На правах рукописи

Ярощук Марина Владимировна

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ОЦЕНИВАНИЕ РАСПРЕДЕЛЕНИЙ В ЗАВИСИМОСТИ ДОЗА-ЭФФЕКТ

05.13.18 – Математическое моделирование, численные методы и комплексы программ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата физико-математических наук

Нижний Новгород – 2011

Работа выполнена в Нижегородском государственном университете им. Н.И. Лобачевского (национальный исследовательский университет).

Научный руководитель: доктор физико-математических наук, профессор
Тихов Михаил Семенович

Официальные оппоненты: доктор физико-математических наук, профессор
Насыров Фарит Сагитович

доктор технических наук, профессор
Лапидус Вадим Аркадьевич

Ведущая организация: Российский Университет Дружбы Народов

Защита состоится «___»___________ 2011 г. на заседании диссертационного совета Д 212.166.13 при Нижегородском государственном университете им. Н.И. Лобачевского по адресу: 603950, г. Нижний Новгород, пр. Гагарина, 23, корпус 2, ауд. 209.

С диссертацией можно ознакомиться в фундаментальной библиотеке Нижегородского государственного университета им. Н.И. Лобачевского.

С текстом автореферата можно ознакомиться на официальном сайте Нижегородского государственного университета им. Н.И. Лобачевского http://www.unn.ru

Автореферат разослан «_____»____________ 2011 г.

Ученый секретарь диссертационного совета
кандидат физико-математических наук,
доцент В.П. Савельев

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Анализ связи между *дозой* и *эффектом* и их количественное определение является одной из актуальных проблем, возникающих при разработке новых лекарственных средств, т.е. веществ, обладающих фармакологической активностью, прошедших клинические испытания и предназначенных для изготовления лекарственных форм. Под *дозой* мы понимаем некоторое значение агента (фактора), которое может изменить состояние исследуемого объекта, а под *эффектом* – наблюдаемый качественный (альтернативный) отклик объекта на введенную дозу. Решение этой задачи представляет большой теоретический интерес и имеет обширные практические приложения во многих областях медицины и биологии.

Основу решения проблемы количественного оценивания связи между введенной дозой и наблюдаемым эффектом составляет *функция эффективности*, под которой понимается зависимость вероятности наблюдения эффекта от введенной дозы (*зависимость доза-эффект*). Задача оценивания функции эффективности по экспериментальным данным, а именно, введенной дозе и наличию или отсутствию эффекта, является главной задачей зависимости доза-эффект.

Функция эффективности имеет очень важное, а иногда и принципиальное значение в фармакологии – при оценке эффективности лекарственных препаратов, в токсикологии и радиологии – при исследовании количественной токсичности ядов и поражающих свойств ионизирующих излучений, в гигиене – при нормировании критических уровней вредных факторов. В экологии анализ зависимости доза-эффект дает возможность определить пределы устойчивости экосистемы.

Построение функции эффективности является статистической задачей, способ решения которой предъявляет соответствующие требования к планированию эксперимента и виду получаемых исходных данных. Биологический эксперимент на завершающем этапе требует методологически обоснованных точных статистических оценок результатов, учитывающих погрешности получения исходных данных и их влияние на конечные результаты. Поэтому с прикладной стороны актуальность темы диссертации не вызывает сомнений.

В некоторых случаях отсутствует понятие дозы в вышеопределенном смысле. Примером могут служить эконометрические модели с дискретными зависимыми переменными при случайной константе регрессионной зависимости[1], оценка инкубационного периода. Тем не менее, такие задачи

---

[1] Тихов, М.С. Эконометрические дискретные модели бинарного выбора / М.С. Тихов // Прикладная статистика в социально-экономических проблемах: материалы межд. конф. (под ред. Стронгиной Н.Р.) – Нижний Новгород: изд-во ННГУ – 2003. – Т.1. – С.104-106.

Tikhov, M.S. Econometric models with discrete dependent variables // 8-th Intern. Vilnius Con. on Probab. Theory and Math. Stat. Abstracts. – TEV. – Vilnius – 2002. – P. 325.

укладываются в рассмотренную в диссертации модель доза-эффект и могут быть решены предложенными в ней методами.

В диссертационной работе нас интересует проблема нахождения функции эффективности и оценка средних или летальных *эффективных доз*[2] $ED_{100\alpha}$ или $LD_{100\alpha}$, соответственно, в широком диапазоне значений $0<\alpha<1$, по результатам наблюдений. Мы строим математическую модель зависимости доза-эффект, где рассматриваем минимальную границу, с которой начинается реакция организма, как латентную случайную величину X. Если нижняя граница чувствительности X и введенная доза U независимы как случайные величины, то функция эффективности будет функцией распределения величины X, однако даже в этом случае для оценки функции эффективности и категорий эффективных доз мы не можем воспользоваться классическими методами математической статистики, поскольку исследуемая величина ненаблюдаема, а вместо нее наблюдаются менее информативные величины: введенные дозы U_i, и индикаторы эффекта $W_i = I(U_i > X_i)$, $i=1,2,\dots,n$. Для оценки функции эффективности мы используем непараметрические методы математической статистики, а именно, ядерные оценки регрессии по выборке $\left\{\left(U_i, W_i\right)\right\}_{i=1}^{n}$ объема n. Потребности практики обуславливают необходимость одновременного определения как эффективных доз от ED_5 до ED_{95}, так и вида самой функции эффективности.

Для построения кривой доза-эффект и оценки дозы ED_{50} обычно применяется официальная методика в модификациях Литчфилда-Вилкоксона[3] (J.T. Litchfield, F.W. Wilcoxon) и Д. Финни[4] (D.J. Finney), а также метод Спирмена–Кербера[5] (E. Spearman, T. Karber) и метод Рида-Менча[6] (L. Reed, H. Muench). Используются также методы сплайн-аппроксимации. Существенные недостатки традиционных методов определения средне-эффективных доз состоят в том, что они позволяют оценивать эффектив-

---

[2]Дозы, гарантирующие наступление эффекта с заданной вероятностью α (например, ED_{50} – средне-эффективная доза – для 50% объектов наблюдается эффект, LD_{90} – 90% летальная доза – 90% объектов, получивших дозу, погибает).

[3]Litchfield, J.T. A simplified method of evaluating dose–effect experiments/ J.T. Litchfield, F.W. Wilcoxon // J. Pharmacol. Exper. Ther. – 1949. – V. 96. – P. 99-113.

[4]Finney, D.J. Probit Analysis / D.J. Finney, 3 ed. – Cambridge: University Press, 1980. – 333 p.

[5] Закс, Л. Статистическое оценивание / Л. Закс. М.: Статистика, 1976. – 598 с

[6] Лакин, Г.Ф. Биометрия / Г.Ф. Лакин. М.: Высшая школа, 1990. – 352 с.

ные дозы в основном в окрестности ED_{50}. Методы сплайн-аппроксимации не позволяют строить доверительные интервалы.

Кроме указанных методов для оценивания зависимости доза-эффект и средне-эффективных доз ED_{50} (называемых еще медианными средне-эффективными дозами) используются модели бинарного выбора – пробит- и логит-модели, основанные на использовании нормальной или логистической функций распределения. Модели бинарного выбора хорошо работают в окрестности медианных средне-эффективных доз. Эти методы реализованы в большинстве современных эконометрических компьютерных программных пакетов (ЭКПП): SPSS, XL STAT–Dose, BioStat 2007, Probit Analysis, Stat-Plus (Статистика+). С помощью этих ЭКПП можно произвести обработку кривой зависимости доза-эффект, вычислить эффективную дозу, а также соответствующие доверительные интервалы, но в окрестности медианы.

Существуют различные модификации пробит- и логит-анализа, которые, имея в своей основе главную идею – преобразование процентов встречаемости эффекта в пробиты, – различаются алгоритмами линеаризации и статистической обработки. Большая часть этих программ основывается на алгоритме метода максимального правдоподобия для регрессионной схемы в модели бинарного выбора. Некоторые авторы (L.S. Miller & M.L. Tainter [7], J.T. Litchfield & F.W. Wilcoxon) используют для этой цели метод наименьших квадратов.

Таким образом, как при пробит-анализе, так и при использовании других моделей, функции распределения рассматриваемых случайных величин хорошо аппроксимируются линейными или нелинейными функциями, но только в окрестности медианы. В диссертации показано, что если истинное распределение отличается от нормального распределения $N(a,\sigma^2)$, то оценки «граничных» эффективных доз могут значительно отличаться от значений квантилей истинного распределения. Кроме того, при практической реализации пробит-анализа или его модификаций отсутствует возможность проведения единичных испытаний (согласно официальной методике испытания должны носить групповой характер), отсутствует возможность построения доверительных интервалов, существующие методы не учитывают погрешность измерений.

В работах М.С. Тихова (1993)[8] и М.С. Тихова, С.В. Криштопенко (1997)[9] был предложен непараметрический метод оценивания функции эффективно-

[7] Miller, L.S. Estimation of the ED50 and its error by means of logariphmic probit graph paper/ L.S. Miller, M.L. Tainter // Proc.Soc.Exp.Biol.Med. – 1944. – V.7. – P. 261-264.

[8]Тихов, М.С. Построение и анализ статистических оценок для неполностью известных семейств распределений: дис. д-ра физ.-мат. наук: 01.01.05/С.-Петербургский гос.ун-т. СПб. 1993. – 377 с.

[9]Криштопенко, С.В. Токсикометрия эффективных доз/ С.В. Криштопенко, М.С. Тихов – Нижний Новгород: Изд-во ННГУ, 1997. – 156 с.; Криштопенко, С.В. Парадоксальная токсичность/ С.В. Криштопенко, М.С. Тихов, Е.Б. Попова – Нижний Новгород: Изд-во НГМА, 2001. – 164 с.; Криштопенко, С.В. Доза–эффект / С.В. Криштопенко, М.С. Тихов, Е.Б. Попова – М.: Медицина, 2008. – 288 с.

сти, который задачу определения функции эффективности сводит к задаче оценивания функции регрессии и использования для этой цели непараметрических (ядерных) оценок регрессии с параметром сглаживания $h-$ ширины окна просмотра данных. Такой подход позволяет по результатам единичных испытаний оценивать средне-эффективную дозу ED_{50} не хуже, чем методы пробит-анализа, а малые и большие дозы, близкие, например, к ED_5 или к ED_{95}, оценивать эффективнее, чем с помощью пробит-анализа, строить доверительные интервалы, достаточно узкие как в середине, так и на краях распределения. С помощью метода, изложенного в этих работах, можно оценивать и немонотонные функции эффективности. Указанный подход позволяет также производить планирование эксперимента. Однако, рассмотренные в работах М.С. Тихова, С.В. Криштопенко оценки функции эффективности (в основном, оценки Надарая–Ватсона) хотя и сходятся к истинному распределению при $n \to \infty$ и $h \to 0, nh \to \infty$, не являются, во-первых, $\sqrt{nh}-$состоятельными (их предельное математическое ожидание отлично от нуля), а во-вторых, предельная дисперсия оценок для случайных доз U с плотностью распределения $g(x) > 0$ пропорциональна величине $\sigma^2(x) = \dfrac{F(x)(1-F(x))}{g(x)}$, где $F(x)-$неизвестная функция распределения минимальной границы, т.е. предельная дисперсия оценок зависит от плотности распределения вводимой дозы в заданной точке x: для $ED_{100\alpha}$ при малых или больших значениях α значение $g(ED_{100\alpha})$ мало, поэтому дисперсия $\sigma^2(x)$ может оказаться довольно большой. В диссертации для оценки функции распределения $F(x)$, вместо оценок Надарая–Ватсона мы используем $kNN-$оценки (оценки k ближайших соседей). Для $kNN-$оценок предельная дисперсия пропорциональна величине $F(x)(1-F(x))$ и не зависит от плотности $g(x)$, т.е. эти оценки равномерно сходятся к функции распределения, однако $kNN-$оценки не являются асимптотически несмещенными. Чтобы сделать их асимптотически несмещенными, мы конструируем двухшаговые оценки, у которых предельная дисперсия такая же, как и у $kNN-$оценок, но предельное математическое ожидание равно нулю.

Кроме того, математическую модель зависимости доза-эффект мы рассматриваем для случая прямых и непрямых наблюдений, т.е. когда вводимая в организм доза измеряется с некоторой ошибкой, а реакция организма (эффект) идет на «чистую» вводимую дозу. Рассмотрены случаи фиксированного плана (вводимая доза выбирается заранее и является неслучайной величиной) и случайного плана эксперимента (вводимая доза является случайной величиной). Таким образом, рассмотренные постановки охватывают возможные практические ситуации, встречающиеся в проблеме доза-эффект. К тому же измерение доз с ошибкой – это типичная практическая ситуация, которая ранее не рассматривалась, а ошибки измерения ранее не учитывались в оценках средне-эффективных доз и оценках среднеквадратичного отклонения.

При изучении вопросов, связанных с конкретным применением рассматриваемых процедур для конечных выборок, возникает проблема выбора оптимального значения параметра сглаживания h, который присутствует в рассматриваемых оценках функции эффективности. Как показывает практика, качество оценок в большей степени зависит от параметра сглаживания, нежели от вида ядерной функции, поэтому так важно выбирать оптимальное значение h. Автоматическому выбору оптимальной ширины окна в задачах оценивания регрессии посвящены работы W.Hardle, P. Hall and J.S. Marron, M. Neumann, R.Eubank and W.Schucany, J.Beran, Y.Feng and S. Heiler, H.-G. Muller and U. Stadtmuller, J.S. Wu and C.K. Chu, J.Rice and M.Rosenblatt, B.W. Silverman и др. В них предлагается использовать процедуру кросс-проверки (cross-validation), метод штрафных функций и метод подстановки. Однако исследования проводились для моделей, отличающихся от модели доза-эффект, и для случая прямых наблюдений. Мы строим комбинированный алгоритм метода подстановки и метода кросс-проверки. Показано, что в условиях непрямых наблюдений этот алгоритм приводит к состоятельным асимптотически нормальным оценкам оптимального значения параметра сглаживания. Указанный комбинированный подход приводит к меньшему риску оценивания, чем метод кросс-проверки или метод подстановки, ориентированные на выбор одного значения параметра h для широкого интервала значений переменной x. Предложенные в диссертации оценки исследованы как теоретическими, так и имитационными численными методами.

Таким образом, поставленные в диссертационном исследовании задачи характеризуются недостаточной научной проработанностью, актуальны и требуют своего решения в широком спектре практического анализа, поскольку они важны для практического применения полученных результатов и являются востребованными в токсикологии, фармакологии, биологии, экологии.

Цель и задачи работы. Целью данной работы является разработка и развитие методов построения $\sqrt{k}$ – состоятельных оценок функции распределения в зависимости доза-эффект, дисперсия которых равномерно мала при больших n, а также систематическое исследование асимптотических свойств построенных оценок. В рамках этого направления требовалось решить задачи непараметрического оценивания как теоретического, так и прикладного характера.

Задачами диссертации являются:

1. Построение оценок, дисперсия которых не зависит от распределения дозы, т.е. равномерно сходящихся к функции распределения.
2. Построение $\sqrt{k}$ – состоятельных оценок с учетом погрешности измерений доз и уменьшение этих погрешностей.
3. Разработка алгоритмов выбора ширины окна и исследование поведения оценок ширины окна просмотра данных.
4. Исследование свойств построенных оценок неизвестной функции распределения имитационными численными методами.

Объект и инструмент исследования. Объектом исследования являются математические конструкции, формализующие исходные объекты зависимости доза-эффект – статистические оценки функции распределения по неполным выборкам. Рассмотрены случайные и фиксированные планы эксперимента в схеме прямых и непрямых наблюдений. Исследования проводились с использованием методов теории вероятностей, математической статистики, теории функций, математического анализа и имитационного моделирования (метод Монте-Карло). Инструментом исследования являются асимптотические методы теории вероятностей и численные методы компьютерного моделирования.

Достоверность полученных в работе результатов обеспечивается строгостью рассуждений, использованием фундаментальных методов теории вероятностей и математической статистики, согласованностью теоретических выводов и численных результатов экспериментов, непротиворечивостью полученных результатов с ранее известными результатами.

Методы исследования. Математический аппарат построения оценок функции распределения по результатам независимых единичных испытаний разработан на основе метода ядерного сглаживания непараметрических оценок регрессии. Построение $kNN-$оценок основывается на использовании порядковых статистик в задаче непараметрического оценивания. Для доказательства состоятельности и асимптотической нормальности оценок использовался аппарат характеристических функций, классические предельные теоремы теории вероятностей и результаты асимптотической теории порядковых статистик.

Для получения оценок с автоматическим выбором оптимальной ширины окна в модели доза-эффект применялась комбинированная процедура кросс-проверки.

Научная новизна и научная значимость. Все основные результаты диссертационной работы являются новыми и впервые опубликованы в работах диссертанта.

Для модели доза-эффект построены непараметрические $\sqrt{k}-$ и $\sqrt{nh}-$состоятельные оценки функции распределения и проведен их асимптотический анализ для случайного и фиксированного планов эксперимента. Доказана асимптотическая нормальность и асимптотическая несмещенность двухшаговых оценок, рассмотрены методы уменьшения погрешности наблюдений для случая непрямых наблюдений. Для фиксированного и случайного планов эксперимента построены комбинированные кросс-проверочные оценки с автоматическим выбором ширины окна. Проведен численный анализ имитационной модели, адаптированный к практической ситуации.

Теоретическая и практическая ценность. *Теоретическая ценность* работы заключается в построении асимптотически несмещенных оценок и исследовании их свойств. *Практическая ценность* исследования состоит в том, что разработано программное обеспечение, которое может быть использовано для оценки уровня биологической активности веществ в опытах на токсичность на основании альтернативных кривых «доза-эффект», при планировании клинических испытаний новых лекарственных средств. Программа позволяет оценивать эффективные дозы уровней от 5% до 95%, в том числе и дозы, близкие к границам интервала распределения, строить оценки в случае, когда значения воздействовавшей дозы измеряются с погрешностью, получать оценки с автоматическим выбором ширины окна по непрямым наблюдениям.

Методы оценивания, разработанные в диссертации, были применены для построения функции распределения и определения категорий эффективных доз при исследовании чувствительности организма к адреналину.

Основные результаты, выносимые на защиту. На защиту выносятся:

1. $kNN-$оценки и двухшаговые $kNN-$оценки для неизвестной функции распределения в зависимости доза-эффект, результаты их теоретического асимптотического анализа в схеме прямых и непрямых наблюдений для модели со случайными и фиксированными планами эксперимента, а также численное исследование построенных оценок.
2. $\sqrt{k}-$ и $\sqrt{nh}-$состоятельные оценки и результаты их асимптотического анализа с учетом погрешности испытанных доз для модели с фиксированным планом эксперимента и способы уменьшения погрешности наблюдений.
3. Адаптивный алгоритм построения оценок с автоматическим выбором ширины окна просмотра данных для модели с фиксированными и случайными планами эксперимента, реализованный в разработанном диссертантом программном комплексе.
4. Численное сравнение оценок функции эффективности в моделях с прямыми и непрямыми наблюдениями.

Апробация работы. Результаты диссертации были представлены на следующих конференциях:

10-я междисциплинарная научная конференция «Нелинейный мир» (Нижний Новгород, 2005 г.);

XII Всероссийская школа-коллоквиум по стохастическим методам (Дагомыс, 2005 г.);

7-й Всероссийский симпозиум по прикладной и промышленной математике (ВСППМ), (Йошкар-Ола, 2006);

Международная конференция RelStat'2006 (Рига, 2006 г.);

XIV Всероссийская школа-коллоквиум по стохастическим методам (Адлер, 2007 г.);

XX Международная конференция «Математические методы в технике и технологиях» (Ярославль, 2007 г.);

8-я международная конференция «Компьютерный анализ данных и моделирование» CDAM'2007 (Минск, 2007г.);

X Всероссийский симпозиум по прикладной и промышленной математике (ВСППМ), (Дагомыс, 2009 г.).

Опубликованность результатов и личный вклад соискателя. По теме диссертации опубликовано 15 работ (общим объемом 9,5 п.л.), из них 6 в изданиях, рекомендованных ВАК Российской Федерации. Две работы, опубликованные в журналах из списка ВАК, написаны без соавторов. Всего без соавторов изданы 5 работ из общего числа объемом 6,1 п.л.

В совместных работах постановка задачи и основные методы исследования принадлежат научному руководителю. Вкладом соискателя являются формулировка и доказательства утверждений, теоретические расчеты, программная реализация и исследование имитационной модели.

Структура и объем диссертации. Диссертация состоит из введения, пяти глав, заключения, списка литературы, содержащего 128 наименований, приложений с численными результатами имитационного моделирования в виде таблиц и графиков. Объем основного текста работы составляет 157 страниц.

СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во **введении** анализируется состояние проблемы оценивания распределений и категорий эффективных доз по неполным данным в зависимости доза-эффект, отмечается, что тема диссертации является актуальной, формулируются цели и задачи исследования, дается обзор содержания глав.

В **главе 1** производится построение математической модели зависимости доза-эффект, используемой в дальнейшем в диссертации, и ставится задача оценивания неизвестной функции распределения для случайных и фиксированных планов эксперимента в случае прямых и непрямых наблюдений.

В параграфе 1.1 излагается математическая модель зависимости доза-эффект, основой которой является следующее представление: в организм вводится доза U; минимальный уровень дозы, с которого начинается реакция организма X, есть латентная переменная. Если $U > X$, то эффект от введенной дозы присутствует, в противном случае, если $U \le X$, то – отсутствует. Определяется величина $W-$ индикатор события $\{U > X\}$, т.е.

$$W = I\left(U > X\right) = \begin{cases} 1, & \text{если} \quad U > X, \\ 0, & \text{если} \quad U \le X. \end{cases}$$

Величина X может принимать различные значения даже при одинаковых условиях эксперимента, что объясняется индивидуальной чувствительностью организма к вводимому препарату, состоянием организма в целом на момент эксперимента. Однако, для однородных групп объектов наблюдения переменная X считается случайной величиной.

Таким образом, мы рассматриваем модель, в которой распределение с.в. X, заданное функцией распределения $F(x) = \mathbf{P}\left(X < x\right)$, неизвестно. При *случайном плане* эксперимента мы считаем, что $U -$случайная величина, а при *фиксированном* – значения вводимых неслучайных доз u_i выбираются заранее (с постоянным или переменным шагом).

Если измерения вводимой дозы U проводятся без ошибок, т.е. измеренная доза есть введенная доза, то мы будем говорить о *прямых наблюдениях*. В экспериментальной практике измерения вводимых доз проводятся, как правило, с погрешностями, иногда весьма значительными. Такие наблюдения мы будем называть *непрямыми*. Пусть измерения вводимой дозы U осуществляются с погрешностью ε, имеющей плотность $q(x)$, тогда вместо с.в. U наблюдается с.в. Y. Если ошибка ε накладывается аддитивно, т.е. $Y = U + \varepsilon$, то плотность условного распределения величины Y равна $q(y-u)$. В общем случае пара (Y, U) имеет совместную и маргинальные плотности распределения $g(y,u)$ и $q(y) = \int g(y,u)\,du$, $g(u) = \int g(y,u)\,dy$ соответственно, а распределение ошибки ε описывается условной плотностью $q(y \mid u)$.

Если случайные величины U и X независимы, то условное математическое ожидание величины W при фиксированном значении дозы U (то есть при $U = x$) будет функцией распределения величины X:

$$\mathbf{E}\big(W \mid U = x\big) = \mathbf{P}\big(W = 1 \mid U = x\big) = \mathbf{P}\,\big(X < U \mid U = x\big) = \mathbf{P}\,\big(X < x\big) = F(x).$$

В таком случае функция распределения здесь есть *регрессия*.

В общем случае условное математическое ожидание с.в. W при $U = x$, т.е. $\mathbf{E}\big(W \mid U = x\big) = \mathbf{P}\big(X < x \mid U = x\big) = F(x \mid x)$ уже может быть немонотонной функцией. Мы назовём $F(x \mid x)$ функцией эффективности. В нашей модели в основном рассматривается случай, когда U и X независимы, поэтому для оценивания функции распределения $F(x)$ используются непараметрические (ядерные) оценки регрессии по выборке $\left\{(U_i, W_i)\right\}_{i=1}^{n}$ (в схеме прямых наблюдений) или по выборке $\left\{(Y_i, W_i)\right\}_{i=1}^{n}$ (в схеме непрямых наблюдений).

В таком случае основная задача формулируется следующим образом: по выборке $\left\{(U_i, W_i)\right\}_{i=1}^{n}$ или по выборке $\left\{(Y_i, W_i)\right\}_{i=1}^{n}$ (в схеме прямых или непрямых наблюдений) построить $\sqrt{nh}$ – состоятельные оценки, дисперсия которых не зависит от плотности распределения дозы, т.е. построить оценки, равномерно сходящиеся к функции распределения $F(x)$ величины X.

Для зависимости доза-эффект в работах М.С. Тихова и С.В. Криштопенко[10] в качестве оценки функции распределения было предложено рассматривать ядерные оценки регрессии типа оценок Надарая–Ватсона, которые в схеме прямых наблюдений имеют вид:

$$F_{NW}(x) = S_{2n}(x) \big/ S_{1n}(x), \text{ если } S_{1n}(x) \neq 0, \text{ и } F_{NW}(x) = 0, \text{ если } S_{1n}(x) = 0, \quad (1)$$

---

[10] Криштопенко, С.В. Статистическое оценивание эффективной дозы зависимости «доза–эффект» с использованием как прямых, так и непрямых наблюдений / С.В. Криштопенко, М.С. Тихов // Вторая Всероссийская школа-коллоквиум по стохастическим методам. – М. – 1995. – С.81-82.

где

$$S_{1n}(x)=\frac{1}{n}\sum_{i=1}^{n}K_h\left(U_i-x\right),\ S_{2n}(x)=\frac{1}{n}\sum_{i=1}^{n}W_iK_h\left(U_i-x\right),\ K_h(x)=\frac{1}{h}K\left(\frac{x}{h}\right). \qquad (2)$$

Мы будем называть их *NW*-оценками. Здесь $K(x)-$ядерная функция.

В случае непрямых наблюдений мы используем *NW*-оценки вида:

$$\breve{F}_n(x)=\breve{S}_{2n}(x)\big/\breve{S}_{1n}(x)\ , \qquad (3)$$

где

$$\breve{S}_{1n}(x)=\frac{1}{n}\sum_{i=1}^{n}K_h\left(Y_i-x\right),\ \breve{S}_{2n}(x)=\frac{1}{n}\sum_{i=1}^{n}W_iK_h\left(Y_i-x\right). \qquad (4)$$

Асимптотическое поведение *NW*-оценок для зависимости доза-эффект изучено в работе М.С. Тихова[11]. В ней показано, что если $h=cn^{-1/5}$ и выполнены некоторые условия регулярности, то *NW*-оценки являются асимптотически нормальными:

$$n^{2/5}(F_{NW}(x)-F(x))\xrightarrow[n\to\infty]{d}N(\nu^2a_1(x),\sigma_1^2(x)), \qquad (5)$$

с асимптотическим смещением

$$a_1(x)=\frac{f'(x)g(x)+2g'(x)f(x)}{g(x)}, \qquad (6)$$

и асимптотической дисперсией

$$\sigma_1^2(x)=\frac{F(x)\left(1-F(x)\right)\|K\|^2}{g(x)}=\frac{\sigma_2^2(x)}{g(x)}, \qquad (7)$$

где

$$\nu^2=\int_{-\infty}^{\infty}x^2K(x)\,dx,\ \|K\|^2=\int_{-\infty}^{\infty}K^2(x)\,dx.$$

Таким образом, если $a_1(x)\neq 0$, то оценка $F_{NW}(x)$ имеет ненулевое асимптотическое смещение, то есть не является $\sqrt{nh}-$состоятельной, а предельная дисперсия оценки $\sigma_1^2(x)$ зависит от значения плотности $g(x)$ (в схеме прямых наблюдений, или $q(x)-$ в схеме непрямых наблюдений). Поэтому, если значение плотности близко к нулю, то предельная дисперсия оценки $F_{NW}(x)$ может оказаться довольно большой. Причина этого состоит в том, что интервал $(x-h,\,x+h)$ имеет фиксированную длину $2h$, и если в него попадает мало значений с.в. U, то оценка $F_{NW}(x)$ может иметь большой разброс.

В диссертации при построении оценок мы предлагаем брать границы интервала $(x-h,\,x+h)$ переменными, чтобы в него попало заданное количество наблюдений. Эти оценки мы получим из оценок Надарая–Ватсона, если возьмем ширину окна просмотра данных h так, чтобы в интервал

[11]Tikhov, M.S. Statistical Estimation on the Basis of Interval–Cencored Data/ M.S. Tikhov// J. Math. Sciences. – 2004. –V . **119**, N. 3. – P. 321 – 335.

$(x-h, x+h)$ попадало ровно k выборочных значений случайной величины U. Тогда параметр h будет случайной величиной. Мы выберем h специальным образом. Именно, пусть для заданного x: $F(x)=\lambda$ $(0<\lambda<1)$ и $\{U_n^{(i)}, 1\le i\le n\}$ есть вариационный ряд, построенный по выборке $U_1, U_2, ..., U_n$, а $\{W_n^{[i]}, 1\le i\le n\}$ есть индуцированные порядковые статистики, т.е. если $U_n^{(i)}=U_j$, то $W_n^{(i)}=W_j$, где $U_n^{(i)}$ есть $i-$ая порядковая статистика. Рассмотрим последовательность ранговых номеров $m=m(n)$, такую что $\frac{m}{n}=\lambda+o\left(\frac{1}{\sqrt{n}}\right)$ при $n\to\infty$. Возьмем $m_1=m-[k/2]$, $m_2=m+[k/2]$, где $[a]$ – целая часть a. Положим $\tilde{h}=U_n^{(m_2)}-U_n^{(m_1)}$.

Определим *kNN*-оценку соотношением

$$\tilde{F}_n(x)=\tilde{S}_{2n}(x)/\tilde{S}_{1n}(x), \tag{8}$$

где

$$\tilde{S}_{1n}(x)=\frac{1}{n}\sum_{i=1}^{n}K_{\tilde{h}}\left(U_n^{(i)}-x\right),\ \tilde{S}_{2n}(x)=\frac{1}{n}\sum_{i=1}^{n}W_n^{[i]}K_{\tilde{h}}\left(U_n^{(i)}-x\right). \tag{9}$$

Пусть теперь величины U_i являются неслучайными, т.е. пусть $u_1<u_2<...<u_n-$выборка с фиксированными упорядоченными значениями. Будем считать, что $a\le u_i\le b$, и, не умаляя общности, рассмотрим случай, когда $a=0, b=1$.

Положим в этом случае

$$S_{1n}(x)=\frac{1}{n}\sum_{i=1}^{n}K_h\left(u_i-x\right),\ S_{2n}(x)=\frac{1}{n}\sum_{i=1}^{n}W_iK_h\left(u_i-x\right), \tag{10}$$

где $W_i=I\left(u_i>X_i\right)-$индикатор события $\{u_i>X_i\}$.

Для неслучайного плана эксперимента с $u_i=i/n$ $NW-$оценку определим следующим образом:

$$F_n(x)=S_{2n}(x)/S_{1n}(x). \tag{11}$$

Для непостоянного шага $\Delta u_i=u_{i+1}-u_i$ в схеме прямых наблюдений в качестве оценки функции распределения наряду с *NW*-оценками мы используем оценки типа оценок Пристли–Чао (*PC*-оценки):

$$F_{PC}(x)=\sum_{i=1}^{n-1}\left(u_{i+1}-u_i\right)W_iK_h\left(u_i-x\right). \tag{12}$$

Для случайного плана эксперимента рассмотрим также *PC*-оценку

$$\hat{F}_{PC}(x)=\sum_{i=1}^{n-1}\left(U_n^{(i+1)}-U_n^{(i)}\right)W_n^{[i]}K_h\left(U_n^{(i)}-x\right), \tag{13}$$

где $U_n^{(i)}-i-$ая порядковая статистика, а $W_n^{[i]}-$ $i-$ая индуцированная порядковая статистика.

В диссертации мы изучаем поведение *NW*- и *PC*-оценок для случайных и неслучайных планов эксперимента в схеме прямых и непрямых наблюдений и доказываем их асимптотическую нормальность.

Для устранения асимптотического смещения *NW*- и *kNN*-оценок, т.е. для получения $\sqrt{nh}-$ и $\sqrt{k}-$состоятельных оценок, мы используем двухшаговую процедуру, которая для оценки плотности была предложена в работе N.W. Hengartner[12]. Мы модифицируем эту процедуру для оценивания регрессии, т.е. для оценивания отношения двух статистик – оценки функции распределения. Модифицированная процедура оценивания состоит в следующем.

1) Задаем $h_0 = C_1 n^{-\alpha}$, где $1/10 < \alpha < 1/5$, и вычисляем ядерные оценки для плотности $g(x)$ и произведения $\varphi(x) = F(x)g(x)$ соответственно в виде:

$$g^*(x) = \frac{1}{n}\sum_{j=1}^{n} K_{h_0}\left(U_j - x\right),\ \varphi^*(x) = \frac{1}{n}\sum_{j=1}^{n} W_j K_{h_0}\left(U_j - x\right). \tag{14}$$

2) Берем $h_1 = C_2 n^{-1/5}$ и оцениваем отношения $\alpha(x) = \dfrac{g(x)}{g^*(x)}$ и $\beta(x) = \dfrac{\varphi(x)}{\varphi^*(x)}$ с помощью статистик

$$\hat{\alpha}(x) = \frac{1}{n}\sum_{j=1}^{n} K_{h_1}(U_j - x)\frac{1}{g^*(U_j)} \text{ и } \hat{\beta}(x) = \frac{1}{n}\sum_{j=1}^{n} W_j K_{h_1}(U_j - x)\frac{1}{\varphi^*(U_j)}. \tag{15}$$

3) Умножаем оценку $\hat{\alpha}(x)$на оценку для плотности $g(x)$ на $g^*(x)$, получаем оценку знаменателя

$$\hat{g}(x) = \hat{\alpha}(x) g^*(x) = \frac{1}{n}\sum_{j=1}^{n} K_{h_1}(U_j - x)\frac{g^*(x)}{g^*(U_j)};$$

умножаем оценку $\hat{\beta}(x)$ на $\varphi^*(x)$, получаем оценку числителя

$$\hat{\varphi}(x) = \hat{\beta}(x)\ \varphi^*(x) = \frac{1}{n}\sum_{j=1}^{n} W_j K_{h_1}(U_j - x)\frac{\varphi^*(x)}{\varphi^*(U_j)}.$$

4) Вычисляем функции

$$V_{2n}^*(x) = \frac{1}{n}\sum_{j=1}^{n} W_j K_{h_1}\left(U_j - x\right)\cdot M_j(x) \text{ и } V_{1n}^*(x) = \frac{1}{n}\sum_{j=1}^{n} K_{h_1}\left(U_j - x\right)\cdot H_j(x) \tag{16}$$

где

$$H_j(x) = \sum_{i=1}^{n} K_{h_0}\left(U_i - x\right) \Big/ \sum_{i=1}^{n} K_{h_0}\left(U_i - U_j\right),$$

[12]Hengartner, N.W. Asymptotic unbiased density estimators / N.W. Hengartner // E. Matzner-Lober // ESAIM. – 2009. – V. 13. – P. 1-14.

$$M_j(x)=\sum_{i=1}^{n}W_iK_{h_0}\left(U_i-x\right)\Big/\sum_{i=1}^{n}W_iK_{h_0}\left(U_i-U_j\right). \tag{17}$$

5) Определяем финальную оценку $F_n^*(x)$ для $F(x)$ как отношение

$$F_n^*(x)=V_{2n}^*(x)\big/V_{1n}^*(x). \tag{18}$$

Доказано, что оценка $F_n^*(x)$ имеет нулевое асимптотическое смещение.

Для непрямых наблюдений $\left\{\left(Y_i, W_i\right)\right\}_{i=1}^{n}$ двухшаговая оценка функции $F(x)$ строится аналогичным образом с заменой U_i на Y_i.

В параграфе 1.3 формулируются основные предположения, при которых исследуется качество построенных оценок функции распределения.

Условия

Условие (К). Функция $K(x)$ является неотрицательной, нормированной, четной, ограниченной, финитной функцией.

Таким образом, в качестве ядер мы берем симметричные плотности. Если выполнены условия (К), то $\nu^2<\infty$, $\|K\|^2<\infty$.

Пусть $h=h(n)-$ ширина окна просмотра данных.

Условие (Н). При $n\to\infty$, $h=h(n)\to 0$ и $nh\to\infty$.

Асимптотически оптимальная ширина окна, для которой выполнены условие (Н) есть $h\sim n^{-1/5}$, при таком h асимптотическое смещение *NW*-оценки и её дисперсия уравновешены.

При построении $kNN-$оценки мы берём $k=[n^{4/5}]$.

При вычислении двухшаговой оценки функции распределения мы используем вспомогательную величину – ширину окна просмотра h_0, скорость сходимости к нулю у которой ниже, чем у основной величины h_1, а именно:

$h_0=C_1n^{-\alpha}$, где $\alpha\in\left(\frac{1}{10},\frac{1}{5}\right)$, $h_1=C_2n^{-1/5}$, C_1,C_2-некоторые константы.

Для выбранных таким образом h_0 и h_1 выполнены соотношения:

$$h_1\to 0,\ h_0\to 0,\ nh_1\to\infty,\ nh_0\to\infty,\ \frac{h_1}{h_0}\to 0 \text{ при } n\to\infty.$$

Для случайных планов эксперимента в схеме прямых наблюдений, где $\left(X_i,U_i\right)$, $i=1,2,...,n$ – независимые одинаково распределенные случайные величины с совместной непрерывной функцией распределения $F(x)G(u),(x,u)\in\mathbf{R}^2$ и плотностью $f(x)g(u)>0$ мы требуем выполнения следующих условий.

Условие (S). Функции $F(x)$ и $g(x)>0$ непрерывны, ограничены и имеет ограниченные производные до третьего порядка включительно.

В схеме непрямых наблюдений определим условные плотности распределения $q(y|u)=g(y,u)/g(u)$, и $g(u|y)=g(y,u)/q(y)$. Обозначим $m(x)=\int\limits_{-\infty}^{\infty}F(u)g(x,u)du$ и $R(x)=\int\limits_{-\infty}^{\infty}F(u)g(u\,|\,x)du=\dfrac{m(x)}{q(x)}$.

Условие (N). Функции $q(y)$ и $m(x)$ непрерывны и ограничены, имеют ограниченные производные до третьего порядка включительно.

Для фиксированных планов эксперимента, когда $U_i=u_i$ есть неслучайная величина, мы предполагаем, что $\max\limits_{i=1,\ldots,n}|\,u_i-u_{i-1}\,|=O\left(n^{-1}\right)$ при $n\to\infty$.

Пусть $V(f)=\sup\limits_{P}\sum\limits_{i=0}^{n-1}\left|f\left(\xi_{i+1}\right)-f\left(\xi_i\right)\right|-$ вариация функции $f(x)$, где $P-$множество всевозможных разбиений отрезка $[a,b]$ точками $a=\xi_0<\xi_1<\ldots<\xi_n=b$.

Условие (F). Функции $f(x)$ и $K(x)$ имеют конечные вариации, т.е. $V(f)<\infty,\ V(K)<\infty$.

В схеме непрямых наблюдений при фиксированных планах эксперимента потребуем, чтобы было выполнено

Условие (G). Функция $g(y)$ четырежды непрерывно дифференцируема и имеет ограниченные производные до четвертого порядка включительно, а функция $m(x)$ трижды непрерывно дифференцируема и имеет ограниченные производные до третьего порядка включительно.

В **главе 2** изучено асимптотическое поведение $kNN-$оценок в схеме прямых наблюдений для случайных планов эксперимента. Доказана состоятельность и асимптотическая нормальность $kNN-$оценок.

Приведем основные результаты главы 2.

Теорема 2.1.3. *Пусть выполнены условия* (K), (H), (S).

Тогда

$$\sqrt{k}(\tilde{F}_n(x)-F(x))\xrightarrow[n\to\infty]{d}N\left(\frac{a_1(x)\nu^2}{2g^2(x)},\sigma_2^2(x)\right).$$

Из теоремы 2.1.3 получаем, что построенные $kNN-$оценки имеют ненулевое асимптотическое смещение. В теореме 2.2.4 показано, что применение двухшаговой процедуры приводит к оценкам с такой же предельной дисперсией, что и у оценок $\breve{F}_n(x)$, но уже с нулевым асимптотическим смещением. Поскольку оценки $\breve{F}_n(x)$ являются асимптотически нормальными, то это позволяет построить для $F(x)$ доверительные границы.

Теорема 2.2.4. *Пусть выполняются условия* (K), (H), (S).

Тогда $\sqrt{k}(F_n^*(x)-F(x))$ *сходится по распределению в точке* x *к нормальной случайной величине с ожиданием, равным нулю, и дисперсией* $F(x)(1-F(x))\|K\|^2$*, т.е.*

$$\sqrt{k}(F_n^*(x)-F(x))\xrightarrow[n\to\infty]{d}N(0,\sigma_2^2(x)).$$

В теореме 2.3.1 доказана состоятельность и асимптотическая нормальность оценки квантиля функции распределения $F(x)$.

Пусть $x_\lambda=F^{-1}(\lambda)$ – квантиль порядка $0<\lambda<1$ функции распределения $F(x)$, где $f(x)>0$. В качестве оценки функции $F^{-1}(\lambda)$, рассмотрим статистику

$$F_n^{*-1}(\lambda)=\frac{1}{n}\int_{-\infty}^{\lambda}\sum_{i=1}^{n}K_{h_0}(\hat{F}_n(U_i)-u)du,$$

где

$$\hat{F}_n(x)=\frac{\sum_{i=1}^{n}W_iK_{h_1}(U_i-x)}{\sum_{i=1}^{n}K_{h_1}(U_i-x)},\ U_i\in R[0,1].$$

Теорема 2.3.1. *Пусть для всех* $F(0)<\lambda<F(1)$ *плотность* $f(x)$ *положительна и выполнены условия* (K), (H), (S).

Тогда

(i) если $\lim\limits_{n\to\infty}\frac{h_0}{h_1}=c$*, то* $\sqrt{nh_0}\,(F_n^{*-1}(\lambda)-b)\xrightarrow[n\to\infty]{d}N\left(0,\frac{\lambda(1-\lambda)\sigma_2^2(\lambda)}{g(x_\lambda)f(x_\lambda)}\right)$,

(ii) если $\lim\limits_{n\to\infty}\frac{h_0}{h_1}=0$*, то* $\sqrt{nh_0}\,(F_n^{*-1}(\lambda)-b)\xrightarrow[n\to\infty]{d}N\left(0,\frac{\lambda(1-\lambda)\|K\|^2}{g(x_\lambda)f^2(x_\lambda)}\right)$,

где $b=x_\lambda+\frac{1}{2}\nu^2h_1^2(F^{-1}(\lambda))''-\frac{1}{2}\nu^2h_0^2\frac{f'(x)g(x)+2f(x)g'(x)}{g(x)}$,

$\sigma_2^2(\lambda)=\iiint K\left(w+cf(x_\lambda)(v-u)\right)K(w)K(u)K(v)dw\,du\,dv$.

Из теоремы 2.3.1 следует, что уже при умеренных объемах выборки оценки эффективных доз близки к истинным значениям.

В **главе 3** диссертации изучается асимптотическое поведение оценок в зависимости доза-эффект при случайном плане эксперимента в схеме непрямых наблюдений. Здесь доказано, что предельное распределение *kNN*-оценок является нормальным. В параграфе 3.2 показано, что двухшаговая процедура, в случае непрямых наблюдений также приводит к $\sqrt{nh}$ – состоятельным оценкам и что двухшаговые оценки являются асимптотически нормальными с той же предельной дисперсией, что и у *kNN*-оценок.

Приведем один из основных результатов этой главы.

Теорема 3.4.1. *Пусть выполнены предположения* (K), (H), (N). *Тогда*

$$\sqrt{k}(\hat{F}_n(x) - R(x)) \xrightarrow[n\to\infty]{d} N(0, R(x)(1-R(x))\|K\|^2),$$

где $R(x) = \int_{-\infty}^{\infty} F(u)g(u\,|\,x)du.$

Из этой теоремы следует, что в случае непрямых наблюдений оценка $\hat{F}_n(x)$ сходится к «средней» функции распределения $R(x)$.

В **главе 4** исследуется поведение оценок функции распределения $F(x)$ для фиксированных планов эксперимента. Это связано с тем, что модели с фиксированным планом эксперимента при наличии погрешности измерения в основном отражают применяемую методику проведения клинических испытаний лекарственных препаратов. Изучены асимптотические свойства оценок Надарая–Ватсона $F_n(x)$ (11) в случае постоянного шага и оценок Пристли–Чао $F_{PC}(x)$ (12) – в случае переменного шага. В схеме прямых и непрямых наблюдений строятся двухшаговые оценки, которые являются $\sqrt{nh}$ – состоятельными оценками. В параграфе 4.4 предлагаются способы уменьшения погрешности наблюдений в случаях, когда распределение ошибки ε нормально. Показано, что оценки по улучшенным выборкам сходятся к истинной функции распределения $F(x)$, а не к усредненной функции распределения $R(x)$. В пункте 4.6 изучены оценки Пристли–Чао (12), и найдено их асимптотическое распределение. В теореме 4.6.2 для фиксированных планов при непостоянном шаге доказана состоятельность и асимптотическая нормальность *PC*-оценок.

В **главе 5** изучение модели доза-эффект производится средствами имитационного моделирования. Программная реализация имитационной модели “Dose-Effect” выполнена на основе Borland Turbo Delphi Explorer. В параграфе 5.1. решается проблема выбора оптимального значения ширины окна просмотра h численными методами.

Показано, что если выбор оптимального значения параметра h производится из дискретного множества Q_n и мощность этого множества $|Q_n|$ не превосходит некоторой степени n, то кросс-проверочная оценка параметра h является состоятельной оценкой ширины окна, полученного при минимизации квадратичного уклонения. Именно, пусть для каждого значения параметра $h \in Q_n$ статистики $F_n^{(h)}(x)$ и $F_{nj}^{(h)}(x)$ есть оценки функции распределения $F(x)$ по выборкам $U^{(n)} = \{(U_1, W_1), \ldots, (U_n, W_n)\}$ и $U_{-j}^{(n-1)} = U^{(n)} \setminus \{(U_j, W_j)\}$ соответственно.

Кросс-проверочное значение параметра h определим из равенства

$$H = H_n = \arg\min_{h \in Q_n} n^{-1} \sum_{j=1}^{n} \left(F_{nj}^{(h)}(U_j) - W_j \right)^2 ,$$

а кросс-проверочную оценку функции распределения получим, полагая

$$F_n^{CV}(x) \equiv F_n^{(H)}(x) .$$

В теореме 5.1.4 показано, что оценка кросс-проверочного значения H является состоятельной оценкой параметра h и при $n \to \infty$ квадратичное уклонение при CV – значении параметра h ведет себя также, как и оптимальное значение $\bar{h}$.

Алгоритм выбора оптимального параметра сглаживания h:

1. По результатам наблюдений $U_{-j}^{(n-1)}$ вычисляем оценку:

$$F_{nj}^{(h)} = \frac{1}{n-1} \sum_{\substack{i=1, \\ i \neq j}}^{n} W_i K_h \left(U_i - x \right) .$$

2. Строим функцию кросс-проверки:

$$CV(h) = \frac{1}{n} \sum_{j=1}^{n} \left(F_{nj}^{(h)}(U_j) - W_j \right)^2 .$$

3. Определяем оптимальное значение параметра h:

$$H = \arg\min_{h} \left(CV(h) \right) .$$

В теореме 5.1.4 показано, что данный алгоритм приводит к асимптотически оптимальным значениям параметра сглаживания.

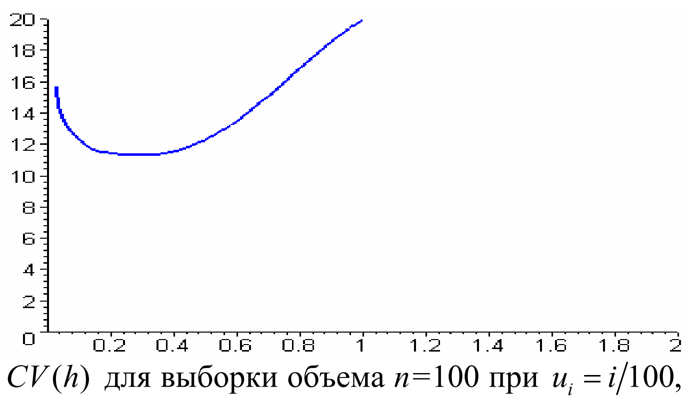
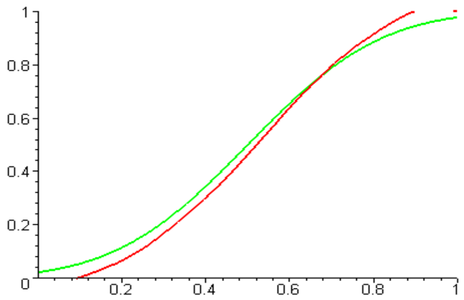


Рис.1. График кривой $CV(h)$ для выборки объема n=100 при $u_i = i/100,\ i = 1,...,100$, с.в. $X \in N(0.5,\ 0.25)$.

В параграфе 5.2 исследование модели при фиксированном плане эксперимента проведено на примерах реальных данных, взятых из литературы, а случайного плана – в параграфе 5.3, на модельных данных. Численные расчеты показали, что двухшаговые оценки лучше оценивают неизвестную функцию распределения $F(x)$.



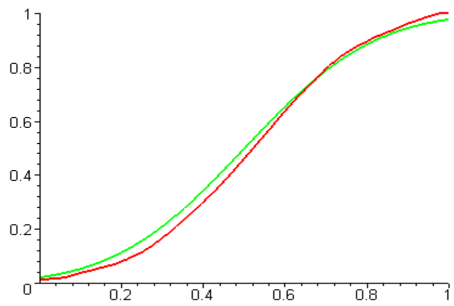


Рис.2. График оценки Надарая-Ватсона и интегральной функции нормального распределения $N(0.5,\ 0.25)$ для выборки объема n=100 при $u_i = i/100,\ i = 1,...,100$, с.в. $X \in N(0.5,\ 0.25)$ кросс-проверочное значение $h = 0.3$.

Рис.3. График *двухшаговой* оценки Надарая-Ватсона и интегральной функции нормального распределения $N(0.5,\ 0.25)$ для выборки объема n=100 при $u_i = i/100,\ i = 1,...,100$,
с.в. $X \in N(0.5,\ 0.25)$, кросс-проверочное значение $h = 0.3$.

В параграфе 5.3 произведен сравнительный анализ нескольких способов получения *kNN*-оценок в программной реализации. Изучено поведение оценок при использовании финитных и нефинитных ядер, в результате чего установлено, что применение нефинитных ядер приводит к оценкам с существенным смещением по сравнению с использованием финитных ядер, причем использование некоторых финитных ядер, например, квартического[13], позволяет уменьшить смещение (в сравнении, например, с ядром Епанечникова[14]). Показано, что при подборе подходящих значений ширины окна для модельных данных, имеющих нормальное или логистическое распределение, оценки и доверительные интервалы получаются не хуже, чем доверительные интервалы, полученные с помощью пакетов прикладных программ SPSS и Probit Analysis, а для распределения Вейбулла они получаются лучше. При этом преимущество оценок, построенных в диссертационной работе, состоит в том, что их можно получать как по результатам единичных независимых испытаний, так и по данным, сгруппированным в однородные группы, тогда как для применения классических методов требуется только наличие однородных групп, причем на каждом уровне требуется испытать не менее 6 доз. Если же рассматриваемое распределение не является нормальным или логистическим, например, рассматривается распределение Коши – распределение с тяжелыми хвостами, то с помощью пробит- и логит-моделей получаются оценки, менее точные по сравнению с нашей моделью, даже для доз, не слишком далеких от 50% (35%, 60%), а 5% и 95% уровни доз оцениваются уже со значительной погрешностью.

---

[13] $K_1(x) = \frac{15}{16}\left(1 - x^2\right)^2 I\left(|x| \le 1\right)$

[14] $K_0(x) = \frac{3}{4}\left(1 - x^2\right) I\left(|x| \le 1\right)$

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИИ

В диссертационной работе проведен статистический и численный анализ различных оценок для функции распределения в зависимости доза-эффект для случайного и фиксированного плана эксперимента как в случае прямых, так и непрямых наблюдений.

Построены *kNN*-оценки и двухшаговые асимптотически несмещенные *kNN*-оценки неизвестной функции распределения при случайных планах эксперимента. Установлено, что построенные оценки равномерно сходятся к функции распределения. Построены $\sqrt{nh}$ – состоятельные оценки для модели с фиксированным планом эксперимента в схеме непрямых наблюдений. Показана их асимптотическая нормальность.

Построены $\sqrt{nh}$ – состоятельные оценки неизвестной функции распределения с учетом погрешности измерений. Предложен метод уменьшения погрешности измерений.

Разработан комбинированный адаптивный алгоритм автоматического выбора оптимальной ширины окна просмотра данных.

Проведен численный анализ построенных оценок. В результате исследования численными методами делается вывод, что предложенные в диссертации оценки функции распределения $F(x)$ более точно оценивают «малые» или «большие» эффективные дозы по сравнению с существующими методами.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах из списка ВАК Российской Федерации:

1. Ярощук, М.В. Статистическое оценивание зависимости доза-эффект с помощью *kNN* – оценок / М.В. Ярощук, М.С. Тихов // Обозрение Прикл. и Пром. Математики – М.: Изд-во ТВП – 2005. – т . 12, в. 3, С. 683-684.
2. Ярощук, М.В. Модифицированные оценки Янга в зависимости доза-эффект / М.В. Ярощук, М.С. Тихов // Обозрение Прикл. и Пром. Математики – М.: Изд-во ТВП – 2006. – т. 13, в. 1, С. 144.
3. Ярощук, М.В. Асимптотическая нормальность *kNN* – оценок в зависимости доза-эффект / М.В. Ярощук, М.С. Тихов // Вестник Нижегородского университета. Серия: Математика. – 2006. Вып.1(4), С. 129-137.
4. Ярощук, М.В. Об оценивании распределений в зависимости доза-эффект / М.В. Ярощук // Обозрение Прикл. и Пром. Математики – М.: Изд-во ТВП – 2007. – т. 14, в. 1, С.178-180.
5. Ярощук, М.В. Статистическое оценивание распределений по интервально цензурированным выборкам в схеме непрямых наблюдений / М.В. Ярощук, М.С. Тихов // Нелинейный мир – М.: изд-во Радиотехника – 2007. – т.5, №1,2, С. 4-8.
6. Ярощук, М.В. Имитационное моделирование зависимости доза-эффект и статистический анализ оценок функции эффективности / М.В. Ярощук // Обозрение Прикл. и Пром. Математики – М.: Изд-во ТВП – 2009. – т. 16, в. 6, С.1148-1150.

Публикации в иных печатных изданиях:

7. Ярощук, М.В. Статистическое оценивание нелинейной зависимости доза-эффект с помощью *kNN* – оценок / М.В. Ярощук, М.С. Тихов // Десятая междисциплинарная научная конференция «Нелинейный мир» – Нижний Новгород: изд-во ННГУ – 2005. – С. 139.
8. Ярощук, М.В. Оценивание распределений в зависимости доза-эффект при фиксированном плане эксперимента / М.В. Ярощук, М.С. Тихов, Д.С. Криштопенко // Статистические методы оценивания и проверки гипотез: межвуз. сб. научных трудов – Пермь: Изд-во Пермского ун-та. – 2006. – С. 66-77.
9. Yarochuk, M.V. Asymptotic normality *k*-Nearest Neighbor estimators in dependence dose-response / M.V.Yarochuk, M.S. Tikhov // Ташкент.: Изд-во Ташкентского ун-та. – 2006. – С. 46-49.
10. Ярощук, М.В. Об асимптотической нормальности *kNN* – оценок в зависимости доза-эффект / М.В. Ярощук. Нижегородский гос. ун-т – Н.Н. –2006. –49 с. – Деп. в ВИНИТИ 1536 – В2006
11. Ярощук, М.В. Предельные распределения *kNN* – оценок в зависимости доза-эффект в схеме непрямых наблюдений / М.В. Ярощук. Нижегородский гос. ун-т – Н.Н. – 2006.– 39 с. – Деп. в ВИНИТИ 1535– В2006
12. Yarochuk, M.V. Statistical estimation of distributions in dose-response dependence / M.V. Yarochuk, M.S. Tikhov // Reliability and Statistics in Transport and Communication: Proc. of the 6th Intern. Conf., Riga, Latvia. – 2006.– P.374-375.
13. Ярощук, М.В. Исследование оценок функции распределения в зависимости доза-эффект / М.В. Ярощук, М.С. Тихов, Д.С. Криштопенко // Математические методы в технике и технологиях – ММТТ-20: сб.трудов XX Междунар. науч. конф., – Ярославль: Изд-во Ярославского гос. техн. ун-та. – 2007. – т.3, С. 171-175.
14. Yarochuk, M.V. Computer-based data analysis in dose-response dependence / M.V. Yarochuk // Computer Data Analysis and Modeling: Proc. of 8th Intern. Conf., Minsk, September 11-15. – 2007. – Vol. 1, P. 186-189.
15. Yarochuk, M.V. Asymptotic normality of the integrated square error at the fixed plan of experiment for indirect observations / M.V.Yarochuk, M.S. Tikhov, D.S. Krishtopenko // Computer Modeling and New Technologies. – 2007. – Vol. 11, No 1, Riga, Latvia, P. 46-56.

Подписано в печать 13.01.11. Формат 64×80 1/16.
Бумага офсетная. Печать офсетная.
Усл. печ. л. 1. Зак. 28. Тир. 100 экз.

Типография Нижегородского госуниверситета
им. Н.И. Лобачевского.
603000, Н. Новгород, ул. Б. Покровская, 37.