апостериорные распределения β при $\mu_n \sim Normal(0, 5)$, $\mu_n \sim Normal(0, 2)$ и $\mu_n \sim Normal(0, 10)$ для обозначенных выше индивидов и атрибутов.

Рис. 3: Апостериорные распределения β при различных априорных распределениях μ_n



Как видно из графиков (Рис.3) апостериорные распределения β при

¹⁰ С помощью `set.seed(219) floor(runif(2, min = 1, max = 297))` и `set.seed(69) floor(runif(2, min = 1, max = 12))`

базовой спецификации не сильно отличаются от модифицированных. Тем не менее проведём тест Колмогорова-Смирнова. За уровень значимости возьмём $\alpha = 0.05$.

Сравним сперва апостериорные распределения при $\mu_n \sim Normal(0, 5)$ и $\mu_n \sim Normal(0, 2)$. Для индивида 41 получаем $p - value = 0.8186$, для индивида 111: $p - value = 0.7947$. Таким образом, при выбранном уровне значимости мы не можем опровергнуть нулевую гипотезу о том, что апостериорные выборки ¹¹ для $\beta_{41,6}$ и $\beta_{111,9}$ при априорном распределении $\mu_n \sim Normal(0, 5)$ и выборки для $\beta_{41,6}$ и $\beta_{111,9}$ при $\mu_n \sim Normal(0, 2)$ были получены из одного распределения. При сравнении модификаций $\mu_n \sim Normal(0, 5)$ и $\mu_n \sim Normal(0, 10)$, для индивида 41: $p - value = 0.01811$, то есть нулевая гипотеза отвергается. Для индивида 111 нулевая гипотеза не отвергается, поскольку $p - value = 0.1392$.

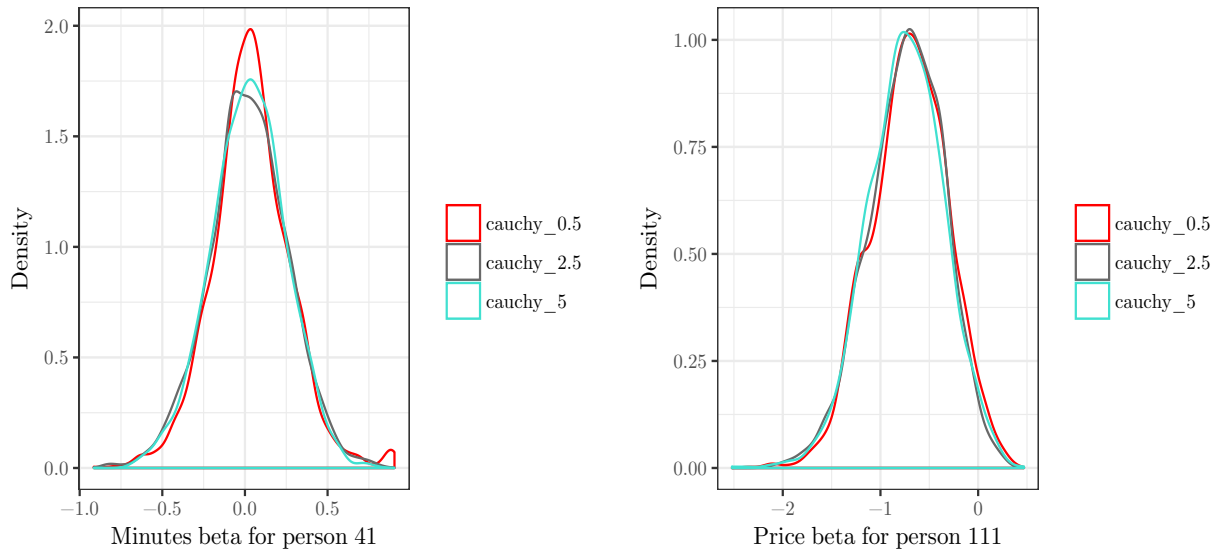
Теперь сравним графически апостериорные распределения β при изменении априорного $\tau_a \sim Cauchy(0, 2.5)$ сперва на $\tau_a \sim Cauchy(0, 0.5)$, затем на $\tau_a \sim Cauchy(0, 5)$.

Графики (Рис.4) опять же не выявляют существенных различий. Вновь используем тест Колмогорова-Смирнова. При сравнении моделей с $\tau_a \sim Cauchy(0, 2.5)$ и $\tau_a \sim Cauchy(0, 0.5)$ для индивидов 41 и 111 получаем $p - value = 0.01219$ и $p - value = 0.1294$, соответственно. Тогда в первом случае на уровне значимости $\alpha = 0.05$ нулевая гипотеза отвергается, во втором - не отвергается. При отклонении от базовой спецификации к $\tau_a \sim Cauchy(0, 5)$, получаем: $p - value_{41} = 0.8632$ и $p - value_{111} = 0.2742$. Нулевые гипотезы не отвергаются.

Перейдём к априорному распределению корреляционной матрицы. Рассмотрим не только $\nu = 2$ (базовая спецификация), но и $\nu = 1$, $\nu = 10$.

¹¹Под апостериорной выборкой подразумевается выборка из 2000 значений из апостериорного распределения параметров

Рис. 4: Апостериорные распределения β при различных априорных распределениях τ_a



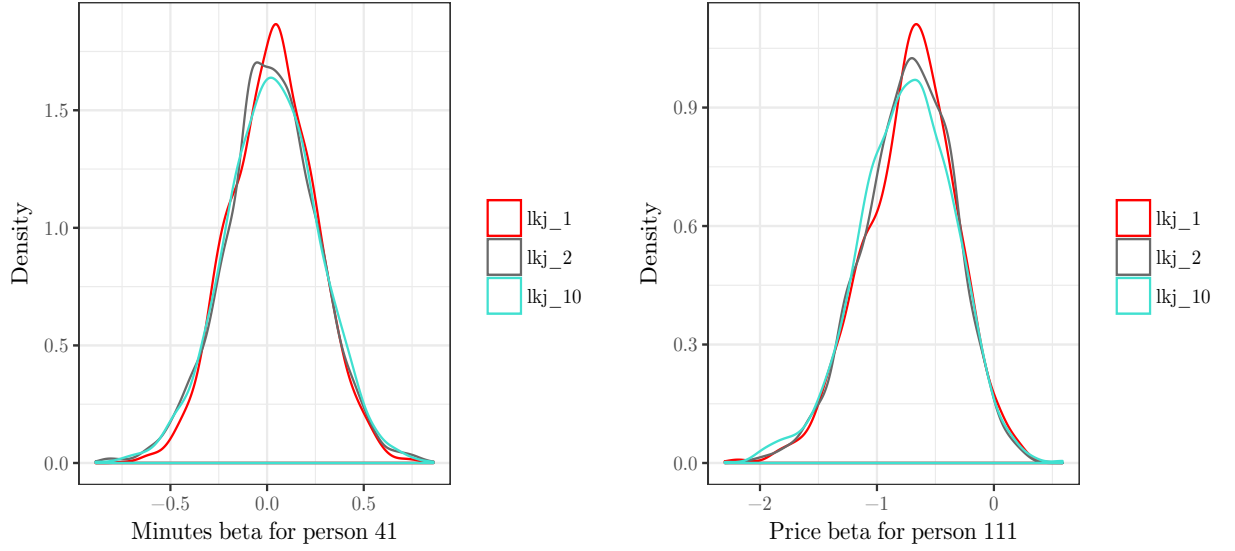
Исходя из полученных графиков (Рис.5), можем сказать, что существенных отличий не наблюдается. Снова посчитаем статистику Колмогорова-Смирнова.

При сравнении $\nu = 1$ и $\nu = 2$ получаем: $p - value_{41} = 0.2262$ и $p - value_{111} = 0.1849$, следовательно, нулевая гипотеза не отвергается для обоих индивидов.

Подводя итог, можем сказать, что полученные результаты не сильно чувствительны к изменениям предпосылок об априорных распределениях параметров. Более того, если бы мы рассматривали уровень значимости $\alpha = 0.01$, гипотезы о том, что апостериорные выборки β были получены из одного распределения, не отвергались бы ни для одного из рассмотренных случаев.

Напомним, что нас интересуют не только полученные оценки коэффициентов, но и предсказания о будущем выборе индивидов. Рассмотрим их чувствительность на примере тех же индивидов. Несмотря на то что

Рис. 5: Апостериорные распределения β при различных априорных распределениях Ω



выбор - величина дискретная, будем изображать непрерывную функцию плотности, ввиду более удобного визуального сравнения, чем в ситуации с гистограммами.

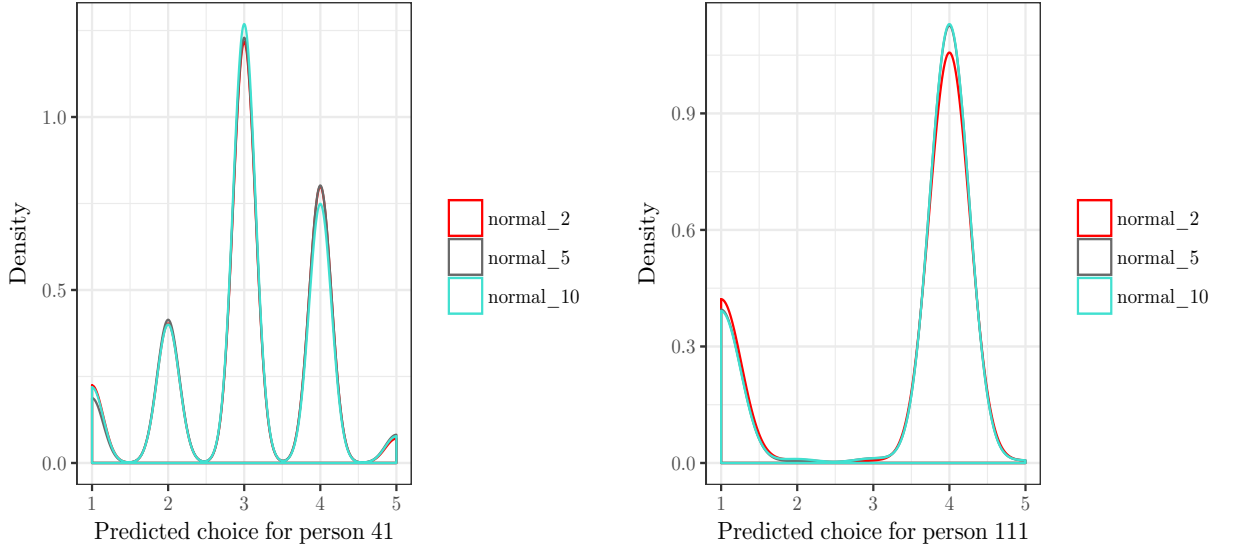
Сперва "пошевелим" распределение μ и посмотрим, как это отразится на предсказанном выборе.

Как можно видеть из графиков Рис.6, изменение априорного распределения μ оказывает малозначительное влияние на апостериорные распределения \tilde{y} ¹². Что касается статистики Колмогорова-Смирнова, при сравнении результатов моделей при $\mu_n \sim Normal(0, 5)$ и $\mu_n \sim Normal(0, 2)$, получаем: $p - value_{41} = 0.9933$ $p - value_{111} = 0.4131$. Если же мы сравниваем предсказанный выбор в моделях с $\mu_n \sim Normal(0, 5)$ и $\mu_n \sim Normal(0, 10)$, то $p - value_{41} = 0.7699$ и $p - value \simeq 1$ ($D_{41} = 0.003$). Значит, все нулевые гипотезы в этом случае не отвергаются.

Теперь поменяем распределение τ_a сперва на $\tau_a \sim Cauchy(0, 0.5)$,

¹²Под апостериорным распределением по-прежнему подразумевается распределение, полученное на основе выборки из 2000 значений, а не истинное апостериорное распределение

Рис. 6: Апостериорное распределение предсказаний выбора при различных априорных распределениях μ_n



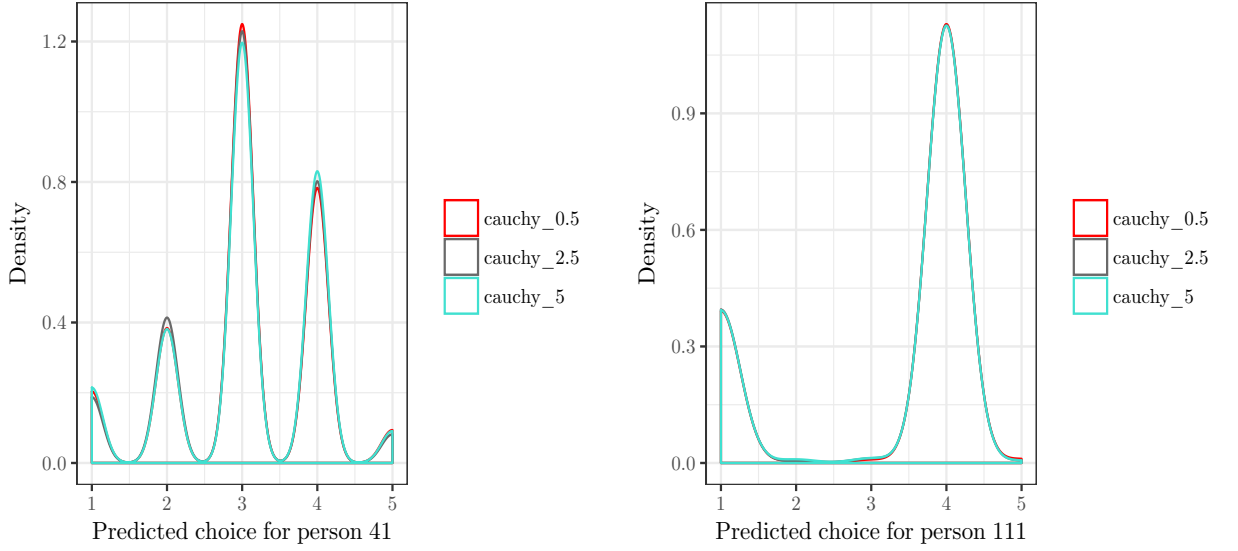
затем на $\tau_a \sim Cauchy(0, 5)$.

Исходя из полученных графиков (Рис.7), можем сделать аналогичный вывод, как и в предыдущем случае. Изменение дисперсии параметра τ_a привело к незначительным колебаниям в терминах предсказанной вероятности выбрать ту или иную альтернативу. При сравнении результатов моделей с $\tau_a \sim Cauchy(0, 2.5)$ и $\tau_a \sim Cauchy(0, 0.5)$ получаем: $p-value_{41} = 1$ ($D_{41} = 0.006$), $p-value_{111} = 1$ ($D_{111} = 0.004$). В ситуации, где $\tau_a \sim Cauchy(0, 2.5)$ и $\tau_a \sim Cauchy(0, 5)$, имеем: $p-value_{41} = 0.9895$ и $p-value_{111} = 1$ ($D_{111} = 0.0015$). Следовательно, нулевые гипотезы не отвергаются.

И последний немаловажный параметр - корреляционная матрица. Проиллюстрируем изменения апостериорного распределения предсказаний выбора индивидов в зависимости от её априорного распределения. Помимо $\nu = 2$, рассмотрим $\nu = 1$ и $\nu = 10$.

Согласно графикам (Рис.8), изменение параметра, отвечающего за форму LKJ распределения для корреляционной матрицы Ω , незначи-

Рис. 7: Апостериорное распределение предсказаний выбора при различных априорных распределениях τ_a

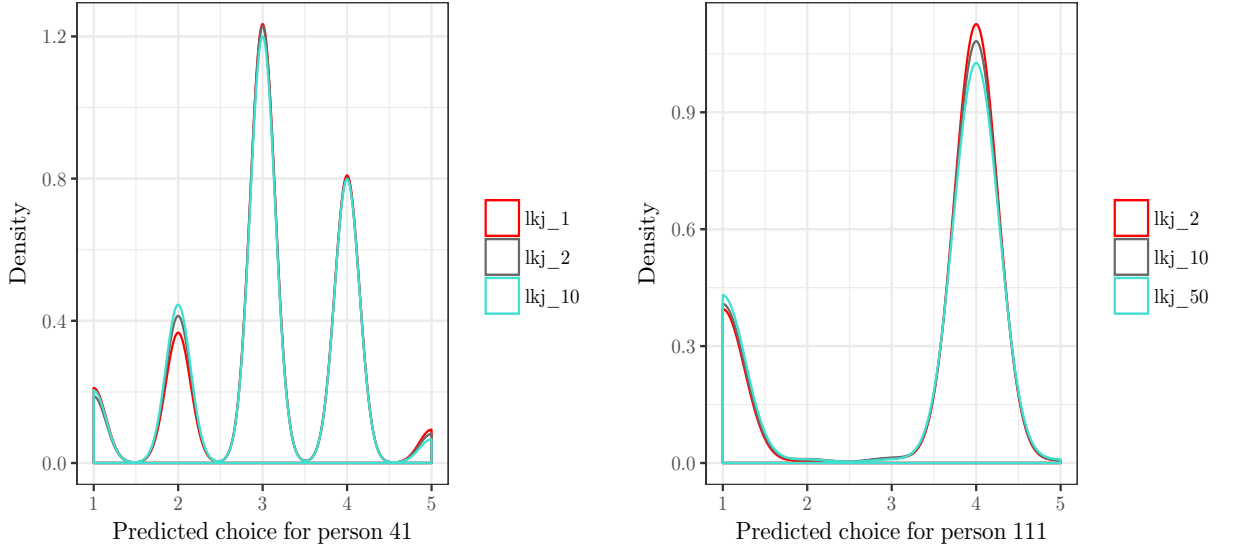


тельно влияет на апостериорную вероятность выбора различных индивидов. При сравнении моделей с $\nu = 2$ и $\nu = 1$, имеем: $p - value_{41} = 1$ ($D_{41} = 0.009$) и $p - value_{111} = 0.7444$. Выявим различия результатов в модификациях с $\nu = 2$ и $\nu = 10$. Получаем следующие значения: $p - value_{41} = 0.9194$, $p - value_{111} = 0.9347$, следовательно, нулевая гипотеза не отвергается.

Проведённый анализ на чувствительность не был особенно глубоким, но всё же есть основания полагать, что различная модификация априорных распределений параметров в рамках, обусловленных в руководстве по языку Stan (Stan Development Team, 2016, версия 2.12.0), не ведёт к значительным переменам в апостериорных распределениях коэффициентов и предсказаний. В связи с этим в дальнейшем мы будем использовать по умолчанию полагать следующее:

$$\mu_n \sim Normal(0, 5)$$

Рис. 8: Апостериорное распределение предсказаний выбора при различных априорных распределениях Ω



$$\tau_a \sim \text{Cauchy}(0, 2.5),$$

$$\Omega \sim \text{LKJcorr}(2),$$

4.4 *Различные спецификации Байесовской иерархической мультиномиальной логит модели*

В [Разделе 4.2](#) было упомянуто о гибкости модели благодаря параметрам L и ll . Возможности делить индивидов на различное число кластеров порождает множество новых вопросов. Основной задачей этого раздела будет попытка сравнить предсказательную силу моделей, отличающихся количеством групп потребителей, объединённых по схожести предпочтений.

В некотором смысле агрегированная и индивидуальная модели - са-

мые простые случаи, поскольку либо предпочтения представляются усреднёнными, либо каждый обладает своим личным вектором коэффициентов. Когда же возникает задача кластеризации, становится совершенно непонятно, по каким критериям объединять людей. Казалось бы, зачем вообще это нужно? Ведь интуитивно создаётся ощущение, что лучше предскажет модель с индивидуальными коэффициентами. Однако мы не станем полагаться лишь на интуицию.

Итак, попробуем удовлетворить исследовательское любопытство ответом на вопрос: действительно ли индивидуальные коэффициенты дают лучшие прогнозы?

Начнём с определения того, какие модели мы будем сравнивать. Помимо агрегированной и полностью индивидуальной мы включим ещё ряд кластерных. Вопрос в том, на какое количество групп разумно было бы разделить респондентов. Мы имеем дело с 11-ти мерным пространством коэффициентов, в котором как-то расположены 296 индивидов. К сожалению, визуально нет возможности ладить с такой размерностью, поэтому мы заинтересованы в её снижении.

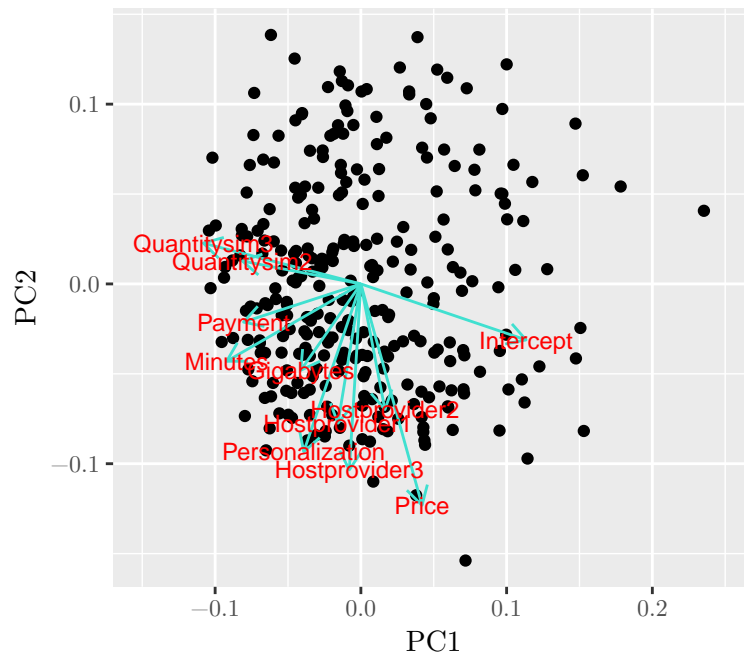
Попробуем сперва визуализировать наших респондентов в осях главных компонент.

К сожалению, глядя на график (Рис.9), трудно судить о количестве групп, в которые можно было бы объединить людей.

Попробуем другой способ снижения размерности, а именно, алгоритм $t - SNE$ (t-distributed stochastic neighbor embedding)¹³. В отличие от метода главных компонент, он представляет собой технику нелинейного снижения размерности. Основным параметром $t - SNE$ является перплексия (perplexity), в зависимости от её значения график может при-

¹³О нём можно прочесть в статье авторов, которые придумали этот метод: Maaten, Hinton, 2008

Рис. 9: Предпочтения респондентов, представленные в осях главных компонент



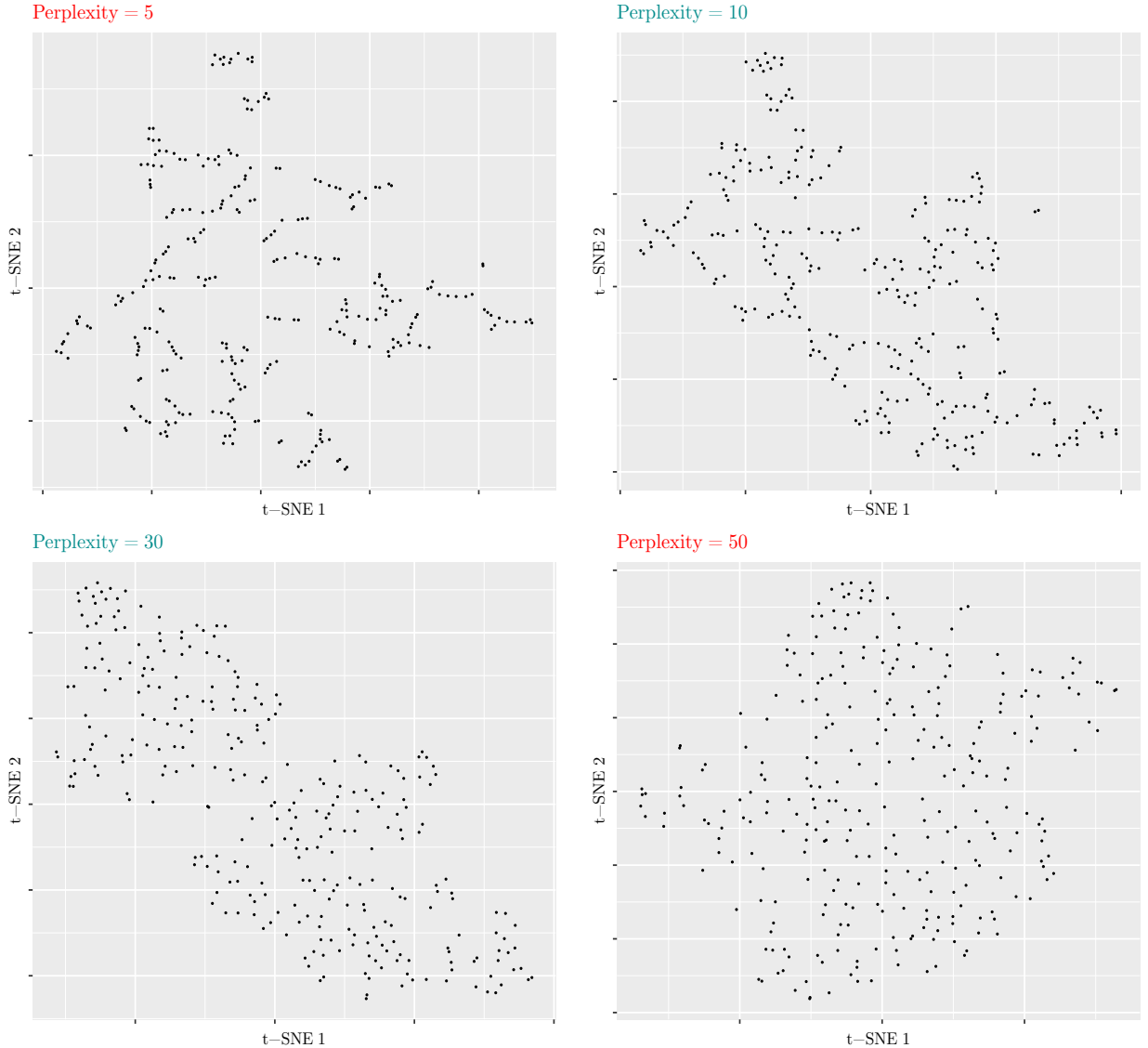
нимать совершенно непохожие обличия. Перплексия интерпретируется как сглаженная оценка количества "соседей" для каждой точки. Авторы (Maaten, Hinton, 2008) рекомендуют выбирать значение этого параметра в пределах от 5 до 50.

Рассмотрим полученный график (Рис.10)¹⁴. При перплексии равной 5 и 10 визуально кажется, что формируются группы из 5-10 людей. Однако мы не можем быть уверены, что сорвали джекпот. Автор статьи (Watternberg 2016) утверждает, что низкий уровень перплексии ведёт к распределению, казалось бы, с достаточно выраженными кластерами. Однако такая картина наблюдается зачастую даже для случайного шума, поэтому мы не можем на неё полагаться.

Итак, поскольку нам не удалось с помощью снижения размерности однозначно определиться с числом кластеров, мы последовательно рассмотрим модели с 3, 30, 60, 90, 120, 150, 180, 210, 240, 270 кластерами.

¹⁴ Был использован `set.seed(78)`

Рис. 10: Предпочтения респондентов, представленные в осях $t - SNE$



Для деления всех респондентов на указанное выше число кластеров возьмём за основу оценки индивидуальных коэффициентов (а именно, среднее), полученные ранее. Будем использовать иерархическую кластеризацию (Hogwood, 1980). Этот метод предполагает, что исходно каждый элемент выборки - это отдельный кластер. Далее два кластера, между которыми минимальное расстояние, объединяются в один. Так происходит до тех пор, пока вся выборка не станет единой. Мы будем использовать метод полной связи (complete-linkage clustering), который опреде-

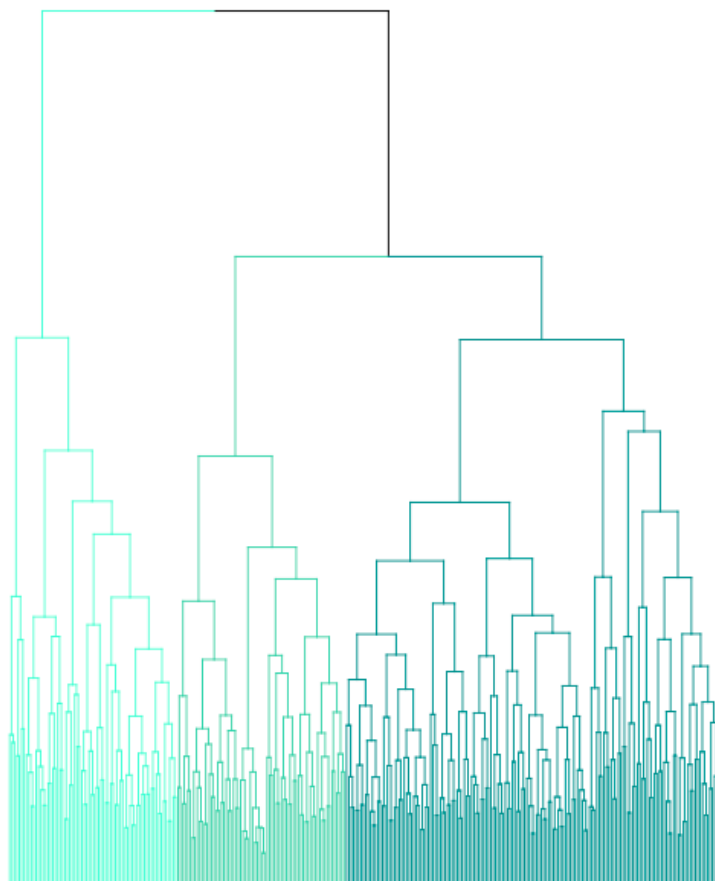
ляет расстояние как максимум из множества расстояний между элементами двух кластеров:

$$D(X, Y) = \max_{x \in X, y \in Y} d(x, y),$$

где $d(x, y)$ - расстояние между x и y .

Для наглядности построим дерево (Рис.11), которое будем использовать для всех наших моделей. Изобразим его в нескольких цветах, соответствующим трём кластерам.

Рис. 11: Иерархическое дерево



Это же дерево будет использоваться для любого другого количества кластеров. Можно представить, что разбиение получается путём пересечения дерева горизонтальной линией столько раз, сколько кластеров

мы желаем.

Итак, получив с помощью иерархической кластеризации вектор ll для каждого из случаев (3, 30, 60, 90, 120, 150, 180, 210, 240, 270), мы можем запускать новые модели.

Осталось условиться, по какому критерию мы будем сравнивать их предсказательные силы. Ошибкой прогноза Байесовских иерархических мультиномиальных моделях можно считать сумму вероятностей, что модель даст неверный прогноз:

$$\sum_{i=1}^{296} (1 - p(\tilde{y}_n = y_n)),$$

где \tilde{y}_n - предсказанный моделью выбор индивида n , а y_n - реальный выбор индивида n . Напомним, что наша тестовая выборка состоит только из одной карточки для каждого респондента, поэтому параметр C не фигурирует в определении выше. Заметим также, что используется просто сумма ошибок прогноза, а не сумма квадратов, поскольку все слагаемые имеют один и тот же знак.

Для удобства будем использовать следующую метрику для сравнения моделей:

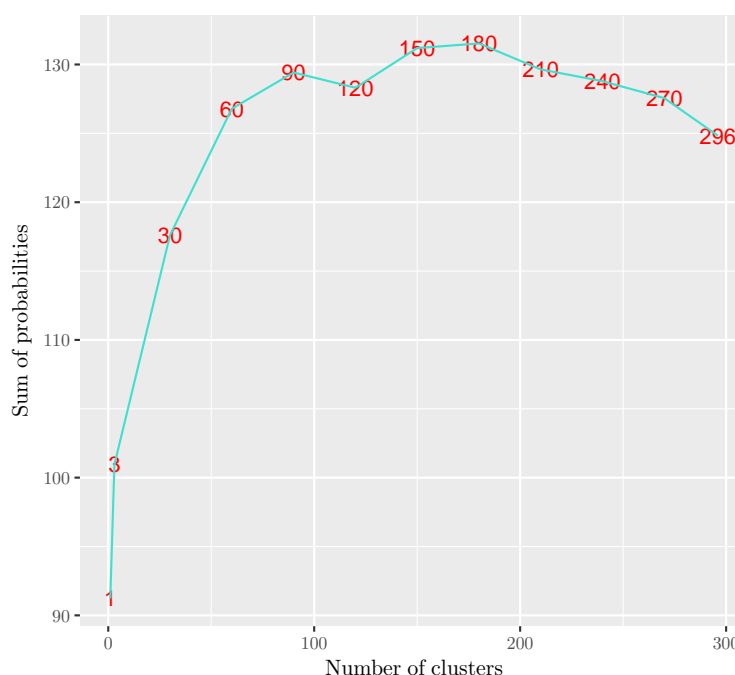
$$\sum_{i=1}^{296} p(\tilde{y}_n = y_n),$$

поскольку она аналогична предыдущей, но более компактна.

Полученные результаты продемонстрируем в виде зависимости выбранной метрики от количества кластеров.

Как мы можем видеть (Рис.12), интуитивное предположение о том, что 296 кластеров приведут к наилучшему прогнозу были ошибочны для нашего набора данных. Это может объясняться тем, что сформирован-

Рис. 12: Предсказательная сила модели в зависимости от числа кластеров



ные малочисленные кластеры включают в себя людей с действительно похожими предпочтениями, и это объединение даёт больше данных, помогая тем самым лучше прогнозировать для этих "близнецов". Также заметим, что при наличии 60 кластеров модель становится лучше, чем при 296, поэтому, быть может, выявление групп по 5 человек алгоритмом t-SNE с низким уровнем перплексии (Рис.10) было не случайно.

В целом мы можем сделать несколько выводов. Во-первых, прежде чем спешить симулировать будущую рыночную ситуацию на основе индивидуальных коэффициентов, имеет смысл рассмотреть кластерные модели. Во-вторых, если кластерные модели дают надёжные предсказания, полученные с их помощью оценки коэффициентов β можно использовать для описания предпочтений потребителей.

5 Заключение

В рамках данного исследования была создана Байесовская иерархическая мультиномиальная логит модель с помощью языков Stan и R. Стоит отметить, что она может оказаться особенно полезной с практической точки зрения, поскольку существующий аналог недоступен большинству исследователей ввиду своей стоимости.

Для проверки качества написанной модели были использованы данные по мобильным тарифам. Сперва были получены выборки из апостериорных распределений индивидуальных коэффициентов и предсказаний выбора. Далее проводилось тестирование апостериорного вывода на чувствительность к априорным распределениям параметров, заложенных в модель. Было выявлено, что распределения полученных выборок не подвержены значительным изменениям при различных модификациях априорных вер. Это, в свою очередь, позволяет будущим исследователям полагаться на использованные нами доопытные распределения параметров.

Кроме того, при анализе различных спецификаций Байесовского иерархического мультиномиального логита было рассмотрено большое число кластерных моделей. Оказалось, что последние могут превосходить по предсказательной силе модель с максимальной гетерогенностью. Это важный результат, поскольку симуляции различных рыночных ситуаций занимают центральное место во многих маркетинговых вопросах. Они позволяют компании получить информацию о том, как отреагируют потребители на новый продукт, без затрат на производство этого продукта и выпуска в продажу. При качественных прогнозах такой подход экономит существенные средства фирм, поэтому наше исследование действительно имеет практическую пользу.

Однако стоит отметить, что даже при таком небольшом количестве наблюдений, требуется несколько часов для подсчёта одной модели. Мы старались писать код оптимально, но, вероятно, есть незамеченные способы ускорить вывод результатов.

Наконец, очертим поле для будущих исследований в данном направлении. Как было упомянуто ранее, наша модель предполагает одинаковое число карточек для всех индивидов. Было бы полезно изменить модель так, чтобы была возможность повысить уровень случайности при разделении выборки на тренировочную и тестовую. Также важно отметить, что теоретически осуществимо написание Байесовской модели, которая самостоятельно бы определяла число кластеров. Однако такая работа очень трудоёмка, поскольку в Stan нереализована возможность задания дискретных априорных распределений параметров.

6 Литература

7 Приложения