

Приближенное оценивание с помощью ускоренного метода наибольшей энтропии.

Часть 2. Исследование свойств оценок*

Ю. А. Дубнов, А. В. Булычев

Федеральное государственное учреждение "Федеральный исследовательский центр "Информатика и управление" Российской академии наук", Москва, Россия

Аннотация. В работе исследуется метод приближенного энтропийного оценивания, предназначенный для ускорения классического метода наибольшей энтропии благодаря применению регуляризации в задаче оптимизации. Данный метод сравнивается с методом наибольшего правдоподобия и байесовским оцениванием как экспериментально, так и с точки зрения теоретических выкладок для некоторых частных случаев. Тестирование методов оценивания проводится на примере задачи линейной регрессии с различными типами ошибок, в том числе с несимметричными распределениями, а также с мультимодальным распределением в виде смеси гауссовских компонент.

Ключевые слова: вероятностная математическая модель, метод наибольшей энтропии, линейная регрессия, регуляризация, распределение ошибок.

DOI 10.14357/20718632230107

Введение

Работа посвящена разработке и исследованию метода энтропийного оценивания с мягкой рандомизацией для восстановления параметров вероятностных математических моделей по имеющимся наблюдениям. Под мягкой рандомизацией понимается техника добавления регуляризации в функционал информационной энтропии с целью упрощения оптимизационной задачи и ускорения обучения. В 1-й части статьи (опубликована в №4 2022 журнала) была представлена концепция метода «мягкого» энтропийного оценивания, включая получение энтропийно-оптимальных функций ПРВ в общем виде. В данной работе представлены результаты тестирования предложенного метода

и экспериментального исследования свойств получаемых оценок в регрессионных задачах.

При решении регрессионных задач принято разделять классический, статистический подход, в котором параметры моделей считаются детерминированными величинами и оцениваются их значения, например, методом наименьших квадратов (МНК) или методом наибольшего правдоподобия (МНП) [1], и второй — вероятностный подход, согласно которому параметры моделей полагаются случайными величинами, а результатом оценивания являются их вероятностные характеристики [2]. Такому подходу соответствуют, например, байесовское оценивание и энтропийно-робастное оценивание — метод наибольшей энтропии

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 20-07-00223).

(МНЭ) [3]. Одно из отличий статистического и вероятностного подхода заключается в том, что последний позволяет использовать вероятностные характеристики и свойства случайных величин с целью уточнения оценок и последующих прогнозов с их использованием.

На практике переход к вероятностному оцениванию наиболее распространен в задачах, где нарушаются предпосылки классического метода МНК и теоремы Гаусса-Маркова [4]. Примерами являются задачи с малым количеством входных данных, когда проверка статистических гипотез становится затруднительна, и задачи с нестандартными видами ошибок или, в общем случае, неструктурированным шумом. Оценки, получаемые вероятностными методами, для таких задач могут оказаться смещенными, однако более эффективными по сравнению с классическими методами, как, например, МНК, в терминах среднеквадратической ошибки оценок [5].

Далее рассмотрим несколько видов регрессионной задачи, начиная с классического случая с нормально-распределенными ошибками. Для каждой задачи будут вычисляться смещение и дисперсия для оценок, полученных различными методами, а именно: методом наибольшего правдоподобия, байесовского оценивания и «мягкого» энтропийного оценивания.

1. Методология сравнительного исследования

В данном разделе приводятся схема эксперимента для исследования и критерии для сравнения результатов различных методов.

Пусть имеется обучающая коллекция данных, состоящая из входных данных (матрица X размером $[n \times m]$) и выходных данных (вектор $y \in R^n$). Можно считать, что n – количество объектов для обучения, а m – количество признаков для этих объектов. В регрессионном анализе вектор y принято называть целевой или зависимой переменной, а матрицу X – предикторами модели. Тогда каждый элемент из y трактуется как одно наблюдение целевой переменной при значениях предикторов $x^{(i)} = \{X_{ij}, j = 1 \dots m\}$.

Предполагается, что между данными X и y существует причинно-следственная связь, ко-

торую необходимо смоделировать. Для этого декларируется модель с параметрами $a \in R^r$, например, следующего вида:

$$y = F(X, a) + \varepsilon, \quad a \in R^r$$

где F – вектор-функция, а ε – остатки модели или шумы измерений.

В наиболее распространенной на практике задаче линейной регрессии функция $F(X, a)$ является линейной комбинацией или скалярным произведением:

$$y_i = \langle X^{(i)}, a \rangle + \varepsilon, \\ y_i = \sum_{j=1}^m X_{ij} a_j, \quad i = 1 \dots n$$

Результатом оценивания будем называть вектор параметров $\hat{a}$ при заданных входных данных X и y .

При сравнении оценок важно понимать, что результат применения любых методов оценивания, в том числе и МНК, является случайным в силу случайности ошибок модели. То есть, оценки являются случайными величинами, обладающими такими характеристиками, как смещение относительно истинных значений и стандартное отклонение. Для метода МНК, например, известно, что в случае нормально-распределенных ошибок и выполнении прочих предпосылок теоремы Гаусса-Маркова, его оценки являются несмещенными и наиболее эффективными в классе линейных несмещенных оценок (Best Linear Unbiased Estimation, BLUE) [6]. Чем ниже дисперсия оценки, тем она считается более эффективной, а при стремлении дисперсии несмещенной (или асимптотически несмещенной) оценки к 0 с ростом объема выборки, оценку называют состоятельной.

2. Схема экспериментов

Свойства несмещенности и состоятельности для таких методов оценивания, как МНК, МНП и даже байесовского вывода при определенных условиях, доказаны аналитически в теории математической статистики [7], однако, не для метода «мягкого» энтропийного оценивания. Поэтому для вычисления свойств оценок метода энтропийного оценивания будем применять схему испытаний Монте-Карло:

1) Фиксируются истинные параметры модели a_0 и генерируется обучающая выборка объемом n точек X и

$$y^{(0)} = F(X, a_0)$$

2) Согласно выбранному распределению ошибок генерируется выборка ε из n точек в качестве аддитивного шума для y :

$$y = y^{(0)} + \varepsilon, \quad y_i = y_i^{(0)} + \varepsilon_i, \quad i = 1 \dots n$$

3) Пункт 2 повторяется для каждого из испытаний Монте-Карло, общим количеством, например, $M = 1000$ раз. В результате имеется выборка из M оценок для параметров модели, по которым можно строить гистограммы значений и вычислять выборочные статистики.

Критериям сравнения методов оценивания будут значения суммарной средней абсолютной ошибки (SMAE) и суммарного среднеквадратического отклонения (SMSE):

$$MAE = \frac{1}{M} \sum_{i=1}^M |\hat{a} - a_0| = |E[\hat{a}] - a_0|,$$

$$MSE = \frac{1}{M} \sum_{i=1}^M (\hat{a} - a_0)^2$$

Префикс «S» означает суммирование по всем параметрам модели. Таким образом, стремление величины MAE к 0 говорит о несмещенности оценки, а величина MSE может служить для сравнения эффективности оценок.

Для построения модели будем использовать не результаты единичного выполнения какого-либо метода оценивания, а результат усреднения по всем испытаниям Монте-Карло. Такая практика зачастую применяется при анализе реальных данных, в частности, на этой идее построена технология бутереп и концепция рандомизированного прогнозирования [8, 9]. В результате мы получим по одной регрессионной модели для каждого из методов оценивания, чтобы сравнить их уже не в терминах свойств оценок, а в терминах качества самой модели.

Универсальным и одним из наиболее распространенных критериев сравнения регрессионных моделей является коэффициент детерминации R^2 :

$$R^2 = 1 - \frac{s^2}{s_y^2},$$

где

$$s^2 = \frac{1}{n-k} \sum_{i=1}^n (\hat{y}_i - y_i)^2,$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Далее приведем кратко описания каждого из сравниваемых методов оценивания, после чего будут представлены и проанализированы результаты численных экспериментов.

3. Методы оценивания

В данном разделе приводятся некоторые теоретические сведения для метода наибольшего правдоподобия, байесовского и энтропийного оценивания.

3.1. Метод наибольшего правдоподобия (МНП)

Правдоподобием выборки называют условную вероятность получения именно такой выборки при заданных значениях параметров. Тогда максимизация правдоподобия относительно значений параметров гарантирует наибольшую вероятность получения именно такой выборки для данной модели. На практике наиболее распространенным примером является случай нормально-распределенных ошибок со средним в нуле. Правдоподобие такой модели записывается в следующем виде:

$$L(X, y|a, \sigma^2) = \prod_{i=1}^n N(y_i - F(X^{(i)}, a), \sigma^2),$$

где $N(\mu, \sigma^2)$ – обозначение для плотности распределения вероятности одномерного нормального распределения, а $X^{(i)}$ – вектор данных для i -ого наблюдения.

Максимизация функции правдоподобия эквивалентна максимизации ее логарифма, поэтому задачу оптимизации решения линейной регрессии методом МНП можно записать так:

$$\begin{aligned} l(X, y|a, \sigma^2) &= \log(L(X, y|a, \sigma^2)) = \\ &= \sum_{i=1}^n \log \left(\frac{1}{\sigma\sqrt{2\pi}} \exp \left(-\frac{(y_i - \langle X^{(i)}, a \rangle)^2}{2\sigma^2} \right) \right) = \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(y_i - \langle X^{(i)}, a \rangle)^2}{2\sigma^2} \Rightarrow \max_{a, \sigma^2} \end{aligned}$$

Известно, что аналитическое решение данной задачи совпадает с формулой оценок по методу МНК, поэтому в классическом случае линейной регрессии с нормально-распределенными ошибками эти методы оценивания эквиваленты. Однако для отличающихся от классического случая условий задачи метод правдоподобия, в отличие от МНК, будет учитывать новую структуру ошибок. Будем использовать МНП в качестве базового метода для сравнения результатов экспериментов.

3.2. Байесовское оценивание

Байесовский подход отличается концепцией восприятия параметров модели. Как упоминалось ранее, параметры модели в вероятностном подходе полагаются случайными, то есть они имеют некоторую функцию распределения вероятности и ПРВ. Знания об этой ПРВ, имеющиеся до эксперимента, называют априорными, а в результате применения байесовского вывода восстанавливается апостериорная функция ПРВ. Согласно теореме Байеса в общем случае для восстановления апостериорной плотности параметров $p(\theta|X)$ используется функция правдоподобия выборки и некоторая априорная плотность $\pi(\theta)$:

$$p(\theta|X) \propto \mathcal{L}(X|\theta) * \pi(\theta),$$

что для задачи линейной регрессии представляется в виде:

$$p(a, \sigma^2|X, y) \propto L(X, y|a, \sigma^2) * \pi(a, \sigma^2).$$

Полученная апостериорная вероятность используется далее для получения статистических характеристик для параметров модели: средних значения или максимума апостериорной вероятности (МАР):

$$\hat{\theta} = \arg \max_{\theta} p(\theta|X).$$

Аналогично методу МНП для применения байесовского оценивания необходимо наличие функции правдоподобия, иными словами – представление о виде ошибок модели. Однако наиболее дискуссионным вопросом в данном случае остается выбор априорных распределений [10].

Принято различать априорные распределения информативного и неинформативного типа. Неинформативные априорные распределения, как

правило, являются, равномерными или неопределенными в общем виде, или же являются решением некоторой задачи оптимизации, в том числе и задачи максимизации энтропии [11]. С другой стороны, информативные распределения имеют некоторую структуру, выбираемую исходя из имеющихся знаний об объекте исследований. Байесовское оценивание, в таком случае, «уточняет» на основе наблюдений знания, заложенные в априорных распределениях.

Рассмотрим два примера априорных распределений:

1) Неинформативное совместное распределение для параметров модели и дисперсии ошибок:

$$\pi(a, \sigma^2) = \frac{1}{\sigma},$$

в таком случае оценки МАР будут эквиваленты оценкам по методу правдоподобия, поскольку задача максимизации апостериорной функции распределения сводится к задаче максимизации правдоподобия.

2) Сопряженное распределение для параметров модели. Сопряженным называется такое априорное распределение, тип которого не меняется в результате применения байесовского вывода [12]. Например, для линейной модели с нормально-распределенными ошибками таким распределением является условное нормальное распределение для параметров модели и обратное гамма-распределение для дисперсии ошибок:

$$\begin{aligned} \pi(a|\sigma^2) &= N(\mu, \sigma^2 V), \\ \pi(\sigma^2) &= IG(A, B), \end{aligned}$$

где вектор математического ожидания μ и матрица V , равно как и значения A, B , становятся гиперпараметрами алгоритма.

В данной работе в качестве μ и V использовались оценки по методу МНК и диагональная матрица их дисперсий, соответственно $A = 3$ и $B = 1$, дополнительная оптимизация гиперпараметров не производилась. Пример построения апостериорной ПРВ приведен на Рис. 1.

3.3. Метод «мягкого» энтропийного оценивания

Классический метод энтропийно-робастного оценивания (метод наибольшей энтропии, МНЭ) базируется на концепции информационной

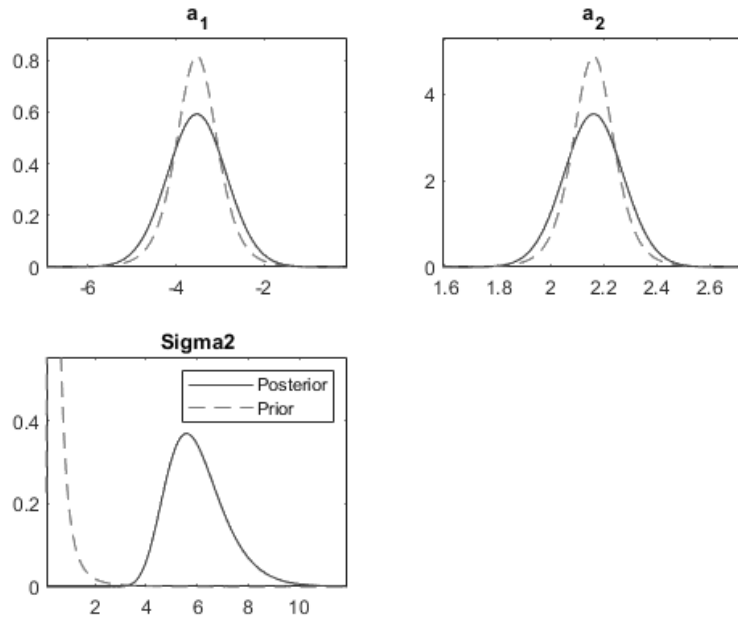


Рис. 1. Байесовское оценивание, восстановление апостериорной ПРВ

энтропии как меры неопределенности, максимизация которой относительно функций ПРВ для параметров модели приводит к наилучшим оценкам в условиях наибольшей неопределенности [13]. Кроме того, функциональная задача оптимизации дополняется интегральными ограничениями на баланс выхода модели в терминах средних значений, что приводит к оптимизационной задаче Ляпуновского типа [14]. Численное решение такой задачи оказывается особенно затруднено в случае многомерных моделей и большого объема входных данных.

Развиваемый в данной работе метод «мягкого» энтропийного оценивания [15] позволяет уйти от необходимости численного решения системы интегральных уравнений благодаря введению регуляризации в энтропийный функционал:

$$J[P(a), Q(\varepsilon)] =$$

$$= H[P(a), Q(\varepsilon)] - \bar{N}_a[P(a)] - \bar{N}_\varepsilon[Q(\varepsilon)] \Rightarrow \max_{P, Q}$$

$$\int_A P(a) da = 1, \quad \int_E Q(\varepsilon) d\varepsilon = 1.$$

Здесь $H[P(a), Q(\varepsilon)]$ – информационная энтропия системы со случайными параметрами $a \in A$ и шумами измерений $\varepsilon \in E$, а $\bar{N}_a[P(a)]$ и $\bar{N}_\varepsilon[Q(\varepsilon)]$ – средние значения для Гельдеровской нормы выхода модели и нормы шума:

$$N_a(a) = \|z(a) - y\|_H, \quad N_\varepsilon(\varepsilon) = \|\varepsilon\|_H.$$

То есть:

$$\bar{N}_a[P(a)] = \mathbb{E}[N_a(a)] = \int_A P(a) N_a(a) da,$$

$$\bar{N}_\varepsilon[Q(\varepsilon)] = \mathbb{E}[N_\varepsilon(\varepsilon)] = \int_E Q(\varepsilon) N_\varepsilon(\varepsilon) d\varepsilon.$$

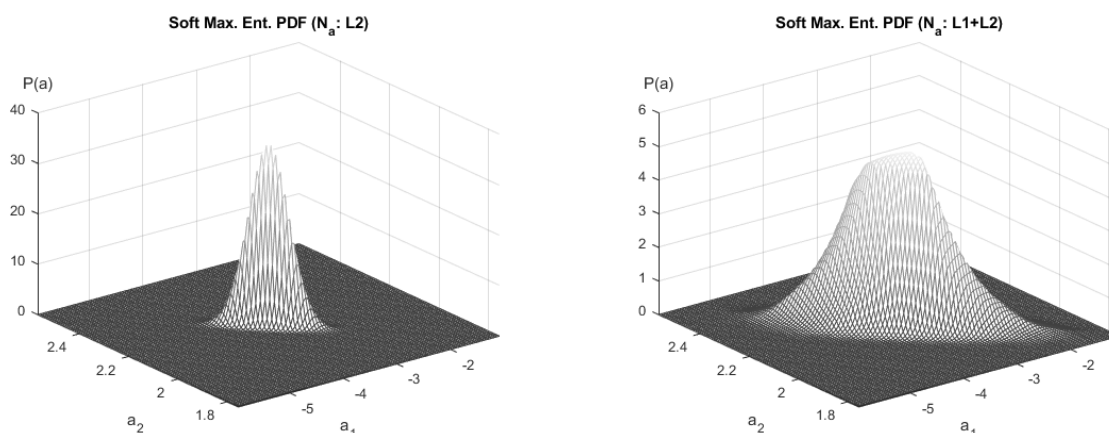
Такой приближенный подход к максимизации энтропийного функционала ускоряет получение оценок и добавляет вариативности в плане выбора степени нормы: L_1 , L_2 и их линейная комбинация.

В 1-й части данной статьи было получено решение оптимизационной задачи для метода «мягкого» энтропийного оценивания в общем виде:

$$P^*(a) = \frac{\exp(-N_a(a))}{\int_A \exp(-N_a(a)) da},$$

$$Q^*(\varepsilon) = \frac{\exp(-N_\varepsilon(\varepsilon))}{\int_E \exp(-N_\varepsilon(\varepsilon)) d\varepsilon}.$$

Функцию $P^*(a)$ можно трактовать как апостериорную функцию распределения параметров и, аналогично, байесовскому подходу использовать средние значения или оценки МАР. Пример энтропийно-оптимальных функций ПРВ приведен на Рис. 2.

Рис. 2. Пример энтропийно-оптимальных функций ПРВ, нормы L_2 (слева) и $L_1 + L_2$ (справа)

Особенностью использования нормы L_2 для линейной регрессии является то, что в таком случае оценки МАР по энтропийно-оптимальной функции $P^*(a)$ будут совпадать с оценками по методу правдоподобия. Этот факт доказывается следующим образом:

$$\begin{aligned}\hat{a}_{SME-MAP} &= \max_a P^*(a) = \max_a [\log P^*(a)] = \\ &= \max_a \left[\log \frac{\exp(-N_a(a))}{\int_A \exp(-N_a(a)) da} \right] = \\ &= \max_a [-N_a(a) - \log C] = \\ &= \max_a \left[-\sum_{i=1}^n (y_i - \langle X^{(i)}, a \rangle)^2 - \log C \right].\end{aligned}$$

Отличие от максимизации правдоподобия состоит лишь в отсутствии дополнительного параметра σ^2 для дисперсии ошибок, поскольку в энтропийном подходе шум оценивается отдельно. Таким образом, в частном случае оценки МАР в методе «мягкого» энтропийного оценивания эквиваленты оценкам по методу МНП при использовании нормально-распределенных ошибок, и, соответственно, также являются состоятельными.

Рассмотрим график дисперсии оценок энтропийного метода при использовании различных норм ошибок. Эксперимент проводится для классической задачи линейной регрессии в зависимости от объема входной выборки (Рис. 3.). В этом эксперименте используется парная регрессия с нормально-распределенными ошибками:

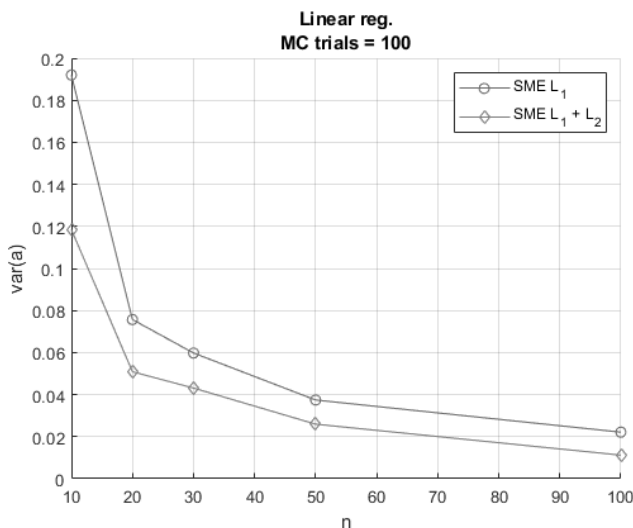


Рис. 3. График зависимости дисперсии энтропийных оценок от объема выборки

$$y = -3 + 2 * x + \varepsilon, \quad \varepsilon \propto N(0,3)$$

Как видно из графика, при увеличении объема выборки, дисперсии оценок монотонно уменьшаются, что может служить признаком состоятельности оценок, соответствующих максимуму энтропийно-оптимальных функций ПРВ.

С другой стороны, в отличие от функции правдоподобия $P^*(a)$ является полноценной функцией ПРВ с выполнением нормировки, поэтому данную функцию можно также использовать для получения среднего значения в качестве оценки параметров:

$$\hat{a}_{SME-MEAN} = \int_A a P^*(a) da$$

Оба вида оценок будут тестироваться по эффективности для сравнения с методом МНП и байесовского оценивания.

4. Экспериментальное сравнение эффективности оценок

Сравнение эффективности методов оценивания производится по описанной выше методике: для каждого нового испытания генерируется новая выборка аддитивных ошибок, после чего последовательно запускаются методы оценивания, результаты экспериментов усредняются для повышения точности.

Рассмотрим три регрессионные задачи: парная линейная регрессия в классическом случае (1); множественная линейная регрессия в случае несимметрично-распределенных ошибок (2), а также ошибок в виде смеси гауссовских компонент (3).

Все данные для экспериментов тщательно подготавливались в течение длительного периода времени, выборки не содержат пропусков и коррелирующих между собой признаков. Все полученные результаты строго воспроизводимы при указании начального положения генератора случайных чисел (rng(2021)) или в пределах стандартных отклонений в общем случае. Для вычислений использовался ПК с 8-ми ядерным процессором Intel(R) Core (TM) i7-9700F CPU @ 3.00 GHz. и 32 Gb оперативной памяти в операционной системе Windows 10 x64.

Моделирование проводилось в среде MATLAB версии R2019b, включающей такие инструменты, как fminsearch() для численной оптимизации и bayeslm() для построения байесовской модели. Последняя входит в инструментарий статистического анализа Econometrics Toolbox. Результаты тестирования приведены в Табл. 1-3.

1) Парная линейная регрессия в классическом случае:

$$y_i = a_1 + a_2 x_i + \varepsilon, \quad \varepsilon \propto N(0, \sigma^2)$$

Объем выборки выбран равным $n = 30$, а истинные значения параметров соответствуют эксперименту, приведенному выше:

$$a = \{-3, 2\}, \quad \varepsilon \propto N(0,3)$$

Как и ожидалось, в классической задаче все методы оценивания показывают высокие результаты, судя по столбцу среднеквадратических ошибок, их эффективность практически одинаковая, а различия лежат в пределах погрешности численного моделирования.

Табл. 1. Результаты тестирования для модели №1

Назв. метода	Ошибка $SMAE$	Ошибка $SMSE$	Коэффициент детерминации R^2
OLS / ML	0.3739	0.1426	0.8308
Bayes, Conjugate priors	0.3737	0.1425	0.8304
Bayes, Entropy priors	0.3780	0.1433	0.8310
Soft Max. Ent. (L1+L2) MAP est.	0.3792	0.1453	0.8310
Soft Max. Ent. (L1+L2) MEAN est.	0.3785	0.1450	0.8309

2) Множественная линейная регрессия в случае несимметрично-распределенных ошибок, генерируемых по распределению хи-квадрат:

$$y_i = \langle X^{(i)}, a \rangle + \varepsilon, \quad \varepsilon \propto \chi^2(k)$$

Размерность задачи $d = 5$, объем выборки $n = 50$ истинные значения параметров:

$$a = \{-2, 1.5, 1.8, -2.1, 2.6\}, \quad \varepsilon \propto \chi^2(2)$$

В этом эксперименте, несмотря на большее количество признаков, они хуже объясняют целевую переменную, то есть, ниже доля объясненной дисперсии в выборке, и, как следствие, коэффициент детерминации.

Стоит обратить внимание на небольшую разницу в значениях $Smae$ и $Smse$. В силу несимметричности ошибок в данном случае возможна ситуация, когда, несмотря на смещенность оценок, их итоговая среднеквадратическая ошибка ниже, то есть такая оценка считается более эффективной. Так в данном эксперименте, метод «мягкого» энтропийного оценивания оказывается немного эффективнее метода наибольшего правдоподобия, а байесовское оценивание с эн-

тропийно-оптимальными априорными вероятностями – эффективнее классического байесовского оценивания с сопряженными априорными распределениями.

3) Модель множественной регрессии из п.2 с ошибками в виде смеси гауссовских компонент (Gaussian mixture model):

$$y_t = a_0 + a_1 * y_{t-1} + a_2 * y_{t-2} + \varepsilon, \\ \varepsilon \propto w_1 N(\mu_1, \sigma_1^2) + w_2 N(\mu_2, \sigma_2^2)$$

Объем выборки $n = 50$, истинные параметры модели и распределения шума, следующие:

$$a = \{-2, 1.5, 1.8, -2.1, 2.6\}, \\ \varepsilon \propto \frac{1}{2} N(-1, 1^2) + \frac{1}{2} N(1.5, 2^2)$$

Последний эксперимент показывает, главным образом, что в случае мультимодального распределения для ошибок оценки среднего могут оказываться эффективнее оценок по методу MAP. В данном случае, «проигрывают» все методы, использующие для оценивания стандартную функцию правдоподобия нормального распределения.

Табл. 2. Результаты тестирования для модели № 2

Назв. метода	Ошибка $SMAE$	Ошибка $SMSE$	Коэффициент детерминации R^2
ML	1.7457	1.2242	0.5492
Bayes, Conjugate priors	1.7440	1.2207	0.5496
Bayes, Entropy priors	1.6913	1.1438	0.5489
Soft Max. Ent. (L1+L2) MAP est.	1.6800	1.1350	0.5490
Soft Max. Ent. (L1+L2) MEAN est.	1.6840	1.1369	0.5494

Табл. 3. Результаты тестирования для модели № 3

Назв. метода	Ошибка $SMAE$	Ошибка $SMSE$	Коэффициент детерминации R^2
ML	1.1404	0.4741	0.4987
Bayes, Conjugate priors	1.1393	0.4737	0.4985
Bayes, Entropy priors	1.1358	0.4754	0.4975
Soft Max. Ent. (L1+L2) MAP est.	1.1229	0.4635	0.4983
Soft Max. Ent. (L1+L2) MEAN est.	1.1153	0.4601	0.4981

Таким образом, если при анализе данных нет возможности установить вид функции правдоподобия, или же он был установлен ошибочно, применение таких методов, как МНП и байесовский вывод, становится затруднительным и менее эффективным.

5. Рекомендации по практическому применению

Предлагаемый метод приближенного энтропийного оценивания обладает некоторыми особенностями в плане реализации и практического применения.

Во-первых, рекомендуется использовать оценки по максимуму энтропийно-оптимальной вероятности ($\hat{a}_{SME-MAP}$) вместо оценок среднего ($\hat{a}_{SME-MEAN}$). В таком случае возможно избежать необходимости численного интегрирования для вычисления знаменателя в функции $P^*(a)$, поскольку, как было показано выше, максимизация логарифма данной функции содержит константу, не влияющую на результат оптимизации. Если же всё-таки необходимо вычислить оценки $\hat{a}_{SME-MEAN}$ для задач большой размерности ($d > 3$), то численное интегрирование рекомендуется реализовывать посредством семплирования подинтегральной функции плотности, аналогично интегрированию по методу Монте-Карло [16]. Это связано с отсутствием в большинстве сред программирования численного интегрирования многомерных величин.

Второй важный практический момент заключается в выборе нормы для оценивания. В данной работе в экспериментах 2 и 3 наибольшую эффективность показала линейная комбинация норм L_1 и L_2 с единичными коэффициентами. Однако логичным шагом будет подобрать значения этих коэффициентов как гиперпараметров всего алгоритма с целью повышения точности оценивания для данной задачи. Аналогичным образом подбираются коэффициенты для лассо-регрессии с целью отбора признаков и для гребневой регрессии (ridge) с целью повышения робастности оценок [17].

В дополнение к данному исследованию различные вариации энтропийно-робастного оценивания применялись для других задач анализа данных, таких как восстановление пропущенных данных и прогнозирования временных ря-

дов. Полученные результаты были представлены в опубликованных публикациях по проекту, приведенных далее.

Заключение

В данной работе исследуется метод приближенного энтропийного оценивания, предназначенный для ускорения классического метода наибольшей энтропии благодаря применению регуляризации в задаче оптимизации. Данный метод сравнивается с методом наибольшего правдоподобия и байесовским оцениванием, как экспериментально, так и с точки зрения теоретических выкладок для некоторых частных случаев. Тестирование методов оценивания проводится на примере задачи линейной регрессии с различными типами ошибок.

Проведенные эксперименты позволяют выделить класс задач, в которых применение предложенного метода оказывается более эффективным по сравнению с аналогами. К таким задачам относятся случаи, когда распределение ошибок модели является несимметричным и значительно отличается от стандартного нормального распределения. Кроме того, эффективность энтропийного оценивания становится особенно заметной в условиях малого объема входных данных.

Литература

1. Huang, David S. Regression and Econometric Methods. New York: John Wiley & Sons. 1970. pp. 127–147.
2. Hazewinkel, Michiel, ed. "Bayesian approach to statistical problems", Encyclopedia of Mathematics, Springer. 2001.
3. Amos Golan, George G. Judge, Douglas Miller. Maximum Entropy Econometrics: Robust Estimation with Limited Data. – John Wiley and Sons Ltd. Chichester, U.K., 1996.
4. Кристофер Доугерти. Введение в эконометрику. — 2-е, пер. с англ. — М.: ИНФРА-М, 2004. — 419 с.
5. Ximing Wu. A Weighted Generalized Maximum Entropy Estimator with a Data-driven Weight // Entropy, 2009. no.11.
6. Theil, Henri (1971). "Best Linear Unbiased Estimation and Prediction". Principles of Econometrics. New York: John Wiley & Sons. pp. 119–124. ISBN 0-471-85845-5.
7. Lehmann, E. L.; Casella, G. (1998). Theory of Point Estimation (2nd ed.). Springer.
8. Efron, B.; Tibshirani, R. (1993). An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall/CRC.
9. Popkov, Y.S.; Dubnov, Y.A.; Popkov, A.Y. New Method of Randomized Forecasting Using Entropy-Robust

- Estimation: Application to the World Population Prediction. // *Mathematics*, 2016, Vol. 4, Iss.1, p.1-16.
10. Andrew Gelman. Bayes, Jeffreys, Prior Distributions and the Philosophy of Statistics // *Statistical Science*, vol.24, No.2, pp.176-178, 2009.
 11. Zellner A. Past and Recent Results on Maximal DataInformation Priors // *Texhcnical Report*, Graduate School of Business, University of Chicago, 1996.
 12. Fink, Daniel (1997). "A Compendium of Conjugate Priors"
 13. R.D. Levin, M. Tribus. The maximum entropy formalism. MIT Press, 1979
 14. Yu. S. Popkov , Yu. A. Dubnov. Entropy-robust randomized forecasting under small sets of retrospective data // *Automation and Remote Control*. 2016, Volume 77, Issue 5, pp 839-854.
 15. Yu. S. Popkov. Soft Randomized Machine Learning // *Doklady Mathematics*, 2018, Vol. 98, No. 3, pp. 646–647.
 16. Fishman, George S. Monte Carlo : concepts, algorithms, and applications. — Springer, 1996.
 17. Rousseeuw, P. J., A. M. Leroy. Robust Regression and Outlier Detection. Wiley, 2003.

Дубнов Юрий Андреевич. Институт системного анализа Федерального государственного учреждения "Федеральный исследовательский центр "Информатика и управление" Российской академии наук" (ФИЦ ИУ РАН), г. Москва, научный сотрудник. Национальный исследовательский университет Высшая школа экономики (НИУ ВШЭ), г. Москва, старший преподаватель. Количество печатных работ: 29. Область научных интересов: байесовское оценивание, машинное обучение, принцип максимума энтропии. E-mail: yury.dubnov@phystech.edu

Булычев Александр Викторович. Институт системного анализа Федерального государственного учреждения "Федеральный исследовательский центр "Информатика и управление" Российской академии наук" (ФИЦ ИУ РАН), г. Москва, ведущий научный сотрудник. Национальный исследовательский университет Высшая школа экономики (НИУ ВШЭ), г. Москва, доцент. Кандидат технических наук. Количество печатных работ: 33. Область научных интересов: анализ данных, байесовские методы в статистике и эконометрике. E-mail: bulytchev.isa.ran@gmail.com

Approximate Estimation Using the Accelerated Maximum Entropy Method. Part 2. Study of the Properties of Estimates

Yu. A. Dubnov, A. V. Boulytchev

Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

Abstract. In this paper, we investigate a method of approximate entropy estimation, designed to speed up the classical method of maximum entropy estimation due to the use of regularization in the optimization problem. This method is compared with the method of maximum likelihood and Bayesian estimation, both experimentally and in terms of theoretical calculations for some special cases. Estimation methods are tested on the example of a linear regression problem with errors of various types, including asymmetric distributions as well as a multimodal distribution in the form of a mixture of Gaussian components.

Keywords: probabilistic mathematical model, maximum entropy method, linear regression, regularization, errors distribution.

DOI 10.14357/20718632230107

References

1. Huang, David S. Regression and Econometric Methods. New York: John Wiley & Sons. 1970. pp. 127–147.
2. Hazewinkel, Michiel, ed. "Bayesian approach to statistical problems", *Encyclopedia of Mathematics*, Springer. 2001.
3. Amos Golan, George G. Judge, Douglas Miller. Maximum Entropy Econometrics: Robust Estimation with Limited Data. — John Wiley and Sons Ltd. Chichester, U.K., 1996.
4. Kristofer Dougerti. Vvedenie v ekonometriku. — 2-e, per. s angl. — M.: INFRA-M, 2004. — 419 s.
5. Ximing Wu. A Weighted Generalized Maximum Entropy Estimator with a Data-driven Weight // *Entropy*, 2009. no.11.
6. Theil, Henri (1971). "Best Linear Unbiased Estimation and Prediction". *Principles of Econometrics*. New York: John Wiley & Sons. pp. 119–124. ISBN 0-471-85845-5.
7. Lehmann, E. L.; Casella, G. (1998). *Theory of Point Estimation* (2nd ed.). Springer.
8. Efron, B.; Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.
9. Popkov, Y.S.; Dubnov, Y.A.; Popkov, A.Y. New Method of Randomized Forecasting Using Entropy-Robust

- Estimation: Application to the World Population Prediction. // *Mathematics*, 2016, Vol. 4, Iss.1, p.1-16.
10. Andrew Gelman. Bayes, Jeffreys, Prior Distributions and the Philosophy of Statistics // *Statistical Science*, vol.24, No.2, pp.176-178, 2009.
 11. Zellner A. Past and Recent Results on Maximal DataInformation Priors // *Texhcnical Report*, Graduate School of Business, University of Chicago, 1996.
 12. Fink, Daniel (1997). "A Compendium of Conjugate Priors"
 13. R.D. Levin, M. Tribus. The maximum entropy formalism. MIT Press, 1979
 14. Yu. S. Popkov, Yu. A. Dubnov. Entropy-robust randomized forecasting under small sets of retrospective data // *Automation and Remote Control*. 2016, Volume 77, Issue 5, pp 839-854.
 15. Yu. S. Popkov. Soft Randomized Machine Learning // *Doklady Mathematics*, 2018, Vol. 98, No. 3, pp. 646–647.
 16. Fishman, George S. Monte Carlo : concepts, algorithms, and applications. — Springer, 1996.
 17. Rousseeuw, P. J., A. M. Leroy. Robust Regression and Outlier Detection. Wiley, 2003.

Dubnov Yu. A. Federal Research Center “Computer Science and Control” of Russian Academy of Sciences (FRC CSC RAS), Moscow, Russia, Researcher. Higher School of Economics (HSE), National Research University, Moscow, Russia, Senior Lecturer. Scientific publications: 29. Research interests: Bayesian estimation, machine learning, principle of maximum entropy. E-mail: yury.dubnov@phystech.edu

Boulytchev A. V. PhD. Federal Research Center “Computer Science and Control” of Russian Academy of Sciences (FRC CSC RAS), Moscow, Russia, Leading Researcher. Higher School of Economics (HSE), National Research University, Moscow, Russia, Assistant Professor. Scientific publications: 33. Research interests: data analysis, bayesian methods in statistics and econometrics. E-mail: boulytchev.isa.ran@gmail.com