САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

На правах рукописи

КОРОБЕЙНИКОВ Антон Иванович

Исследование специальных моделей кривых дожития в условиях неполных данных

05.13.18 – Математическое моделирование,
численные методы и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Санкт-Петербург – 2010

Работа выполнена на кафедре статистического моделирования математико-механического факультета Санкт-Петербургского государственного университета.

Научный руководитель:	доктор физико-математических наук, профессор ЕРМАКОВ Сергей Михайлович
Официальные оппоненты:	доктор физико-математических наук, профессор НЕВЗОРОВ Валерий Борисович (Санкт-Петербургский государственный университет)
	кандидат физико-математических наук ГОРМИН Анатолий Андреевич, инженер (Mirantis, Inc.)
Ведущая организация:	Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

Защита состоится «____» ____________ 2010 г. в ____ часов на заседании совета Д 212.232.51 по защите докторских и кандидатских диссертаций при Санкт-Петербургском государственном университете, расположенном по адресу: 198504, Санкт-Петербург, Петергоф, Университетский пр., 28.

С диссертацией можно ознакомиться в Научной библиотеке им. М. Горького Санкт-Петербургского государственного университета, расположенной по адресу: 199034, Санкт-Петербург, Университетская наб., 7/9.

Автореферат разослан «____» ____________ 2010 г.

Ученый секретарь диссертационного совета,

доктор физ.-мат. наук, проф. Даугавет И.К.

Общая характеристика работы

Актуальность темы. Анализ данных типа времени жизни является одной из активно развивающихся областей современной прикладной статистики. Данные такого вида возникают не только в медико-биологических задачах, но и во многих других областях, где требуется изучение «времени до наступления некоторого события», например, при анализе демографических, экономических, финансовых, эпидемиологических показателей, в теории надежности и социологии.

Несмотря на бурное развитие в последние годы непараметрических и семипараметрических методов, специальные (параметрические) модели кривых дожития по-прежнему являются основным инструментом для исследования данных типа времени жизни. Этот факт обусловлен, в первую очередь, высокой информативностью параметрических моделей и возможностью их интерпретации экспериментаторами.

В силу ряда объективных причин для анализа данных типа времени жизни требуются специальные статистические методы. Одной из таких причин, отделяющих эту область от других областей прикладной статистики, является наличие так называемого *цензурирования*: в процессе сбора данных вместо интересуемой случайной величины наблюдается другая, менее информативная. Таким образом, при анализе данных типа времени жизни имеет место проблема неполной информации о выборке. Механизмы цензурирования могут быть достаточно сложными и, вследствие этого, требовать отдельного подхода. Стандартные методы анализа данных, как правило, просто не могут быть адекватно применены к случаю цензурирования.

Задача оценивания параметров и выбора специальной (параметрической) модели кривых дожития рассматривалась в работах многих авторов [2, 5, 8, 13]. Как правило, большинство полученных результатов предполагают наличие так называемого случайного правого цензурирования. Однако, такая модель не очень часто встречается при анализе реальных данных [4], и может рассматриваться только как достаточно простая аппроксимация; вопрос адекватности полученных в таком предположении результатов остается открытым. Вместо этой модели в приложениях более подходящей представляется модель интервального цензурирования [11], специальным случаем которой является вышеупомянутое случайное правое цензурирование.

В связи со сложностью этой модели и ее специальной структурой необходимо развитие специальных методов для оценивания параметров и выбора адекватных параметрических моделей (теория для случайного правого цензурирования не применима здесь по крайней мере без серьезной доработки).

Настоящая работа призвана в известной степени заполнить обозначенный пробел и посвящена построению оценок параметров специальных моделей кривых дожития, исследованию асимптотических свойств полученных оценок, а также развитию методологии сравнения различных параметрических моделей в условиях интервального цензурирования.

Целью работы являются:

1. построение оценок параметров для специальных моделей кривых дожития в условиях интервального цензурирования и изучение асимптотических свойств полученных оценок;

2. разработка методологии выбора адекватной параметрической модели посредством модификации информационных критериев типа Акайке на случай интервального цензурирования;

3. разработка численных методов и систем программ, позволяющих производить оценивание параметров в указанных моделях.

Общая методика работы. В работе применяются методы статистического моделирования, теории вероятностей и математической статистики (оценки максимального правдоподобия, законы больших чисел и центральные предельные теоремы, теория эмпирических процессов), функционального анализа (теория Фредгольмовых операторов), линейной алгебры. Программирование осуществлялось в статистическом пакете `R`.

Научная новизна. В данной работе впервые получены достаточные условия строгой состоятельности оценок типа максимального правдоподобия в условиях интервального цензурирования и исследованы их асимптотические свойства. Помимо этого, были предложены робастные оценки параметров в случае, когда предполагаемая параметрическая модель не точна. Показано, что эти оценки являются оптимальными с точки зрения расстояния Кульбака-Лейблера между предполагаемой параметрической моделью и истинным распределением данных. При помощи построенных оценок информационные критерии типа Акайке выбора адекватной параметрической модели впервые были распространены на случай интервального цензурирования.

Теоретическая и практическая ценность. В работе математически обоснована применимость (строгая состоятельность, асимптотическое распределение) двух классов оценок к анализу данных в случае интервального цензурирования. Созданы программы, в которых эффективно реализованы разработанные методы оценивания параметров. Методология информационных критериев может быть успешно использована экспериментаторами для подбора адекватной параметрической модели данных.

Апробация работы. Основные результаты диссертации докладывались и обсуждались на семинаре кафедры статистического моделирования математико-механического факультета СПбГУ, а также на конференциях:

- II Всероссийская научно-практическая конференция с международным участием «Высокотехнологичные методы диагностики и лечения заболеваний сердца, крови и эндокринных органов», Федеральный центр сердца им. В.А. Алмазова, г. Санкт-Петербург, 20 – 22 Мая 2008 г.

- 18th Population Approach Group in the Europe (PAGE) Meeting, Saint Petersburg, 23 – 26 June, 2009.

- 6th Saint Petersburg Workshop on Simulation, Saint Petersburg, June 28 – July 4, 2009.

Публикации. По теме диссертации опубликованы работы [A1, A2, A3, A4]. Статья [A1] опубликована в журнале, входящем в перечень ВАК по специальности 05.13.18. Статья [A2] написана в соавторстве, в ней автору принадлежит доказательство теоремы об асимптотических свойствах оценок параметров специальной модели кривой дожития.

Структура и объем диссертации. Диссертация состоит из введения, 5 глав, заключения, списка литературы и 2 глав приложения. Библиография содержит 87 наименований. Общий объем работы 144 страницы.

Содержание работы

Через X обозначена случайная величина с неотрицательным носителем. В приложениях реализациям случайных величин подобного рода часто придают смысл времени до наступления какого-либо события. Само событие обычно называется *отказом*, а случайная величина X — временем отказа. Мы предполагаем, что распределение X описывается параметрической моделью с функцией распределения F_ϑ, где параметр ϑ принадлежит некоторому метрическому пространству Θ.

На практике анализ данных типа «времени жизни» сопряжен с определенным трудностями. Как правило, значение случайной величины X известно лишь с точностью до некоторого интервала, которому она принадлежит. Например, пусть X — время до возникновения рецидива некоторого заболевания. Ввиду невозможности на практике осуществить непрерывный контроль состояния пациента, определить факт возникновения рецидива можно только лишь в определенные моменты наблюдения (тем самым момент возникновения рецидива наблюдается с точностью до промежутка между отдельным наблюдениями за состоянием).

Введем модель *цензурирования*, задающую наблюдаемую величину Y. Пусть K — положительная целочисленная случайная величина. Через T обозначен набор случайных величин $\{T_{k,j}, j = 0 \dots k+1, k = 1 \dots, +\infty\}$, таких, что $0 = T_{k,0} < T_{k,1} < T_{k,2} < \dots < T_{k,k} < T_{k,k+1} = +\infty$. Отметим, что вообще говоря, случайные величины X и (K, T) могут быть зависимыми. Определим случайную величину $Y = (\Delta_K, T_K, K)$, где T_k — k-я строка треугольного массива T, $\Delta_k = (\Delta_{k,1}(X), \dots, \Delta_{k,k+1}(X))$ и $\Delta_{k,j}(X) = 1_{(T_{k,j-1}, T_{k,j}]}(X)$. Таким образом, Y описывает разбиение вещественной полуоси $[0, +\infty)$ на $K+1$ (случайный) подинтервал и определяет интервал, содержащий X.

Описанная модель цензурирования известна как *модель интервального цензурирования смешанного типа* [10] и характерна для реальных задачах.

Рассмотрим модель повторных наблюдений: пусть $X_1, \dots, X_n$ — набор независимых одинаково распределенных случайных величин с одинаковой функцией распределения F_ϑ. Наблюдаемые цензурированные случайные величины обозначены через $Y_1, \dots, Y_n$ с $Y_i = \left(\Delta_{K^{(i)}}(X_i), T_{K^{(i)}}, K^{(i)}\right)$. В диссертации рассматривается задача построения оценок параметра ϑ по выборке $Y_1, \dots, Y_n$ в случае неизвестного механизма цензурирования (K, T).

Содержание по главам

Во **введении** обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана практическая значимость полученных результатов, представлены выносимые на защиту научные положения и делается обзор существующих результатов.

В **первой главе** изучаются свойства оценок типа максимального правдоподобия. В первом параграфе вводится модель интервального цензурирования смешанного типа. Во втором параграфе выполняется построение оценок типа максимального правдоподобия. Через Q обозначено распределение с.в. Y и через $\mathbb{Q}_n$ — эмпирическое распределение с.в. $Y_1, \ldots, Y_n$. (Частная) логарифмическая функцию правдоподобия для θ вводится как:

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K^{(i)}+1} \Delta_{K,j}^{(i)} \log \left[F_\theta(T_{K,j}^{(i)}) - F_\theta(T_{K,j-1}^{(i)}) \right] = \int m_\theta \, d\mathbb{Q}_n, \qquad (1)$$

где функция m_θ определяется следующим образом:

$$m_\theta(\delta_k, t_k, k) = \sum_{j=1}^{k+1} \delta_{k,j} \log \left[F_\theta(t_{k,j}) - F_\theta(t_{k,j-1}) \right].$$

Отметим, что когда X и (K, T) независимы, имеет место т.н. *неинформативное цензурирование*. В таком случае m_θ является логарифмом плотности величины Y относительно совместного распределения (K, T_K) и функция l_n в (1) является логарифмом полной функции правдоподобия для параметра θ. Всюду далее мы будем опускать определение «частная» для функции l_n.

Оценкой максимального правдоподобия $\hat{\theta}_n$ назовем такое значение параметра θ, что

$$\int m_{\hat{\theta}_n} \, d\mathbb{Q}_n = \sup_{\theta \in \Theta} \int m_\theta \, d\mathbb{Q}_n. \qquad (2)$$

В некоторых случаях супремум в (2) может не достигаться, но позволяет сколь угодно точное приближение последовательностью оценок $\hat{\theta}_n$. В связи с этим расширим класс оценок максимального правдоподобия, включив такие последовательности $\hat\theta_n$, для которых выполняется

$$\sup_{\theta\in\Theta}\int m_\theta\, d\mathbb{Q}_n - \int m_{\hat\theta_n}\, d\mathbb{Q}_n \xrightarrow{\mathbf{P}} 0, \quad n\to\infty.$$

В литературе оценки такого характера носят название приближенных оценок максимального правдоподобия [7]. Свойства именно таких оценок исследуются в первой главе диссертации.

В третьем параграфе исследуется состоятельность оценок $\hat\theta_n$. Введем множество Θ_0 точек максимума предельной функции правдоподобия:

$$\Theta_0 = \left\{\theta^*\in\Theta : \int m_{\theta^*}\, dQ = \sup_\theta \int m_\theta\, dQ\right\}.$$

В диссертации доказана следующая теорема о сходимости оценок $\hat\theta_n$.

Теорема 1. *Пусть $\mathbf{E}(K)<\infty$ и $F_\theta(x)$ непрерывна по θ для почти всех x. Предположим, что для любого достаточно малого шара $B\subset\Theta$ и почти всех $x<y$ функция $(x,y)\mapsto \sup_{\theta\in B}\log\left[F_\theta(y)-F_\theta(x)\right]$ измерима. Тогда для любого $\varepsilon>0$ и компакта $K\subset\Theta$ выполняется*

$$\mathbf{P}\left(\operatorname{dist}\left(\hat\theta_n,\Theta_0\right)\geqslant\varepsilon,\, \hat\theta_n\in K\right)\to 0, \quad n\to\infty.$$

Достаточные условия идентифицируемости модели (т.е. условия того, что $\Theta_0=\{\vartheta\}$) даны следующей теоремой

Теорема 2. *Введем меру μ на Борелевской σ-алгебре $\mathcal{B}$ подмножеств $\mathbb{R}$:*

$$\mu(B)=\sum_{k=1}^{+\infty}\mathbf{P}(K=k)\sum_{j=1}^{k}\mathbf{P}\left(T_{k,j}\in B\,|\,K=k\right), \quad B\in\mathcal{B}.$$

Пусть для функции распределения $F_\theta(x)$ и для любых $\theta_1,\theta_2\in\Theta$ выполняется

$$F_{\theta_1}=F_{\theta_2}\ \mu - \text{п.н.} \Rightarrow \theta_1=\theta_2.$$

Тогда $\Theta_0=\{\vartheta\}$. Более того, в случае компактного Θ в условиях теоремы 1 для оценки $\hat\theta_n$ имеет место сходимость к ϑ п.н.

Теоремы 1 и 2 доказаны в предположении о неинформативном механизме цензурирования, то есть независимости с.в. X и (K, T). Далее в параграфе рассматривается случай информативного цензурирования и обсуждаются его отличия от случая неинформативного цензурирования с точки зрения состоятельности оценок.

В четвертом параграфе рассматривается асимптотическое распределение оценок $\hat{\theta}_n$. Действительно, имея состоятельную оценку $\hat{\theta}_n$, возможно применить дельта-метод и стандартные условия регулярности М-оценок [3] для установления достаточных условий асимптотической нормальности. Методы теории эмпирических процессов позволили существенно ослабить эти условия в случае оценивания по выборке с интервальным цензурированием.

Теорема 3. *Предположим, что пространство параметров* Θ *евклидово. Пусть функция* $(x, y) \mapsto \log[F_\theta(y) - F_\theta(x)]$ *дифференцируема в окрестности* ϑ *для* $\mu \times \mu$*-п.в. пар* (x, y) *с* $x < y$*, и градиент принадлежит* $L^2(Q)$*. Пусть функция* $\theta \mapsto \int m_\theta \, dQ$ *в точке* ϑ *допускает разложение по Тейлору до второго члена с невырожденой матрицей вторых производных* Σ_ϑ*. Обозначим* $\psi_\theta = \frac{\partial}{\partial\theta} m_\theta$*. Тогда для состоятельной оценки* $\hat{\theta}_n$ *выполняется:*

$$\sqrt{n}\left(\hat{\theta}_n - \vartheta\right) \xrightarrow[n\to\infty]{d} N\left(0, \Sigma_\vartheta^{-1} \int \psi_\vartheta \psi_\vartheta^T \, dQ \, \Sigma_\vartheta^{-1}\right).$$

Обозначим через $I(f_\theta\ f_0)$ *расстояние Кульбака-Лейблера* между плотностями f_θ и f_0:

$$I(f_\theta; f_0) = \int \log \frac{f_0}{f_\theta} \, dF_0, \tag{3}$$

где F_0 — функция распределения, соответствующая плотности f_0. В случае оценивания по полным данным известно, что обычные оценки максимального правдоподобия $\hat{\theta}_n$ обладают свойством минимизации расстояния Кульбака-Лейблера:

$$I(f_{\hat{\theta}_n}, f_0) \to \min_{\theta\in\Theta} I(f_\theta, f_0), \quad n \to \infty. \tag{4}$$

Это свойство оказывается важным в случае, когда предполагаемая параметрическая модель не точна, так как свойство (4) не зависит от того, найдется ли $\vartheta \in \Theta$, что $f_0 = f_\vartheta$. Однако, оценки типа максимального правдоподобия (2) при оценивании по выборке с интервальным цензурирование свойством минимизации расстояния между распределениями (4) не обладают.

Во **второй главе** изучаются оценки, доставляющие минимум расстоянию Кульбака-Лейблера. В первом параграфе выполняется их построение. Пусть $\hat{F}_n(x)$ — непараметрическая оценка [10] функции распределения $F_0(x)$ случайной величины X. Тогда величина $\hat{I}_n(f_\theta, f_0)$, определенная как

$$\hat{I}_n(f_\theta, f_0) = \int \log \frac{f_0}{f_\theta}\, d\hat{F}_n,$$

является естественной оценкой расстояния (3).

Оценка $\tilde{\theta}_n$ вводится как точка минимума величины $\hat{I}_n(f_\theta, f_0)$:

$$\tilde{\theta}_n = \operatorname*{argmin}_{\theta\in\Theta} \hat{I}_n(f_\theta, f_0) = \operatorname*{argmax}_{\theta\in\Theta} \int f_\theta(x)\, d\hat{F}_n(x). \tag{5}$$

Несложно видеть, что оценки $\hat{\theta}_n$ и $\tilde{\theta}_n$ совпадают при оценивании по полным данным, так как в таком случае $\hat{F}_n$ является обыкновенной эмпирической функцией распределения для $X_1, \ldots, X_n$.

Следует заметить, что процедура оценивания, вообще говоря, не предполагает точности параметрической модели. В случае, если модель не точна, то $\tilde{\theta}_n$ по прежнему будет «адекватной» оценкой, доставляя минимум расстоянию Кульбака-Лейблера между предполагаемым параметрическим семейством и истинным распределением величин $X_1, \ldots, X_n$.

Основной проблемой при изучении асимптотических свойств оценок типа (5) является отсутствие в литературе предельных теорем для оценивания по выборке с интервальным цензурированием линейных функционалов относительно мер вида

$$K(F) = \int c(x)\, dF(x).$$

Задача оценивания функционалов такого вида рассматривалась в работе [6] лишь для случая интервального цензурирования второго типа (то есть для $K \equiv 2$). Во втором параграфе эти результаты обобщаются на случай интервального цензурирования смешанного типа. Основной результат заключен в следующей теореме, доказанной в диссертации.

Теорема 4. *При наложении некоторых условий регулярности на параметрическое семейство $\{f_\theta, \theta \in \Theta\}$ и модель интервального цензурирования смешанного типа (K, T) оценки функционалов вида*

$$\int \log f_\theta \, d\hat{F}_n(x)$$

асимптотически эффективны:

$$\sqrt{n} \int \log f_\theta \, d\left(\hat{F}_n(x) - F_0(x)\right) \xrightarrow[n\to\infty]{d} N\left(0, \sigma^2(\theta)\right).$$

Теорема 4 позволяет естественным образом распространить классические результаты о состоятельности и асимптотическом распределении оценок максимального правдоподобия на оценки $\tilde{\theta}_n$. Этому вопросу посвящен третий параграф. Теорема о состоятельности $\tilde{\theta}_n$ выглядит следующим образом

Теорема 5. *Пусть выполняются условия теоремы 4 и кроме этого:*

1. *Для любого достаточно малого шара $B \subset \Theta$:*

$$\int \sup_{\theta \in B} \log f_\theta(x) \, dF_0(x) < \infty;$$

2. *Θ — компакт и из $f_{\theta_1} = f_{\theta_2}$ п.н. следует $\theta_1 = \theta_2$.*

Тогда для оценки $\tilde{\theta}_n$ выполняется:

$$\tilde{\theta}_n = \operatorname*{argmin}_{\theta\in\Theta} \hat{I}_n(f_0, f_\theta) \xrightarrow[n\to\infty]{} \theta^* = \operatorname*{argmin}_{\theta\in\Theta} I(f_0, f_\theta) \text{ п.н.}$$

При этом, если найдется ϑ: $f_0 = f_\vartheta$ п.н., то $\theta^* = \vartheta$.

Теорема о асимптотической нормальности оценки $\tilde{\theta}_n$ аналогична теореме 3 и доказывается в четвертом параграфе.

Расстояние Кульбака-Лейблера (3) можно использовать для сравнения различных параметрических моделей и выбора наилучшей (в смысле минимизации этого расстояния). Такая процедура в литературе носит название информационного критерия Акайке [1].

В третьей главе производится построение информационных критериев в случае интервального цензурирования. Рассмотрим два семейства моделей $\mathcal{G}_1 = \{g_{\theta_1}, \theta_1 \in \Theta_1\}$ и $\mathcal{G}_2 = \{g_{\theta_2}, \theta_2 \in \Theta_2\}$. Пусть $\bar{\theta}_n^{(1,2)}$ — некоторые оценки параметров θ_1, θ_2, а $\hat{F}_n$ — оценка функции распределения F_0. Тогда задача сравнения двух параметрических моделей $\mathcal{G}_1$ и $\mathcal{G}_2$ с точки зрения расстояния Кульбака-Лейблера фактически сводится [1] к сравнению величин $\int g_{\bar{\theta}_n^{(1)}}\, d\hat{F}_n$ и $\int g_{\bar{\theta}_n^{(2)}}\, d\hat{F}_n$.

Однако, известно [1], что $\int g_{\bar{\theta}_n^{(1,2)}}\, d\hat{F}_n$ является *смещенной* оценкой величины $\int g_{\bar{\theta}_n^{(1)}}\, dF_0$. Более того, это смещение зависит от параметрических классов $\mathcal{G}_{1,2}$ и способа оценивания $\bar{\theta}_n$. Поэтому без оценивания и коррекции этого смещения использовать для сравнения моделей величины $\int g_{\bar{\theta}_n^{(1,2)}}\, d\hat{F}_n$ нельзя.

В первом параграфе это смещение оценивается для случая оценивания при помощи ОМКЛ $\tilde{\theta}_n$. В частном случае интервального цензурирования первого типа выражение для смещения получается в явном виде.

Во втором параграфе рассматривается случай построения оценок при помощи ОМП $\hat{\theta}_n$. Как и для $\tilde{\theta}_n$, явное выражение для смещения оказывается возможным получить только в случае интервального цензурирования первого типа. В остальных случаях для оценки смещения применяется комбинация процедур бутстреп и складного ножа (jack-knife).

В четвертой главе результаты, полученные в предыдущих главах, проверяются на модельных выборках. Рассмотрены типичные модели, используемые в литературе для анализа данных типа времени жизни: распределения, связанные с экспоненциальным (Вейбулла, обощенное гамма), модель Гомперца-Макегама [9], модель ExpCos А.Г. Барта [12].

В первом параграфе приводится обзор используемых моделей и механизма цензурирования.

Во втором параграфе исследуются асимптотические свойства оценок $\hat{\theta}_n$ (состоятельность и скорость сходимости к предельному распределению) при различных механизмах цензурирования, отличающихся величиной «потери информации о выборке». В третьем параграфе аналогичное исследование проводится для оценок $\tilde{\theta}_n$.

Далее, в четвертом параграфе производится сравнение оценок $\hat{\theta}_n$ и $\tilde{\theta}_n$. В качестве критерия для сравнения использовались дисперсия оценок и среднеквадратическое отклонение. В частности, здесь были получены следующие результаты: в том случае, когда предполагаемая параметрическая модель $\{f_\theta, \theta \in \Theta\}$ точна, оценка $\hat{\theta}_n$ по сравнению с $\tilde{\theta}_n$ обладает меньшей дисперсией и смещением. Ситуация противоположна, когда предполагаемая параметрическая модель не точна (то есть не существует такого $\vartheta \in \Theta$, что $f_0 = f_\vartheta$). В этом случае оценка $\hat{\theta}_n$ обладает существенно большей дисперсией по сравнению с $\tilde{\theta}_n$.

Пятая глава посвящена рассмотрению предложенных оценок на примере анализа реальных данных из стоматологии, кардиологии, фармакологии. При этом производится сравнение с известными, но полученными с игнорированием процесса цензурирования при сборе данных и оценивании, результатами.

В **заключении** подводятся итоги диссертационного исследования и формулируются основные результаты работы.

В **приложение** вынесены доказательства некоторых технических теорем из главы 2.

Список литературы

1. *Akaike H.* Information theory and an extension of the maximum likelihood principle // Second International Symposium on Information Theory / Ed. by B. Petrov, B. Csaki. Academiai Kiado: Budapest, 1973. Pp. 267–281.

2. *Andersen P. K., Borgan Ø., Gill R. D., Keiding N.* Statistical Models Based on Counting Processes. Springer, 1993. 784 pp.

3. *Bickel P. J., Klaassen C. A. J., Ritov Y., Wellner J. A.* Efficient and Adaptive Estimation for Semiparametric Models. Springer, 1998. 588 pp.

4. *Finkelstein D., Wolfe R. A.* Isotonic regression for interval-censored survival data using an E-M algorithm // *Comm. Statist.: Theory & Methods.* 1986. Vol. 15. Pp. 2493–2505.

5. *Fleming T. R., Harrington D. P.* Counting Processes and Survival Analysis. Wiley-Blackwell, 2005. 448 pp.

6. *Geskus R. B., Groeneboom P.* Asymptotically optimal estimation of smooth functionals for interval censoring, case 2 // *The Annals of Statistics.* 1999. Vol. 27, no. 2. Pp. 627–674.

7. *Huber P. J.* The behavior of Maximum Likelihood Estimates under nonstandard conditions // Proc. 5th Berkeley Sympos. Math. Statist. Probab., Univ. Calif. 1967. Pp. 221–233.

8. *Kalbfleisch J. D., Prentice R. L.* The Statistical Analysis of Failure Time Data. Wiley-InterScience, 2002. 462 pp.

9. *Marshall A. W., Olkin I.* Gompertz and Gompertz-Makeham Distributions // Life Distributions. Springer New York, 2007. Pp. 363–398.

10. *Schick A., Yu Q.* Consistency Of The GMLE With Mixed Case Interval-Censored Data // *Scand. J. Statist.* 1998. Vol. 27. Pp. 45–55.

11. *Sun J.* The Statistical Analysis of Interval-censored Failure Time Data (Statistics for Biology and Health). Springer, 2006. 406 pp.

12. *Барт А. Г., Бондаренко Б. Б., Бойко В. И.* Математический анализ течения ХГН // Гломерулонефрит. М.: Наука, 1980. С. 213–215.

13. *Кокс Д. Р., Оукс Д.* Анализ данных типа времени жизни. Москва: Финансы и статистика, 1988. 192 с.

Список публикаций автора

Статьи в журналах, рекомендованных ВАК:

A1. *Коробейников А. И.* Сравнение оценок параметров специальной модели кривой дожития для выборки с интервальным цензурированием // *Вестник С.-Петербургского университета, сер. 10.* 2009. Т. 2. С. 36–47.

Остальные публикации:

A2. *Барт А. Г., Коробейников А. И.* Об оценке параметров специальной модели кривой дожития // Математические модели. Теория и приложения / Под ред. М. К. Чиркова. 2007. Т. 8. С. 15–25.

A3. *Коробейников А. И.* Методы и программное обеспечение задач оценивания параметров в специальном случае модели кривых дожития // Математические модели. Теория и приложения / Под ред. М. К. Чиркова. 2009. Т. 10. С. 28–42.

A4. *Korobeynikov A.* On the Consistency of ML-estimates for the Special Model of Survival Curves with Incomplete Data // Proc. of 6th St. Petersburg Workshop on Simulation / Ed. by S. M. Ermakov, V. B. Melas, A. N. Pepelyshev. 2009. Pp. 1039–1045.