Страница 1

Stat Comput
DOI 10.1007 / s11222-014-9464-2

Быстрое разделение параметров сглаживания в многомерных обобщенные P-сплайны: алгоритм SAP

Мария Хосе Родригес-Альварес · Дэ-Джин Ли · Томас Кнайб · Мария Дурбан · Пол Эйлерс

Получено: 30 июля 2013 г. / принято: 22 марта 2014 г.

Аннотация Новый вычислительный алгоритм для оценки параметры сглаживания многомерной штрафной сплайновая обобщенная линейная модель с анизотропным штрафом: представлен. Это новое предложение основано на смешанной модели. представление многомерного P-сплайна, в котором параметр сглаживания для каждой ковариаты выражается в терминах компонент дисперсии. На основании оштрафованных квази-методы правдоподобия, выражения в закрытой форме для оценки получены сопряжения компонентов дисперсии. Эта формула-lation приводит к эффективной реализации, которая значительно снижает вычислительную нагрузку. Предлагаемый алгоритм можно рассматривать как обобщение алгоритма Шалла (1991) - для оценки компонентов дисперсии - чтобы иметь дело с нестандартные структуры ковариационной матрицы случайных дом эффекты. Практическая эффективность предложенного алгоритма ритм оценивается с помощью моделирования и сравнения альтернативными методами производятся на основе среднего критерий квадратичной ошибки и время вычислений. Наконец, мы проиллюстрируем наше предложение анализом двух реальных наборов данных: двухмерный пример исторических записей ежемесячных данные об осадках в США и трехмерный данные о смертности от респираторных заболеваний в зависимости от возраста при смерти - год смерти и месяц смерти.

Ключевые слова Сглаживание · P-сплайны · Тензорное произведение · Анизотропный штраф · Смешанные модели

Электронный дополнительный материал Онлайн-версия этого статья (doi:10.1007 / s11222-014-9464-2) содержит дополнительные материал, доступный авторизованным пользователям.

МХ Родригес-Альварес (B)
Департамент статистики и исследований операций, Университет of Vigo, Campus Lagoas-Marcosende s / n, 36310 Виго, Испания
электронная почта: mxrodriguez@uvigo.es

Д.-Дж. Ли
CSIRO Computational Informatics, Клейтон, Вирджиния, Австралия

Д.-Дж. Ли
BCAM - Баскский центр прикладной математики, Бильбао, Испания

Т. Кнейб
Заведующий статистикой Гёттингенского университета Георга Августа, Гёттинген, Германия

М. Дурбан
Департамент статистики Мадридского университета Карлоса III, Леганес, Испания

П. Эйлерс
Медицинский центр Эразмус, Роттердам, Нидерланды

1. Введение

Сглаживание штрафа шероховатости стало наиболее популярным. Личный метод выполнения непараметрической регрессии. Как-Тем не менее, эта методология зависит от ключевого шага: выбора параметр сглаживания, который контролирует торговлю отключение между точностью данных и сглаживанием. Есть два основных подхода к выбору параметров сглаживания: один на основе оптимизации некоторых критериев, таких как Akaike информационные критерии или обобщенная перекрестная проверка (GCV) (в контексте штрафных шлицев см., например, Эйлерс и Маркс 1996 ; Дерево 2004, 2008), и тот, в котором гладкая функция считается случайным, а оценка параметров сглаживания связаны максимальной вероятностью (ML) ил[redacted]ным максимумом вероятность (REML) (К[redacted]ан [redacted]l. 2004 ; Ruppert et al . 200[redacted]3 ; [redacted]а модель включает несколь[redacted] функций, рас-чет-нагрузка быстро увеличивается с увеличением числа сглаживающих параметры, которые необходимо выбрать, и процедура минимизации может стать нестабильным. Было разработано несколько алгоритмов для достичь численной стабильности и улучшить вычислительную

Страница 2

время. Большинство этих алгоритмов находятся в рамках GCV, некоторые из них основаны на матричных факторизациях (Дерево 2004), и тоже использовать полный метод Ньютона (Дерево 2008), а не итеративный повторный взвешенный метод наименьших квадратов. Совсем недавно Вуд (2011) предоставил стабильный метод вложенных итераций для REML или ML, которые

и Чжан 1999). Имитационное исследование оценивает практические производительность предложенного алгоритма в разд. 5. Мы иллю-Изучите наш метод в разд. 6, используя два реальных примера, и закончим обсуждение в Разд. 7. Некоторые технические детали были добавлены в виде приложений. Подробные доказательства и расширенные

доказал, что превосходит предыдущие подходы в этом контексте .
Когда дело дошло до расширения вышеуказанной арроа чеса
к оценке многомерных поверхностей взаимодействия ,
малоранговые сглаживатель тензорного произведения стали вообще
подход (Эйлерс и Маркс, 2003 г.; Дерево 2006b). Их поп-
Улярность в первую очередь обусловлена гибкостью, что тензорное произведение
сглаживателей обеспечивают, в основном, есть incorpora тин
анизотропные штрафы. Однако в этом контексте возникает проблема
с задачей сделать оценку выполнимой с помощью ком-
предполагаемая точка зрения. Кроме того, для REML / ML на основе
подходы к оценке также сталкиваются с тем, что оценка
мация из компонентов дисперсии не может быть accomoda Ted
с использованием стандартного программного обеспечения смешанной модели. Хотя Estima Тион
можно сделать путем численной максимизации (ограниченного)
журнал - вероятность (Дерево 2006b, 2011 г.), у него есть недостаток
требовательности к вычислениям, особенно для больших
наборы данных. Совсем недавно Wood et al. (2013) и Ли и др.
(2013) предложили альтернативный метод оценки -
сглаживание тензорного произведения с анизотропными штрафами .
Оба подхода основаны на разложении муль-
трехмерный гладкий член в разных терминах каждый из них
зависит от одного параметра сглаживания. Хотя оба
подходы оказались полезными, развитие в
эффективные и быстрые алгоритмы для работы с правильной анизотропией
штрафы по-прежнему актуальны, особенно за более чем два
ковариаты.

В данной статье представлен новый алгоритм оценки
параметры сглаживания многомерного тензорного произведения -
Обобщенная линейная модель uct со штрафом (P-сплайн) с
анизотропные штрафы на основе смешанной модели для
муляция. Следуя идеям, изложенным в Харвилле (1977 г.)
и Шалл (1991 г.) выводим отдельные выражения в замкнутой форме.
сии для оценки каждого компонента дисперсии, которые являются
затем встраивается в итеративную процедуру. Algor ithm
поэтому легко реализовать на практике. Более-
При этом некоторые характеристики полученных выражений могут быть
используется для дальнейшего ускорения вычислений.

Остальная часть этого документа организована следующим образом: в
краткое введение в тензорное произведение низкого ранга P-splines mod-
els и его представление в виде смешанной модели .
Для наглядности, мы в первую очередь сосредоточим ATTEN Тион
на двумерном (2D) обобщенном P-сплайне. Тем не менее ,
сглаживание в более чем двух измерениях может быть также сопро-
модифицированный. После того, как необходимая фон и обозначения имеют
были представлены, мы подробно описываем наш подход в Разд . 3.
В разд. 4, мы представляем некоторые расширения предложенного алгоритма
ритм. В частности, мы описываем трехмерное (3D)
случай и обобщенная аддитивная смешанная модель (ГАММ, Lin

моделирования доступны в дополнительном онлайн-матче-
риал.

2 Двумерные сглаживатели тензорного произведения низкого ранга

Рассмотрим двумерную задачу *обобщенной* регрессии в
которые наблюдения за *i*- й из *n* независимых единиц состоят
переменной одномерного отклика y_i и двумерного вектора ковариации
$\mathbf{x}_{я} = (x_{я1}, x_{я2})_m$

$$g(E[y_i \mid \mathbf{x}_i]) = g(\mu_i) = \eta_i = f(x_{i1}, x_{i2}), \qquad (1)$$

где *f*- гладкая и неизвестная функция, а *g* -
монотонная функция связи. Здесь мы предполагаем, что y_i следует
экспоненциальное семейное распределение, где Yar $(y_i \mid \boldsymbol{x}_i) =$
$\varphi v(\mu_i)$, где *v* - функция дисперсии, которая определяется
по экспоненциальному семейству переменная отклика принадлежит
к, а φ- параметр дисперсии, который может быть известен или
неизвестный.

В рамках P-сплайна неизвестная поверхность
$f(x_1, x_2)$можно аппроксимировать тензорным произведением двух
сы низкого ранга (см., например, Эйлерс и Маркс

$$f(x_1, x_2) \text{знак равно} \sum_{j=1} \sum_{k=1} \theta_{jk} B_{1j}(x_1) B_{2k}(x_2),$$

где B_{1j} и B_{2k} - одномерные базисные функции x_1
и x_2, соответственно (как, например, B-splines, de Boor 2001, или тонкий
регрессионные шлицы пластины , Wood 2003) и θ_{jk} - вектор
коэффициенты регрессии . Обозначим $\boldsymbol{B}_1$ и $\boldsymbol{B}_2$ маргинальные
модельные матрицы для ковариантных значений $\boldsymbol{x}_1 = (x_{11}, \ldots, x_{n1})_t$
и $\boldsymbol{x}_2 = (x_{12}, \ldots, x_{n2})_t$ соответственно. Тогда в матричных обозначениях
ция, модель (2) можно выразить как

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{B\theta}, \qquad (2)$$

где $\boldsymbol{B} = \boldsymbol{B}_2$ D $\boldsymbol{B}_1$ - матрица полной регрессии (с
ющий «построчное» произведение Кронекера, Eilers et al.
$\mu_1, \ldots, \mu_n)_t$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)_t$ и $\boldsymbol{\theta} =$
$_1, \ldots, \theta_{c_1 c_2}$.

В контексте P-шлицев гладкость достигается за счет
наложение штрафа на коэффициенты регрессии $\boldsymbol{\theta}$ в виде
$\boldsymbol{\theta}_t \breve{P} \boldsymbol{\theta}$, где $\breve{\boldsymbol{P}}$ - матрица штрафа. Для сглаживания P-сплайна
более чем в одном измерении, каждый сталкивается с решением
либо предполагать одинаковую величину сглаживания для всех
ковариатами (изотропное наказание) или разрешить различные
гладкость по каждой ковариате (анизотропный штраф).
В то время как изотропия могла быть оправдана при моделировании, для

например, гладкая функция широты и долготы, это
не всегда бывает, когда ковариаты, например, x_1 и x_2 , равны
измеряется в разных единицах.

В этой статье мы предполагаем анизотропную пенализацию, т.е.
разная степень сглаживания для x_1 и x_2 . Соответственно,
тогда матрица штрафов имеет вид (см., например, Eilers et al.
2006)

$$\breve{\boldsymbol{P}} = \lambda_1 \boldsymbol{I}_{c_2} \otimes \breve{\boldsymbol{P}}_1 + \lambda_2 \breve{\boldsymbol{P}}_2 \otimes \boldsymbol{I}_{c_1}, \qquad (3)$$

где $\otimes$ обозначает произведение кронекера, $\boldsymbol{I}_k$ - тождество
матрица размерности *k* , λ_d - параметр сглаживания, который
управляет степенью сглаживания по ковариате x_d ,
и $\breve{\boldsymbol{P}}_d$ - положительные полуопределенные матрицы $c_d \times c_d$ ранга
$(c_d - q_d)$, элементы которого зависят от выбранного базиса сплайна.
Например, в случае B-сплайнов эти штрафные матрицы
можно выразить как $\breve{\boldsymbol{P}}_d = \boldsymbol{D}_{t\,d} \boldsymbol{D}_d$, где $\boldsymbol{D}_d$ - матрица,
образует разности порядка q_d ($d = 1, 2$) (см. Эйлерс и Маркс
1996 для подробностей).

2.1 Представление смешанной модели

Оценить модель (2) подлежит наказанию, определенному в
(3), мы принимаем здесь эквивалентность P-сплайнов и
обобщенные линейные смешан (ГЛММ) (Линь и Чжан

матрица, обратная ковариационной матрице случайных эффектов $\boldsymbol{G}$ в (4)
становится блочно-диагональной матрицей

$$\text{грамм}^{-1} = \begin{pmatrix} \frac{1}{\tau_2^2} \breve{}_2 \otimes \boldsymbol{I}_{q_1} & & & \\ & \frac{1}{\tau_1^2} \boldsymbol{Я}_{q_2} \otimes \breve{}_1 & & \\ & & \frac{1}{\tau_2^2} \breve{}_2 \otimes \boldsymbol{I}_{c_1 - q_1} + \frac{1}{\tau_1^2} \boldsymbol{I}_{c_2 - q_2} \otimes \breve{}_1 \end{pmatrix},$$

где $\breve{}_d$ - подматрица матрицы $_d$ с ненулевым sin-
расчетные значения, $\tau_1^2 = \frac{\varphi}{\lambda_1}$ и $\tau_2^2 = \frac{\varphi}{\lambda_2}$. Как можно заметить,
в этой новой конфигурации параметр сглаживания λ_d равен
задается отношением компонент дисперсии, т. е. $\lambda_d = \varphi \tau_d^{-2}$,
$(d = 1, 2)$. Обратите внимание на взаимосвязь между каждым блоком $\boldsymbol{G}$
и каждый блок матрицы случайной модели $\boldsymbol{Z}$, определенный в
(5). Каждая компонента дисперсии τ_d^2 (как и $\breve{}_d$) появляется в
грамм^{-1} всякий раз, когда матрица $\boldsymbol{Z}_d$ находится в соответствующем блоке
из $\boldsymbol{Z}$. Это соответствие может быть полезно для лучшего понимания
стоять, как $\boldsymbol{G}^{-1}$ построен в трехмерном корпусе, который будет
представлены в разд. 4 (или, расширенно, в d-мерном
дело). В связи с этим соответствием блок-структура
от $\boldsymbol{Z}$ также приводит к очень интересному разложению
штрафная часть двумерной поверхности *f* в (1) в трех
различные термины: (а) член, связанный с $\boldsymbol{Z}_2$ D $\boldsymbol{X}_1$, который соединяет
tains гладкого *Основной* эффект от $й_2$ и $(\partial_1 - 1)$ Различного

1999 ; Карри и Дурбан 2002 ; Ж[redacted]од этим подход, матрица плана B и вектор [redacted]

коэффициенты θ в (2) переформулированы таким образом, что

$g(\mu) = \eta = B\theta = X\beta + Z\alpha$, где $\alpha \sim N(0, G)$, (4)

где X и Z - модельные матрицы, а β и α - коэффициенты фиксированных и случайных эффектов обобщенного линейного смешанная модель соответственно. Случайные эффекты ковари- матрица G , которая зависит от двух компонент дисперсии τ^2_1 и τ^2_2 .

Для получения смешанного модельного представления (4), мы следуем предложение Ли (2010), Ли и Дурбан (2011 г.). Их подход основан на сингулярном разложении (SVD) предельных штрафов $\breve{P}_d$, участвующих в (3) для $d = 1, 2$. Пусть $\breve{P}_d = U_d \Sigma_d U_d^t$, где U_d - матрица сингулярных векторов. tors, Σ_d - диагональная матрица сингулярных чисел. Давайте также обозначим U_{ds} подматрицу матрицы U_d, содержащую сингулярную ларные векторы, соответствующие $(c_d - q_d)$ ненулевым сингулярным значения. Затем смешанные модельные матрицы для модели (4) находятся

$X = [X_2 \square X_1]$,

$$Z = [Z_2 \square X_1 \mid X_2 \square Z_1 \mid Z_2 \square Z_1], \quad (5)$$

где $X_d = \left[1_n \mid x_d \mid \cdots \mid x_d^{(q_d - 1)}\right]$ и $Z_d = B_d U_{ds}$, причем 1_n - вектор-столбец единиц длины n и x_d^r обозначает поэлементного r -порядок многочлен вектора - столбца ковариантные значения x_d (для $d = 1, 2$). Обратите внимание, что матрица X_d содержит столько столбцов, сколько нулевых сингулярных значений матрицы штрафов $\breve{P}_d$, т.е. q_d . Согласно этому представлению,

члены коэффициентов (Хасти и Тибширани 1993) с x_2 переменными- плавно с x_1 $\left(f_2(x_2) + \sum_{j=1}^{q_1-1} x_1^j h_{j2}(x_2)\right)$, (б) а член, связанный с $X_2 \square Z_1$ с гладким *основным* эффектом из $\breve{u}_1$ и $(q_2 - 1)$ переменного коэффициент членов с $\breve{u}_2$ варьировавшими плавно с x_1 $\left(f_1(x_1) + \sum_{j=1}^{q_2-1} x_2^j h_{j1}(x_1)\right)$; а также, (с) член *чисто гладкого* взаим[redacted]связанный с $Z_2 \square Z_1$ $f_{1|2}(x_1, x_2)$ (Ли и др . 20[redacted]

Что касается оценки любо[redacted]оценка модели (4) включает два взаимосвязанных этапа: (а) фиксированный и случайный оценка коэффициентов эффектов (β и α); и (б) дисперсия оценка компонентов (τ^2_1 , τ^2_2 и, возможно, дисперсия параметр φ). В нашем контексте и для фиксированных значений компоненты дисперсии, оценка фиксированных и случайные эффекты не представляют проблемы. Их можно получить с использованием штр[redacted]ов *квази-правдоподобия* (PQL) (Stiratelli и дру[redacted] это очень простой мет[redacted]GLMM, и он может легко реализовать путем итеративной подгонки *рабочего* линейного смешанная модель к *рабочей* зависимой переменной z , на основе алгоритма оценки Фишера, который включает матрицу весов W, который обновляется на каждой итерации (мы описываем этот момент подробно в Приложении 1). Однако оценка τ^2_1 , τ^2_2 и φ не может быть адаптирован с использованием стандартных процедур для оценка компонентов дисперсии в смешанных моделях (или, более точно, стандартное программное обеспечение смешанной модели, как, например, `R` - функции `lme` и `lme4` , или процедура `PROC MIXED` в SAS®), поскольку ковариационная матрица случайных эффектов G (см. (15) в Приложении 1) имеет нестандартный вид, с

блок, включающий обе компоненты дисперсии τ^2_1 и τ^2_2 . В следующем разделе мы представляем вычислительную эффективность Эффективный алгоритм оценки составляющих дисперсии. Следовать- ING Харвилл (1977 г.) и Шалл (1991), мы вывели закрытые сформированные выражение для оценок компонент дисперсии что, в свою очередь, избавляет от необходимости использовать численную оптимизацию методы и, таким образом, обеспечивают очень хорошее время вычислений.

3 Оценка компонентов дисперсии: алгоритм SAP

В этом разделе мы представляем основной результат статьи. С, на основе PQL оценка модели (4) реализуется путем повторной оценки *работающей* линейной смешанной модели (см. Приложение 1), здесь мы сосредоточимся на оценке компоненты дисперсии в каждой из этих итераций. Согласие- Таким образом, при небольшом злоупотреблении терминологией мы будем ссылаться на к производным выражениям для компонент дисперсии как ML или REML оценивает, хотя, строго говоря, это только применяется для нормально распределенных ответов с идентификационной ссылкой функция.

Ключевым моментом нашего предложения является вывод первого - частные производные порядка приближенных (ограниченных) log- правдоподобия отдельно для каждой компоненты дисперсии τ^2_1 а также τ^2_2 . Это разделение позволяет нам оценивать отдельно каждый компонент дисперсии, приравнивая частную производную первого порядка - Активы к нулю. Соответственно итерационный алгоритм, представленный в этой статье можно рассматривать как обобщение алгоритма по Schall (1991) для работы с нестандартными конструкциями ковариационная матрица G . Однако некоторые другие итерационные процедуры dures также могли использоваться в этом контексте. Однажды получены отдельные частные производные первого порядка, частные производные второго порядка также могут быть получены, если использование градиентных алг[redacted]их как Ньютон-Рафсон или оценка Фишера (Harville 197[redacted]шенные версии эти, как средняя информация (Г[redacted]995 г.). Учитывая, что упомянутое разделение является ключевой новинкой В статье мы обозначили алгоритм как SAP: Separation анизотропных штрафов.

В целях иллюстрации в этом разделе мы ограничиваем наше внимание к оценке компонентов дисперсии на основе REML. Однако оценки ML также могут быть легко получено по тем же рассуждениям, которые будут использоваться для

инверсия A , а A / M обозначает диагональную матрицу, образованную от поэлементного частного от деления A и M .

Теорема *На каждой итерации алгоритма Фишера-Скоринга REML оценки компонент дисперсии* τ^2_d $(d = 1, 2)$ *и, когда неизвестно,* φ *задаются формулами*

$$\tau^2_d = \frac{\alpha^t \Lambda_d \alpha}{ed_d},$$

$$\varphi = \frac{\left(z - X\beta - Z\alpha\right)^t \breve{W} \left(z - X\beta - Z\alpha\right)}{n - \sum_{d=1}^{2} ed_d - \text{ранг}(X)}, \quad (6)$$

с участием

$$ed_d = \text{след}\left(Z^t PZG \frac{\Lambda_d}{\tau^2_d} G\right)$$

где $P = V^{-1} - V^{-1} X \left(X^t V^{-1} X\right)^{-1} X^t V^{-1}$ *с* $V = \varphi \breve{W}^{-1} + ZGZ^t$, $\breve{W} = \varphi W$ *и*

$$\Lambda_1 = \begin{pmatrix} \breve{\Sigma}_2 \otimes I_{q_1} & & \\ & 0_{q_2(c_1-q_1)} & \\ & & \breve{\Sigma}_2 \otimes I_{c_1-q_1} \end{pmatrix},$$

$$\Lambda_2 = \begin{pmatrix} 0_{q_1(c_2-q_2)} & & \\ & I_{q_2} \otimes \breve{\Sigma}_1 & \\ & & I_{c_2-q_2} \otimes \breve{\Sigma}_1 \end{pmatrix},$$

где 0_q *- квадратная матрица нулей порядка q.*

Доказательство теоремы приведено в Приложении 1. Сканирование Как видно из доказательства, $ed_1 + ed_2$ соответствует действующему Размер штрафной части подобранной модели. Этот эффективное измерение (плюс размер нераскрытого часть), можно, как обычно, интерпретировать как меру гладкости размер подогнанной поверхности взаимодействия. Тем не менее было бы интересно прояснить интерпретацию ed_d в этом контексте текст. Может возникнуть соблазн интерпретировать эти величины как мера гладкости в соответствующей ковариате (как, например, в аддитивном случае или в многомерном множестве ting, в P-сплайн ANOVA, предложенном Lee et al . 2013). Тем не менее, детальная оценка того, как эти значения сочетаются предполагаемый приносит совершенно другой и, возможно, удивительный, результат. *Вычисление* следа $Z^t PZG \frac{\Lambda_d}{\tau^2_d} G$, учитывая это $\frac{\Lambda_d}{\tau^2_d} G$ - диагональная матрица (поскольку Λ_d и G таковы), может быть получено как

REML. Соответствующие выражения в замкнутой форме для ML оценки компонентов дисперсии описаны в Дополнительные материалы онлайн.

Для удобства чтения мы будем использовать следующие обозначения: для обозначения операций над диагональными матрицами: пусть $\boldsymbol{A}$ и $\boldsymbol{M}$ - диагональные матрицы, $\boldsymbol{A}$ - вектор, содержащий диагональные элементы $\boldsymbol{A}$, 2 обозначает диагональную матрицу, диагональ которой образована поэлементным квадратом $\boldsymbol{A}$, $1 / \boldsymbol{A}$ обозначает диагональную матрицу, образованную поэлементным

$$\text{след}\left(\boldsymbol{Z}_t \boldsymbol{PZG} \frac{d}{\tau_d^2} \text{грамм} \right) \text{знак равно} \sum_{j=1}^{c_1 c_2 - q_1 q_2} \gamma_j \phi_j^d , \quad (7)$$

где γ_j - j- й элемент диагонали $\boldsymbol{Z}_t \boldsymbol{PZG}$, а ϕ^d_j - j- й элемент диагонали $\frac{d}{\tau_d^2} \boldsymbol{G}$. Учитывая, что след из $\boldsymbol{Z}_m \boldsymbol{PZG}$ соответствует следу шлема матрицы штрафная часть (см. доказательство теоремы), выражение (7) может можно интерпретировать как разложение *эффективной размерности*

подобранной модели на компоненты, относящиеся к каждой ковариате x_d согласно значениям ϕ^d_j. Взглянув на $\frac{d}{\tau_d^2}$ **грамм** матрица, мы имеем

$$\tau_2^2 \frac{2}{} \boldsymbol{G} = \begin{pmatrix} \boldsymbol{Я}_{q_1 (c_2 - q_2)} & & \\ & \boldsymbol{0}_{q_2 (c_1 - q_1)} & \\ & & \frac{\frac{1}{\tau_2^2} \tilde{\ }^2 \otimes \boldsymbol{I}_{c_1 - q_1}}{\frac{1}{\tau_2^2} \tilde{\ }^2 \otimes \boldsymbol{I}_{c_1 - q_1} + \frac{1}{\tau_1^2} \boldsymbol{I}_{c_2 - q_2} \otimes \tilde{\ }^1} \end{pmatrix},$$

а также,

$$\tau_1^2 \frac{1}{} \boldsymbol{G} = \begin{pmatrix} \boldsymbol{0}_{q_1 (c_2 - q_2)} & & \\ & \boldsymbol{Я}_{q_2 (c_1 - q_1)} & \\ & & \frac{\frac{1}{\tau_1^2} \boldsymbol{I}_{c_2 - q_2} \otimes \tilde{\ }^1}{\frac{1}{\tau_2^2} \tilde{\ }^2 \otimes \boldsymbol{I}_{c_1 - q_1} + \frac{1}{\tau_1^2} \boldsymbol{I}_{c_2 - q_2} \otimes \tilde{\ }^1} \end{pmatrix}.$$

В результате первые $q_1 (c_2 - q_2)$ элементов диагонали $\boldsymbol{Z}_t \boldsymbol{PZG}$ распределяются по ковариате x_2 , следующие $q_2 (c_1 - q_1)$ в x_1 , а последние $(c_1 - q_1)(c_2 - q_2)$ элементы являются распределены между x_1 и x_2 в соответствии с весами ϕ^d_j, которые обратно пропорциональна компоненту дисперсии, связанной с соответствующей ковариатой. Однако альтернатива интерпретация может быть предоставлена выражением этих весов как

$$\frac{\tau_1^2 \tilde{\ }^2 \otimes \boldsymbol{I}_{c_1 - q_1}}{\tau_1^2 \tilde{\ }^2 \otimes \boldsymbol{I}_{c_1 - q_1} + \tau_2^2 \boldsymbol{I}_{c_2 - q_2} \otimes \tilde{\ }^1} \quad \text{а также}$$

$$\frac{\tau_2^2 \boldsymbol{I}_{c_2 - q_2} \otimes \tilde{\ }^1}{\tau_1^2 \tilde{\ }^2 \otimes \boldsymbol{I}_{c_1 - q_1} + \tau_2^2 \boldsymbol{I}_{c_2 - q_2} \otimes \tilde{\ }^1} .$$

Отсюда следует, что последние $(c_1 - q_1)(c_2 - q_2)$ элементы диагонали $\boldsymbol{Z}_t \boldsymbol{PZG}$ распределяются на x_1 в соответствии с весами которые *прямо* пропорциональны компоненту дисперсии ассо- связанный с x_2 (и то же самое верно для x_2).

Соответственно, принимая во внимание трехчленное разложение двумерной поверхности f на (1) объяснил в разд. 2.1, каждое ed_d может быть получено как сумма двух компоненты, которые можно интерпретировать следующим образом: тот, который собирает количество сглаживания вдоль x_d (своего рода *внутри* гладкость), и другой, который собирает, насколько плавный эффект x_d изменяется по другой ковариате (*между* гладкость).

3.1 Алгоритм оценки

В этом разделе мы резюмируем алгоритм оценки модели (4):

Инициализировать. Установите начальные значения для фиксированных и случайных моделей. эффекты и компоненты дисперсии. Так , например, $\beta^{(0)}_k$ знак равно $\alpha^{(0)}_л = 0 \ (1 \leq k \leq q_1 q_2 ,\ 1 \leq l \leq (c_1 c_2 - q_1 q_2))$ и $m^{2(0)}_1 = T^{2(0)}_2 = 1$. В тех случаях, когда φ равен неизвестно, установите начальное значение для этого параметра, например, $\varphi_{(0)} = 1$. Положим $к = 0$

Шаг 1. Учитывая начальные *оценки* фиксированных и случайных значений модели. dom, построим *рабочую* переменную z и матрица весов $\boldsymbol{W}$ в следующем виде

$$z_i = g(\mu_i^{(k)}) + (y_i - \mu_i^{(k)}) g'(\mu_я^{(k)}),$$

$$w_{ii} = \left\{ \Phi_{(к)} \left[g'(\mu_я^{(k)}) \right]^2 \nu(\mu_я^{(k)}) \right\}^{-1},$$

$$\text{с } \boldsymbol{\mu}_{(k)} = g^{-1} \left(\boldsymbol{X\beta}^{(k)} + \boldsymbol{Z\alpha}_{(к)} \right) .$$

Шаг 1.1. Учитывая первоначальные *оценки* дисперсии компо- элементов, оценим $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$, решив линейную систему (12). Пусть $\boldsymbol{\alpha}$ и $\boldsymbol{p}$ будут эти оценки.

Шаг 1.2. Оцените компоненты дисперсии как

$$\tau^2_d = \frac{\boldsymbol{\alpha}_m \, d \, \boldsymbol{\alpha}}{ред^{(k)}_d} ,$$

и, при необходимости,

$$\varphi = \frac{\left(\boldsymbol{z} - \boldsymbol{X\beta} - \boldsymbol{Z\alpha} \right)_t \tilde{\ } \boldsymbol{W} \left(\boldsymbol{z} - \boldsymbol{X\beta} - \boldsymbol{Z\alpha} \right)}{n - \sum_{d=1}^{2} ед.^{(k)}_d - n} ,$$

с участием

$$ред^{(k)}_d = \text{след} \left(\boldsymbol{Z}_t \boldsymbol{P}^{(k)} \boldsymbol{ZG}^{(k)} \frac{d}{T^{2(к)}_d} \text{грамм}^{(k)} \right) ,$$

где $\boldsymbol{P}^{(k)}$ и $\boldsymbol{G}^{(k)}$ обозначим соответствующие $\boldsymbol{P}$ и $\boldsymbol{G}$ матрицы, полученные на основе *исходных* оценок.

Шаг 1.3. Повторите Шаг 1.1 и Шаг 1.2 с $m^{2(к)}_1$, $T^{2(к)}_2$, и, если они будут обновлены, φ заменяется m^2_1 , τ^2_2 , и φ соответственно, пока критерий сходимости

$$| \Phi - \varphi_{(к)} | + \frac{\sum_{д=1}^{2} | m^2_д - \tau^{2(к)}_д |}{3} \leq \varsigma,$$

где ς - небольшой порог (допуск на критерий вергентности), например, 1×10^{-6}

Шаг 2. Повторите шаг 1. с фиксированными и случайными значениями модели. компоненты эффектов и дисперсии заменяются на те полученный на последней итерации шагов 1.1 – Шаг 1.3, пока критерий сходимости

$$\frac{\| \boldsymbol{\eta}_{(k+1)} - \boldsymbol{\eta}_{(k)} \|_2}{\| \boldsymbol{\eta}_{(к+1)} \|_2} \leq \upsilon,$$

где υ - малый порог.

3.2 Вычислительные аспекты

Мы представляем здесь некоторые вычислительные аспекты, которые могут быть используется для быстрой реализации алгоритма оценки представлены в разд. 3.1. В частности, мы фокусируемся на вычислительных Изменение компонент дисперсии (шаг 1.2). Также должно быть отметил, что, когда данные находятся в структуре массива, общие - Модель линейного массива (GLAM) Currie et al. (2006) может использоваться для построения модельных матриц задействованных в линейной системе (12), тем самым повышая скорость алгоритм оценки.

Оценка компонентов дисперсии с использованием выражение, данное в (6) требует вычисления следа из $Z_m PZG \frac{\partial}{\tau_d^2} G$, который включает вычисление и манипулирование использование нескольких матриц размера $n \times n$. Как указывалось ранее, это вычисления можно упростить, если учесть, что оба, G и d - диагональные матрицы, поэтому $G_d G$ является также диагональная матрица. потом

- $G_d G = \mathrm{diag}(G * d * G)$, где $*$ обозначает элемент-мудрое векторное произведение.
- Для вычисления первой трассы требуется только компьютер-ция диагонали $Z_t PZ$.

Более того, по выражению (5.3) в Харвилле (1977) у нас есть

$$Z_t PZ = \left[\mathbf{0}_{(c_1 c_2 - q_1 q_2) \times q_1 q_2} \mid Я_{(c_1 c_2 - q_1 q_2)} \right] C^{-1} [X \mid Z]_t WZ,$$

с C^{-1} является обратной матрицей C, определенной в (12). Кор-соответственно, вычисление его диагонали можно проводить путем добавления по столбцам

$$\left(\left[\mathbf{0}_{(c_1 c_2 - q_1 q_2) \times q_1 q_2} \mid Я_{(c_1 c_2 - q_1 q_2)} \right] C^{-1} \odot \begin{bmatrix} X_t WZ \\ Z_t WZ \end{bmatrix} \right)_m, \quad (8)$$

где обозначает произведение Адамара или поэлементной матрицы. uct. Для простоты обозначений обозначим ζ этот диагональный вектор, и $\xi_d = G * d * G$. Из этого следует, что:

$$след \left(Z_t PZG \frac{d}{\tau_d^2} грамм \right) знак равно \frac{1}{\tau_d^2} \sum_{j=1}^{c_1 c_2 - q_1 q_2} \zeta_j \xi_{dj}.$$

Обратите внимание, что для оценки не требуется вычислять новые матрицы. выражение (8), поскольку все они уже вычислены для оценки p и $\& alpha$; .

$$грамм^{-1} = \begin{pmatrix} \frac{d_{1u}}{\tau_1^2} & & & & & & \\ & \frac{d_{2u}}{\tau_2^2} & & & & & \\ & & \frac{d_{3u}}{\tau_3^2} & & & & \\ & & & \frac{d_{11}}{\tau_1^2} + \frac{d_{21}}{\tau_2^2} & & & \\ & & & & \frac{d_{12}}{\tau_1^2} + \frac{d_{31}}{\tau_3^2} & & \\ & & & & & \frac{d_{22}}{\tau_2^2} + \frac{d_{32}}{\tau_3^2} & \\ & & & & & & \frac{\partial_{1m}}{\tau_1^2} + \frac{d_{2m}}{\tau_2^2} + \frac{d_{3m}}{\tau_3^2} \end{pmatrix},$$

4 Некоторые расширения

В этом разделе мы представляем некоторые расширения алгоритма SAP. рифм представлен в разд. 3. Как будет видно, ключевой момент состоит в том, чтобы определить матрицу вариации-ковариации G случайного dom, а также его производные по переменной компоненты. В частности, единственным требованием будет чтобы указать форму матрицы, входящей в выражение оценки каждой компоненты дисперсии [см. (6)]. Этот функция делает, например, прямое расширение алгоритм SAP для работы с декомпозицией ANOVA-типа. позиция, представленная в Ли и Дурбане (2011 г.). Поэтому мы сфокусируйтесь здесь на представлении более сложных расширений. Мы сначала представить обобщение алгоритма SAP на три размерный случай, а затем мы покажем, как алгоритм может также должны быть включены в оценку GAMM (Lin и Zhang 1999[redacted]

4.1 Распространение на трехмерный случай

Рассмотрим трехмерную обобщенную регрессионную задачу. лем

$$g(E[y_i \mid x_i]) = g(\mu_i) = \eta_i = f(x_{i1}, x_{i2}, x_{i3}),$$

где f - гладкая неизвестная функция. Что касается биди В частном случае мы моделируем функцию f тензорным произведением сплайновые базисные функции, и мы предполагаем анизотропную штрафную ization

$$\breve{P} = \lambda_1 \breve{P}_1 \otimes I_{c_2} \otimes I_{c_3} + \lambda_2 I_{c_1} \otimes \breve{P}_2 \otimes I_{c_3} + \lambda_3 I_{c_1} \otimes I_{c_2} \otimes \breve{P}_3.$$

Следуя той же процедур[redacted]зд. 2.1 для бид[redacted] случай (см. Lee 2010 ; L[redacted]n 2011 для дал[redacted] детали), получаем матр[redacted] смешанной модели

$$X = [X_1 \square X_2 \square X_3]$$

$$Z = [Z_1 \square X_2 \square X_3 \mid X_1 \square Z_2 \square X_3 \mid X_1 \square X_2 \square Z_3 \mid Z_1 \square Z_2 \square X_3 \mid Z_1 \square X_2 \square Z_3 \mid X_1 \square Z_2 \square Z_3 \mid Z_1 \square Z_2 \square Z_3],$$

и инверсия ковариационной матрицы случайных эффектов G

Стр.7

где

$d_{1u} = \breve{}_1 \otimes I_{q_2} \otimes I_{q_3}$, $d_{2u} = I_{q_1} \otimes \breve{}_2 \otimes I_{q_3}$, $d_{3u} = I_{q_1} \otimes I_{q_2} \otimes \breve{}_3$,

$d_{11} = \breve{}_1 \otimes I_{c_2 - q_2} \otimes I_{q_3}$, $d_{12} = \breve{}_1 \otimes I_{q_2} \otimes I_{c_3 - q_3}$, $d_{21} = I_{c_1 - q_1} \otimes \breve{}_2 \otimes I_{q_3}$,

$d_{22} = I_{q_1} \otimes \breve{}_2 \otimes I_{c_3 - q_3}$, $d_{31} = I_{c_1 - q_1} \otimes I_{q_2} \otimes \breve{}_3$, $d_{32} = I_{q_1} \otimes I_{c_2 - q_2} \otimes \breve{}_3$,

$d_{1t} = \breve{}_1 \otimes I_{c_2 - q_2} \otimes I_{c_3 - q_3}$, $d_{2t} = I_{c_1 - q_1} \otimes \breve{}_2 \otimes I_{c_3 - q_3}$, $\partial_{3m} = я_{c_1 - \partial_1} \otimes I_{гр_2 - \partial_2} \otimes \sim_3$.

Как показано в разд. 3 и Приложение 1, ковариационная матрица G и его производные по компонентам дисперсии τ_d^2 $(d = 1, 2, 3)$ легко получить

$$\frac{\partial G}{\partial \tau_d^2} знак равно \frac{1}{\tau_d^4} G_d G,$$

с участием

$$_1 = \mathrm{diag}(d_{1u}, \mathbf{0}_{q_1 q_3 (c_2 - q_2)}, \mathbf{0}_{q_1 q_2 (c_3 - q_3)}, d_{11}, d_{12}, \mathbf{0}_{q_1 (c_2 - q_2)(c_3 - q_3)}, d_{1t}),$$

с X_l и Z_l $(l = 1, ..., P)$, как определено в разд. 2.1 и $\breve{X}_p = \left[x_p \mid \cdots \mid Икс_p^{(q_p - 1)} \right]$, где вектор единиц был удален из X_p для обеспечения идентифицируемости $(p = 3, ..., P)$. Наконец, U_j $(j = 1, ..., c)$ - матрицы случайных эффектов связанные с *собственно* случайными эффектами u_j . Это прямо-вперед, чтобы показать, что ковариационная матрица G случайного эффекты $(\alpha_t, u_t)_t$ становится

$$\begin{pmatrix} \breve{G} & \\ & \tau_3^2 \end{pmatrix}$$

$_2$ = диаг $(\mathbf{0}_{q_2 q_3 (c_1 - q_1)}, \mathbf{d}_{2u}, \mathbf{0}_{q_1 q_2 (c_3 - q_3)},$
$\mathbf{d}_{21}, \mathbf{0}_{q_2 (c_1 - q_1)(c_3 - q_3)}, \mathbf{d}_{22}, \mathbf{d}_{2t}),$

$_3$ = диаг $(\mathbf{0}_{q_2 q_3 (c_1 - q_1)}, \mathbf{0}_{q_1 q_2 (c_2 - q_2)}, \mathbf{d}_{3u},$
$\mathbf{0}_{q_3 (c_1 - q_1)(c_2 - q_2)}, \mathbf{d}_{31}, \mathbf{d}_{32}, \mathbf{d}_{3t}).$

Наконец, получены оценки компонент дисперсии.
согласно выражению (6).

$$\mathbf{G} = \begin{pmatrix} \tilde{}_3 & & \\ & \cdots & \\ & & \tau^2_p \tilde{}_p \end{pmatrix},$$

где $\tilde{\mathbf{G}}$ определено в (15), а τ^2_p - дисперсия компоненты, связанные с гладкой функцией f_p ($p = 3, ..., \Pi$).

4.2 Расширение обобщенных аддитивных смешанных моделей

Рассмотрим обобщенную аддитивную смешанную модель

$$g(E[y_i | \mathbf{x}_i, \mathbf{u}]) = g(\mu_i) = \eta_i = f_{(1,2)}(x_{i1}, x_{i2}) + \sum_{p=3}^{n} f_p(x_{ip}) + \mathbf{U}_{m\,i1}\mathbf{u}_1 + \cdots + \mathbf{U}_{t\,ic}\mathbf{u}_c, \quad (9)$$

где $f_{(1,2)}$ и f_p ($p = 3, ..., P$) - гладкие функции.
ции, $\mathbf{u}_j$ - $k_j \times 1$ векторов случайных эффектов, таких что $\mathbf{u} = (\mathbf{y}_{m1}, ..., \mathbf{u}_{mc}) \sim N(\mathbf{0},)$, где $=$ диаг $(\sigma^2_1 \mathbf{1}_{K_1}, ..., \sigma^2_c \mathbf{1}_{K_c})$ И $\mathbf{U}_{IJ}$ известны векторы *ковариат* свя-
связано со случайными эффектами.

Оценить модель (9), каждое f_p ($p = 3, ..., P$) приближенно соединены сплайновым базисом низкого ранга (с матрицей штрафов $\lambda_p \tilde{\mathbf{P}}_p$), и, $f_{(1,2)}$, как показано в Разд. 2, тензорным произведением два одномерных сплайна по базису и анизотропный штраф. Более-выше, мы также принимаем здесь эквивалентность между (9) и GLMM. На основе SVD матриц штрафов $\tilde{\mathbf{P}}_j$ ($j = 1, ..., P$), получаем матрицы моделей смешанной модели

$$\mathbf{X} = [\mathbf{X}_2 \,\mathrm{D}\, \mathbf{X}_1 \,|\, \tilde{\mathbf{X}}_3 \,|\, \cdots \,|\, \tilde{\mathbf{X}}_P]$$

$$\mathbf{Z} = [\mathbf{Z}_2 \,\mathrm{D}\, \mathbf{X}_1 \,|\, \mathbf{X}_2 \,\mathrm{D}\, \mathbf{Z}_1 \,|\, \mathbf{Z}_2 \,\mathrm{D}\, \mathbf{Z}_1 \,|\, \mathbf{Z}_3 \,|\, \cdots \,|\, \mathbf{Z}_P \,|\, \mathbf{U}_1 \,|\, \cdots \,|\, \mathbf{U}_c],$$

Замкнутые выражения для оценок дисперсии компоненты τ^2_l ($l = 1, ..., P$) и σ^2_j ($j = 1, ..., c$) на основе на REML / ML можно получить, используя ту же процедуру, что и представлены в разд. 3. Как указывалось ранее, нам просто нужно рассчитать матрица, входящая в производную $\mathbf{G}$ с по каждому компоненту дисперсии. В случае τ^2_1 а также τ^2_2, эти матрицы эквивалентны матрицам, определенным в (14), но с подматрицей нулей, соответствующей этим блокам матрица $\mathbf{G}$, где τ^2_p и σ^2_j появляются. Причем для каждого τ^2_p, это легко показать, что

$$_p = \mathrm{diag}(\mathbf{0}_{(c_1 c_2 - q_1 q_2)}, \mathbf{0}_{(c_3 - q_3)}, ..., \tilde{}_p, ..., \mathbf{0}_{(c_P - q_P)}, \mathbf{0}_K), \text{ при } K = \sum_{j=1}^{c} K_j,$$

и, поскольку компоненты дисперсии σ^2_j ($j = 1, ..., c$) есть обеспокоены, мы получаем

$$_j = \mathrm{diag}(\mathbf{0}_{(c_1 c_2 - q_1 q_2)}, \mathbf{0}_{(c_3 - q_3)}, ..., \mathbf{0}_{(c_P - q_P)}, \mathbf{0}_{k_1}, ..., \mathbf{1}_{K_j}, ..., \mathbf{0}_{K_c}).$$

5 Имитационное исследование

В этом разделе сообщается о результатах имитационного исследования кон-
направлен на изучение эмпирической эффективности оценки процедура, описанная в разд. 3 выше. В частности, цели этого исследования были двоякими: (а) для оценки практического поведения iour на основе среднеквадратичной ошибки (RMSE); а также б) изучить эффективность алгоритма с точки зрения вычислительное время.

Для этих целей мы сравнили алгоритм SAP с метод, приведенный в Вуде (2011) совместно с анизотропным подход тензорного произведения, представленный в Вуде (2006b). В лесу (2011) автор представляет быстрый и стабильный подход к оценка параметров сглаживания GAM на основе на ML или REML. Этот подход превосходит - с точки зрения MSE, сбоев конвергенции и вычислительных затрат - предыдущие подходы в этом контексте (см Вуда 2011 другие детали), и поэтому он был выбран в качестве с mark для наших симуляций. Более того, метод реализуется в `гам ()` функция `R` -package `mgcv` (версия 1.7-22) (Дерево 2006a). `Mgcv` пакет в последние годы стал эталонным `R`- пакетом для оценка GAM, благодаря своей универсальности, простоте в использовании интерфейс и хорошая и стабильная работа. Обратите внимание, что `R` -package `mgcv` также включает в себя несильно `BAM ()` специально предназначен для работы с очень большими наборами данных, которые, в свою очередь, могут быть намного быстрее, чем `gam ()`. Мы знаем, что оценка доработка предложенного алгоритма в части вычислительной время будет более точным и справедливым с соблюдать `бам ()` функцию. Однако предварительное моделирование-исследования показали, что в некоторых случаях это функция представляет собой серьезные проблемы сходимости, таким образом, 343 время расчета около 30 минут для небольшой выборки размеры. Более того, для средних размеров выборки (как те, которые использовались в это исследование) время вычислений может быть даже больше, чем с использование `gam ()`. По всем этим причинам в этой модели исследования, мы ограничили сравнения нашего подхода к `гам ()` функция.

2. $y \sim$ *Бернулли* (p), где $p = \exp(\tilde{\eta}) / \exp(1 + \tilde{\eta})$,
где $\sim \eta = (\eta + 0.2) / 0.5$,

где коэффициенты масштабирования, которые появляются в случае Бернулли использовались для контроля отношения сигнал / шум. Для каждого мар-
был выбран 14-мерный базис, $R = 500$ реплицирования.
Каты были выполнены.

Затем на основе предыдущего сценария мы оценили влияние увеличения размера выборки, и, как следствие, Изучите базовое измерение на время вычислений. Здесь, моделирование проводилось исходя из размера выборки 1000, и только $\sigma = 0.5$ рассматривался для гауссова случая. $R = 100$ повторностей, 30-мерный маргиналы были выбраны.

Наконец, мы также провели небольшое исследование с использованием моделирования. три коварианты. Пятьсот значений коварианты x_1, x_2 и x_3 моделировались независимо от равномерного распределения на интервале [0, 1], и ответ был сгенерирован из

$$y = 1.5 \text{ опыта} \left(-\frac{(x_1 - 0.2)^2}{5} - \frac{(x_2 - 0.5)^2}{3} - \frac{(x_3 - 0.9)^2}{4} \right) + 0.5 \text{ опыта} \left(-\frac{(x_1 - 0.3)^2}{4} - \frac{(x_2 - 0.7)^2}{2} - \frac{(x_3 - 0.4)^2}{6} \right) + \exp\left(-\frac{(x_1 - 0.1)^2}{5} - \frac{(x_2 - 0.3)^2}{5} - \frac{(x_3 - 0.7)^2}{4} \right) + \varepsilon,$$

где $\varepsilon \sim N(0, \sigma^2)$ (см. Wood 2006b). Что касается первого Исследование, различные уровни шума были рассмотрены $(c_2 \in \{0.1; 0.5; 1\})$, $R =$ были выполнены 500 повторов, но только 7-мерные маргиналы использовались, определяя базовый размер из 343.

Для обоих, алгоритм SAP и `гам ()` функция, базисные функции кубических B-сплайнов с разностью второго порядка. штраф ($q_d = 2$) были выбраны для получения предельного матриц модели, а для оценки использовался критерий REML. Формирование компонентов дисперсии. Для `GAM ()` функ-
ции тензорное произведение маргинальных базисов ( )
использовалась и предполагалась анизотропия (Вуд 2 ием

5.1 Сценарии и настройка

В первом исследовании 200 значений ковариат x_1 и x_2 были моделируется независимо от равномерного распределения на interval [0 , 1], и был рассмотрен следующий сценарий:

$$\eta = f(x_1, x_2) = \cos\left(2\pi\sqrt{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}\right).$$

Обратите внимание, что этот сценарий также использовался в Lee et al. (2013). Данные ответа y были затем сгенерированы под двумя разными раздачи:

1. $y = \eta + \varepsilon$, где $\varepsilon \sim N(0, \sigma^2)$ где $\sigma \in \{0.1; 0.5; 1\}$.

что касается числовых вариантов процесса подгонки, для алгоритм SAP, допуск на критерий сходимости терион компонентов дисперсии и оценка Фишера алгоритм был установлен на 1×10^{-6} , а начальные значения компоненты дисперсии, фиксированные и случайные эффекты были установлены на 1 и 0 соответственно. Что касается `GAM ()` функции обеспокоен, числовые варианты были установлены по умолчанию. Оценка практической работы оба подхода оценивались на основе RMSE, com- рассчитано на наблюдаемых ковариатных значениях. Для гауссовых данных истинный линейный предиктор был выбран в качестве цели. Тем не мение, в случае двоичных данных RMSE вычислялась на шкала отклика (вероятность). Наконец, что касается оценка времени вычислений, для алгоритма SAP

Стр.9

указанное время включает время вычислений, необходимое для (a) построение матриц, участвующих в алгоритме; и (б) сам алгоритм. Все расчеты вы[redacted]4-бит) R 3.0.1 (R Core Team 2013[redacted]е 2,40 ГГц Компьютер с процессором i5 с[redacted] Windows® 7 Операционная система.

5.2 Результаты

Рисунок 1а показывает результаты с точки зрения RMSE для двух размерный случай, гауссово распределение и выборка размер $n = 200$. На рисунке показан журнал $_{10}$ (RMSE) обоих подходов (левая ось Y), а также разница между $\log_{10}$ (RMSE) алгоритма SAP и метода Вуда (2011) (правая ось Y). Таким образом, в последнем случае значения ниже чем ноль указывает на лучшее поведение нового предложения. Как можно заметить, алгоритм SAP показал лучшую производительность. во всех случаях. Однако различия между ними приближения уменьшаются по мере уменьшения отношения сигнал / шум. Рисунок 1b изображает поведение обоих подходов до тех пор, пока что касается эффективного измерения. Для простоты меж- Мы включили в этот рисунок соотношение эффективный размер Дерева (2011) к Алгоритм SAP (правая ось Y). Алгоритм SAP предусматривает, в в общем, более низкий эффективный размер, чем указанный метод в лесу (2011), хотя, опять же, эти различия уменьшаются. при уменьшении отношения сигнал / шум. На основе как среднеквадратичное отклонение, так и результаты эффективного изме размер, что для высокого отношения сигнал / шум, метод, приведенный в Дерево (2011) имеет тенденцию к недостаточной плавности по сравнению с подход. Рис. 1с, результаты в отношении вычислений время работы обоих подходов, а также их Соотношение (в лог - $_{10}$ масштабе). Опять же, для алгоритма SAP com- Время установки зависит от отношения сигнал / шум. В чем ниже отношение сигнал / шум, тем медленнее сходимость алгоритм. Например, медиана (диапазон) числа количество итераций было 12 (7, 23), 17 (8, 41) и 19 (9, 61), для $\sigma = 0.1$, $\sigma = 0.5$ и при $\sigma = 1$ соответственно. Несмотря на в этом отношении наше предложение превосходит Wood (2011 г.) метод, требуя, в среднем, от 13 . 0 (при $\sigma = 0.1$) и 7 . 12 (для $\sigma = 1$) раз меньше времени вычислений.

Результаты для двумерного случая, Бернулли распределение и $n = 200$, показаны на рис. 2. Опять же, наш метод превосходит Wood (2011 г.) с точки зрения обоих, RMSE и время вычисления. Однако отличия в этом случае не так заметны, как в гауссовском случае. Для Например, требуемое время вычисления нашего алгоритма составляло, в медиане, 3 . В 19 раз меньше, чем у Wood (2011 г.) метод. Еще раз, эффективное измерение алгоритма SAP был ниже, чем эффективный размер, полученный с помощью `gam ()` функция. Наконец, что касается количества запросов PQL. Для достижения конвергенции медиана (диапазон) была 5 (4, 7).

На рисунке 3 показаны результаты для двумерного случая. и $n = 1,000$, и для RMSE и вычислений время соответственно. Как указывалось ранее, для гауссовского только случай $\sigma = 0.5$ считалось. Что касается вычислительных- время, как можно заметить при сравнении Сопоставляя эти результаты с результатами, представленными на рис. 1с и 2с для $n = 200$, поведение алгоритма SAP относительно Wood (2011 г.) метод улучшается по мере увеличения размера выборки, и поэтому размер основания увеличивается. В этом случае наши метод необходим 10 . 55 и 4 . В 00 раз меньше, чем метод автор: Вуд (2011 г.), для распределения Гаусса и Бернулли соответственно.

Наконец, результаты для трехмерного случая таковы: изображенный на рис. 4. По той же схеме, что и в предыдущих исследованиях отображается (результаты для эффективного измерения не показаны). Что касается времени вычислений, алгоритм SAP требовал, в медиана, 10, 4,66 и 3,65 (для $\sigma = 0.1$, $\sigma = 0.5$ и $\sigma = 1$ соответственно) раз меньше, чем метод, приведенный в Дерево (2011).

Отметим, что, несмотря на то, что в сравнение исследований с помощью моделирования между обоими подходами сделано максимально справедливо (с точки зрения основы, сглаживания критерий выбора параметра,…), имеется несколько символов теристика алгоритма SAP и Wood (2011) метод это могло бы объяснить наблюдаемые различия. В этом смысле, главное несоответствие состоит в том, что оба подхода используют разные текущая параметризация модели (2) как смешанная модель. Помимо этого факта, некоторые другие характеристики орудия- обсуждение обоих предложений может также повлиять на результаты, такие как выбранный критерий сходимости или начальный значения.

6 Применение к реальным данным

В этом разделе мы проиллюстрируем полезность вычислительной алгоритм представлен в разд. 3 на двух реальных примерах. В первый пример показывает эффективность нашего подхода в простейший случай, 2D случай, но с довольно большим размером выборки что требует относительно большого количества внутренних узлов. С участием Во втором примере мы проиллюстрируем, как алгоритм может также использоваться в трехмерном случае с негауссовым ответом. Более того, поскольку данные в этом случае находятся в структуре массива, мы также воспользуйтесь возможностью использования GLAM в этом контекст.

6.1 Данные об осадках

Этот набор данных содержит записи наблюдений за погодой, собранные в Соединенных Штатах Америки (США). Данные поступают из Национальный центр климатических данных США и кон- Ежемесячное общее количество осадков (в миллиметрах) с января С 1895 г. по декабрь 1997 г.

а)

(б)

(с)

Рис.1 Сравнение производительности алгоритма SAP и Дерево (2011) для двумерного случая. Коробчатые сюжеты показать результаты для распределения Гаусса, различных уровней шума $\sigma \in \{0.1; 0.5; 1\}$, размер выборки $n = 200$ и $R = 500$ реплик. Кейтс. Сверху вниз: бревно $_{10}$ (СКО), **б** измерение эффективного и **c** Время вычисления (с). В каждом случае две левые диаграммы показывают $\log_{10}$ (RMSE), эффективное измерение или достигнутое время вычислений при каждом подходе (*левая* ось y), в то время как один из правых (*правая* ось y) показывает: **a** журнал SAP $_{10}$ (RMSE) минус журнал $_{10}$ (RMSE) gam () функция, **b** отношение эффективных размеров дерева (2011 г.) метод к алгоритму SAP; и **c** отношение времени вычисления гам () функция алгоритма SAP (в лог - $_{10}$ масштабе)

наш анализ по оценке пространственного распределения осадков на апрель 1948 г. в США. Этот ограниченный набор данных может быть найдено в спаме R- пакета под названием USprecip , доступно на сайте cran.r-project.org (Основная команда R 2013). В частности, набор данных включает в себя 11 918 записи. Для каждой записи долгота-широта предусмотрены станции мониторинга, вместе с ежемесячной суммой осадки в миллиметрах и стандартизация этого сырья наблюдение, называемое *аномалией* (см. Johns et al . 2003). Из из этих 11 918 записей, только 5 906 соответствуют станциям, где

а)

(б)

(с)

Рис.2 Сравнение производительности алгоритма SAP и Дерево (2011) для двумерного случая. Коробчатые сюжеты показывают результаты для распределения Бернулли, размер выборки $n = 200$ и $R = 500$ повторностей. Слева направо: бревно $_{10}$ (СКО), **б** Действуют время, достигаемое каждым подходом (*левая* ось y), а время правого (*правая* ось Y) показывает: **a** журнал SAP $_{10}$ (RMSE) минус журнал $_{10}$ (RMSE) функции gam () , **b** отношение эффективных размеров Дерево (2011 г.) к алгоритму SAP; и **c** отношение измерение и **с** Время вычисления (с). В каждом случае два левых ящика-графики показывают логарифм 10 (RMSE), эффективное измерение или вычисление время вычисления функции `gam ()` алгоритму SAP (в журнале 10 шкала)

Рис. 3 Сравнение производительность SAP алгоритм и Вуд (2011) метод двумерного дело. Коробчатые диаграммы показывают результаты для выборки размером $n = 1,000$ и $R = 100$ реплицирует. Сверху донизу: распределение Gaussian и **б** Распределение Бернулли. В каждом случае два левых коробчатых графика показывают журнал 10 (RMSE) или время вычислений достигается за счет каждый подход (*левая* ось Y), в то время как один из правых (*правых* ось y) показывает SAP log 10 (RMSE) минус log 10 (RMSE) гаммы () функция или соотношение время вычисления Леса (2011) в SAP алгоритм (в логарифмической шкале 10)

а)

(б)

наблюдались значения суммы осадков за месяц, а остальные 6012 соответствуют отсутствующим осадкам на станции значения, заполненные с использованием пространственной статистики (Джонс и др . … истинные рекорды.

Рисунок 5 а показывает необработанные данные о месячных осадках. аномалии в США за апрель 1948 года. построил двухмерную P-сплайн-модель с долготой и широтой как ковариаты,

атес, штрафы второго порядка и 40 внутренних узлов для каждого марша. Гинальный кубический B-шлицевой базис. Таким образом, модель имела основу измерения 1936 года. Допуск сходимости вари- … использовался REML. Встроенна поверхность показана на рис. 5б. Эффективное измерение для долготы и широты было 302,656 и 408,757 соответственно. тивно. Что касается времени вычислений, алгоритм взял 5,76 мин.

Стр. 12

а)

(б)

Рис. 4 Сравнение журнала 10 (RMSE) и времени вычисления (в секунд) производительность для алгоритма SAP и Wood (2011 г.) метод для трехмерного случая. Коробчатые диаграммы показывают результаты для гауссова распределения разные уровни шума $\sigma \in \{0.1; 0.5; 1\}$, размер выборки $n = 500$ и $R = 500$ повторностей. Верхняя фигура: журнал 10 (RMSE). Нижний рисунок: время вычисления (с). В каждом случае два левых прямоугольных графика показывают журнал 10 (RMSE) или время вычисления для каждого подход (*левая* ось Y), в то время как один из правых (*правая* ось Y) показывает: (*вверху*) журнал SAP 10 (RMSE) минус журнал 10 (RMSE) гаммы () функция; и (*внизу*) отношение времени вычисления Wood (2011 г.) к алгоритму SAP (в логарифмической шкале 10)

Для сравнения мы также проанализировали этот набор данных. используя `гам ()` и `бац ()` функции в `R` -package `mgcv` . Как указывалось ранее, функция `bam ()` была специально предназначен для работы с очень большими наборами данных. Однако поскольку при моделировании наблюдалась серьезная проблема сходимости поэтому при использовании этой функции этот набор данных также был ана- анализируется с помощью функции `gam ()` . Как и раньше, тензорное произведение сглаживания, а также штрафы второго порядка и 40 внутренних узлов для каждого маргинального кубического базиса B-сплайна, и критерий REML (`метод = «REML»` и `метод = «FREML»` для `gam ()` и `bam ()` соответственно) было выбрано sen для автоматического выбора параметров сглаживания (Дерево 2011). В обоих случаях числовые параметры процесс подгонки был установлен по умолчанию, а подгонка тинговые процессы сошлись. Что касается результатов использования `bam ()` подобранная модель имела эффективный размер 774,50, и время вычислений, достигнутое этим подход составил 22,17 мин, что примерно в 3,8 раза больше, чем с используя наш алгоритм. Что касается функции `gam ()` , эффективное измерение было 796,1, а время вычислений увеличено до 48,22 мин, что в 8,4 раза больше, чем при использовании продолжил подход.

6.2 Респираторные данные

В этом примере используются американские данные о количестве смертей. от респираторных заболеваний (Currie et al . 2006 г.). Набор данных содержит количество смертей в зависимости от возраста на момент смерти (от 1 до 105), календарный год смерти (от 1959–1998) и месяц смерти (от 1 до 12). В набор данных также содержит количество дней в месяце и году. В частности, набор данных представляет собой структуру массива размерных размеров. 105 × 40 × 12, всего 50 400 наблюдений. Эта функция предлагает нам возможность использования в сочетании Используя наш подход, GLAM для вычисления модельные матрицы, участвующие в (12).

Рис.5 Месячное количество осадков аномалии в США за апрель 1948. **a** Исходные данные. **b** Установлен поверхность

а)

(б)

Следуя статье Currie et al. (2006 г.) мы смоделировали количество смертей с 3D P-сплайновой моделью (с возрастом, год и месяц как ковариаты) с ошибкой Пуассона и лог-ссылкой. Логарифм количества дней в месяце использовался как смещение. Для всех анализов штрафы второго порядка вместе с 11, 6 и 3 внутренними узлами для краевого кубического B-сплайна возраст, год и месяц, соответственно, доходность - базисный размер 1050. Поскольку количество смертей в 1972 году был крайне неудачным, мы удалили в этом году из анализ, присвоив ему нулевой вес (см. Currie et al. 2006 г.). Чтобы ускорить время вычислений, начальная оценка из $(\beta_t, \alpha_t)_t$ было получено в предположении *лог* $\{(y+0.5)/d\}$ как начальную оценку $X\beta + Z\alpha$, где $\boldsymbol{y}$ и $\boldsymbol{d}$ - векторы, содержащие количество смертей и количество дней в месяц соответственно. При подгонке модели с использованием функции `bam ()` и `gam ()` , начальная оценка μ (аргумент `mustart`) был получен на основании начального предварительная оценка (β_t, α_t) объяснено ранее. Взяв во внимание предлагаемый алгоритм, допуск по критерию сходимости компонентов дисперсии и алгоритма оценки Фишера. ритм был установлен на 1×10^{-6}. Что касается анализов с использованием `R` -package `mgcv` касается, численные параметры для процесс подгонки был установлен по умолчанию. Чтобы сделать сравнения между нашим подходом и теми, кто использует `R` -package `mgcv` fair, мы подогнали модель по нашему алгоритму с GLAM и без.

Что касается численных результатов, то эффективная размерность для возраста, года и месяца составила 62,86, 194,34 и 209,55, соответственно, как для алгоритма SAP с, так и для без использования GLAM, что дает общий эффективный размер 474,75 (включая размер не расчетная или параметрическая часть). Что касается вычислений времени алгоритм занял 4,33 и 1,70 мин без и с использование GLAM соответственно. Как можно заметить, и как Ожидается, что использование GLAM в процессе оценки влияет только на время вычислений, сокращаясь за примерно в 2,6 раза. Для подходов `gam ()` и `bam ()` общий эффективный размер составлял 638,50 и 639,50 соответственно. Соответственно, время вычислений увеличилось до 105,39 мин. в случае функции `gam ()` и 27,50 для `bam ()` , т.е. примерно в 24,3 и 6,34 раза больше, чем с нашим алгоритмом. Несмотря на то, что общий эффективный размер мод- els подобраны с использованием предложенного алгоритма и алгоритмов, использующих `R` -package `mgcv` различаются в значительной степени, подогнанные значения про- При этом оба подхода были очень похожи. Фактически, среднее различий между подобранными значениями (на ответ а) (б) (с)

Рис. 6 Наблюдаемые (*кружок*) и сглаженные (*сплошная линия*) числа журнала (смертей / день) по возрасту, году и месяцу. *Черная линия* : предлагаемый алгоритм. *Красная линия* : `gam ()` функция. январь 1959; **б** возраст 53 года, январь; и **с** возраста 53 лет, 1959. (цвет фигуры онлайн)

scale), предоставляемые функцией `gam ()`, и полученные с нашим алгоритмом было 0,00712. Более того, 2.5 и 97,5% эмпирических квантилей этих различий составили -3 . 238 и 3,410 соответственно, что довольно мало, если учесть учтите, что наблюдаемое количество смертей колебалось между 0 и 1605. На рис. 6 расчетная логарифмическая смертность от показаны возраст, год и месяц для различных значений ковариаты. Как можно заметить, оба подхода дали одинаковые результаты. полученные результаты.

7 Обсуждение

В данной статье мы рассмотрели оценку гладкого параметры многомерного тензорного произведения, генерирующего модель P-сплайна с анизотропным штрафом. На основании смешанного модельного представления P-сплайна и использования методов PQL, замкнутые выражения для оценок составляющих дисперсии были получены на основе как примерные ML и REML. Помимо простого достижимого выражения оценок, исключающие необходимость использования численных методов оптимизации, мы также представили некоторые вычислительные аспекты, которые можно использовать для быстрого внедрения Формирование предложенного алгоритма. Для данных, размещенных в многомерные сетки, методы GLAM также могут быть модифицированы, что еще больше увеличивает время вычислений. В Кроме того, предложенная процедура может быть легко интегрирована в оценка GAMM с наборами независимых случайных эффекты. Для ясности мы сосредоточились здесь на GAMM заданные в терминах одномерных эффектов совместно с двумерным интер поверхность действия. Тем не менее, алгоритм SAP также может быть легко расширяется, чтобы иметь дело с поверхностными взаимодействиями. Следует отметить, что, хотя представленная методика в этой статье может иметь дело с любыми d -мерными обобщенными P-сплайн, для размеров больше $d = 5$ расчетный время, необходимое для алгоритма, может привести к его применению невыполнимо. В этом контексте P-сплайн ANOVA-типа inter- модели действия, предложенные Ли и Дурбаном (2011 г.) представлять альтернатива полностью многомерным моделям. На основании результатов, представленных в разд. 4, расширение SAP алгоритм работы с такими моделями прост, поэтому позволяя оценивать модели P-сплайнов, включая унифицированные варьировать основные эффекты, а также двухсторонние и / или трехсторонние анизотропные взаимодействия.

Результаты моделирования показали хорошую производительность. В основе предлагаемого метода - как среднеквадратичная ошибка, и время вычислений по сравнению с установленным подходы. Однако следует отметить, что нежелательный Свойство нашего метода состоит в том, что на него влияет сигнал-шум. По мере уменьшения отношения сигнал / шум различия между новым предложением и методом про- поставил Вуд (2011 г.) стать меньше. Хотя в Имитационное исследование наш метод превзошел Вуд (2011) метод во всех случаях, это область, которая требует дальнейшего изучение.

Как в моделировании, так и в данных об осадках начальные оценки фиксированных и случайных эффектов модели были установлены равными нулю, а компоненты дисперсии - единице. Мы знаем, что более подходящие первоначальные оценки могут даже улучшить поведение предложенного алгоритма, давая лучшее время вычислений, а также предотвращение сбоев конвергенции присутствуют в процедуре оценки. Что касается фиксированных и запущенных Это связано с тем, что наш опыт показывает, что специфика если исходная оценка $\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha}$ на основе вектор ответа $\boldsymbol{y}$ и функция связи $g(\cdot)$, как это сделано в пример респираторных данных, обычно обеспечивает хорошее стартовое значение. UES для $(\boldsymbol{\beta}^t, \boldsymbol{\alpha}^t)^t$. Фактически, мы применили алгоритм SAP к два примера, представленные во введении к Wood (2008 г.), указав в качестве начальных оценок те, которые установлены по умолчанию в `гам ()` функции ($\boldsymbol{\eta} = g((\boldsymbol{y} + 0.5)/2)$ для обоих яйцо скумбрии и смоделированные данные). В обоих случаях SAP

алгоритм сходился и полученные результаты были похожи на те, кто использует Wood (2011 г.) метод.

Хорошо известно, что методы [redacted] страдать от серьезных предвзятость (Бреслоу и Клейтон [redacted] и Бреслоу 1996), особенно для кластеризованных [redacted]анных, когда размер кластера маленький. Следовательно, можно ожидать, что метод про- изложенное в этой статье, также наследует это поведение. Расширение- Проведено комплексное имитационное исследование (результаты не показаны) оценить практическую производительность алгоритма SAP в разных сценариях, давая в целом хорошие результаты.

старшего редактора и двух рецензентов за их ценные комментарии и предложения, которые способствовали существенному улучшению этого бумага.

Приложение 1: Коэффициенты фиксированных и случайных эффектов оценка

Для заданных значений компонент дисперсии τ^2_d $(d = 1, 2)$ и φ - оценка коэффициентов фиксированных и случайных эффектов

Тем не мене[redacted] по
и Чжан (199[redacted]
встроены в алгоритм SAP. Изучение ком-
предположительно эффективные способы включения скорректированных смещения
процедуры в этой обстановке остается интересной областью
исследовать.

Когда дело дошло до представления расширения про-
введенная процедура в структуру GAMM, наборы независимых
предполагались постоянные случайные эффекты. Этот случайный эффект
структура подразумевает диагональную ковариационную матрицу дисперсии
случайные эффекты, что позволяет немедленно включить
соотношение алгоритма SAP в этом контексте. Хотя
такой структуры случайных эффектов может быть достаточно в широком
область реальных приложений, например, в многоуровневых исследованиях.
Таким образом, текущее направление исследований сосредоточено на изучении
возможность применения алгоритма SAP в продольном
окончательные исследования с возможно коррелированными случайными перехватами и
склоны.

Возможный недостаток модели P-сплайна тензорного произведения
состоит в том, что он предполагает гладкую поверхность, т.е. плавный переход.
Эффект по всей поверхности. На некоторой практике
Однако более сложные ситуации могут
возникают с эффектами, которые могут не измениться в некоторых регионах
поверхности, но быстро меняется в других регионах. В
В этих обстоятельствах предположение о единой гладкой
Параметр для каждой ковариаты может быть недостаточным
чтобы уловить [redacted]-сплайны (Lang
и Брезгер 2004[redacted]
доставлено. В адаптивных P-сплайнах глоба[redacted]метр сглаживания
заменены локально адаптивными параметрами сглаживания,
таким образом обеспечивая большую гибкость. Расширение алгоритма SAP
Ритм адаптивных анизотропных P-сплайнов представляет собой текущую линию
исследовать.

Наконец, R– код, используемый для моделирования, а также
R пакет реализации алгоритма SAP для 2D
и 3D-кейсы можно скачать с https://bitbucket.org/mxrodriguez/sap.

Благодарности Авторы выражают благодарность
за поддержку, полученную от Министерства экономики Испании
и гранты на повышение конкурентоспособности MTM2011-28285-C02-01 и MTM2011-
28285-C02-02. Исследование Дэ-Джин Ли финансировалось Национальным институтом здравоохранения.
грант для Суперфонда по смесям металлов, биомаркерам и нейроразведке
Проект улучшения 1PA2ES016454-01A2. Мы также благодарны ассо-

[redacted]правдоподобие со штрафом (см.
(6) в Бреслоу и Клейтоне, 1993 г[redacted]

$$-\frac{1}{2\varphi}\sum_{я=1}^{n} Dev_i(y_i,\mu_i) - \frac{1}{2}\boldsymbol{\alpha}_t \boldsymbol{G}^{-1}\boldsymbol{\alpha},$$

где Dev_i обозначает отклонение. Эта максимизация может
выполняться на основе алгоритма Fisher-Scoring,
включая рабочую зависимую переменную и матрицу весов,
который следует обновлять на каждой итерации. В частности, на
$(k+1)$ -я итерация Fisher-Scoring, рабочий вектор z равен
получено как

$$z_i = g(\mu_i^{(k)}) + (y_i - \mu_i^{(k)})\, g'(\mu_я^{(k)}),$$

и затем оцениваются фиксированные и случайные эффекты модели.
в виде

$$\hat{\boldsymbol{\beta}}^{(к+1)} \text{ знак равно } \left(\boldsymbol{X}_t \boldsymbol{V}^{-1} \boldsymbol{Икс}\right)^{-1} \boldsymbol{X}_t \boldsymbol{V}^{-1} \boldsymbol{z}, \tag{10}$$

$$\hat{\boldsymbol{A}}_{(K+1)} = \boldsymbol{ГЗ}_m \boldsymbol{V}^{-1}\left(\boldsymbol{z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(к+1)}\right)$$

$$= \boldsymbol{GZ}_t \boldsymbol{Pz}, \tag{11}$$

где

$$\boldsymbol{V} = \boldsymbol{W}^{-1} + \boldsymbol{ZGZ}_t,$$

[redacted] $\boldsymbol{X}_t \boldsymbol{V}^{-1}$,

а $\boldsymbol{W}$ - диагональная матрица весов с элементами $w_{ii} = \{\varphi[g'(\mu_i^{(k)})]^2 v(\mu_я^{(k)})\}^{-1}$.

С вычислительной точки зрения более удобный
Способ совместного получения $\boldsymbol{\beta}$; и $\boldsymbol{\alpha}$; являет[redacted] из
линейная система (см. уравнение (9) в Breslow and Clayton[redacted]

$$\begin{bmatrix} \boldsymbol{X}_t \boldsymbol{WX} & \boldsymbol{X}_t \boldsymbol{WZG} \\ \boldsymbol{Z}_t \boldsymbol{WX} & \underbrace{\boldsymbol{I} + \boldsymbol{Z}_t \boldsymbol{WZG}}_{\boldsymbol{C}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}^{(к+1)} \\ \hat{\boldsymbol{б}}^{(к+1)} \end{bmatrix} \text{ знак равно } \begin{bmatrix} \boldsymbol{X}_t \boldsymbol{Wz} \\ \boldsymbol{Z}_t \boldsymbol{Wz} \end{bmatrix}, \tag{12}$$

где $\hat{\boldsymbol{б}}^{(к+1)} = \boldsymbol{G}^{-1}\hat{\boldsymbol{\alpha}}_{(K+1)}$. Обратите внимание, что (12) соответствует
нормальные уравнения наилучшей линейной несмещенной оценки
из $\boldsymbol{p}$ и лучшего линейного несмещенного предсказания $\boldsymbol{\alpha}$; в соответствии с

Стр.16

рабочая линейная смешанная модель

$\boldsymbol{z} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{\epsilon}$, где $\boldsymbol{\alpha} \sim N(\boldsymbol{0}, \boldsymbol{G})$
и $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{W}^{-1})$.

Приложение 2: Доказательство теоремы

Доказательство. Пренебрегая зависимостью $\boldsymbol{W}$ от τ_d ($d = 1, 2$),
приблизительная ограниченная логарифмическая правдоподобие *рабочей* линейной
смешанная модель дается формулой (Бреслоу и Клейтон 1993)

$$л^* = -\frac{1}{2}\text{журнал}\,|\boldsymbol{V}| - \frac{1}{2}\text{журнал}\,|\boldsymbol{X}_t \boldsymbol{V}^{-1}\boldsymbol{X}| - \frac{1}{2}(\boldsymbol{Г} - \boldsymbol{X}\hat{\boldsymbol{\beta}})_m \boldsymbol{V}^{-1}(\boldsymbol{г} - \boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

REML-оценки компонентов дисперсии тогда
получается обычным способом, максимизируя это количество.
Производные по компонентам дисперсии
τ_d^2 ($d = 1, 2$), получаем (см. дополнительные материалы онлайн
для подробностей)

$$\frac{\partial l^*}{\partial \tau_d^2} = -\frac{1}{2}\text{след}\left(\boldsymbol{Z}_t \boldsymbol{PZ}\frac{\partial \boldsymbol{G}}{\partial \tau_d^2}\right) + \frac{1}{2}\hat{\boldsymbol{\alpha}}_m \text{грамм}^{-1}\frac{\partial \boldsymbol{G}}{\partial \tau_d^2}\text{грамм}^{-1}\hat{\boldsymbol{\alpha}} \tag{13}$$

Применяя свойства матричного дифференцирования, имеем

$$\frac{\partial \boldsymbol{G}}{\partial \tau_2^2} = -\boldsymbol{G}\frac{\partial \boldsymbol{G}^{-1}}{\partial \tau_d^2}\boldsymbol{G} = \frac{1}{\tau_d^4}\boldsymbol{G}\boldsymbol{G}_d\boldsymbol{G}, \tag{14}$$

где

участвуют в знаменателях компонентов дисперсии
оценки соответствует эффективному измерению штрафных санкций.
размерная часть (или случайная часть) подобранной модели

$$\text{след}\left(\boldsymbol{Z}_t \boldsymbol{PZG}\frac{1}{\tau_1^2}\text{грамм}_1\right) + \text{след}\left(\boldsymbol{Z}_t \boldsymbol{PZG}\frac{1}{\tau_2^2}\text{грамм}_2\right)$$

$$= \text{след}\left(\boldsymbol{Z}_t \boldsymbol{PZG}\right)$$

$$= \text{след}\left(\boldsymbol{ЗГЗ}_m \boldsymbol{П}\right)$$

$$= \text{трассировка}\,(\boldsymbol{H}_{Случайно}),$$

где $\boldsymbol{H}_{Random}$ обозначает шляпную матрицу (Хасти и Тибширани
1990) случайной части [см. (10)].

Наконец, оценка φ получается, как и раньше, если взять
производные приближенного ограниченного логарифмического правдоподобия с
относительно φ

$$\frac{\partial l^*}{\partial \varphi} = -\frac{1}{2}\text{след}\left(\boldsymbol{п}\frac{\partial \boldsymbol{V}}{\partial \varphi}\right) + \frac{1}{2}(\boldsymbol{Г} - \boldsymbol{X}\hat{\boldsymbol{\beta}})_m \boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \varphi}\boldsymbol{V}^{-1}(\boldsymbol{г} - \boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

Во-первых, по формуле. (5.2) в Харвилле (1977 г.) имеем $\boldsymbol{V}^{-1}(\boldsymbol{z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{W}(\boldsymbol{г} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{Z}\hat{\boldsymbol{\alpha}})$. Более того, учитывая, что $\boldsymbol{V}$ зависит от
на φ через $\boldsymbol{W}^{-1}$ которое можно переписать в виде $\boldsymbol{W} = \frac{1}{\varphi}\tilde{\boldsymbol{W}}$,
где $\tilde{\boldsymbol{W}}$ - диагональная матрица с элементами $\tilde{w}_{ii} = \{[g'(\mu_i)]^2 v(\mu_i)\}^{-1}$, и снова игнорируя зависимость
$\tilde{\boldsymbol{W}}$ на φ, тогда

$$2\frac{\partial l^*}{\partial \varphi} = -\frac{1}{\varphi}\text{след}\left(\boldsymbol{PW}^{-1}\right)$$

$$G = \text{диагональ}\left(\tau^{2}_{2}/\partial^{*}_{2},\ \tau^{2}_{1}/\partial^{*}_{1},\ 1/(\partial^{*}_{2}/\tau^{2}_{2} + \partial^{*}_{1}/\tau^{2}_{1})\right), \qquad (15)$$

$$_1 = \text{diag}\,(\mathbf{0}_{q_1(c_2-q_2)}, \boldsymbol{d}_1, \boldsymbol{d}^{*}_{1}),$$

$$_2 = \text{diag}\,(\boldsymbol{d}_2, \mathbf{0}_{q_2(c_1-q_1)}, \boldsymbol{d}^{*}_{2}),$$

где $\mathbf{0}_r$ - вектор нулей длины r , а $\boldsymbol{d}_1 = \boldsymbol{I}_{q_2} \otimes \breve{}_1$, $\boldsymbol{d}_2 = \breve{}_2 \otimes \boldsymbol{I}_{q_1}$, $\boldsymbol{d}^{*}_1 = \boldsymbol{I}_{c_2-q_2} \otimes \breve{}_1$, $\boldsymbol{d}^{*}_2 = \breve{}_2 \otimes \boldsymbol{I}_{c_1-q_1}$.

По выражению плагина (14) в (13) получаем, что первая-частные производные порядка приближенных ограниченных лог-вероятность стать

$$2\frac{\partial l^{*}}{\partial \tau^{2}_{d}} = -\frac{1}{\tau^{2}_{d}}\ \text{след}\left(\boldsymbol{Z}^{t}\boldsymbol{PZG}\frac{d}{\tau^{2}_{d}}\ \text{грамм}\right) + \frac{1}{\tau^{4}_{d}}\ \boldsymbol{\alpha}^{m}\ {}_{d}\ \boldsymbol{\alpha}\,. \qquad (16)$$

Тогда REML-оценки компонент дисперсии τ^{2}_{d} (d = 1 , 2) находятся приравниванием выражения (16) к нулю, что дает

$$\tau^{2}_{d} = \frac{\boldsymbol{\alpha}^{m}\ {}_{d}\ \boldsymbol{\alpha}}{\text{след}\left(\boldsymbol{Z}^{t}\boldsymbol{PZG}\frac{d}{\tau^{2}_{d}}\ \text{грамм}\right)}.$$

Прежде чем приступить к оценке φ - если неизвестно - важно отметить, что сумма количеств

$$+\frac{1}{\varphi^{2}}(\boldsymbol{г} - \boldsymbol{X\beta} - \boldsymbol{Z\alpha})^{m}\sim\boldsymbol{Ш}(\boldsymbol{г} - \boldsymbol{X\beta} - \boldsymbol{Z\alpha}).$$

Приравнивая приведенное выше выражение к нулю, получаем

$$\varphi = \frac{(\boldsymbol{Г} - \boldsymbol{X\beta} - \boldsymbol{Z\alpha})^{m}\sim\boldsymbol{Ш}(\boldsymbol{г} - \boldsymbol{X\beta} - \boldsymbol{Z\alpha})}{\text{след}\left(\boldsymbol{PW}^{-1}\right)},$$

где [см. (5.3) в Харвилле 1977 г. и выр [redacted] (10) и (12)]

$$\begin{aligned}
\text{след}\left(\boldsymbol{PW}^{-1}\right) &= \text{след}\left(\boldsymbol{W}^{-1}\boldsymbol{n}\right)\\
&= \text{след}\left(\boldsymbol{I}_n - [\,\boldsymbol{X}\,|\,\boldsymbol{ZG}\,]\,\boldsymbol{C}^{-1}\begin{bmatrix}\boldsymbol{X}^{m}\boldsymbol{Bm}\\ \boldsymbol{Z}^{t}\boldsymbol{W}\end{bmatrix}\right)\\
&= \text{след}\left(\boldsymbol{I}_n - [\,\boldsymbol{X}\,|\,\boldsymbol{ZG}\,]\begin{bmatrix}(\boldsymbol{X}^{t}\boldsymbol{V}^{-1}\,\boldsymbol{Икс})^{-1}\,\boldsymbol{X}^{t}\boldsymbol{V}^{-1}\\ \boldsymbol{Z}^{t}\boldsymbol{P}\end{bmatrix}\right)\\
&= n - \text{след}\left((\boldsymbol{Икс}\,\boldsymbol{X}^{t}\boldsymbol{V}^{-1}\,\boldsymbol{Икс})^{-1}\,\boldsymbol{X}^{t}\boldsymbol{V}^{-1}\right)\\
&\quad - \text{след}\left(\boldsymbol{ЗГЗ}^{m}\boldsymbol{П}\right)\\
&= n - \text{ранг}\,(\boldsymbol{X}) - \sum_{d=1}^{2} ed_{d}\,.
\end{aligned}$$

Стр.17

Обратите внимание, что $\boldsymbol{H} = [\,\boldsymbol{X}\,|\,\boldsymbol{ZG}\,]\,\boldsymbol{C}^{-1}[\,\boldsymbol{X}\,|\,\boldsymbol{Z}\,]^{t}\boldsymbol{W}$ соответствует матрица подобранной модели, след которой, как показано, может быть разлагается как сумма следов шляпных матриц нештатная (или фиксированная) часть и штрафная (или случайная) часть. ⊓⊔

Рекомендации

Бреслоу, Н. Э., Клейтон, Д. Г.: Аппроксимационный вывод в обобщенном линейные смешанные модели. Варенье. Стат. Доц. **88** , 9–25 (1993)

Карри, И., Дурбан, М., Эйлерс, PHC: Обобщенные модели линейных массивов с приложениями к многомерному сглаживанию. JR Stat. Soc. Сер. B **68** , 259–280 (2006)

Карри, И., Дурбан, М .: Гибкое сглаживание с помощью P-шлицев: унифицированное подход. Стат. Модель. **4. C.** 333–349 (2002).

де Бур, Калифорния: Практическое руководство по сплайнам. Исправленное издание. Спрингер, Нью-Йорк (2001)

Eilers, PHC, Marx, BD: гибкое сглаживание с помощью B-шлицев и штрафы. Стат. Sci. **11. C.** 89–121 (1996).

Eilers, PHC, Marx, BD: многомерная калибровка по температуре Взаимодействие происходит с помощью двумерной регрессии сигнала со штрафами. Chemom. Intell. Лаборатория. Syst. **66** , 159–174 (2003)

Эйлерс, PHC, Карри, И., Дурбан, М .: Быстрое и компактное сглаживание на большие многомерные сетки. Comput. Стат. Data Anal. **50** , 61–76 (2006)

Фармейр, Л., Кнейб, Т., Ланг, С.: Штрафные санкционированные структурные аддитивные регрессии Анализ пространственно-временных данных: байесовская перспектива. Стат. Грех. **14** , 715–745 (2004)

Гилмор, AP, Томпсон, Р., Каллис, Б.Р .: Средняя информация REML: эффективный алгоритм оценки параметра дисперсии в линейные смешанные модели. **51** , 1440–1450 (1995).

Хасти, Т.Дж., Тибширани, Р.Дж .: Обобщенные аддитивные модели. Чепмен и Холл, Лондон (1990)

Хасти, Т.Дж., Тибширани, Р.Дж .: Модели с переменным коэффициентом. JR Stat. Soc. Сер. B **55** , 757–796 (1993)

Харвилл, Д.А.: Подходы максимального правдоподобия к компоненту дисперсии оценка и связанные с ней проблемы. Варенье. Стат. Доц. **72** , 320–338 (1977)

Джонс, К., Ничка, Д., Киттель, Т., Дейли, К.: Заполнение скудных записей о пространственные поля. Варенье. Стат. Доц. **98** , 796–806 (2003).

Кривобокова Т., Крайничану С.М., Кауэрманн Г .: Быстрая адаптивная штрафные шлицы. J. Comput. График. Стат. **17** , 1–20 (2008)

Ланг, С., Брезгер, А .: Байесовские P-сплайны. J. Comput. Grap. Стат. **13** , 183–212 (2004)

Ли, Д.-Дж .: Смешанная модель сглаживания для пространственного и пространственно-временного данные. Кандидатская диссертация, Департамент статистики, Университет Карлоса III де Мадрид, Испания (2010)

Ли, Д.-Дж., Дурбан, М .: P-сплайн модели взаимодействия типа ANOVA для пространственно-временное сглаживание. Стат. Модель. **11. C.** 49–69 (2011).

Ли, Д.-Дж., Дурбан, М., Эйлерс, PHC: Эффективное двумерное сглаживание смешанными моделями P-сплайна ANOVA и вложенными базами. Comput. Стат. Data Anal. **61** , 22–37 (2013)

Лин, Х., Бреслоу, NE: Коррекция смещения в обобщенных линейных смешанных модели с несколькими компонентами дисперсии. Варенье. Стат. Доц. **91** , 1007–1016 (1996).

Линь, Х., Чжан, Д .: Вывод в обобщенных аддитивных смешанных моделях с помощью сглаживающих сплайнов. JR Stat. Soc. Сер. B **61** , 381–400 (1999)

Павитан, Ю.: По всей вероятности: статистическое моделирование и вывод Использование правдоподобия. Oxford University Press, США (2001)

R Core Team. R: язык и среда для статистических вычислений. ing. R Фонд статистических вычислений, Вена, Австрия. URL http://www.R-project.org/ (2013)

Рупперт Д., Ванд М. П., Кэрролл Р. Дж.: Полупараметрическая регрессия. Издательство Кембриджского университета, Кембридж (2003)

Шалл, Р .: Оценка в обобщенных линейных моделях со случайными эффектами. Биометрика **78** , 719–721 (1991).

Стирателли, Р., Лэрд, Н.М., Уэр, Дж. Х .: Модели случайных эффектов с серийные наблюдения с бинарными ответами. Биометрия **40** , 719–727 (1984)

Wand, MP: Сглаживающие и смешанные модели. Comput. Стат. **18** , 223–249 (2003)

Дерево, SN: тонкие пластинчатые регрессионные шлицы. JR Stat. Soc. Сер. B **65** , 95–114 (2003)

Wood, SN: стабильная и эффективная оценка параметров множественного сглаживания. ция для обобщенных аддитивных моделей. Варенье. Стат. Доц. **99** , 673–686 (2004)

Вуд, С. Н.: Обобщенные аддитивные модели. Знакомство с R. Чепмен и Холл / CRC, Бока-Ратон (2006a)

Вуд, С. Н.: Масштабно-инвариантное тензорное произведение низкого ранга сглаживает обобщенные аддитивные модели. JR Stat. Soc. Сер. B **70** , 495–518 (2006b)

Дерево, SN: быстрая стабильная прямая подгонка и выбор гладкости для генераторов. обобщенные аддитивные модели. Биометрия **62** , 1025–1036 (2008)

Wood, SN: Быстрая стабильная ограниченная максимальная вероятность и маргинальность оценка правдоподобия полупараметрических обобщенных линейных моделей. JR Stat. Soc. Сер. B **73** , 3–36 (2011)

Вуд, С. Н., Шайпл, Ф., Фарауэй, Дж. Дж .: Простой промежуточный уровень ранговое сглаживание тензорного произведения в смешанных моделях. Стат. Comput. **23** , 341–360 (2013)