

ОБРАБОТКА ИНФОРМАЦИИ

УДК 519.2

В.А. Дёмин, Е.В. Чимитова, В.Ю. Щеколдин

ИССЛЕДОВАНИЕ МЕТОДА ВЫБОРА ОПТИМАЛЬНОГО ПАРАМЕТРА СГЛАЖИВАНИЯ ПРИ НЕПАРАМЕТРИЧЕСКОМ ОЦЕНИВАНИИ РЕГРЕССИОННЫХ МОДЕЛЕЙ НАДЕЖНОСТИ

Рассматривается один из популярных подходов к непараметрическому оцениванию регрессионной модели надежности, предложенный Р. Бераном, позволяющий оценить условную функцию надёжности при заданном значении объясняющей переменной. Представлены результаты исследования статистических свойств оценки Берана на основе предложенного метода выбора оптимального параметра сглаживания. Выявлены факторы, влияющие на точность оценки Берана, и сформулированы рекомендации по применению предложенного метода выбора оптимального параметра сглаживания.

Ключевые слова: функция надёжности; регрессионная модель; непараметрическая оценка Берана; параметр сглаживания; параметр размытости; робастное оценивание.

Одной из важнейших задач в теории надежности является построение модели зависимости функции надежности от объясняющих переменных (ковариат), в качестве которых обычно выступают воздействия различного типа, такие как температура, давление, напряжение, механические нагрузки и др. На сегодняшний день существует широкий спектр методов, позволяющих решать данную задачу, отличающихся друг от друга спецификой используемых данных, видом привлекаемой априорной информации и степенью точности получаемых результатов. В последнее время большую популярность завоевали непараметрические методы, которые не требуют знания априорной информации о виде регрессионной функции и законе распределения результатов измерений. Однако в большинстве пакетов прикладных программ статистического анализа данных для решения задачи непараметрического оценивания функции надежности привлекаются лишь оценки Каплана – Мейера, которые не позволяют учесть влияние объясняющих переменных на изменение функции надежности. В этом смысле представляются актуальными разработка и развитие более широкого класса методов, учитывающих влияние ковариат. К таким методам относится предложенное в 1981 г. Р. Бераном обобщение оценки Каплана – Мейера на случай построения регрессионных моделей надежности [1].

Несмотря на преимущества оценки Берана как непараметрического метода [8], до сих пор существует весьма ограниченное число публикаций, посвященных исследованию статистических свойств данной оценки, среди которых можно выделить работы [2–5], в которых эти свойства изучаются для случайного плана эксперимента, когда значение ковариаты не фиксировано, а в [6] рассматриваются свойства оценки для неслучайного плана, когда значения ковариаты определяются заранее.

Важнейшую роль при построении непараметрических оценок играет выбор параметра сглаживания, от которого существенно зависит точность получаемых оценок. В случае оценивания классических регрессионных моделей разработано множество методов определения оптимальных значений параметра сглаживания, таких как метод кросс-валидации, метод минимума интегральной среднеквадратической ошибки, методы асимптотической оценки и др. [7, 8]. Применяя тот или иной метод выбора оптимального параметра сглаживания, важно понимать, что смысловая интерпретация этого параметра в оценке Берана существенно отличается от его интерпретации для классических моделей

регрессии, поскольку на его основе могут быть получены лишь весовые коэффициенты, которые впоследствии используются при построении оценки. Среди публикаций на эту тему можно отметить [9], где предлагается решение данной задачи при помощи бутстреп-метода в предположении, что ковариата является случайной величиной и определяется некоторым законом распределения. Однако для теории надёжности такой вариант совершенно не характерен, поскольку трудно представить ситуацию, в которой объясняющие переменные имеют стохастическую природу. Преодоление этой проблемы представляет собой отдельную тему исследований, некоторые направления которых были упомянуты в предыдущих работах авторов [5, 9]. В частности, в [10] предложен метод выбора оптимального параметра сглаживания для оценки Берана, основанный на минимизации среднеквадратического отклонения времен отказов от непараметрической оценки обратной функции надёжности. Выяснено, что оценки Берана, построенные с использованием оптимального параметра сглаживания, оказались точнее, чем для фиксированного значения параметра сглаживания. Тем не менее вопросы о выборе вида ядерной функции, метода определения параметра размытости для непараметрической оценки обратной функции надёжности и типа весовых коэффициентов требуют проведения дополнительных исследований. Целью данной работы являются изучение статистических свойств оценки Берана и разработка практических рекомендаций по применению предложенного метода выбора оптимального параметра сглаживания.

1. Непараметрическая оценка Берана и выбор оптимального параметра сглаживания

Обозначим через T_x время безотказной работы исследуемых технических изделий, которое зависит от скалярной ковариаты x , диапазон значений которой определяется условиями эксперимента и представляет собой отрезок числовой прямой. Без потери общности будем считать этот отрезок равным $[0;1]$. Функция надёжности определяется соотношением

$$S(t|x) = P(T_x \geq t) = 1 - F(t|x), \quad (1)$$

где $F(t|x)$ – условная функция распределения случайной величины T_x .

Главной особенностью данных типа времени жизни является наличие цензурированных справа наблюдений, которые можно представить в виде

$$(Y_1, x_1, \delta_1), (Y_2, x_2, \delta_2), \dots, (Y_n, x_n, \delta_n),$$

где n – объем выборки, x_i – значение ковариаты для i -го объекта, Y_i – время наработки до момента отказа или цензурирования, δ_i – индикатор цензурирования, который принимает значение 1, если наблюдение полное, и 0, если цензурированное.

Оценка Берана, согласно [1], имеет следующий вид:

$$\tilde{S}_{h_n}(t|x) = \prod_{Y_{(i)} \leq t} \left\{ 1 - \frac{W_n^i(x; h_n)}{1 - \sum_{j=1}^{i-1} W_n^j(x; h_n)} \right\}^{\delta_i}, \quad (2)$$

где x – значение ковариаты, для которой оценивается функция надёжности; $W_n^i(x; h_n)$, $i = 1, \dots, n$, – веса Надарая – Уотсона, которые [5, 8] определяются как

$$W_n^i(x; h_n) = K\left(\frac{x - x_i}{h_n}\right) / \sum_{j=1}^n K\left(\frac{x - x_j}{h_n}\right), \quad (3)$$

где $K(\cdot)$ – ядерная функция, удовлетворяющая условиям регулярности

$$K(y) = K(-y), 0 \leq K(y) < \infty, \int_{-\infty}^{\infty} K(y) dy = 1,$$

$h_n > 0$ – параметр сглаживания такой, что $\lim_{n \rightarrow \infty} h_n = 0$, $\lim_{n \rightarrow \infty} n h_n = \infty$.

В [10] предложен алгоритм выбора оптимального параметра сглаживания h_n для оценки Берана, основанный на минимизации среднеквадратического отклонения времен отказов $Y_1, Y_2, \dots, Y_n$ от непараметрической оценки обратной функции надежности $S_x^{-1}(p)$. В данной работе будем определять значения оптимального параметра как решение следующей оптимизационной задачи:

$$h_n^{\text{opt}} = \arg \min_{h_n} \sum_{i=1}^n \delta_i \cdot |\hat{g}(\hat{p}_i | x_i) - Y_i|, \quad (4)$$

где

$$\hat{g}(\hat{p}_i | x_i) = \sum_{j=1}^n \omega_j(\hat{p}_i) \cdot Y_j. \quad (5)$$

Вероятности $\hat{p}_i$ вычисляются с использованием оценки Берана, веса $\omega_j(\hat{p}_i)$ можно рассчитать, используя различные весовые функции. Далее будем рассматривать веса первого порядка Надарая – Уотсона

$$\omega_j^{(1)}(\hat{p}_i) = W_n^j(\hat{p}_i; b_n) = K\left(\frac{\hat{p}_i - \hat{p}_j}{b_n}\right) / \sum_{k=1}^n K\left(\frac{\hat{p}_i - \hat{p}_k}{b_n}\right)$$

и веса второго порядка Пристли – Чао [1]:

$$\omega_j^{(2)}(\hat{p}_i) = n(\hat{p}_{(i)} - \hat{p}_{(i-1)}) K\left(\frac{\hat{p}_i - \hat{p}_j}{b_n}\right),$$

где b_n – параметр размытости.

Таким образом, для вычисления оптимального параметра сглаживания для оценки Берана необходимы непараметрические оценки вида (5), для которых, в свою очередь, нужно выбирать параметр b_n .

2. Выбор весовой функции, способа вычисления параметра b_n и ядерных функций

Исследуем статистические свойства оценки Берана при использовании оптимального параметра сглаживания (4). Очевидно, что то, насколько хорошо будет подобран оптимальный параметр сглаживания h_n^{opt} , напрямую зависит от качества получаемых оценок (5). Сравним различные веса $\omega_j(\hat{p}_i)$ для ядерной оценки $\hat{g}(\hat{p}_i | x_i)$, а также различные методы выбора параметра размытости b_n с точки зрения точности получаемых оценок Берана.

Поскольку рассматривается задача, предполагающая использование ядерного сглаживания, можно воспользоваться уже разработанными подходами для выбора оптимального параметра сглаживания ядерной оценки регрессии. Рассмотрим следующие методы:

1. Метод кросс-валидации, который чаще всего рассматривается как наиболее точный, однако требующий значительных вычислительных затрат [8].

2. Метод минимума средней интегральной ошибки, в соответствии с которым параметр размытости вычисляется по формуле [8]:

$$b_{NS} = \left[\frac{8\pi^{1/2} R(K)}{3\mu_2(K)^2 n} \right]^{1/5} \hat{\sigma}, \quad (6)$$

где $\mu_2(K) = \int y^2 K(y) dy$, $R(K) = \int K^2(y) dy$. Величина $\hat{\sigma}$ – оценка среднеквадратического отклонения, которую можно вычислять различными способами, чаще всего для этого применяется несмещенная выборочная дисперсия

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{p}_i - \bar{\hat{p}})^2. \quad (7)$$

Однако оценка (7), во-первых, не является робастной, а во-вторых, обладает «хорошими» статистическими свойствами лишь в случае законов распределения, близких к нормальному. Поэтому в данной работе будем также рассматривать робастную оценку среднеквадратического отклонения вида [11]:

$$\hat{\sigma}_{\text{robust}} = 1,4826 \operatorname{med}_{i=1 \dots n} \left| \hat{p}_i - \operatorname{med}_{j=1 \dots n, k=j+1 \dots n} \left(\frac{\hat{p}_j + \hat{p}_k}{2} \right) \right|. \quad (8)$$

Такая оценка представляет собой комбинацию известной робастной оценки Ходжеса – Леманна (Hodges – Lehmann) для параметра сдвига и робастной оценки Русьё (Rousseeuw), называемой медианой абсолютных отклонений, для параметра масштаба.

Исследование свойств получаемых оценок проводилось с помощью методов статического моделирования. В качестве отклонения оценки Берана от истинной условной функции надёжности использовалась статистика:

$$D_{h_n} = \sup_{j=1 \dots k, t < \infty} \left| \tilde{S}_{h_n}(t | x_j) - S_{x_j}(t) \right|, \quad (9)$$

при этом истинной моделью надёжности являлась параметрическая модель пропорциональных интенсивностей Кокса [9]:

$$S_x(t) = (S_0(t))^{r(x;\beta)}, \quad (10)$$

с функцией от ковариат вида $r(x;\beta) = \ln(1 + e^{\beta x})$ и базовой функцией надёжности, соответствующей логнормальному распределению с функцией плотности

$$f_0(t) = \frac{1}{\sqrt{2\pi\theta_1 t}} \exp \left(-\frac{1}{2\theta_1^2} \ln^2 \left(\frac{t}{\theta_2} \right) \right) \quad (11)$$

с параметрами $\theta_1 = 21,5$; $\theta_2 = 1,6$.

Для используемых при моделировании весовых функций, методов оценки дисперсии и истинного значения параметра регрессии β примем следующие обозначения: 1 – веса Пристли – Чао, робастная оценка дисперсии (8), $\beta = 2$; 2 – веса Пристли – Чао, неробастная оценка дисперсии (7), $\beta = 2$; 3 – веса Надарая – Уотсона, робастная оценка дисперсии (8), $\beta = 2$; 4 – веса Надарая – Уотсона, неробастная оценка дисперсии (7), $\beta = 2$; 5 – веса Пристли – Чао, робастная оценка дисперсии (8), $\beta = 5$; 6 – веса Пристли – Чао, неробастная оценка дисперсии (7), $\beta = 5$; 7 – веса Надарая – Уотсона, робастная оценка дисперсии (8), $\beta = 5$; 8 – веса Надарая – Уотсона, неробастная оценка дисперсии (7), $\beta = 5$.

На рис. 1–3 приведены усредненные значения отклонения (9), оптимального параметра сглаживания и параметра размытости b_{NS} , соответственно, при объемах выборок $n = 100, 200, 300$. Ковариата x принимала значения из множества $\{0; 0,11; 0,22; 0,33; 0,44; 0,56; 0,67; 0,78; 0,89; 1\}$, $m = 10$. Количество наблюдений, соответствующее различным значениям ковариаты, во всех экспериментах считалось одинаковым.

Из рис. 1 видно, что использование весовой функции Пристли – Чао позволяет получить более точную оценку Берана. При объеме выборки $n = 100$ выигрыш в точности оценок Берана с использованием весов Пристли – Чао в выражении (5) составляет примерно 3% от средней величины отклонения (9) в случае использования весов Надарая – Уотсона, при $n = 200$ выигрыш около 8%, а при $n = 300$ – 11%. Ухудшение точности оценки Берана при использовании весов Надарая – Уотсона может быть вызвано тем, что данная весовая функция не предполагает упорядочивания весов, в результате чего при построении оценки Берана используются «лишние» наблюдения и соответствующая дисперсия увеличивается. Использование робастной оценки дисперсии при вычислении параметра размытости повышает точность оценки Берана, которая также растёт с увеличением объема выборки.

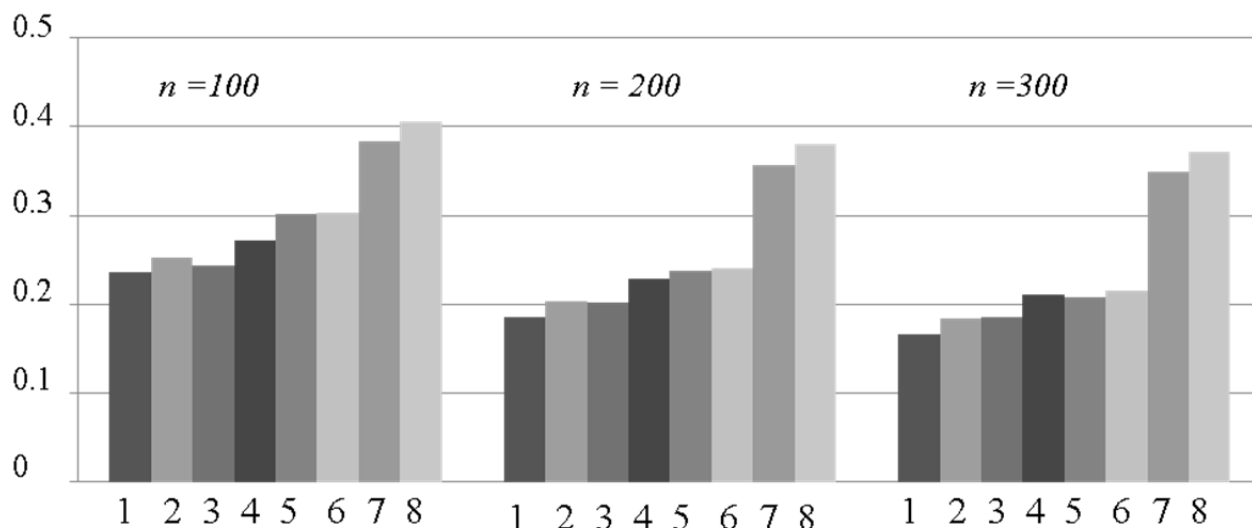


Рис. 1. Усреднённое отклонение между оценкой Берана и истинной условной функцией надежности при различных объёмах выборки

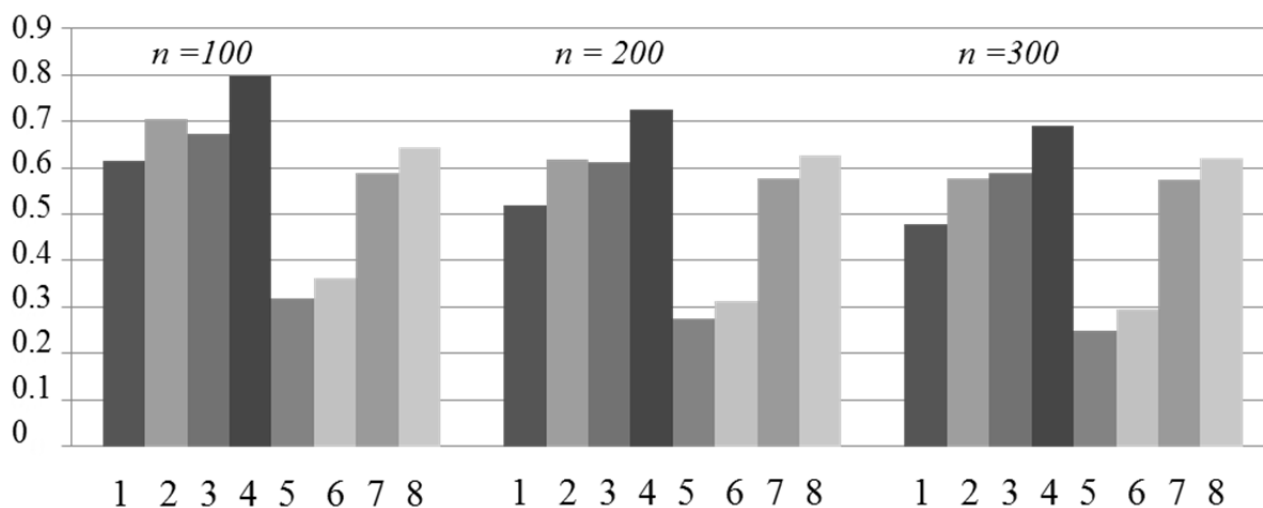


Рис. 2. Усреднённый параметр сглаживания h_n^{opt} при различных объёмах выборки

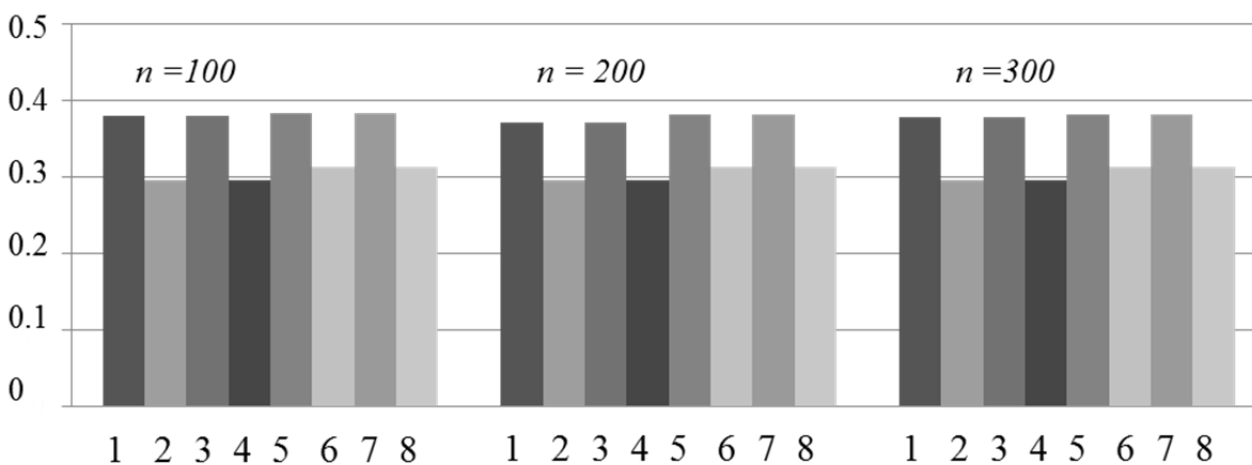


Рис. 3. Усреднённое значение параметра размытости b_{NS} при различных объёмах выборки

Из рис. 2 видно, что с увеличением объёма выборки параметр сглаживания h_n^{opt} уменьшается, что вполне логично, так как число наблюдений при разных значениях ковариаты увеличивается, а значит, растёт и число «лишних» наблюдений. Важно отметить, что с увеличением параметра β в 2,5 раза (т.е. при увеличении влияния ковариаты на функцию надёжности) значение h_n^{opt} уменьшилось почти в два раза при использовании весов Пристли – Чао. Однако в случае весов Надарая – Уотсона такой ситуации не наблюдается, поэтому в данном случае точность оценок Берана с использованием весов Надарая – Уотсона существенно ухудшается.

Как видно из рис. 3, усреднённое значение параметра размытости практически не изменяется при увеличении объёма выборки. Такой результат согласуется с выводами, полученными в работе [12], согласно которым значение параметра b_{NS} не зависит от вида весовой функции.

В рассматриваемых условиях дискретный план эксперимента будет зависеть от числа значений m , которые может принимать ковариата. Следующая часть исследований посвящена изучению зависимости искомой оценки ее свойств от m .

На рис. 4–6 представлены результаты моделирования при фиксированном объёме выборки $n = 200$, при различном числе значений ковариаты m .

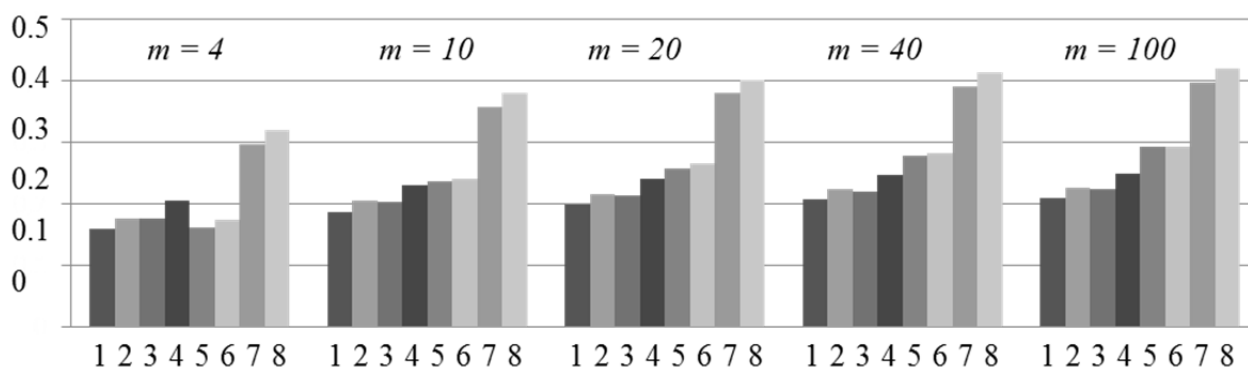


Рис. 4. Усреднённое отклонение между оценкой Берана и истинной условной функцией надёжности при различном числе значений ковариаты

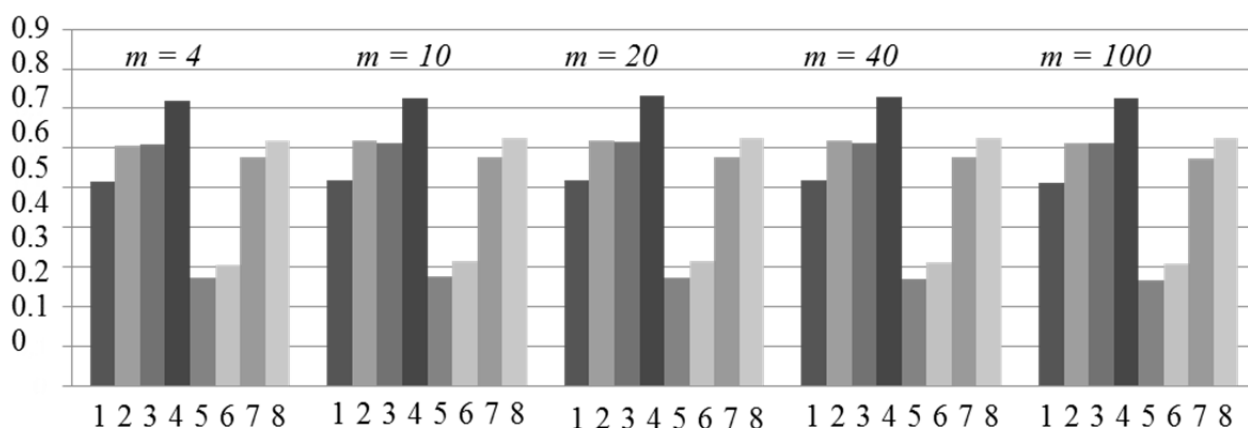


Рис. 5. Усреднённый параметр сглаживания h_n^{opt} при различном числе значений ковариаты

Как видно из рис. 4, с увеличением числа возможных значений ковариаты точность оценки Берана снижается. В частности, при использовании весов Пристли – Чао и робастной оценки дисперсии при $m=10$ относительно $m=4$ ухудшение составляет около 17%, а при $m=20$ относительно $m=10$ – 6%. Этот результат является вполне логичным, поскольку количество наблюдений, соответ-

ствующих фиксированному значению ковариаты, уменьшается, следовательно, объем информации о каждом наблюдении также становится меньше, что приводит к потере точности.

Анализ значений параметра сглаживания на основе рис. 5 показывает, что оптимальные значения параметра практически не изменяются. Аналогичная ситуация наблюдалась при изучении усреднённых значений параметра размытости b_{NS} . Данное свойство оптимальных параметров сглаживания h_n^{opt} и размытости b_{NS} распространяется и на случай других значений регрессионного параметра (при увеличении степени влияния ковариаты на функцию надёжности).

Исследования свойств оценки Берана при различных условиях моделирования были также проведены для экспоненциальной модели Кокса. Выявленные закономерности оказались практически такими же, поэтому конкретные численные результаты для этого случая опускаются.

В большинстве исследований, посвященных непараметрическим методам, отмечается, что выбор ядерной функции практически не влияет на точность получаемых оценок [7, 8]. Однако в данной постановке задачи это свойство может нарушаться. Для проверки этого предположения было проведено исследование точности оценки Берана при использовании различных ядерных функций при решении оптимизационной задачи (4) и оценивании параметра размытости (6). Результаты представлены в таблице.

Усреднённое отклонение (9) при использовании различных ядерных функций

$b_{NS} \backslash h_n$	Квартическое	Епанечникова	Прямоугольное	Гаусса
Квартическое	0,20	0,20	0,21	0,23
Епанечникова	0,21	0,22	0,23	0,26
Прямоугольное	0,25	0,26	0,29	0,30
Гаусса	0,32	0,35	0,36	0,38

Как видно из таблицы, наилучшая точность оценки Берана достигается при использовании квартического и ядра Епанечникова как при оценивании параметра сглаживания, так и при вычислении параметра размытости. Таким образом, изменение в постановке задачи, т.е. при применении ядерных функций для вычисления весовых коэффициентов в непараметрической оценке (2), приводит к тому, что выбор формы ядра оказывает существенное влияние на свойства оценок Берана. Конечно такое влияние обуславливается наличием взаимосвязи между параметрами размытости и сглаживания, определяющими значения оптимизируемого функционала в (4).

Заключение

В настоящей работе методами статистического моделирования проведено исследование свойств оценки Берана на основе разработанного метода выбора оптимального параметра сглаживания. Показано, что на точность оценки Берана оказывают влияние объем выборки, число значений ковариаты, выбор вида весовой функции, способ оценивания параметра размытости, а также вид ядерных функций, используемых при оценивании параметра сглаживания и вычислении параметра размытости.

Полученные результаты позволяют сформулировать ряд рекомендаций по оцениванию условной функции надежности с помощью оценки Берана. На наш взгляд, наилучшие результаты могут быть достигнуты при использовании весовой функции Пристли – Чао с параметром размытости b_{NS} , вычисляемом на основе медианы абсолютных отклонений. Среди рассмотренных вариантов ядерных функций можно рекомендовать квартическое и ядро Епанечникова, поскольку они позволяют получить наиболее точную оценку Берана с точки зрения усредненного отклонения (9).

Следует отметить, что рассмотренные в данной работе методы не охватывают всего многообразия подходов к построению непараметрических оценок регрессионных моделей надежности. В частности, перспективным представляется привлечение техники бутстрепа и разработка адаптивных алгоритмов для выбора оптимальных значений параметра сглаживания.

ЛИТЕРАТУРА

1. *Beran R.* Nonparametric regression with randomly censored survival data // Technical report. Department of Statistics. Berkeley : University of California, 1981.
2. *Dabrowska D.M.* Nonparametric quantile regression with censored data // *Sankhya Ser. A.* 1992. V. 54. P. 252–259.
3. *Gonzalez M.W., Cadarso S.C.* Asymptotic properties of a generalized Kaplan-Meier estimator with some application // *J. Nonparametric Statistics.* 1994. No. 4. P. 65–78.
4. *McKeague I.W., Utikal K.J.* Inference for a nonlinear counting process regression model // *Ann. Statist.* 1990. V. 18. P. 1172–1187.
5. *Van Keilegom I., Akritas M.G., Veraverbeke N.* Estimation of the conditional distribution in regression with censored data: a comparative study // *Computational Statistics & Data Analysis.* 2001. V. 35. P. 487–500.
6. *Akritas M.G.* Nearest neighbor estimation of a bivariate distribution under random censoring // *Ann. Statist.* 1994. V. 22. P. 1299–1327.
7. *Расин Д.* Непараметрическая эконометрика: вводный курс // *Квантиль.* 2008. № 4. С. 7–26.
8. *Хардле В.* Прикладная непараметрическая регрессия. М. : Мир, 1993. 349 с.
9. *Gang L., Somnath D.* A bootstrap approach to nonparametric regression for right censored data // Technical report #99-8. 1999. P. 6–10.
10. *Демин В.А., Чимитова Е.В.* Выбор оптимального параметра сглаживания для непараметрической оценки регрессионной модели надежности // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика.* 2013. № 1(22). С. 59–65.
11. *Rousseeuw P.J., Verboven S.* Robust estimation in very small samples // *Journal Computational Statistics & Data Analysis.* 2002. V. 40(4). P. 741–758.
12. *Кошкин Г.А.* Основы страховой математики : учеб. пособие. Томск : Томский государственный университет, 2002. 116 с.

Демин Виктор Андреевич. E-mail: vicdemin@gmail.com

Чимитова Екатерина Владимировна, канд. техн. наук. E-mail: ekaterina.chimitova@gmail.com

Щеколдин Владислав Юрьевич. E-mail: raix@mail.ru

Новосибирский государственный технический университет

Поступила в редакцию 2 февраля 2014 г.

Demin Viktor A., Chimitova Ekaterina V., Schekoldin Vladislav Yu. (Novosibirsk State Technical University, Novosibirsk, Russian Federation).

The research for optimal choice method of bandwidth parameter in nonparametric estimation of reliability regression models.

Keywords: reliability function; regression model; nonparametric Beran estimator; bandwidth parameter; smoothing parameter; robust estimation.

In the paper, we consider one of the most popular nonparametric estimators of regression reliability models proposed by R. Beran. Such estimator allows to evaluate the conditional reliability function with the given values of covariates by the following formula:

$$\tilde{S}_{h_n}(t|x) = \prod_{Y_{(i)} \leq t} \left\{ 1 - \frac{W_n^i(x; h_n)}{1 - \sum_{j=1}^{i-1} W_n^j(x; h_n)} \right\},$$

where x is the value of covariate in the reliability function $S(t|x)$; $Y_{(i)}$ is the element of variational series; $W_n^i(x; h_n)$, $i = 1, \dots, n$, are the Nadaraya – Watson weights, i.e.,

$$W_n^i(x; h_n) = K\left(\frac{x - x_i}{h_n}\right) / \sum_{j=1}^n K\left(\frac{x - x_j}{h_n}\right).$$

It is well-known, the quality of the Beran estimator essentially depends on the chosen value of the bandwidth parameter h_n . In our previous paper, the method of selecting the optimal bandwidth parameter was proposed, which is based on the minimization of

the distance of failure times with kernel estimation for the inverse reliability function. Here, we consider the modification of this method by solving such optimization problem:

$$h_n^{\text{opt}} = \arg \min_{h_n} \sum_{i=1}^n |\hat{g}(\hat{p}_i | x_i) - Y_i|,$$

where

$$\hat{g}(\hat{p}_i | x_i) = \sum_{j=1}^n \omega_j(\hat{p}_i) \cdot Y_j.$$

The probabilities $\hat{p}_i$ are calculated by using the instrumentality of the Beran estimators, $\omega_j(\hat{p}_i)$ are certain weights which can be calculated with various weight functions.

We investigate the statistical properties of the Beran estimators by Monte Carlo simulations. It is shown that the accuracy of this estimators depend on the sample size, the number of covariates' values, the selection of the weight function's form, the method of smoothing parameter estimation and the type of kernel functions used in the smoothing parameter estimation and the bandwidth parameter calculation.

The obtained results allow us to formulate recommendations for estimating the conditional reliability function by the Beran estimator. In our opinion, the most appropriate results are achieved by the Priestley – Chao weight function

$$\omega_j^{(2)}(\hat{p}_i) = n(\hat{p}_{(i)} - \hat{p}_{(i-1)}) K\left(\frac{\hat{p}_i - \hat{p}_j}{b_{NS}}\right),$$

with the smoothing parameter

$$b_{NS} = \left[\frac{8\pi^{1/2} R(K)}{3\mu_2(K)^2 n} \right]^{1/5} \hat{\sigma},$$

where $\mu_2(K) = \int y^2 K(y) dy$, $R(K) = \int K^2(y) dy$. We recommend the robust standard deviation estimator based on the mixing method using the median absolute deviation and the Hodges – Lehmann estimator:

$$\hat{\sigma}_{\text{robust}} = 1.4826 \text{med}_{i=1 \dots n} \left| \hat{p}_i - \text{med}_{j=1 \dots n, k=j+1 \dots n} \left(\frac{\hat{p}_j + \hat{p}_k}{2} \right) \right|.$$

Also, we note that the quartic and Epanechnikov kernel functions lead to the most accurate Beran estimators.

REFERENCES

1. Beran R. *Nonparametric regression with randomly censored survival data. Technical report.* Department of Statistics, University of California, Berkeley. 1981.
2. Dabrowska D.M. Nonparametric quantile regression with censored data. *Sankhya Ser. A*, 1992, no. 54, pp. 252-259.
3. Gonzalez M.W., Cadarso S.C. Asymptotic properties of a generalized Kaplan-Meier estimator with some application. *Journal of Nonparametric Statistics*, 1994, no. 4, pp. 65-78. DOI: 10.1080/10485259408832601
4. McKeague I.W., Utikal K.J. Inference for a nonlinear counting process regression model. *Annals of Statistics*, 1990, no. 18, pp. 1172-1187.
5. Van Keilegom I., Akritas M.G., Veraverbeke N. Estimation of the conditional distribution in regression with censored data: a comparative study. *Computational Statistics & Data Analysis*, 2001, vol. 35, pp. 487-500. DOI: 10.1016/S0167-9473(00)00025-6
6. Akritas M.G. Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics*, 1994, no. 22, pp. 1299-1327.
7. Racine J.S. *Neparametricheskaya ekonometrika: vvodnyy kurs [Nonparametric econometrics: a primer]*. *Quantile*, 2008, no. 4, pp. 7-56.
8. Hardle W., Malyutov M.B. *Prikladnaya neparametricheskaya regressiya [Applied nonparametric regression]*. Moscow: Mir Publ., 1993. 349 p.
9. Gang L., Somnath D. A bootstrap approach to nonparametric regression for right censored data. *Technical report*, 1999, no. 99-8, pp. 6-10. DOI: 10.1023/A:1014644700806
10. Demin V.A., Chimitova E.V. Choice of optimal smoothing parameter for nonparametric estimation of regression reliability model. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*, 2013, vol.1(22), pp. 59-65. (In Russian).
11. Rousseeuw P.J., Verboven S. Robust estimation in very small samples. *Journal Computational Statistics & Data Analysis*, 2002, vol. 40(4), pp. 741-758. DOI: 10.1016/S0167-9473(02)00078-6
12. Koshkin G.M. *Osnovy strakhovoy matematiki [Basics of insurance mathematics]*. Tomsk: Tomsk State University Publ., 2002, p. 116.