Приближенное оценивание с помощью ускоренного метода наибольшей энтропии. Часть 1. Постановка задачи и реализация для задачи регрессии*

Ю. А. Дубнов, А. В. Булычев

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

Аннотация. Работа посвящена разработке 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 по сравнению с традиционным методом наибольшей энтропии. В данной работе была разработана концепция метода энтропийного оценивания с мягкой рандомизацией, включая получение энтропийно-оптимальных функций ПРВ в общем виде. В ходе экспериментов были протестированы несколько типов регуляризации модели на примере классической задачи регрессионного анализа.

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

DOI 10.14357/20718632220407

Введение

Работа посвящена проблеме оценивания вероятностных параметрических моделей и нацелена на развитие метода рандомизированного энтропийно-робастного оценивания, основанного на максимизации функционала информационной энтропии.

Наиболее распространенными методами оценивания математических моделей являются метод наименьших квадратов (МНК), метод наибольшего правдоподобия (МНП) и байесовское оценивание. Несмотря на то, что МНК является непараметрическим методом, его эффективность и свойства получаемых оценок существенным образом зависят от функции правдоподобия для ошибок модели [1, 2]. Информация о функции правдоподобия также необходима для МНП и байесовского оценивания. С другой стороны, существует ряд прикладных задач, при решении которых отсутствуют априорные знания о функции правдоподобия, или нарушаются иные предпосылки классических методов оценивания, такие как эргодичность и стационарность данных, что становится особенно актуальным при анализе временных рядов [3, 4]. Примерами таких задач являются анализ биржевых индексов и макроэкономических показателей из области экономики и анализ записей ЭКГ и ЭЭГ из области медицины.

Решить проблему отсутствия функции правдоподобия позволяет развиваемый в последние несколько лет метод энтропийно-робастного оценивания [5-7], который базируется на

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

оптимизации энтропийного функционала в соответствии с ограничениями на согласованность выхода модели с реальными данными. Согласованность выхода модели определяется выполнением, так называемых, балансовых ограничений, представляющих собой систему интегральных уравнений, решение которой требует существенных вычислительных ресурсов. Новая идея «мягкой» рандомизации заключается в использовании приближенной согласованности модели с реальными данными вместо строгого выполнения балансовых ограничений.

1. Метод энтропийно-робастного оценивания

Приведем кратко формулировку метода энтропийного оценивания в традиционной форме с выполнением балансовых ограничений (с «жесткой» рандомизацией).

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 (характеристик) этих объектов. В регрессионном анализе вектор y принято называть целевой 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, которую необходимо смоделировать. Для этого декларируется модель с параметрами $a \in R^r$, например, следующего вида:

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

где F – вектор-функция, а ε – остатки модели или шумы измерений. В методе энтропийного оценивания параметры модели и шумы полагаются случайными величинами, в общем случае непрерывными, определенными на некоторых интервалах и характеризующимися функциями ПРВ:

$$a \in A = [a^-, a^+], \qquad \varepsilon \in E = [\varepsilon^-, \varepsilon^+].$$

Целью энтропийно-робастного оценивания является восстановление оптимальных функций ПРВ $P(a)$ и $Q(\varepsilon)$ посредством максимизации функционала информационной энтропии

$$\mathrm{H}[P(a), Q(\varepsilon)] = -\int_A P(a) \ln P(a) da - \int_E Q(\varepsilon) \ln Q(\varepsilon) d\varepsilon \Rightarrow \max_{P,Q}$$

при выполнении ограничений на функции ПРВ (нормировка):

$$\int_A P(a) da = 1, \qquad \int_E Q(\varepsilon) d\varepsilon = 1;$$

и на средний выход модели (балансовые ограничения):

$$\mathbb{E}[\hat{y}_i] = \mathbb{E}\left[F\left(x^{(i)}, a\right) + \varepsilon_i\right] = \int_A P(a) F\left(x^{(i)}, a\right) da + \int_E Q(\varepsilon) \varepsilon_i d\varepsilon = y_i,$$
$$i = 1, \dots, n.$$

Данная задача является вариационной задачей Ляпуновского типа и имеет аналитическое решение, параметризованное множителями Лагранжа:

$$P^*(a, \lambda) = \frac{\exp\left(-\sum_{i=1}^n \lambda_i F\left(x^{(i)}, a\right)\right)}{\int_A \exp\left(-\sum_{i=1}^n F\left(x^{(i)}, a\right)\right) da},$$

$$Q^*(\varepsilon, \lambda) = \frac{\exp\left(-\sum_{i=1}^n \lambda_i \varepsilon_i\right)}{\int_E \exp\left(-\sum_{i=1}^n \lambda_i \varepsilon_i\right) d\varepsilon},$$

где множители Лагранжа $\lambda_1, \dots, \lambda_n$ вычисляются из системы уравнений:

$$\int_A P^*(a,\lambda) F\big(x^{(i)}, a\big) da + \int_E Q^*(\varepsilon, \lambda)\varepsilon_i d\varepsilon = y_i, \qquad i = 1, \dots, n.$$

Таким образом, при восстановлении энтропийно-оптимальных функций ПРВ необходимым этапом является численное решение системы интегральных уравнений, что может потребовать существенных вычислительных ресурсов, особенно для анализа больших данных, и ограничивает область применения энтропийно-робастного оценивания в целом.

Максимизация функционала информационной энтропии, как меры неопределенности системы, позволяет получить в некотором смысле наилучшие оценки в условиях наибольшей неопределенности. Эта логическая взаимосвязь была декларирована в [8] и развивалась в работах [9, 10].

В практическом плане энтропийное оценивание оказывается эффективным в условиях малого объема входных данных и при наличии несимметричности шумов измерений [11]. Существуют примеры применения данной технологии оценивания как в задачах регрессионного анализа для прогнозирования экономических показателей [12, 13], так и в задачах машинного обучения для классификации [14, 15].

2. Метод «мягкого» энтропийного оценивания

В данном разделе рассматривается альтернативный подход к восстановлению энтропийно-оптимальных функций ПРВ, нацеленный на ускорение вычислений.

Балансирование выхода модели $\hat{y} = z(a)$ и реальных данных y будем осуществлять приближенно в терминах одной из гельдеровых норм [16]:

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

Первое выражение характеризует расстояние между векторами выхода модели и реальных данных, второе – норму вектора шума. Обе величины – это функции случайных переменных a и ε. Определим их математические ожидания, т.е. эмпирический риск [17]:

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

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

Для приближенного («мягкого») оценивания введем синтетический функционал $J[P(a), Q(\varepsilon)]$, агрегирующий энтропийную меру неопределенности с эмпирическим риском и средней нормой вектора ошибок. Тогда алгоритм «мягкого» оценивания приобретает следующий вид:

$$J[P(a), Q(\varepsilon)] = \mathrm{H}[P(a), Q(\varepsilon)] - \overline{N}_a[P(a)] - \overline{N}_\varepsilon[Q(\varepsilon)] \Rightarrow \max\nolimits_{P,Q}$$

при условии нормировки функций ПРВ:

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

Данная задача имеет несколько иную форму функционала для оптимизации и не требует выполнения балансовых ограничений. Ее решение аналогично решению задачи классического энтропийного оценивания посредством введения функционала Лагранжа:

$$\mathrm{L}[P(a), Q(\varepsilon)] = J[P(a), Q(\varepsilon)] + \nu\left(1 - \int_A P(a) da\right) + \mu\left(1 - \int_E Q(\varepsilon) d\varepsilon\right).$$

Здесь ν, μ – скалярные переменные. Условия оптимальности формулируются в терминах множителей Лагранжа и производных Гато [18]. Поскольку ПРВ $P(a), Q(\varepsilon)$ принадлежат классу $\mathbb{C}_1$ непрерывно-дифференцируемых функций, представим их вариацию в следующем виде:

$$P(a) = P^*(a) + \alpha p(a), \qquad Q(\varepsilon) = Q^*(\varepsilon) + \beta q(\varepsilon),$$

где α, β – вещественные переменные, а функции $p(a)$, $q(\varepsilon)$ также принадлежат классу $\mathbb{C}_1$. Тогда производные Гато функционала $\mathrm{L}[P(a), Q(\varepsilon)]$ и условия оптимальности имеют вид:

$$\left.\frac{dL}{d\alpha}\right|_{\alpha=0} = \frac{d}{d\alpha}\left[-\int_A \big(P^*(a)+\alpha p(a)\big)\ln\big(P^*(a)+\alpha p(a)\big)\,da - \int_A \big(P^*(a)+\alpha p(a)\big)N_a(a)da\right]\Bigg|_{\alpha=0}$$

$$= \left[-\int_A \Big(\mathrm{p(a)}\ln\big(P^*(a)+\alpha p(a)\big) + p(a) + p(a)N_a(a)\Big)\,da\right]\Bigg|_{\alpha=0}$$

$$= \int_A p(a)\big(\ln P^*(a) + 1 + N_a(a)\big)da = 0,$$

и аналогично для шума:

$$\left.\frac{dL}{d\beta}\right|_{\beta=0} = \frac{d}{d\beta}\left[-\int_E \big(Q^*(\varepsilon)+\beta q(\varepsilon)\big)\ln\big(Q^*(\varepsilon)+\beta q(\varepsilon)\big)\,d\varepsilon - \int_E \big(Q^*(\varepsilon)+\beta q(\varepsilon)\big)\,N_\varepsilon(\varepsilon)d\varepsilon\right]\Bigg|_{\beta=0}$$

$$= \left[-\int_A \big(q(\varepsilon)\ln\big(Q^*(\varepsilon)+\beta q(\varepsilon)\big) + q(\varepsilon) + q(\varepsilon)\,N_\varepsilon(\varepsilon)\big)d\varepsilon\right]\Bigg|_{\beta=0}$$

$$= \int_E \mathrm{q}(\varepsilon)\big(\ln Q^*(\varepsilon) + 1 + N_\varepsilon(\varepsilon)\big)d\varepsilon = 0.$$

Для того, чтобы эти уравнения обращались в ноль при любых функциях $p(a), q(\varepsilon)$, необходимо и достаточно обращение в ноль подынтегральных выражений. Отсюда получаем решение исходной экстремальной задачи в общем виде:

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

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

Таким образом, задача «мягкого» энтропийного оценивания порождает экспоненциальное семейство функций ПРВ. Однако их детальная структура определяется характеристиками эмпирического риска и нормы ошибок. Примуществом здесь является отсуствие параметризации множителями Лагранжа, вычисление которых было бы необходимым для выполнения балансовых ограничений.

Вообще говоря, поскольку величина эмпирического риска $N_a(a)$ зависит не только от параметров модели, но и от имеющихся измерений, то функцию $P^*(a)$ можно трактовать как аналог функции правдоподобия при имеющейся выборке наблюдений. Следовательно, в задачах оценивания при остутствии информации о виде распределения ошибок и настоящей функции праводоподобия технология «мягкого» энтропийного оценивания позволяет восстановить ее альтернативу. Тогда точечные оценки параметров могут быть либо точкой максимума энтропийно-оптимальных функций ПРВ, либо математическим ожиданием соответствующих случайных величин.

3. Задача линейной регрессии

Описанный метод «мягкого» энтропийного оценивания открывает широкий спектр возможностей для подбора наиболее обобщающих моделей, позволяя варьировать вид эмпирического риска и/или норму ошибок модели.

В данном разделе рассмотрим пример классической задачи регрессионного анализа – восстановление коэффициентов линейной регрессии. В качестве эмпирического риска тестируется 3 вида функций: нормы L_1, L_2 и их линейная комбинация.

Итак, пусть имеется вектор наблюдений $Y = \{y_i, i = 1, \dots, n\}$ и матрица предикторов $X = \{X_{ij}, i = 1, \dots, n; j = 1, \dots, m\}$, которые связаны уравнением линейной регрессии:

$$Y = Xb + \varepsilon,$$

где $b = \{b_j, j = 1, \dots, m\}$ – вектор параметров регрессии, и $\varepsilon = \{\varepsilon_i, i = 1, \dots, n\}$ – остатки модели, в общем случае неизвестного распределения.

Определим интервалы принадлежности параметров модели и шумов измерений как

$$B = [b^-, b^+], \qquad E = [\varepsilon^-, \varepsilon^+],$$

и соответствующие им функции ПРВ как $P(b)$ и $Q(\varepsilon)$. В общем случае, как параметры модели, так и остатки для разных измерений могут иметь разные интервалы принадлежности. Определение данных интервалов является предварительным этапом настройки алгоритма энтропийного оценивания.

Степень согласованности выхода модели с реальными данными будем определять в следующих трёх вариантах:

$$N_1(b) = \|\hat{y} - y\|_1 = \sum_{i=1}^{n} |\hat{y}_i - y_i| = \sum_{i=1}^{n} \left|\langle X^i, b\rangle - y_i\right|;$$

$$N_2(b) = \|\hat{y} - y\|_2 = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} \left(\langle X^i, b\rangle - y_i\right)^2;$$

$$N_3(b) = \mathrm{c}_1 \|\hat{y} - y\|_1 + \mathrm{c}_2 \|\hat{y} - y\|_2$$

В общем случае, $N_b(b)$ и, аналогично, $N_\varepsilon(\varepsilon)$. Софрмулируем задачу «мягкого» энтропийного оценивания:

$$J[P(b), Q(\varepsilon)] = \mathrm{H}[P(b), Q(\varepsilon)] - \overline{N}_b[P(b)] - \overline{N}_\varepsilon[Q(\varepsilon)] \Rightarrow \max_{P,Q}$$

при условии нормировки функций ПРВ:

$$\int_B P(b)db = 1, \qquad \int_E Q(\varepsilon)d\varepsilon = 1.$$

Решение для энтропийно-оптимальных функций ПРВ в общем виде приведено в предыдущем разделе, представим его для одного из трех выбранных типов регуляризации на примере функции $P(b)$:

$$P^*(\mathrm{b}) = \frac{1}{C} \exp\left(-\sum_{i=1}^{n} \left(\langle X^i, b\rangle - y_i\right)^2\right), \qquad C = \int_B \exp\left(-\sum_{i=1}^{n} \left(\langle X^i, b\rangle - y_i\right)^2\right) db$$

Как видно, при расчете эмпирического риска в терминах нормы L_2, вид энтропийно-оптимальных решений соответствует с точностью до константы функции правдоподобия для параметров линейной регрессии при нормально-распределенных остатках. Следовательно, при расчете оценки как точки максимума энтропийно-оптимальной функции ПРВ, эта оценка будет совпадать с оценками по методу МНК. В остальных случаях энтропийно-оптимальные функции ПРВ также принадлежат экспоненциальному классу распределений, но имеют несколько иную форму.

4. Результаты экспериментов

Для апробации предложенного метода была проведена серия экспериментов с модельными данными. Целью экспериментов являлась проверка работоспособности метода «мягкого» энтропийного оценивания на примере классической задачи линейной регрессии.

4.1. Описание эксперимента

Методология экспериментов включала формирование массивов входных данных, соответствующих выбранным коэффициентам для парной регрессии с последующим добавлением случайного шума. Все параметры генерации данных, включая распределение шума и состояние генератора случайных чисел, фиксировались для обеспечения строгой воспроизводимости всех результатов.

После генерации набора данных по имеющимся наблюдениям производилось оценивание параметров регрессии методами наименьших квадратов (OLS), классического энтропийно-робастного оценивания (Max Ent.) и «мягкого» энтропийного оценивания (Soft Ent.). Результаты оценивания сравниваются в терминах коэффициента детерминации модели:

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

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

Коэффициент детерминации R^2 принимает значения не более 1 и является одной из наиболее распространенных метрик качества модели линейной регрессии.

Пример восстановления коэффициентов модели линейной регрессии представлен на Рис. 1. Стоит отметить, что поскольку шум в данных имеет случайную природу, и сами оценки являются случайными величинами, об устойчивости результатов следует судить по итогу множества испытаний, в данном случае проводилось $mc = 100$ испытаний Монте-Карло с последующим усреднением результатов.

В эксперименте, продемонстрированном на Рис. 1, использовалась выборка из $n = 50$ измерений и несимметричные ошибки по распределению «хи»-квадрат. Для «мягкого» энтропийного оценивания использовалась норма L_1. Как видно, получаемые всеми тремя методами модели имеют сравнимые значения коэффициента детерминации. Данный эксперимент демонстрирует работоспособность предложенного метода «мягкого» энтропийного оценивания.

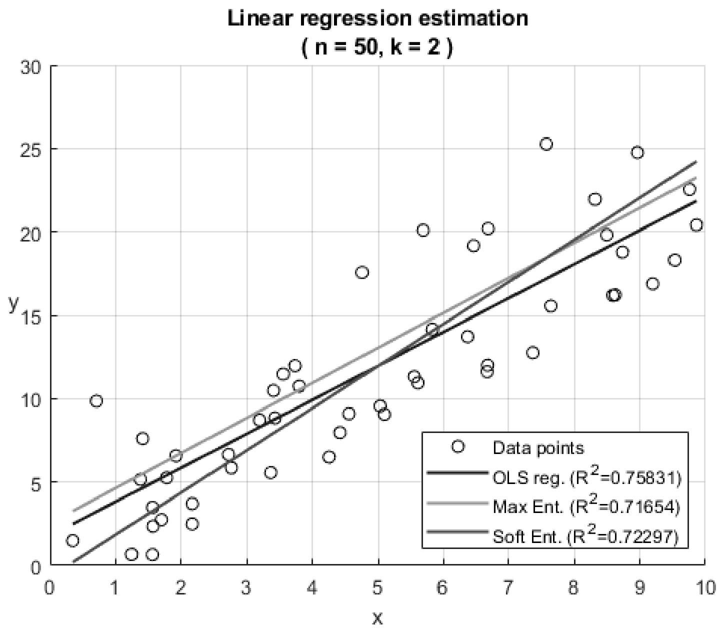

Рис. 1. Пример восстановления линейной модели

Отметим, что наибольшее значение коэффициента R^2 достигается именно для метода наименьших квадратов, а не энтропийного оценивания. Этот факт также напрямую следует из выражения для R^2, в котором в явном виде участвует сумма квадратов отклонений выхода модели от реальных измерений. Не следует, однако, по этим значениям в полной мере судить об эффективности методов оценивания. Для более основательного сравнения необходимо исследовать отклонение получаемых оценок от оригинальных значений коэффициентов, потому, как в случае ошибок, отличных от нормально-распределенных, метод МНК дает смещенные оценки.

4.2. Восстановление функций ПРВ

Далее, на Рис. 2-4 представлены примеры восстановленных в ходе экспериментов энтрпоийно-оптимальных функций ПРВ при использовании норм L_1, L_2 и их комбинации соответственно. Можно отметить характерный пик на вершине графика для нормы L_1 (Рис. 2), и, наоборот, плавность функции ПРВ для расчета по норме L_2 (Рис. 4).

Данные рисунки демонстрирют, почему точки максимума восстановленных функций ПРВ и их средние могут приводить к отличающимся оценкам. По аналогии с методом наибольшего правдоподобия, в проведенных экспериментах наилучшую точность показали оценки по точкам максимума энтропийно-оптимальных решений.

4.3. Эксперимент для малого объема данных

На Рис. 5 приведены результаты сравнения эффективности метода наименьших квадратов и «мягкого» энтропийного оценивания в условиях малого объема входных данных.

В данном эксперименте использовалась выборка из $n = 20$ точек, что приводит к менее точным оценкам. Однако падение точности для метода энтрпоийного оценивания оказывается менее значительным, чем для метода МНК.

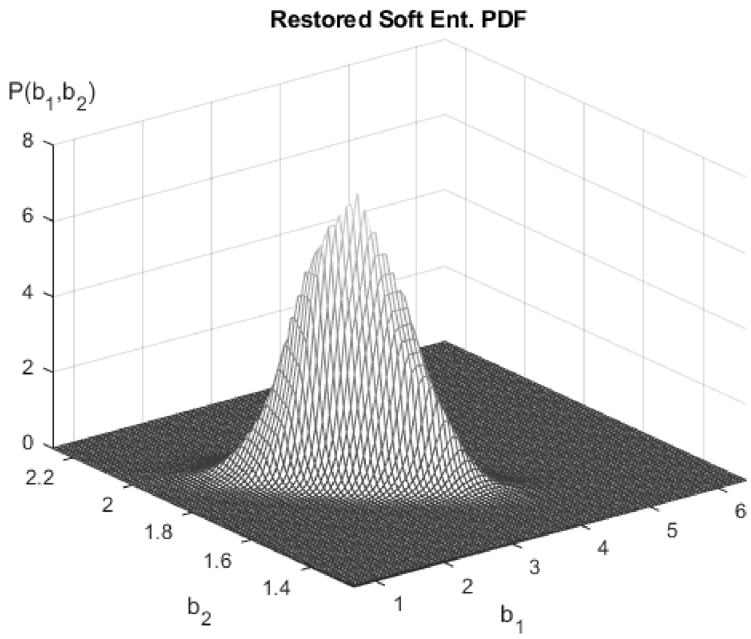

Рис. 2. Вид энтропийно-потимальной функции ПРВ для нормы L_1.

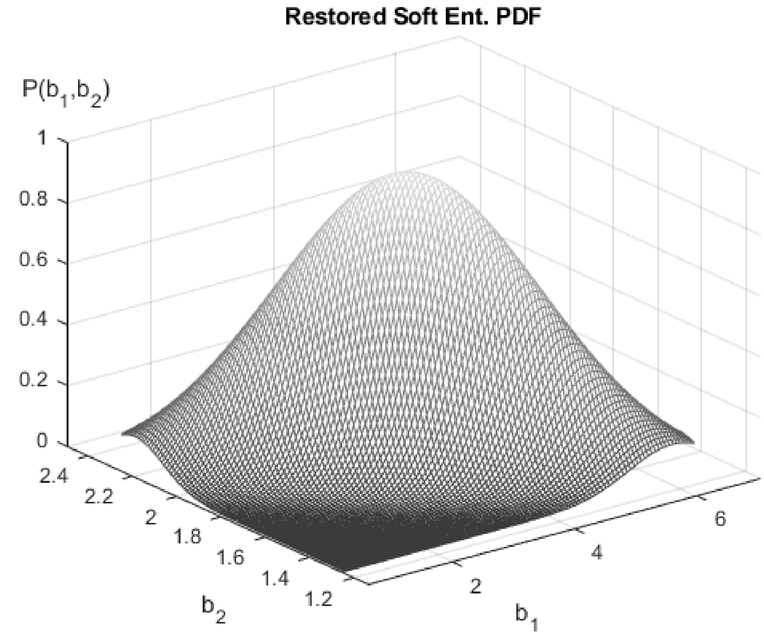

Рис. 3. Вид энтропийно-потимальной функции ПРВ для нормы L_2.

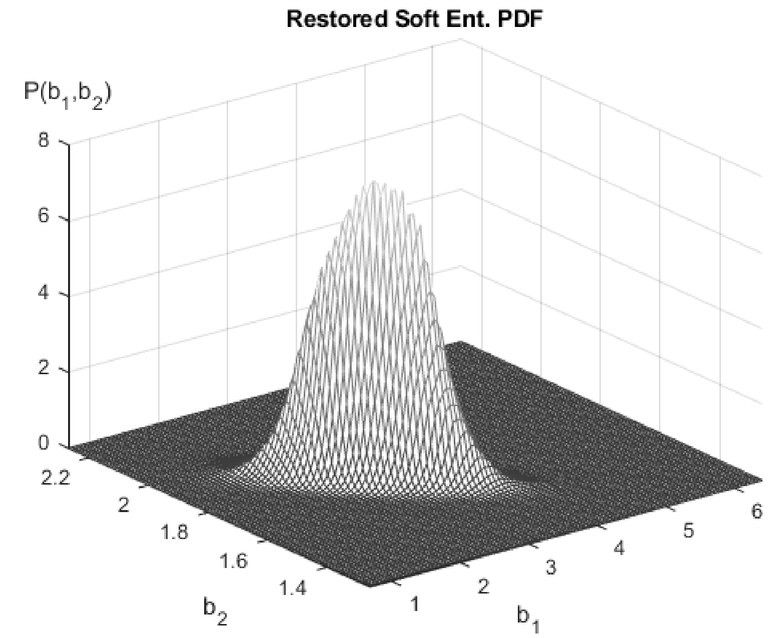

Рис. 4. Вид энтропийно-потимальной функции ПРВ для комбинации норм L_1 и L_2

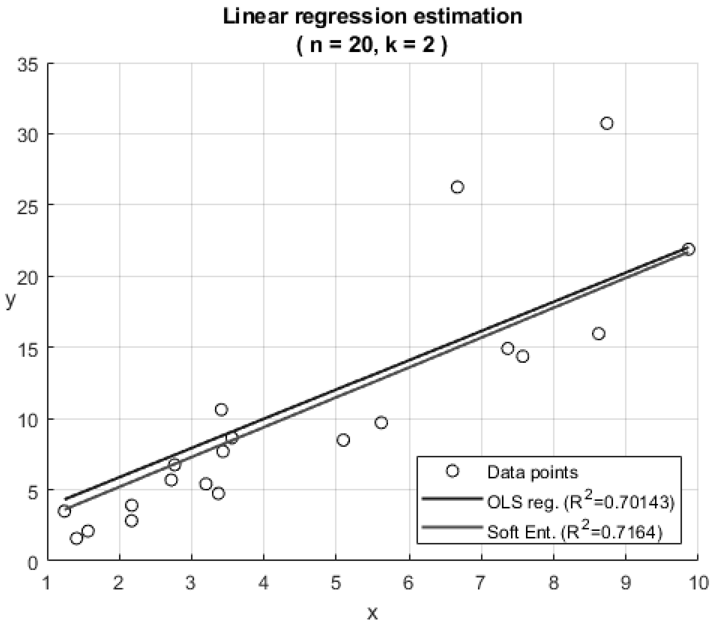

Рис. 5. Пример восстановления линейной модели для малого объема входных данных

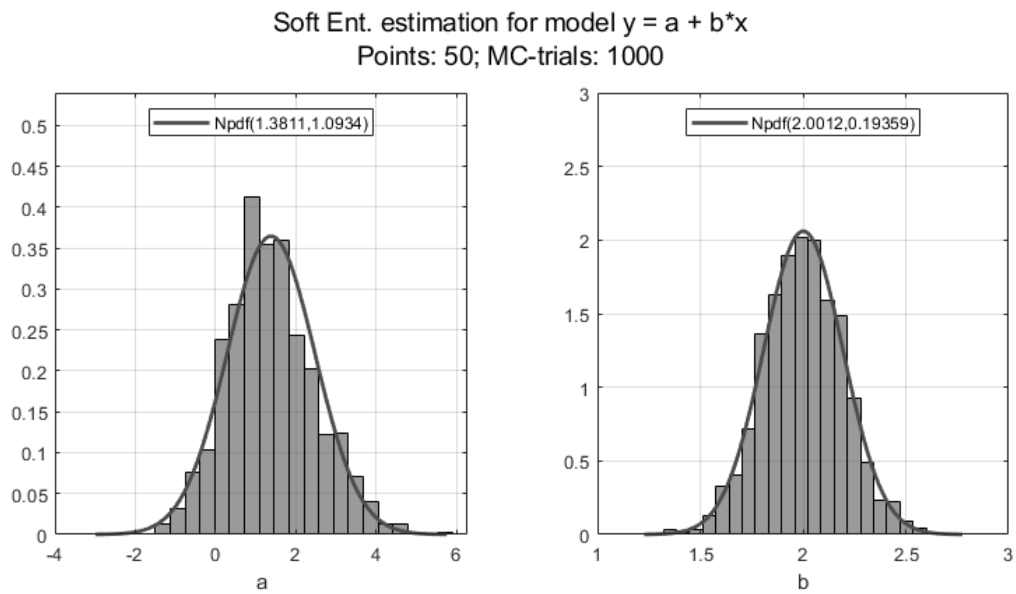

Рис. 6. Гистограммы оценок при использовании нормы L_1

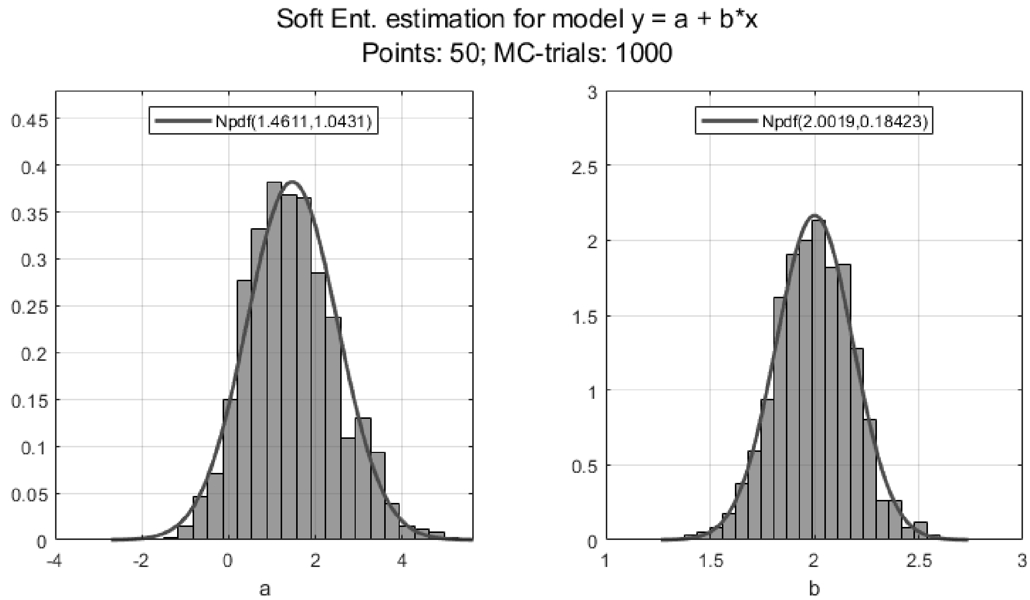

Рис. 7. Гистограммы оценок при использовании комбинации норм L_1 и L_2

4.4. Распределение оценок

Наконец, последняя серия экспериментов была посвящена изучению некоторых свойств оценок по методу «мягкого» энтропийного оценивания. Для этого строилась гистограмма значений коэффициентов линейной модели, полученных в результате оценивания. В данном случае использовалось $mc = 1000$ повторений. Полученные гистограммы приведены на Рис. 6 для оценивания по норме L_1 и на Рис. 7 – для комбинации норм L_1 и L_2.

Как видно из Рис. 6 и 7, в обоих случаях получаемые гистограммы приближенно соответствуют некоторому нормальному распределению, что может говорить об асимптотической нормальности оценок.

Заключение

В рамках данного исследования была разработана концепция метода энтропийного оценивания с мягкой рандомизацией, включая получение решения для энтропийно-оптимальных функций ПРВ в общем виде. В ходе экспериментов на примере задачи линейной регрессии были протестированы несколько типов регуляризации модели. Дополнительно была исследована асимптотическая нормальность оценок и поведение предложенного метода в условиях малого объема входных данных. В дальнейшем планируется детально исследовать свойства получаемых оценок и поведение предложенного метода в условиях различных объемов обучающих выборок и различных типов шумов измерений.

Литература

1. Björck, Å. Numerical Methods for Least Squares Problems. SIAM. 1996.
2. Huang, David S. Regression and Econometric Methods. New York: John Wiley & Sons. 1970. pp. 127–147.
3. Мишулина О. А. Статистический анализ и обработка временных рядов. — М.: МИФИ, 2004. — С. 180.
4. Woodward, W. A., Gray, H. L. & Elliott, A. C. Applied Time Series Analysis, CRC Press. 2012.
5. Amos Golan, George G. Judge, Douglas Miller. Maximum Entropy Econometrics: Robust Estimation with Limited Data. – John Wiley and Sons Ltd. Chichester, U.K., 1996.
6. 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.

7. 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.
8. Jaynes E.T. Information Theory and Statistical Mechanics // Physics Review Notes. 1957. V. 106. P. 620–630.
9. R.D. Levin, M. Tribus. The maximum entropy formalism. MIT Press, 1979.
10. Jaynes E.T. Probability Theory. The logic and science. Cambrige Univ. Press, 2003.
11. Ximing Wu. A Weighted Generalized Maximum Entropy Estimator with a Data-driven Weight // Entropy, 2009. no.11.
12. Ю. С. Попков, А. Ю. Попков, Ю. А. Дубнов. Элементы рандомизированного прогнозирования и его применение для предсказания суточной электрической нагрузки энергетической системы", Автомат. и телемех., 2020, № 7, 148–172
13. Popkov, Y.S.; Popkov, A.Y.; Dubnov, Y.A.; Solomatine, D. Entropy-Randomized Forecasting of Stochastic Dynamic Regression Models. Mathematics 2020, 8, 1119.
14. Yuri S. Popkov, Zeev Volkovich, Yuri A. Dubnov, Renata Avros and Elena Ravve. Entropy "2"-Soft Classification of Objects // Entropy, 2017, Vol. 19, Iss. 4, No.178.
15. Дубнов Ю.А. Энтропийное оценивание в задачах классификации // «Автомат. и телемех.», 2019, № 3, 138–151.
16. Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. М., Наука, 1984.
17. Цыпкин Я.З. Основы теории обучающихся систем. М., Наука, 1970.
18. Kaashoek M.A., Seatzu S., van der Mee C. Recent Advances in Operator Theory and its Application. 2006, Springer, p.478.

Дубнов Юрий Андреевич. Институт системного анализа Федерального государственного учреждения "Федеральный исследовательский центр "Информатика и управление" Российской академии наук" (ФИЦ ИУ РАН), г. Москва, Россия, Научный сотрудник. Национальный исследовательский университет Высшая школа экономики (НИУ ВШЭ), г. Москва, Россия, 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, Россия, Ведущий научный сотрудник. Национальный исследовательский университет Высшая школа экономики (НИУ ВШЭ), г. Москва, Россия, 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 1. Problem Statement and Implementation for the Regression Problem

Yu. A. Dubnov, A. V. Boulytchev

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

Abstract The work is devoted to the development of an entropy estimation method with "soft" randomization for restoring the parameters of probabilistic mathematical models from the available observations. Soft randomization refers to the technique of adding regularization to the information entropy functional to simplify the optimization problem and speed up learning process compared to the traditional maximum entropy method. In this work, the concept of the soft randomization entropy estimation method was developed, including obtaining entropy-optimal PDF functions in general form. During the experiments, several types of model regularization were tested on the example of a classical regression analysis problem.

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

DOI 10.14357/20718632220407

References

1. Björck, Å. Numerical Methods for Least Squares Problems. SIAM. 1996.
2. Huang, David S. Regression and Econometric Methods. New York: John Wiley & Sons. 1970. pp. 127–147.
3. Mishulina O. A. Statisticheskij analiz i obrabotka vremennyh ryadov. — M.: MIFI, 2004. — S. 180.

4. Woodward, W. A., Gray, H. L. & Elliott, A. C. Applied Time Series Analysis, CRC Press. 2012.
5. Amos Golan, George G. Judge, Douglas Miller. Maximum Entropy Econometrics: Robust Estimation with Limited Data. – John Wiley and Sons Ltd. Chichester, U.K., 1996.
6. 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.
7. 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.
8. Jaynes E.T. Information Theory and Statistical Mechanics // Physics Review Notes. 1957. V. 106. P. 620–630.
9. R.D. Levin, M. Tribus. The maximum entropy formalism. MIT Press, 1979.
10. Jaynes E.T. Probability Theory. The logic and science. Cambrige Univ. Press, 2003.
11. Ximing Wu. A Weighted Generalized Maximum Entropy Estimator with a Data-driven Weight // Entropy, 2009. no.11.
12. Popkov, Y., Popkov, A. & Dubnov, Y. Elements of Randomized Forecasting and Its Application to Daily Electrical Load Prediction in a Regional Power System // Automation and Remote Control. 2020. V. 81, Iss.. 7. Pp. 1286–1306.
13. Popkov, Y.S.; Popkov, A.Y.; Dubnov, Y.A.; Solomatine, D. Entropy-Randomized Forecasting of Stochastic Dynamic Regression Models. Mathematics 2020, 8, 1119.
14. Yuri S. Popkov, Zeev Volkovich, Yuri A. Dubnov, Renata Avros and Elena Ravve. Entropy “2”-Soft Classification of Objects // Entropy, 2017, Vol. 19, Iss. 4, No.178.
15. Yu. A. Dubnov. Entropy-Based Estimation in Classification Problems // Automation and Remote Control. 2019, V. 80, Iss. 3, pp 502-512.
16. Voevodin V.V., Kuznecov Yu.A. Matricy i vychisleniya. M., Nauka, 1984.
17. Cypkin Ya.Z. Osnovy teorii obuchayushchihsya sistem. M., Nauka, 1970.
18. Kaashoek M.A., Seatzu S., van der Mee C. Recent Advances in Operator Theory and its Application. 2006, Springer, p.478.

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: bulytchev.isa.ran@gmail.com