ТОМСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

Г.М.Кошкин

ОСНОВЫ СТРАХОВОЙ МАТЕМАТИКИ

Учебное пособие

Томск - 2002

УДК 519.25
ББК 32.973
Р 47

Кошкин Г.М.

Р 47 ОСНОВЫ АКТУАРНОЙ МАТЕМАТИКИ: Учебное пособие / Томск: Томский государственный университет, 2002. 116 с.

В учебном пособии дается элементарное введение в страховую (актуарную) математику, которая вместе с соответствующими экономическими и юридическими дисциплинами является одной из теоретических основ страхового бизнеса. Пособие предназначено для студентов Международного факультета управления (специальность "Государственное муниципальное управление") и факультета прикладной математики и кибернетики (специальность "Применение математических методов и исследование операций в экономике"). В конце каждой главы приводятся контрольные задания, которые могут использоваться при проведении практических занятий.

Учебное пособие может быть полезно также специалистам и аспирантам, которые интересуются приложениями математической статистики при разработке и исследовании математических моделей страховой математики.

1 ВВЕДЕНИЕ В СТРАХОВОЕ ДЕЛО

1.1 Предмет актуарной математики

В учебном пособии даются начальные сведения о *математических методах и моделях*, используемых в страховании. Общепринятое название этого научного направления – **актуарная математика** (actuarial mathematics) и происходит от actuary – актуарий, статистик страхового общества. Вместе с соответствующими экономическими и юридическими дисциплинами актуарная математика образует более широкую область знаний – **актуарную науку** (actuarial science), которая является теоретической основой страхового бизнеса.

Хотя актуарная математика широко использует методы теории вероятностей и математической статистики, она является самостоятельным научным направлением со своими *предметом*, *методами* и *сферой применений*.

Актуарное образование в мире имеет вековые традиции. Однако в нашей стране с почти 70-летним отсутствием свободных рыночных отношений актуарное образование и актуарная наука практически отсутствовали до 90-х годов XX века. В настоящее время появился огромный интерес к этой сфере деятельности, что связано с большой потребностью в специалистах-актуариях со стороны страховых компаний, число которых в России уже составляет несколько тысяч.

Самый высокий мировой стандарт профессиональной подготовки в области актуарной науки дает, по-видимому, программа Общества Актуариев (США). Отметим, что Общество Актуариев довольно жестко предписывает при изучении курса "Актуарная математика"использовать фундаментальную монографию [1].

К настоящему времени имеется литература и на русском языке [2]–[6], правда, труднодоступная студентам, а также всем желающим изучить страховую математику.

При написании пособия использовалась фундаментальная монография [1], книги [2]–[4], а также другие работы, ссылки на которые приводятся по мере необходимости.

Различают актуарную математику в имущественном и личном страховании. Под *имущественным страхованием* (non-life insurance) понимаются все виды страховой деятельности, не связанные с личным страхованием (страхование жилья, автомобилей, предприятий, банковских капиталов и т.п.). Под *личным страхованием* (life insurance) понимается страхование жизни, здоровья, пенсий и т.п.

Определение 1.1.1 *Денежные суммы или величины* $p_1, p_2, \ldots, p_n$, *которые страхуемые платят страховой компании, называются* страховыми премиями (premiums) *.*

Определение 1.1.2 *Величины* $b_1, b_2, \ldots, b_\nu$, $\nu \leq n$, *которые платит компания в результате наступления* ν страховых случаев *называются* страховыми выплатами (benefits) *.*

Ясно, что величины $b_j >> p_j$, иначе никто не будет страховаться. Купив за p рублей страховой полис, застрахованный избавил себя от риска финансовых потерь, связанных с неопределенностью страхового случая. Этот *риск* (risk) приняла на себя страховая компания, для которой риск заключается в случайности *иска* (claim), который может быть ей предъявлен.

В актуарной математике особое внимание уделяется стандартизации терминологии и обозначений. Для того, чтобы упростить общение между актуариями, облегчить внедрение результатов научных исследований и т.п., еще в 1898 году международный актуарный конгресс решил стандартизировать терминологию и обозначения основных величин, встречающихся в страховой математике. В пособии встретится большое число таких обозначений, порой довольно необычных. Свободное владение этими понятиями и обозначениями является важной неотъемлемой частью профессиональных знаний актуария. Поэтому необходимо обратить серьезное внимание на проблему запоминания терминов и обозначений.

1.2 Простейшая модель страховой компании

В страховой математике решаются следующие основные проблемы.

1. Нахождение "правильного"соотношения между премией p и выплатой b, $p << b$. Сюда входят, например, расчет нетто-премий, брутто-премий, расчет выплат, которые может позволить себе страховая компания и т.д. Заметим, что нетто-премия соответствует нулевой средней прибыли компании.

2. Расчет вероятности разорения, которая служит основой для принятия важнейших решений. Если обозначить через U капитал компании или ее резерв, а через

$$S = b_1 + b_2 + \cdots + b_\nu$$

— сумму выплат, где b_j — иск j-го страхуемого к компании, то грубо вероятность разорения равна $\mathbf{P}(S > U)$, а вероятность не разорения равна $\mathbf{P}(S \le U)$. Понятно, что

$$\mathbf{P}(S > U) + \mathbf{P}(S \le U) = 1.$$

3. Расчет резервов страховой компании.

Решение указанных проблем проиллюстрируем на примере простейшей модели работы страховой компании, построенной с использованием лишь центральной предельной теоремы теории вероятностей.

Рассмотрим следующую идеализированную схему. Пусть в начале года в фирме застраховалось n мужчин возраста $x = 26$ лет. Будем считать, что каждый клиент платит премию p, и, следовательно, фирма получила суммарный доход $p \cdot n$, который и будет составлять ее резерв $U = pn$. Обозначим через $b_i = 1$ иски, предъявляемые к фирме, если в течение года i-ый клиент умрет.

Воспользовавшись таблицей смертности из Приложения 1, для любого $i = 1, \dots, n$ находим

$$\mathbf{P}(b_i = 1) = q_{26} = 0,00293, \quad \mathbf{P}(b_i = 0) = p_{26} = 1 - q_{26} = 0,99707.$$

Вероятность того, что компания не разорится, равна

$$\mathbf{P}\left\{\sum_{i=1}^{n} b_i \le U\right\} = \mathbf{P}\left\{\frac{\sum_{i=1}^{n} b_i - \mathbf{E}\sum_{i=1}^{n} b_i}{\sqrt{\mathbf{D}\sum_{i=1}^{n} b_i}} \le \frac{U - \mathbf{E}\sum_{i=1}^{n} b_i}{\sqrt{\mathbf{D}\sum_{i=1}^{n} b_i}}\right\}. \qquad (1.2.1)$$

Здесь и всюду в дальнейшем символы $\mathbf{E}\xi$ и $\mathbf{D}\xi$ обозначают математическое ожидание (среднее) и дисперсию случайной величины ξ.

Учитывая дискретность случайных величин b_i, для любых $i = 1, \dots, n$ имеем

$$\mathbf{E}b_i = 1 \cdot q_{26} + 0 \cdot p_{26} = q_{26} = 0,00293,$$

$$\mathbf{D}b_i = \mathbf{E}b_i - (\mathbf{E}b_i)^2 = 0,00293 - (0,00293)^2 \approx 0,00292.$$

Так как иски b_i обычно на практике независимы, то формула (1.2.1) преобразуется к виду

$$\mathbf{P}\left\{\frac{\sum_{i=1}^{n} b_i - n\cdot 0,00293}{\sqrt{n\cdot 0,00292)}} \le \frac{U - n\cdot 0,00293}{\sqrt{n\cdot 0,00292}}\right\}.$$

Для числа застраховавшихся $n = 3071$, получаем

$$n\mathbf{E}b_1 = 3071\cdot 0,00293 \approx 9, \quad n\mathbf{D}b_1 = 3071\cdot 0,00292 \approx 9,$$

т.е. среднее и дисперсия случайной величины S совпадает, что, вообще говоря, указывает на пуассоновость распределения S. Для простоты воспользуемся центральной предельной теоремой, в силу которой

$$\mathbf{P}\left\{\frac{\sum_{i=1}^{n} b_i - 9}{3} \le \frac{U-9}{3}\right\} \approx \Phi\left(\frac{U-9}{3}\right),$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e{-}u^2/2du$$

– интеграл вероятностей (функция распределения стандартного нормального закона).

Пусть руководство компании устраивает вероятность неразорения 0,95. Тогда, решая относительно $\frac{U-9}{3}$ уравнение

$$\Phi\left(\frac{U-9}{3}\right) = 0,95,$$

имеем $\frac{U-9}{3} = x_{0,95}$, где $x_{0,95} = 1,645$ – квантиль уровня 0,95 стандартного нормального закона.

Следовательно, страховая компания в этом случае должна иметь резерв

$$U = 3\cdot 1,645 + 9 \approx 13,935,$$

а плата за страховку (премия) должна составить 0,00453 часть от иска $b = 1$, так как

$$p = \frac{U}{n} = \frac{13,935}{3071} = 0,00453.$$

Если $b = 100000$ руб., то премия

$$p = 100000 \cdot 0,00453 \approx 453 \text{ руб. в год},$$

или приблизительно 38 рублей в месяц, что могут позволить себе клиенты и с небольшим доходом.

1.3 Контрольные задания

Задание 1.3.1

Составьте различные риски для:
1) студента,
2) шахтера,
3) водителя автомобиля,
4) семьи,
5) организации,
6) области,
7) страны,
8) человечества в целом.

Задание 1.3.2

Почему нельзя получить страховку для защиты от риска застрелить завтра одну из собак вашего соседа?

Какие риски страхуются частично?

Какие риски страхуются обязательно?

Задание 1.3.3

Какие из составленных Вами в задании 1.3.1 рисков:
1) страхуемые,
2) нестрахуемые,
3) частично страхуемые,
4) обязательно подлежащие страхованию.

Задание 1.3.4

По методологии раздела 1.2 найдите премии p при вероятностях разорения 0,05 и 0,01 для женщин возраста $x = 26$ лет. Проведите сравнительный анализ полученных для женщин результатов с результатами для мужчин из раздела 1.2.

Задание 1.3.5

Постройте модель работы страховой компании, считая, что распределение суммарных выплат S является пуассоновским. Убедитесь, что в этом случае в условиях раздела 1.2 $p \approx 114$ руб. при вероятности разорения 0,05 и $p \approx 130$ руб. при вероятности разорения 0,01.

2 ВЕРОЯТНОСТНЫЕ ХАРАКТЕРИСТИКИ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ

2.1 Функция выживания (survival function)

Прежде всего отметим, что рассматриваемые в страховании жизни методы и модели применимы и к другим видам страхования. Пусть X обозначает продолжительность жизни, X_i – продолжительность жизни i -го индивидуума. В других видах страхования под X можно, например, понимать:

1) время до наступления заболевания (клещевой энцефалит, болезнь Лайма);

2) время безаварийной работы автомобиля;

3) время до причинения ущерба имуществу и т.п.

Неопределенность или непредсказуемость момента смерти, заболевания, аварии является основным источником случайности при страховании, что позволяет использовать случайные события, величины, процессы при математическом анализе различных аспектов страхования жизни, здоровья, автомобиля и т.п.

Очевидно, что относительно момента смерти конкретного человека, как правило, трудно сказать что-либо определенное. Однако, если рассматривается достаточно большая *однородная* группа людей, то для нее уже будут справедливы закономерности, присущие массовым случайным явлениям, например, устойчивость частот, сходимость к нормальному или пуассоновскому законам распределения и т.д. Поэтому, привлекая терминологию теории вероятностей, можно говорить о продолжительности жизни как о случайной величине X, причем $X \geq 0$.

Исчерпывающей характеристикой случайной величины X является функция распределения

$$F(x) = \mathbf{P}(\mathbf{X} \leq \mathbf{x}).$$

В актуарной математике обычно используют вместо функции распределения *функцию выживания*

$$\boxed{s(x) = \mathbf{P}(X > x) = 1 - F(x)}\ , \tag{2.1.1}$$

которая есть вероятность того, что человек доживет до возраста x лет. Индивидуума в возрасте x лет в актуарной математике обозначают (x).

Формулу для функции выживания мы поместили в рамку, так как $s(x)$ является одним из основных терминов и обозначений актуарной математики. В дальнейшем формулы, на которые следует обратить особое внимание будут обрамляться подобно (2.1.1).

Функция выживания как дополнительная функция к функции распределения обладает свойствами:

1) $s(x)$ убывает (нестрого);
2) $s(0) = 1$, $s(\infty) = 0$;
3) $s(x)$ непрерывна справа.

Однако для реального процесса смертности свойства 1) и 3) несколько видоизменяются. В самом деле, функция выживания должна *строго убывать*, так как существовал бы некоторый период в жизни человека, например, $\Delta x = x_2 - x_1$, в течение которого он не умирает, и $s(x)$ должна быть *непрерывной*, иначе существовал бы такой момент $x_\circ$ в жизни человека, в который он умирал бы с отличной от нуля вероятностью $\Delta\mathbf{P} = s(x_{\circ -}) - s(x_\circ)$, $s(x_{\circ -}) = \lim\limits_{x \to x_{\circ -}} s(x)$, $s(x_\circ) = \lim\limits_{x \to x_{\circ +}} s(x)$.

В какой-то степени о характере реальных зависимостей функций выживания от возраста x можно судить по приведенным ниже таблицам для мужчин и женщин СССР (1984–1985 г.), а также для населения США.

МУЖЧИНЫ (СССР, 1984–1985 г.) [5, с.86]

x	0	14	20	30	40	50	60	70	80	90	100	110
$s_1(x)$	1,000	0,954	0,947	0,922	0,878	0,795	0,651	0,434	0,188	0,003	0	0

ЖЕНЩИНЫ (СССР, 1984–1985 г.) [чет, с.86]

x	0	14	20	30	40	50	60	70	80	90	100	110
$s_2(x)$	1,000	0,964	0,961	0,953	0,939	0,908	0,841	0,700	0,417	0,008	0	0

НАСЕЛЕНИЕ США [2, с.7]

x	0	10	20	30	40	50	60	70	80	90	100	110
$s_3(x)$	1,000	0,983	0,977	0,965	0,949	0,915	0,837	0,682	0,432	0,142	0,012	0

Так как реальная продолжительность жизни X ограничена *предельным возрастом* (limiting age) $\omega = 100 - 120$ лет, то

$$s(x) = 0, \quad x > \omega.$$

В связи с этим таблицы продолжительности жизни (ТПЖ) составляются обычно для целых $x \le \omega$. Однако для функций $s(x)$, задаваемых аналитически, время жизни, как правило, неограничено, причем параметры $s(x)$ подбирают так, чтобы вероятность $\mathbf{P}\{X > \omega\}$ была достаточно малой.

Выясним, как функция выживания $s(x)$ связана с одной из основных характеристик l_x общих ТПЖ (Приложение 1). Рассмотрим достаточно большую группу из $l_\circ$ новорожденных (для удобства берут $l_\circ = 10000$ или $l_\circ = 100000$) и будем фиксировать их продолжительности жизни или моменты смерти: $X_1, X_2, \ldots, X_{l_\circ}$. Введем для события A индикатор

$$I(A) = \{1, \text{если событие } A \text{ произошло}; \ 0, \text{в противоположном случае}\}.$$

Тогда число представителей этой группы, доживших до возраста x, есть случайная величина

$$L(x) = \sum_{i=1}^{l_\circ} I(X_i > x),$$

математическое ожидание которой и определяет величину l_x :

$$l_x = \mathbf{E} l_x = \sum_{i=1}^{l_\circ} \mathbf{E} I(X_i > x) = \sum_{i=1}^{l_\circ} \mathbf{P}(X_i > x) = \sum_{i=1}^{l_\circ} s(x) = l_\circ s(x),$$

т.е.

$$\boxed{l_x = l_\circ \cdot s(x).} \tag{2.1.2}$$

Из формулы (2.1.2) следует, что

1) кривая l_x изменяется в зависимости от возраста x аналогично функции выживания $s(x)$ с точностью до множителя-константы $l_\circ$;

2) $s(x) = l_x / l_\circ$ — это средняя доля доживших до возраста x из рассматриваемой группы новорожденных.

Согласно ТПЖ, представленной в Приложении 1, $l_{14} = 95,438$ для мужчин обозначает, что из 100000 новорожденных до возраста 14 лет доживает в среднем 95438 мальчиков, а средняя доля доживших до возраста 14 лет из данной группы новорожденных равна $s(14) = 0,95438$.

2.2 Кривая смертей (the curve of deaths)

Рассмотрим данные ТПЖ для населения США предыдущего параграфа и проинтерпретируем их по другому. Если взять $l_\circ = 1000$, то в первые 10 лет жизни умрет примерно 17 человек, от 10 до 20 лет – 6 человек, от 20 до 30 лет – 12 человек, от 30 до 40 лет – 16 человек, от 40 до 50 лет – 34 человека, от 50 до 60 лет – 78 человек, от 60 до 70 лет – 155 человек, от 70 до 80 лет – 250 человек, от 80 до 90 лет – 290 человек, от 90 до 100 лет – 130 человек, от 100 до 110 лет – 12 человек.

Очевидно, что эти данные, разбитые на интервалы, более наглядно характеризуют смертность, по сравнению с данными, представленными в упомянутой таблице. Например, мы можем выделить первое десятилетие, в котором смертность втрое выше, чем во втором, самом безопасном десятилетии, или период от 70 до 90 лет, за который умирает 540 человек, что составляет более половины исходной группы в $l_0 = 1000$ человек.

Теперь становится понятным введение для интервала возрастов $(x, x+t)$ случайной величины

$$
{}_tD_x = L(x) - L(x+t) = \sum_{i=1}^{l_\circ} I(x < X_i \le x+t), \qquad (2.2.1)
$$

которая равна числу умерших в возрасте от x до $x+t$ из фиксированной группы $l_\circ$ новорожденных. Математическое ожидание этой случайной величины определяет еще одну из основных характеристик ${}_td_x$ общих таблиц продолжительности жизни (Приложение 1):

$$
\boxed{{}_td_x = \mathbf{E}\,{}_tD_x.}
$$

Очевидно, что согласно (2.2.1)

$$
{}_td_x = \mathbf{E}[L(x) - L(x+t)] = l_x - l_{x+t} = l_0[s(x) - s(x+t)],
$$

где $s(x) - s(x+t) = \mathbf{P}(x < X_i \le x+t)$ – вероятность смерти в промежутке $(x, x+t]$.

Отметим, что индекс 1 в обозначении ${}_1d_x$ обычно опускается, поэтому

$$
\boxed{d_x = l_x - l_{x+1},}
$$

т.е. d_x выражается через l_x и l_{x+1}, имеющиеся в таблицах продолжительности жизни. Тем не менее, величина d_x также приводится в этих таблицах в качестве основной.

Функция $f(x) = F^{'}(x) = -s^{'}(x)$ называется плотностью распределения случайной величины X, и за ней в актуарной математике закреплен термин *кривая смертей* (the curve of deaths).

Покажем, что кривая d_x от переменной возраста x изменяется приближенно как кривая смертей $f(x)$ с точностью до множителя-константы $l_\circ$, т.е.

$$
\boxed{d_x \approx l_\circ f(x).} \qquad (2.2.2)
$$

В самом деле, воспользовавшись формулой Тейлора (Теорема П3, Приложение 4), имеем

$$
s(x) = s(x+1) - s^{'}(\xi),
$$

где $\xi \in (x, x+1)$. Так как $s(x)$ почти не меняется в течение года, то

$$
d_x = -l_\circ s^{'}(\xi) \approx -l_\circ s^{'}(x) = l_\circ f(x).
$$

Формула (2.2.2) доказана.

Резюмируя, можно сказать, что кривая смертей в каком-то смысле является более тонкой характеристикой по сравнению с функцией выживания.

2.3 Функция интенсивности смертности (force of mortality)

В свою очередь, по сравнению с кривой смертей более тонкой характеристикой является функция интенсивности смертности. Найдем вероятность смерти человека, дожившего до x лет, в течение ближайших t лет, т.е. в промежутке $(x, x+t]$:

$$\mathbf{P}(x < X \leq x+t | X > x) = \frac{\mathbf{P}(x < X \leq x+t \cap X > x)}{\mathbf{P}(X > x)} =$$

$$= \frac{\mathbf{P}(x < X \leq x+t)}{\mathbf{P}(X > x)} = \frac{s(x) - s(x+t)}{s(x)} = \frac{F(x+t) - F(x)}{1 - F(x)}. \qquad (2.3.1)$$

Применяя формулу Тейлора к функции $F(x+t)$ в (2.3.1), получаем

$$F(x+t) - F(x) = F(x) + F^{'}(\xi) \cdot t - F(x) = f(\xi) \cdot t, \quad \xi \in (x, x+t). \quad (2.3.2)$$

Если величина t мала или функция $f(\xi)$ слабо изменяется в промежутке $(x, x+t)$, то согласно (2.3.1) и (2.3.2) справедливо приближенное равенство:

$$\mathbf{P}(x < X \leq x+t | X > x) \approx \frac{f(x)}{1 - F(x)} \cdot t. \qquad (2.3.3)$$

Отношение в правой части (2.3.3)

$$\boxed{\mu_x = \frac{f(x)}{1 - F(x)} = \frac{f(x)}{s(x)}} \qquad (2.3.4)$$

называется *функцией интенсивности смертности* и является важной характеристикой страховой математики. Согласно (2.3.4) и (2.3.3) величина $\mu_x \cdot t$ приближенно равна вероятности смерти человека возраста x в интервале $(x, x+t)$.

Важность функции μ_x подтверждается и тем, что ее приближение q_x приводится в ТПЖ (см. Приложение 1) также в качестве основной:

$$q_x = \frac{d_x}{l_x} \approx \frac{l_\circ f(x)}{l_\circ s(x)} = \mu_x.$$

Отметим, что в теории надежности функция μ_x называется *функцией отказов* (hazard rate function).

Функция интенсивности смертности обладает следующими свойствами:

I) $$\boxed{\mu_x \geq 0.} \qquad (2.3.5)$$

$$\text{II)} \quad \boxed{\int_0^{\infty} \mu_u du = +\infty.} \tag{2.3.6}$$

В самом деле, из условий, что $s(+\infty) = 0$, а $s(0) = 1$, получаем

$$\int_0^{\infty} \mu_u du = -\int_0^{\infty} \frac{ds(u)}{s(u)} = -\ln s(u))\big|_0^{\infty} = +\infty.$$

$$\text{III)} \quad \boxed{s(x) = e^{-\int_0^x \mu_u du},} \quad \boxed{F(x) = 1 - e^{-\int_0^x \mu_u du}.} \tag{2.3.7}$$

Выведем соотношения (2.3.7). Равенство $\mu_u = -s^{'}(u)/s(u)$ можно трактовать как дифференциальное уравнение. Умножим обе части данного уравнения на дифференциал du

$$\mu_u du = -\frac{s^{'}(u)du}{s(u)} = -\frac{ds(u)}{s(u)},$$

затем проинтегрируем обе части полученного равенства от нуля до x

$$\int_0^x \mu_u du = -\int_0^x \frac{ds(u)}{s(u)} = -\ln s(u)\big|_0^x =$$

$$= -(\ln s(x) - \ln s(0)) = -(\ln s(x) - \ln 1) = -\ln s(x),$$

откуда

$$s(x) = 1 - F(x) = e^{-\int_0^x \mu_u du}.$$

Соотношения (2.3.7) доказаны.

Из свойств I)–III) следует, что функция интенсивности смертности может быть использована как основная характеристика продолжительности жизни наряду с функцией распределения, функцией выживания и плотностью распределения.

2.4 Теоремы о моментах неотрицательных случайных величин

Приведем с доказательствами две полезных теоремы из ([1, с.62–64]), с помощью которых удобно находить моменты неотрицательных случайных величин как непрерывного, так и дискретного типа. Здесь и в дальнейшем конец доказательств теорем и лемм будем отмечать зачерненным указателем ♠.

Теорема 2.4.1 *(непрерывный случай). Пусть X – непрерывная случайная величина с функцией распределения $F(x)$ и соответствующей плотностью распределения $f(x) = F^{'}(x)$, причем*

1) $F(0) = 0$,

2) *функция $z(x)$ неотрицательна, монотонна и дифференцируема,*

3) *среднее $\mathbf{E}z(X) = \int_0^{\infty} z(x)f(x)dx < \infty$. Тогда*

$$\boxed{\mathbf{E}z(X) = z(0) + \int_0^{\infty} z^{'}(t)[1 - F(t)]dt.} \tag{2.4.1}$$

Д о к а з а т е л ь с т в о. Интегрируя по частям, имеем

$$\int_0^x z(t)f(t)dt = -\int_0^x z(t)d[1 - F(t)] =$$

$$= -z(t)[1 - F(t)]|_0^x + \int_0^x [1 - F(t)]z^{'}(t)dt.$$

Формула (2.4.1) имеет место, если $\lim\limits_{t\to\infty} z(t)[1 - F(t)] = 0$. Рассмотрим два случая:

А. Если неотрицательная функция $z(t)$ не возрастает, то

$$\lim_{t\to\infty} z(t)[1 - F(t)] \leq z(0) \lim_{t\to\infty} [1 - F(t)] = 0,$$

так как $\lim_{t\to\infty} F(t) = 1$.

Б. Если неотрицательная функция $z(t)$ не уменьшается, то

$$0 \leq z(t)[1 - F(t)] = z(t) \int_t^{\infty} f(s)ds \leq \int_t^{\infty} z(s)f(s)ds.$$

Но согласно условию 3) доказываемой теоремы несобственный интеграл $\int_0^{\infty} z(x)f(x)dx$ сходится, поэтому

$$\lim_{t\to\infty} \int_t^{\infty} z(s)f(s)ds = 0,$$

и, следовательно, тем более

$$\lim_{t\to\infty} z(t)[1 - F(t)] = 0. \quad \spadesuit$$

Теорема 2.4.2 *(дискретный случай). Пусть – дискретная случайная величина, принимающая неотрицательные целые значения, характеризуемая функцией распределения $F(k)$ и вероятностями* $p(k) = \mathbf{P}(Y = k) = F(k) - F(k-1) = \Delta F(k-1)$, *причем*

1) *функция $z(k)$ неотрицательна и монотонна,*

2) *среднее* $\mathbf{E}z(K) = \sum_{k=0}^{\infty} z(k)p(k) < \infty$. *Тогда*

$$\boxed{\mathbf{E}z(K) = z(0) + \sum_{k=0}^{\infty}[1 - F(k)]\Delta z(k).} \qquad (2.4.2)$$

Д о к а з а т е л ь с т в о. Суммируя по частям, имеем

$$\sum_{j=0}^{k-1} z(j)p(j) = -\sum_{j=0}^{k-1} z(j)\Delta[1 - F(j-1)] =$$

$$= -z(j)[1 - F(j-1)]|_0^k + \sum_{j=0}^{k-1}[1 - F(j)]\Delta z(j).$$

Формула (2.4.2) имеет место, если $\lim_{k\to\infty} z(k)[1-F(k-1)] = 0$. Рассмотрим два случая:

А. Если неотрицательная функция $z(k)$ не возрастает, то

$$\lim_{k\to\infty} z(k)[1 - F(k-1)] \leq z(0) \lim_{k\to\infty}[1 - F(k-1)] = 0,$$

так как $\lim_{k\to\infty} F(k-1) = 0$.

Б. Если неотрицательная функция $z(k)$ не уменьшается, то

$$0 \leq z(k)[1 - F(k-1)] = z(k)\sum_{j=k}^{\infty} p(j) \leq \sum_{j=k}^{\infty} z(j)p(j).$$

Но согласно условию 2) доказываемой теоремы ряд $\sum_{k=0}^{\infty} z(k)p(k)$ сходится абсолютно, поэтому

$$\lim_{k\to\infty}\sum_{j-k}^{\infty} z(j)p(j) = 0,$$

и, следовательно,

$$\lim_{k\to\infty} z(k)[1 - F(k-1)] = 0. \ \spadesuit$$

2.5 Среднее время жизни, его дисперсия, коэффициент асимметрии и эксцесс

Известно, что средняя продолжительность жизни случайной величины X

$$\boxed{\overset{\circ}{e}_{\circ} = \mathbf{E}X = \int_0^{\infty} x f(x) dx,} \tag{2.5.1}$$

является одним из важнейших показателей, с помощью которого сравнивают качество жизни населения различных стран. Наряду с $\overset{\circ}{e}_{\circ}$ важными практическими макрохарактеристиками случайной величины X также являются ее дисперсия

$$\boxed{\mathbf{D}X = \mathbf{E}(X - \overset{\circ}{e}_{\circ})^2 = \mathbf{E}X^2 - (\overset{\circ}{e}_{\circ})^2 = \int_0^{\infty} x^2 f(x) dx - \left(\overset{\circ}{e}_{\circ}\right)^2,} \tag{2.5.2}$$

коэффициент асимметрии

$$\boxed{\gamma = \frac{\mathbf{E}(X - \overset{\circ}{e}_{\circ})^3}{(\mathbf{D}X)^{3/2}},} \tag{2.5.3}$$

и эксцесс

$$\boxed{oe = \frac{\mathbf{E}(X - \overset{\circ}{e}_{\circ})^4}{(\mathbf{D}X)^2} - 3.} \tag{2.5.4}$$

Например, величина $\overset{\circ}{e}_{\circ}$ дает среднюю продолжительность жизни наудачу выбранного новорожденного, $\mathbf{D}X$ – средний квадрат разброса его продолжительности жизни относительно $\overset{\circ}{e}_{\circ}$, коэффициент асимметрии $\gamma > 0$ указывает на длинный хвост в правой части распределения случайной величины X, а если эксцесс $oe \approx 0$, то можно считать, что X распределено приближенно по нормальному закону (для нормальной случайной величины $oe = 0$).

Здесь же важно отметить, что знание, по крайней мере, первых четырех моментов или их оценок позволяет приближать неизвестные распределения случайной величины X различными параметрическими семействами (Пирсон, Джонсон, Тьюки).

С помощью теоремы 2.4.1 легко выразить формулы (2.5.1)–(2.5.4) в терминах функции выживания. Например, для $\overset{\circ}{e}_{\circ} = \mathbf{E}X$ функция $z(x) = x$, $z(0) = 0$, $z^{'}(x) = 1$ и, таким образом,

$$\boxed{\mathbf{E}X = \int_0^{\infty} (1 - F(x)) dx = \int_0^{\infty} s(x) dx.} \tag{2.5.5}$$

Аналогично для $\mathbf{E}X^2$ получаем $z(x) = x^2$, $z(0) = 0$, $z^{'}(x) = 2x$ и

$$\boxed{\mathbf{E}X^2 = 2\int_0^\infty x(1-F(x))dx = 2\int_0^\infty xs(x)dx.} \qquad (2.5.6)$$

2.6 Аналитические законы смертности: модели де Муавра, Гомпертца, Мэйкхама, Вейбулла и Эрланга

При теоретическом анализе процессов смертности, первоначальном и упрощенном изучении реальных ситуаций используют, как правило, стандартные вероятностные модели, позволяющие выявить основные закономерности, интересующие исследователя. К тому же, некоторые реальные процессы смертности достаточно хорошо аппроксимируются рассматриваемыми ниже аналитическими законами.

Модель де Муавра (de Moivre)

Один из основоположников теории вероятностей А.Муавр в 1729 г. предложил считать, что время жизни распределено равномерно на интервале $(0,\omega)$, где параметр ω, полностью определяющий закон равномерного распределения, называется *предельным возрастом.* Понятно, что для этой модели при $0 < x < \omega$ имеем

$$f(x) = \frac{1}{\omega},\ F(x) = \frac{x}{\omega},\ s(x) = 1 - \frac{x}{\omega},\ \mu_x = \frac{f(x)}{s(x)} = \frac{1}{\omega - x}.$$

Закон де Муавра не отражает многие характерные особенности, связанные с продолжительностью жизни человека. Например, в рассматриваемой модели кривая смертей является горизонтальной прямой, а эмпирические кривые имеют максимум в районе 80 лет.

Модель Гомпертца (Gompertz)

В этой модели (1825 г.) интенсивность смертности задается формулой

$$\mu_x = \frac{f(x)}{s(x)} = B\,e^{\alpha x},$$

где $\alpha > B > 0$ – некоторые параметры. Здесь функция выживания

$$s(x) = \exp\left[-\int_0^x \mu_u du\right] = \exp\left[-\int_0^x B\,e^{\alpha u} du\right] =$$
$$= \exp\left[-B\left(e^{\alpha x} - 1\right)/\alpha\right],$$

кривая смертей равна

$$f(x) = \mu_x s(x) = B\exp\left[\alpha x - B\left(e^{\alpha x} - 1\right)/\alpha\right]$$

и имеет максимум в точке $x = (\ln\alpha - \ln B)/\alpha$ (см. Задание 2.6.2).

Эту информацию можно использовать при нахождении оценок параметров α и B. Так, если из каких-то посылок нам известно, что максимальная смертность наблюдается для возраста 78,3 года, а квартиль функции распределения Гомпертца $x_{0,25}$ равен 33,4 года, то система уравнений для нахождения оценок принимает вид:

$$\left\{ \begin{array}{c} (\ln\alpha - \ln B)/\alpha = 78,3, \\ 1 - \exp\left[-B\,(e^{\alpha 33,4} - 1)/\alpha\right] = 0,25. \end{array} \right. \tag{2.6.1}$$

Систему нелинейных уравнений (2.6.1) можно решить на компьютере численными методами.

Модель Мэйкхама (Makeham)

Позднее, в 1860 г., Мэйкхам предложил приближать интенсивность смертности более общей функцией вида

$$\mu_x = A + B\,e^{\alpha\,x},$$

где параметр A учитывает риски, связанные с несчастными случаями, а слагаемое $B\,e^{\alpha\,x}$ учитывает влияние возраста на смертность. Здесь

$$s(x) = \exp\left[-\int_0^x (A + B\,e^{\alpha u})du\right] = \exp\left[-Ax - B\,(e^{\alpha\,x} - 1)/\alpha\right],$$

$$f(x) = -s^{'}(x) = [A + B\,e^{\alpha\,x}]\exp\left[-Ax - B\,(e^{\alpha\,x} - 1)/\alpha\right].$$

Из моделей, приведенных выше, закон Мэйкхама наиболее подходит для изучения процесса смертности человека, так как в нем учитывается, что для малых возрастов преобладающую роль в смертности играют несчастные случаи, а с увеличением возраста их роль ослабевает.

Модель Вейбулла (Weibull)

В 1939 г. Вейбулл в качестве простого приближения интенсивности смертности стал использовать степенную функцию

$$\mu_x = k\,x^n,$$

вид которой определяет функцию выживания

$$s(x) = \exp\left[-\int_0^x ku^n du\right] = \exp\left[-\frac{k}{n+1}x^{n+1}\right]$$

и кривую смертей

$$f(x) = -s^{'}(x) = kx^n \exp\left[-\frac{k}{n+1}x^{n+1}\right]$$

с максимумом в точке $x = (n/k)^{1/(n+1)}$ (см. Задание 2.6.2).

Модель Эрланга

Рассмотрим модель Эрланга 2-го порядка, для которой кривая смертей описывается формулой

$$f(x) = \frac{x}{a^2}\, e^{-\frac{x}{a}}, \quad x \geq 0.$$

В этом случае функция выживания

$$s(x) = \frac{x+a}{a}\, e^{-\frac{x}{a}},$$

а интенсивность смертности

$$\mu_x = \frac{x}{a(x+a)}.$$

В заключение отметим, что определенным преимуществом аналитических законов является то, что для них вероятностные характеристики продолжительности жизни можно быстро вычислять по небольшому числу параметров. Это может оказаться важным также и в случаях, когда доступные данные немногочисленны.

2.7 Контрольные задания

Задание 2.7.1

Характер зависимости функции выживания $s(x)$ от возраста x дают следующие таблицы:

МУЖЧИНЫ (СССР, 1984–1985 г.) [4, с.86]

x	0	14	20	30	40	50	60
$s_1(x)$	1,000	0,954	0,947	0,922	0,878	0,795	0,651

70	80	90	100	110
0,434	0,188	0,003	0	0

ЖЕНЩИНЫ (СССР, 1984–1985 г.) [4, с.86]

x	0	14	20	30	40	50	60
$s_2(x)$	1,000	0,964	0,961	0,953	0,939	0,908	0,841

70	80	90	100	110
0,700	0,417	0,008	0	0

НАСЕЛЕНИЕ США [1, с.7]

x	0	10	20	30	40	50	60
$s_3(x)$	1,000	0,983	0,977	0,965	0,949	0,915	0,837

70	80	90	100	110
0,682	0,432	0,142	0,012	0

1) Исходя из смысла функции выживания проанализируйте данные таблицы.

2) Привлекая порядковые статистики, постройте параметрические оценки функций выживания для модели де Муавра, воспользовавшись приведенными таблицами.

3) Опишите несколько способов построения параметрических оценок функций выживания для модели Вейбулла на основе приведенных таблиц (в модели Вейбулла интенсивность смертности μ_x приближается степенной функцией вида kx^n).

Задание 2.7.2

В модели Гомпертца интенсивность смертности μ_x приближается показательной функцией вида $Be^{\alpha x}$, где $\alpha > B > 0$ – некоторые параметры.

1) Найти точку максимума для кривой смертей в модели Гомпертца.

2) Где можно использовать полученный результат?

3) Найти точку максимума для кривой смертей в модели Вейбулла.

Задание 2.7.3

Пусть имеются две функции выживания:

$$s_1(x) = e^{-x^3/12}, \quad x \geq 0;$$

$$s_2(x) = \left(1 - \frac{x}{\omega}\right)^{\alpha}, \quad 0 \leq x < \omega, \quad \alpha > 0.$$

Функция выживания $s_2(x)$ при $\alpha = 1$ описывает модель де Муавра.

1) Найти интенсивности смертности μ_{ix}, кривые смертей $f_i(x)$ и функции распределения $F_i(x)$, соответствующие функциям выживания $s_i(x)$, $i = 1, 2$.

2) Найти вероятности того, что умрут в промежутке от 10 до 30 лет:

а) случайно выбранный человек,

б) застрахованный в возрасте 10 лет.

Задание 2.7.4

Рассмотрим
$$f(x) = \frac{x}{a^2} e^{-x/a}.$$

1) Покажите, что $f(x)$ может рассматриваться как кривая смертей.

2) Определите вид соответствующих функции выживания $s(x)$ и интенсивности смертности μ_x, а также соответствующую среднюю продолжительность жизни $\overset{\circ}{e}_\circ$.

3) Проанализируйте степень соответствия предложенного аналитического описания реальным данным с помощью таблицы продолжительности жизни (сравните среднее время жизни с точкой максимума кривой смертей).

4) Найдите дисперсию продолжительности жизни.

Задание 2.7.5

Рассмотрим три таблицы значений функций выживания задания 2.18.1. Подсчитайте среднее и дисперсию числа представителей исходной группы в l_0 новорожденных, которые умрут в возрасте от 50 до 70 лет.

Задание 2.7.6

Аппроксимируйте выбранную из задания 2.18.1 какую-либо функцию выживания $s_i(x)$, $i = 1, 2, 3$, используя модель Вейбулла для целых $n = 2, 4$.

1) Найдите параметр k с помощью МНК.

2) Постройте графики полученных аппроксимаций при $n = 2, 4$, выясните какая из построенных моделей наилучшая в смысле некоторого критерия, выбором которого распорядитесь самостоятельно.

3) Как найти оптимальное значение n?

4) Можно ли провести одновременно оптимизацию по параметрам k и n?

Задание 2.7.7

Предположим, что кривая смертей описывается формулой
$$f(x) = \frac{x}{a^2} e^{-x/a}, \quad x \geq 0.$$

1) Найдите функцию распределения $F_x(t)$ остаточного времени жизни $T(x) = X - x$ (X, x – момент смерти и возраст индивида, соответственно), $F_x(t) = P(T(x \leq t))$.

2) Покажите, что плотность распределения остаточного времени жизни $f_x(t) = \dfrac{d}{dt} F_x(t)$ является взвешенной суммой экспоненциальной плотности $\dfrac{1}{a} e^{-t/a}$ и эрланговской плотности $\dfrac{t}{a^2} e^{-t/a}$.

3) Найдите вероятность $P(T(x) > t)$, предел $\lim\limits_{x\to\infty} P(T(x) > t)$ и выясните, можно ли применять такую аппроксимацию кривой смертей для больших возрастов x.

3 ОСТАТОЧНАЯ ПРОДОЛЖИТЕЛЬНОСТЬ ЖИЗНИ

3.1 Остаточное время жизни (time-untill-death), его распределение

Если человек дожил до возраста x лет, то страховую компанию, в силу специфики ее деятельности, интересует, вообще говоря, уже не его общая продолжительность жизни X, а остаточное время жизни $T(x) = X - x$. Найдем функцию распределения случайной величины $T(x)$, которая является условной функцией распределения величины $X - x$ при условии, что $X > x$:

$$F_x(t) = \mathbf{P}(T(x) \le t) = \mathbf{P}(X - x \le t | X > x) =$$

$$= \mathbf{P}(X \le x + t | X > x) = {}_tq_x = \frac{\mathbf{P}(X \le x + t \bigcap X > x)}{\mathbf{P}(X > x)} =$$

$$= \frac{\mathbf{P}(x < X \le x + t)}{\mathbf{P}(X > x)} = \frac{F(x+t) - F(x)}{1 - F(x)}. \tag{3.1.1}$$

Если известны табличные значения l_x и l_{x+t}, то через них функция распределения $F_x(t)$ выражается следующим образом:

$$F_x(t) = \frac{s(x) - s(x+t)}{s(x)} = \frac{l_x/l_\circ - l_{x+t}/l_\circ}{l_x/l_\circ} = \frac{l_x - l_{x+t}}{l_x}.$$

Воспользовавшись (3.1.1), найдем теперь плотность $f_x(t)$ случайной величины $T(x)$:

$$f_x(t) = \frac{d}{dt}F_x(t) = \frac{f(x+t)}{1 - F(x)}, \quad 0 \le t < \infty.$$

В качестве примеров рассмотрим законы де Муавра и Мэйкхама.

Модель де Муавра. Введем для удобства обозначение

$$I_x(a,b] = \begin{cases} 1, & \text{если } x \in (a,b] \\ 0, & \text{если } x \notin (a,b] \end{cases}.$$

Пусть распределение продолжительности жизни описывается законом де Муавра, для которого плотность распределения и функция выживания, соответственно, задаются формулами:

$$f(x) = \frac{I_x(0,\omega)}{\omega}, \quad s(x) = I_x(-\infty,\omega) - \frac{x\, I_x(0,\omega)}{\omega}.$$

Так как $0 < X < \omega$, то $0 < T(x) < \omega - x$, и

$$F_x(t) = \frac{t}{\omega - x} I_t(0, \omega - x) + I_t[\omega - x, \infty),$$

а для $t \in (0, \omega - x)$

$$f_x(t) = \frac{d}{dt}\left(\frac{t}{\omega - x}\right) = \frac{1}{\omega - x}, \quad t \in (0, \omega - x], \tag{3.1.2}$$

т.е. остаточное время жизни $T(x)$ также равномерно распределено, однако на промежутке $(0, \omega - x)$.

Модель Мэйкхама. Для этого закона

$$\mu_t = A + B\,e^{\alpha t}, \quad s(x) = \exp\left[-\int_0^x \mu_t dt\right] = \exp\left[-Ax - B\,(e^{\alpha\,x} - 1)/\alpha\right],$$

$$f(x) = -s^{'}(x) = [A + B\,e^{\alpha\,x}]\exp\left[-Ax - B\,(e^{\alpha\,x} - 1)/\alpha\right],$$

$$f(x+t) = [A + B\,e^{\alpha\,(x+t)}]\exp\left[-A(x+t) - B\,(e^{\alpha\,(x+t)} - 1)/\alpha\right].$$

Поэтому плотность распределения остаточного времени жизни

$$f_x(t) = \frac{f(x+t)}{s(x)} = [A + B\,e^{\alpha\,x}e^{\alpha\,t}]\exp\left[-At - B\,e^{\alpha\,x}(e^{\alpha\,t} - 1)/\alpha\right],$$

откуда следует, что остаточное время жизни $T(x)$ также имеет распределение Мэйкхама с такими параметрами: $A_x = A$, $B_x = B\,e^{\alpha\,x}$, $\alpha_x = \alpha$.

3.2 Величины, связанные с $T(x)$: ${}_tq_x$, ${}_tx$, q_x, p_x, ${}_{t|u}q_x$, ${}_{t|}q_x$

В актуарной математике вероятность $\mathbf{P}(T(x) \le t)$ и дополнительная вероятность $\mathbf{P}(T(x) > t)$, обозначаемые символами ${}_tq_x$ и ${}_tx$, в соответствии с (3.1.1) определяются следующими формулами:

$${}_tq_x = \mathbf{P}(T(x) \le t) = \frac{s(x) - s(x+t)}{s(x)}, \tag{3.2.1}$$

$${}_tx = \mathbf{P}(T(x) > t) = 1 - {}_tq_x = \frac{s(x+t)}{s(x)}. \tag{3.2.2}$$

Величина ${}_tq_x$ выражает вероятность смерти человека возраста x лет в промежутке времени $(x, x+t]$, а ${}_tx$ — вероятность, что такой человек доживет до возраста $x + t$.

При $t = 1$ передние индексы у ${}_tq_x$ и ${}_t{}_x$ опускаются, и из общих формул (3.2.1) и (3.2.2) получаем

$$q_x = \mathbf{P}(T(x) \le 1) = \frac{s(x) - s(x+1)}{s(x)} = \frac{l_x - l_{x+1}}{l_x}, \tag{3.2.3}$$

$$_x = \mathbf{P}(T(x) > 1) = \frac{s(x+1)}{s(x)} = \frac{l_{x+1}}{l_x}. \tag{3.2.4}$$

Переменные q_x и ${}_x$ чаще всего используются на практике; величина q_x равна вероятности смерти индивидуума возраста x лет в течение ближайшего года, а p_x — вероятности, что он проживет еще по крайней мере один год.

Учитывая формулу (3.2.3), представим

$$\boxed{q_x = \frac{d_x}{l_x}.}$$

Величина q_x является третьей основной характеристикой, которая входит в ТПЖ (Приложение 1). Так как $l_x = l_\circ\, s(x)$, $d_x \approx l_\circ\, f(x)$, то $q_x \approx \dfrac{f(x)}{s(x)} = \mu_x$, т.е. форма кривой q_x приближенно совпадает с формой кривой функции интенсивности смертности.

Таким образом, мы выяснили смысл трех основных величин, содержащихся в ТПЖ:

$l_x = l_\circ\, s(x)$ связана через мультипликативный множитель $l_\circ$ с функцией выживания $s(x)$ и повторяет ее характер изменения в зависимости от возраста x;

$d_x \approx l_\circ\, f(x)$ связана приближенным равенством через мультипликативный множитель $l_\circ$ с кривой смертей $f(x)$ и повторяет приближенно ее характер изменения в зависимости от возраста x;

$q_x \approx \dfrac{f(x)}{s(x)} = \mu_x$ совпадает приблизительно с интенсивностью смертности μ_x.

В страховом деле возникает необходимость рассматривать и более сложные случаи. Например, определим вероятность того, что человек возраста x проживет t лет, но умрет на протяжении следующих u лет. Эту вероятность обозначают

$${}_{t|u}q_x = \mathbf{P}(t < T(x) \le t + u), \tag{3.2.5}$$

и она может быть выражена или через функцию ${}_tq_x$, или ${}_t{}_x$, или, принимая во внимание соотношения (3.2.1) и (3.2.2), через основную функцию страховой математики – функцию выживания $s(x)$:

$${}_{t|u}q_x = \mathbf{P}(t < T(x) \le t + u) =$$

$$= \mathbf{P}(T(x) \le t+u) - \mathbf{P}(T(x) \le t) = {}_{t+u}q_x - {}_tq_x, \tag{3.2.6}$$

$${}_{t|u}q_x = \mathbf{P}(t < T(x) \le t+u) =$$

$$= \mathbf{P}(T(x) > t) - \mathbf{P}(T(x) > t+u) = {}_tp_x - {}_{t+u}p_x, \tag{3.2.7}$$

$${}_{t|u}q_x = \frac{s(x+t) - s(x+t+u)}{s(x)}. \tag{3.2.8}$$

Снова случай $u = 1$ наиболее интересен для практики страхования жизни, и, как обычно, этот индекс опускается. Согласно формулам (3.2.6), (3.2.7) и (3.2.8) имеем:

$${}_{t|}q_x = {}_{t+1}q_x - {}_tq_x = {}_tp_x - {}_{t+1}p_x = \frac{s(x+t) - s(x+t+1)}{s(x)}. \tag{3.2.9}$$

3.3 Среднее остаточное время жизни, его дисперсия. Коэффициент асимметрии и эксцесс

В актуарной математике среднее остаточного времени жизни человека возраста x обозначается

$$\overset{\circ}{e}_x = \mathbf{E}T(x);$$

эту характеристику часто используют при анализе ситуации на рынке страховых услуг. Понятно, что $\mathbf{E}T(0) = \mathbf{E}X = \overset{\circ}{e}_\circ$, и, следовательно, средняя продолжительность жизни $\overset{\circ}{e}_\circ$ больше среднего остаточного времени жизни $\overset{\circ}{e}_x$ для любого $x > 0$.

Так как $T(x)$ является неотрицательной случайной величиной, то, воспользовавшись теоремой 2.4.1, получаем

$$\overset{\circ}{e}_x = \mathbf{E}T(x) = \int_0^\infty t\,dF_x(t) = \int_0^\infty t\,d\mathbf{P}(T(x) \le t) =$$

$$= \int_0^\infty \mathbf{P}(T(x) > t)\,dt = \int_0^\infty {}_tp_x\,dt.$$

Применив формулу (3.2.2) и преобразовав последний интеграл

$$\int_0^\infty {}_tp_x\,dt = \frac{1}{s(x)} \int_0^\infty s(x+t)\,dt = \frac{1}{s(x)} \int_x^\infty s(u)\,du,$$

приходим к соотношению

$$\boxed{\overset{\circ}{e}_x = \frac{1}{s(x)} \int_x^\infty s(u)\,du.} \tag{3.3.1}$$

Рассуждая аналогично, для второго начального момента случайной величины $T(x)$ имеем:

$$\mathbf{E}[T(x)]^2 = 2\int_0^\infty t\,\mathbf{P}(T(x) > t)\,dt = 2\int_0^\infty t\,{}_tp_x\,dt =$$

$$= \frac{2}{s(x)}\int_0^\infty t\,s(x+t)\,dt,$$

откуда

$$\mathbf{D}T(x) = \frac{2}{s(x)}\int_0^\infty t\,s(x+t)\,dt - \left(\overset{\circ}{e}_x\right)^2.$$

Для нахождения коэффициента асимметрии и эксцесса необходимо определить $\mathbf{E}[T(x)]^3$ и $\mathbf{E}[T(x)]^4$:

$$\mathbf{E}[T(x)]^3 =$$

$$= 3\int_0^\infty t^2\mathbf{P}(T(x) > t)\,dt = 3\int_0^\infty t^2{}_tp_x\,dt = \frac{3}{s(x)}\int_0^\infty t^2 s(x+t)\,dt,$$

$$\mathbf{E}[T(x)]^4 = \frac{4}{s(x)}\int_0^\infty t^3 s(x+t)\,dt.$$

Пример 3.3.1. Найти среднее $\overset{\circ}{e}_x$, дисперсию $\mathbf{D}T(x)$ и среднее квадратическое отклонение $\sigma(T(x)) = \sqrt{\mathbf{D}T(x)}$ остаточного времени жизни $T(x)$ в модели де Муавра.

Р е ш е н и е. Для этой модели

$$\mathbf{P}(T(x) > t) = {}_tp_x = \frac{s(x+t)}{s(x)} = \frac{1-(x+t)/\omega}{1-x/\omega}, \quad 0 \le t \le \omega - x.$$

Таким образом,

$$\overset{\circ}{e}_x = \int_0^{\omega-x}\frac{\omega-x-t}{\omega-x}\,dt = \frac{-1}{\omega-x}\int_{\omega-x}^0 u\,du = \frac{\omega-x}{2}. \qquad (3.3.2)$$

Проще формулу (3.3.2) получить из (3.3.1):

$$\overset{\circ}{e}_x = \frac{\omega}{\omega-x}\int_x^\omega\frac{\omega-t}{\omega}\,dt = \frac{-1}{\omega-x}\int_{\omega-x}^0 u\,du = \frac{\omega-x}{2}.$$

Также нетрудно показать, что

$$\mathbf{D}T(x) = \frac{(\omega-x)^2}{12}.$$

3.4 Смешанное страхование. Частичная остаточная продолжительность жизни

Рассмотрим *n-летнее страхование на дожитие* (n-year endowment insurance) или смешанное страхование, суть которого состоит в том, что клиент заключает договор страхования на n лет, и выплата производится:

либо в момент смерти застрахованного, если она наступила до окончания n-летнего периода,

либо в конце n-летнего периода, если застрахованный остался жив.

Момент страховой выплаты выражается формулой $\min(T(x), n)$ и называется *частичной продолжительностью жизни*, а соответствующее математическое ожидание — *частичной средней продолжительностью жизни* и обозначается $\overset{\circ}{e}_{x:\overline{n}|}$:

$$\overset{\circ}{e}_{x:\overline{n}|} = \mathbf{E}\min(T(x), n).$$

Чтобы применить теорему 2.4.1, необходимо найти вероятность $\mathbf{P}(\min(T(x), n) > t)$. Так как для $t < n$ событие $\min(T(x), n) > t$ равносильно событию $T(x) > t$, то

$$\mathbf{P}(\min(T(x), n) > t) = \begin{cases} {}_tp_x, & \text{если } 0 \le t < n \\ 0, & \text{если } t \ge n. \end{cases}$$

Привлекая формулу (3.2.2), получаем:

$$\overset{\circ}{e}_{x:\overline{n}|} = \int_0^n {}_tp_x\,dt = \frac{1}{s(x)}\int_0^n s(x+t)\,dt = \frac{1}{s(x)}\int_x^{x+n} s(u)\,du.$$

Для дисперсии частичной продолжительности жизни имеем

$$\mathbf{D}\min(T(x), n) = 2\int_0^n t\,{}_tp_x\,dt - (\overset{\circ}{e}_{x:\overline{n}|})^2 = \frac{2}{s(x)}\int_0^n t\,s(x+t)\,dt - (\overset{\circ}{e}_{x:\overline{n}|})^2.$$

Пример 3.4.1. Найти среднее и дисперсию частичной продолжительности жизни в модели де Муавра.

Р е ш е н и е. Прежде всего отметим, что $x + n < \omega$, поэтому

$$\overset{\circ}{e}_{x:\overline{n}|} = \frac{1}{s(x)}\int_x^{x+n} s(u)\,du = \frac{1}{1 - x/\omega}\int_x^{x+n}(1 - u/\omega)\,du =$$

$$= \frac{1}{\omega - x}\int_x^{x+n}(\omega - u)\,du = \frac{2n(\omega - x) - n^2}{2(\omega - x)} = n - \frac{n^2}{2(\omega - x)}. \qquad (3.4.1)$$

Из формулы (3.4.1) следует, что

1) $\overset{\circ}{e}_{x:\overline{n}|} < n$;

2) если $n = \omega$ и $x = 0$, то $\overset{\circ}{e}_{0:\overline{\omega}|} = \overset{\circ}{e}_{\circ} = \omega/2$;

3) если $x = \omega - n$, то $\overset{\circ}{e}_{(\omega-n):\overline{n}|} = n/2$.

Также нетрудно получить выражение для дисперсии:

$$\mathbf{D}\min(T(x), n) = \frac{n^3}{3(\omega - x)} - \frac{n^4}{4(\omega - x)^2}. \tag{3.4.2}$$

Из формулы (3.4.2) следует, что при $n = \omega$ и $x = 0$

$$\mathbf{D}\min(T(x), \omega) = \frac{\omega^2}{12},$$

т.е. получаем известный результат для дисперсии равномерного закона распределения на интервале $(0, \omega)$.

Рассмотрим для $\omega = 90$ лет два конкретных случая при $n = 5$ и $n = 10$ лет. Значения средних и дисперсий частичных средних продолжительностей жизни, вычисленные по формулам (3.4.1) и (3.4.2), для ряда возрастов и $n = 5, 10$ сведены в следующую таблицу:

x	0	10	20	30	40	50	60	70	80	85	90
$\overset{\circ}{e}_{x:\overline{5}|}$	4,853	4,844	4,812	4,792	4,750	4,688	4,583	4,375	3,750	2,500	—
$\mathbf{D}\min(T(x), 5)$	0,444	0,496	0,563	0,651	0,771	0,943	1,215	1,632	3,604	2,080	—
$\overset{\circ}{e}_{x:\overline{10}|}$	9,445	9,375	9,286	9,167	9,000	8,750	8,333	7,500	5,000	—	—
$\mathbf{D}\min(T(x), 10)$	2,777	3,836	4,252	4,861	5,666	6,771	8,333	13,542	8,333	—	—

3.5 Округленное остаточное время жизни, его распределение, среднее и дисперсия

В актуарной математике наряду с остаточным временем жизни $T(x)$ рассматривают его целую часть $K(x) = [T(x)]$, которую называют *округленной остаточной продолжительностью жизни* (curtate-future-lifetime). Это связано со следующими причинами:

1) человек обычно считает свой возраст в целых годах;

2) договоры страхования жизни, как правило, заключаются на целое число лет;

3) в ТПЖ приводятся данные для возрастов в целых годах.

Если $T(x) = 10$ лет 9 месяцев $= 10,75$ лет, то $K(x) = 10$ лет. Таким образом, случайная величина $K(x)$ является дискретной случайной величиной, принимающей целые значения. Как известно, исчерпывающей характеристикой такой случайной величины является набор вероятностей

$$\mathbf{P}(K(x) = k), \quad k = 0, 1, 2, \ldots$$

Понятно, что

$$\mathbf{P}(K(x) = k) = \mathbf{P}(k \le T(x) < k + 1).$$

Так как $T(x)$ — непрерывная случайная величина, то

$$\mathbf{P}(T(x) = k) = \mathbf{P}(T(x) = k + 1) = 0,$$

и поэтому

$$\mathbf{P}(K(x) = k) = \mathbf{P}(k < T(x) \le k + 1) =$$

$$= \frac{s(x+k) - s(x+k+1)}{s(x)} = {}_kp_x - {}_{k+1}p_x.$$

Теперь, учитывая, что $X = T(0)$, можно определить распределение и *округленного времени жизни* $K(0) = [X]$:

$$\mathbf{P}(K(0) = k) = \frac{s(k) - s(k+1)}{s(0)} = s(k) - s(k+1) = \frac{l_k - l_{k+1}}{l_\circ} = \frac{d_k}{l_\circ}.$$

Но так как $d_x \approx l_\circ f(x)$, то

$$\mathbf{P}(K(0) = k) \approx f(k), \tag{3.5.1}$$

где $f(x)$ — плотность случайной величины X, причем в правой части приближенного равенства (3.5.1), вообще говоря, правильнее писать $f(k) \cdot 1$ год, так как в его левой части находится безразмерная величина.

Поэтому равенство 3.5.1 подразумевает, что

$$\mathbf{P}(K(0) = k) \approx f(k) \cdot 1\text{год},$$

откуда видно, что кривая смертей тесно связана с распределением округленного времени жизни.

Среднее случайной величины $K(x)$ называется *средней округленной остаточной продолжительностью жизни* , обозначается

$$e_x = \mathbf{E}\, K(x),$$

и согласно определению математического ожидания для дискретной случайной величины

$$e_x = \sum_{k=1}^{\infty} k\,\mathbf{P}(K(x) = k).$$

Так как

$$\mathbf{P}(K(x) = k) = \frac{s(x+k) - s(x+k+1)}{s(x)}$$

и

$$\sum_{k=1}^{\infty} k[s(x+k) - s(x+k+1)] = 1 \cdot s(x+1) + 2 \cdot s(x+2) + \cdots$$

$$\cdots - 1 \cdot s(x+2) - 2 \cdot s(x+3) - \cdots = \sum_{k=1}^{\infty} s(x+k),$$

то

$$\boxed{e_x = \frac{1}{s(x)} \sum_{k=1}^{\infty} s(x+k).}$$

Аналогично находится второй начальный момент:

$$\mathbf{E}\,[K(x)]^2 = \sum_{k=0}^{\infty} k^2\,\mathbf{P}(K(x) = k) =$$

$$= \frac{1}{s(x)} \left\{ \sum_{k=0}^{\infty} k^2\, s(x+k) - \sum_{k=0}^{\infty} k^2\, s(x+k+1) \right\} =$$

$$= \frac{1}{s(x)} \left\{ \sum_{k=0}^{\infty} k^2\, s(x+k) - \right.$$

$$\left. - \sum_{k=0}^{\infty} (k+1)^2\, s(x+k+1) + \sum_{k=0}^{\infty} (2k+1)\, s(x+k+1) \right\} =$$

$$= \frac{1}{s(x)} \sum_{k=1}^{\infty} (2k-1)\, s(x+k) = \frac{2}{s(x)} \sum_{k=1}^{\infty} k\, s(x+k) - e_x,$$

так как

$$\sum_{k=0}^{\infty} k^2\, s(x+k) - \sum_{k=0}^{\infty} (k+1)^2\, s(x+k+1) + \sum_{k=0}^{\infty} (2k+1)\, s(x+k+1) =$$

$$= 1^2 \cdot s(x+1) + 2^2 \cdot s(x+2) + 3^2 \cdot s(x+3) + \cdots$$
$$\cdots - 1^2 \cdot s(x+1) - 2^2 \cdot s(x+2) - 3^2 \cdot s(x+3) - \cdots$$
$$\cdots + 1 \cdot s(x+1) + 3 \cdot s(x+2) + 5 \cdot s(x+3) + \cdots = \sum_{k=1}^{\infty}(2k-1)s(x+k).$$

Теперь легко получаем дисперсию

$$\boxed{\mathbf{D}\, K(x) = \mathbf{E}\,[K(x)]^2 - e_x^2 = \frac{2}{s(x)} \sum_{k=1}^{\infty} k\, s(x+k) - e_x - e_x^2.}$$

Все числовые характеристики округленного времени жизни выражаются через функции выживания, поэтому, в силу равенства (2.1.2), их также можно находить и по данным ТПЖ. В частности,

$$e_x = \frac{1}{l_x} \sum_{k=1}^{\infty} l_{x+k}, \quad e_\circ = \frac{1}{l_\circ} \sum_{k=1}^{\infty} l_k,$$

$$\mathbf{E}\,[K(x)]^2 = \frac{2}{l_x} \sum_{k=1}^{\infty} k\, l_{x+k} - e_x, \quad \mathbf{E}\,[K(0)]^2 = \frac{2}{l_\circ} \sum_{k=1}^{\infty} k\, l_k - e_\circ.$$

Пример 3.5.1. По данным ТПЖ (Приложение 1) отдельно для мужчин и женщин найдите:

1) средние округленные продолжительности жизни e_{89}, e_{88}, e_{84};

2) дисперсии округленных продолжительностей жизни $\mathbf{D}K(89)$, $\mathbf{D}K(88)$.

Р е ш е н и е. Для мужчин:

$$e_{89} = \frac{290}{1449} = 0,2, \quad \mathbf{D}K(89) = \frac{2 \cdot 290}{1449} - 0,2 - (0,2)^2 = 0,16;$$

$$e_{88} = \frac{1449 + 290}{3623} = 0,48,$$

$$\mathbf{D}K(88) = \frac{2(1 \cdot 1449 + 2 \cdot 290)}{3623} - 0,48 - (0,48)^2 = 0,41;$$

$$e_{84} = \frac{9063 + 7546 + 6037 + 3623 + 1449 + 290}{10735} = 2,6.$$

Для женщин:

$$e_{89} = \frac{806}{4030} = 0,2, \quad \mathbf{D}K(89) = \frac{2 \cdot 806}{4030} - 0,2 - (0,2)^2 = 0,16;$$

$$e_{88} = \frac{4030 + 806}{10075} = 0,48,$$

$$\mathbf{D}K(88) = \frac{2(1 \cdot 4030 + 2 \cdot 806)}{10075} - 0,48 - (0,48)^2 = 0,41;$$

$$e_{84} = \frac{24265 + 20988 + 16791 + 10075 + 4030 + 806}{27665} = 2,75.$$

3.6 Контрольные задания

Задание 3.6.1

Предположим, что кривая смертей описывается формулой

$$f(x) = \frac{x}{a^2} e^{-x/a}, \quad x \geq 0.$$

1) Найдите функцию распределения $F_x(t)$ остаточного времени жизни $T(x) = X - x$.

2) Покажите, что плотность распределения остаточного времени жизни $f_x(t) = \dfrac{d}{dt} F_x(t)$ является взвешенной суммой экспоненциальной плотности $\dfrac{1}{a} e^{-t/a}$ и эрланговской плотности $\dfrac{t}{a^2} e^{-t/a}$.

3) Найдите вероятность $P(T(x) > t)$, предел $\lim\limits_{x \to \infty} P(T(x) > t)$ и выясните, можно ли применять такую аппроксимацию кривой смертей для больших возрастов x.

Задание 3.6.2

Используя ТПЖ Приложения 1 оцените вероятности того, что индивидуум (21):

1) доживет до 70, 80, 90 лет;
2) умрет до 70, 80, 90 лет;
3) умрет от 60 до 70, от 70 до 80, от 80 до 90 лет.

Задание 3.6.3

Раскройте смысл следующих обозначений актуарной математики:

$$p_{21}, \quad {}_5p_{21}, \quad q_{21}, \quad {}_5q_{21}, \quad {}_{1|}q_{21}, \quad {}_{3|}q_{25}, \quad {}_{3|4}q_{29}, \quad \frac{q_{60}}{q_{20}}, \quad \frac{q_{80}}{q_{20}}.$$

Используя ТПЖ Приложения 1 оцените приведенные выше величины.

Задание 3.6.4

Доказать, что

$${}_{t|u}q_x = {}_tp_x \cdot {}_uq_{x+t}.$$

Задание 3.6.5

Используя выбранную из Задания 2.18.1 какую-либо таблицу для функции выживания определить вероятность того, что остаточное время жизни (20) лежит в промежутке от 30 до 40 лет, от 40 до 50 лет, от 70 до 80 лет.

Задание 3.6.6

Найдите дисперсию частичной продолжительности жизни в модели де Муавра, постройте ее графики в зависимости от возраста страхующегося для $n = 5$ и $n = 10$ лет.

Задание 3.6.7

Пусть функции выживания задаются формулами:

$$s_1(x) = \frac{x+a}{a} e^{-x/a}, \quad x \geq 0; \quad s_2(x) = \sqrt{1 - \frac{x}{110}}, \quad 0 \leq x \leq 110.$$

Найдите:

1) среднюю продолжительность жизни $\overset{\circ}{e}_{\circ}$;

2) полную вероятную продолжительность жизни $\overset{\circ}{e}_x$;

3) дисперсию среднего остаточного времени жизни $\mathbf{D}T(x)$;

4) частичную среднюю продолжительность жизни $\overset{\circ}{e}_{x:\overline{n}|}$;

5) дисперсию частичной продолжительности жизни $\mathbf{D}\{\min(T(x), n)\}$.

Изобразите графически полученные зависимости, проведите сравнительный анализ.

Задание 3.6.8

По данным ТПЖ (Приложение 1) отдельно для мужчин и женщин найдите:

1) средние округленные продолжительности жизни $e_{\circ}$, e_{10}, e_{21}, e_{40}, e_{70};

2) дисперсии округленных продолжительностей жизни $\mathbf{D}K(0)$, $\mathbf{D}K(21)$, $\mathbf{D}K(70)$, $\mathbf{D}K(84)$.

Обсудит

4 ДРОБНАЯ ПРОДОЛЖИТЕЛЬНОСТЬ ЖИЗНИ

4.1 Сплайновые аппроксимации для дробных возрастов (fractional ages)

Реальная статистика доступна обычно только для целых значений x (в годах), что обусловлено как определенными традициями и удобством сбора статистических данных, так и формой их представления в ТПЖ, где аргументы x, как правило, принимают значения $0, 1, 2, \dots$. Так как большинство клиентов в свой день рождения в страховую компанию не приходят, то для работы с конкретными индивидуумами нужно уметь находить приближения некоторых вероятностных характеристик для дробных возрастов по их известным значениям для целых x.

Рассмотренная задача есть типичная задача интерполяции, причем можно ограничиться ее решением только для функции выживания $s(x)$, так как другие величины можно выразить через $s(x)$.

В актуарной математике эту задачу решают с помощью так называемых сплайнов. Рассмотрим три постулата, дающие различные приближения:

1. равномерное распределение смертей,
2. постоянная интенсивность смертности,
3. предположение Балдуччи (Balducci).

Равномерное распределение смертей

В этом случае функция выживания интерполируется линейной функцией вида $s(x) = a_n + b_n x$ при $n \le x \le n+1$. Так как $s(n)$ и $s(n+1)$ известны (например, из ТПЖ), составляем уравнения

$$\begin{aligned} a_n + b_n n &= s(n) \\ a_n + b_n(n+1) &= s(n+1) \end{aligned}$$

и определяем неизвестные a_n и b_n (вычитаем из второго уравнения первое уравнение):

$$\begin{aligned} b_n &= s(n+1) - s(n), \\ a_n &= s(n)(n+1) - s(n+1)n. \end{aligned}$$

Отметим, что $b_n < 0$.

Следовательно, на отрезке $n \le x \le n+1$ функция $s(x)$ аппроксимируется линейным сплайном вида

$$\boxed{s(x) = (n+1-x)s(n) + (x-n)s(n+1),} \qquad (4.1.1)$$

$n \le x \le n+1$. Отсюда для кривой смертей $f(x)$ и интенсивности смертности μ_x получаем соответственно:

$$f(x) = -s^{'}(x) = s(n) - s(n+1), \quad n < x < n+1,$$

$$\mu_x = \frac{f(x)}{s(x)} = \frac{s(n) - s(n-1)}{(n+1-x)s(n) + (x-n)s(n+1)} =$$

$$= \frac{s(n) - s(n+1)}{(n+1)s(n) - ns(n+1) - x[s(n) - s(n+1)]}. \tag{4.1.2}$$

Таким образом,

$$\boxed{f(x) = -b_n = s(n) - s(n+1),}$$

$n < x < n+1$.

С помощью ранее введенной величины $q_n = \dfrac{s(n) - s(n+1)}{s(n)}$, равной вероятности того, что человек в возрасте n лет умрет в течении ближайшего года, преобразуем формулу (4.1.2) к более удобному виду:

$$\mu_x = \frac{s(n) - s(n+1)}{s(n) + (x-n)(s(n+1) - s(n))} = \frac{q_n}{1-(x-n)q_n},$$

$n < x < n+1$. Видим, что такое приближение влечет возрастание интенсивности смертности

$$\boxed{\mu_x = \frac{q_n}{1-(x-n)q_n}} \tag{4.1.3}$$

между узлами интерполяции ($n < x < n+1$), а плотность распределения $f(x) = \text{const}$ не меняется, причем в целочисленных точках $f(x)$ и μ_x не определены.

Заметим, что $q_n > s(n) - s(n+1)$, так как $s(n) < 1$.

Постоянная интенсивность смертности

Приблизим функцию $s(x)$ на отрезке $n \le x \le n+1$ убывающей показательной функцией $a_n e^{-b_n x}$. При этом уравнения принимают вид

$$a_n e^{-b_n n} = s(n)$$
$$a_n e^{-b_n (n+1)} = s(n+1),$$

и, разделив второе из них на первое, получаем

$$-b_n = \ln \frac{s(n+1)}{s(n)}, \quad a_n = s(n) e^{b_n n} = s(n) \left(\frac{s(n+1)}{s(n)} \right)^{-n},$$

т.е.

$$b_n = -\ln p_n, \quad a_n = s(n) p_n^{-n},$$

где p_n есть вероятность того, что человек в возрасте n лет проживет по крайней мере еще один год.

В этом случае

$$s(x) = a_n e^{-b_n x} = s(n) p_n^{-n} e^{(\ln p_n)x} = s(n) p_n^{x-n}, \quad n \le x \le n+1,$$

и основные вероятностные характеристики приближенно выражаются следующим образом:

$$\boxed{s(x) = s(n) p_n^{x-n},} \quad n \le x \le n+1,$$

$$\boxed{f(x) = -s'(x) = -s(n) p_n^{x-n} \ln p_n,} \quad \boxed{\mu_x = \frac{f(x)}{s(x)} = -\ln p_n,}$$

$n < x < n+1$, т.е. между узлами интерполяции $\mu_x = \mathrm{const}$.

Заметим, что $-\ln p_n > -s(n) \ln p_n$, так как $s(n) < 1$.

Предположение Балдуччи

В этом случае линейной функцией вместо $s(x)$ интерполируется $s^{-1}(x)$. Формально заменяя в (4.1.1) $s(x)$ на $\dfrac{1}{s(x)}$, немедленно приходим к соотношению

$$\frac{1}{s(x)} = \frac{n+1-x}{s(n)} + \frac{x-n}{s(n+1)}, \quad n \le x \le n+1,$$

откуда

$$s(x) = \frac{s(n)s(n+1)}{(n+1-x)s(n+1) + (x-n)s(n)} = \frac{s(n+1)}{(n+1-x)p_n + x - n} =$$

$$= \frac{s(n+1)}{p_n + (x-n)q_n}, \quad n \le x \le n+1,$$

$$f(x) = -s'(x) = \frac{s(n)s(n+1)(s(n) - s(n+1))}{(s(n+1)(n+1-x) + (x-n)s(n))^2} =$$

$$= \frac{s(n+1)q_n}{(p_n + (x-n)q_n)^2}, \quad n < x < n+1,$$

$$\mu_x = \frac{f(x)}{s(x)} = \frac{q_n}{p_n + (x-n)q_n}, \quad n < x < n+1.$$

Пример 4.1.1. Подсчитайте вероятность того, что мужчина СССР (80) середины 80-х годов при предположении о равномерном распределении смертей умрет в возрасте от $80\dfrac{1}{2}$ до $81\dfrac{1}{2}$ лет.

Р е ш е н и е. Воспользуемся формулой

$$s(x) = (n+1-x)s(n) + (x-n)s(n+1), n \le x \le n+1,$$

согласно которой

$$s\left(80\frac{1}{2}\right) = \left(81 - 80\frac{1}{2}\right)s(80) + \left(80\frac{1}{2} - 80\right)s(81) = 0,5(s(80)+s(81)),$$

$$s\left(81\frac{1}{2}\right) = 0,5(s(81)+s(82)).$$

Таким образом,

$$\mathbf{P}\left(\frac{1}{2} < \mathbf{T}(80) < 1\frac{1}{2}\right) = \frac{s\left(80\frac{1}{2}\right) - s\left(81\frac{1}{2}\right)}{s(80)} =$$

$$= 0,5\frac{s(80)+s(81)-s(81)-s(82)}{s(80)} =$$

$$= 0,5\left(1 - \frac{s(82)}{s(80)}\right) = 0,5\left(1 - \frac{s(82)}{s(81)}\frac{s(\dot{8}1)}{s(80)}\right) =$$

$$= 0,5(1 - p_{81}p_{80}) = 0,5(1-(1-q_{81})(1-q_{80})) =$$

$$= 0,5(1-(1-0,12548)(1-0,11672)) =$$

$$= 0,5(1-0,87452\dot{0},88328) = 0,11378.$$

4.2 Распределение дробного возраста

Введем случайную величину $\tau = \{X\}$, где $\{X\}$ обозначает дробную часть величины X. Теперь продолжительность жизни X можно представить суммой целой и дробной частей: $X = K(0) + \tau$, где $K(0) = [X]$ — округленное время жизни. Понятно, что величина τ описывает момент смерти внутри года. Найдем условное распределение τ при условии, что смерть наступила в возрасте n лет:

$$\mathbf{P}\{\tau \le t | K(0) = n\} = \mathbf{P}\{X - K(0) \le t | K(0) = n\} =$$

$$= \mathbf{P}\{X \le t+n | n \le X < n+1\} = \frac{\mathbf{P}\{X \le t+n, n \le X < n+1\}}{\mathbf{P}\{n \le X < n+1\}} =$$

$$= \frac{\mathbf{P}\{n \le X \le n+t\}}{\mathbf{P}\{n \le X < n+1\}} = \frac{s(n)-s(n+t)}{s(n)-s(n+1)}, \quad 0 < t < 1. \qquad (4.2.1)$$

Величина $s(n+t)$, точнее l_{n+t}, когда n — целое, а $0 < t < 1$, в ТПЖ отсутствует, поэтому при ее нахождении воспользуемся приближениями для дробных возрастов.

При постулате равномерного распределения смертей формула (4.2.1), если взять в $s(x)$ аргумент $x = n+t$, преобразуется следующим образом:

$$\mathbf{P}\{\tau \le t | K(0) = n\} =$$

$$= \frac{s(n) - [(n+1-n-t)s(n) + (n+t-n)s(n+1)]}{s(n)-s(n+1)} =$$

$$= \frac{t(s(n)-s(n+1))}{s(n)-s(n+1)} = t, \quad 0 < t < 1.$$

$$\mu_{t|\cdot} = \frac{f(t|\cdot)}{1-F(t|\cdot)} = \frac{1}{1-t}$$

Итак, при этой интерполяции

1) смерть в любой день между двумя днями рождений человека равновероятна;

2) условное распределение $\mathbf{P}\{\tau \le t | K(0) = n\}$ не зависит от n и поэтому совпадает с безусловным распределением $\mathbf{P}(\tau \le t)$;

3) случайные величины $K(0)$ и τ независимы.

Для постулата постоянной интенсивности смертности аналогично имеем:

$$\mathbf{P}\{\tau \le t | K(0) = n\} = \frac{s(n)-s(n+t)}{s(n)-s(n+1)} =$$

$$= \frac{s(n)-s(n)p_n^{n+t-n}}{s(n)-s(n+1)} = \frac{s(n)(1-p_n^t)}{s(n)-s(n+1)} = \frac{1-p_n^t}{1-p_n}, \quad 0 < t < 1.$$

$$f(t|\cdot) = F'(t|\cdot) = \frac{-p_n^t \ln p_n}{1-p_n},$$

$$s(t|\cdot) = 1 - F(t|\cdot) = 1 - \frac{1-p_n^t}{1-p_n} = \frac{1-p_n-1+p_n^t}{1-p_n} = \frac{p_n(p_n^{t-1}-1)}{1-p_n},$$

$$\mu_{t|\cdot} = \frac{f(t|\cdot)}{s(t|\cdot)} = \frac{-p_n^t \ln p_n}{p_n(p_n^{t-1}-1)} = \frac{-p_n^{t-1}\ln p_n}{p_n^{t-1}-1}.$$

Рассуждая аналогично, также можно получить формулы для $f(t|\cdot)$, $s(t|\cdot)$, $\mu_(t|\cdot)$ и при постулате Балдуччи.

Замечание. Вообще говоря, формулы при постоянной интенсивности смертности удобнее выразить через $q_n = 1 - p_n$, так как величина q_n имеется в ТПЖ.

4.3 Среднее и дисперсия дробного возраста

Найдем среднее дробного возраста τ при условии, что смерть наступила в возрасте n лет:

$$a(n) = \mathbf{E}\{\tau|K(0) = n\} = \int_0^1 \mathbf{P}\{\tau > t|K(0) = n\}dt.$$

Очевидно, что

$$\mathbf{P}\{\tau > t|K(0) = n\} = 1 - \mathbf{P}\{\tau \leq t|K(0) = n\} =$$

$$= 1 - \frac{s(n) - s(n+t)}{s(n) - s(n+1)} = \frac{s(n) - s(n+1) - sn) + s(n+t)}{s(n) - s(n+1)} =$$

$$= \frac{s(n+t) - s(n+1)}{s(n) - s(n+1)}.$$

Отсюда

$$a(n) = \frac{1}{s(n) - s(n+1)} \int_0^1 [s(n+t) - s(n+1)]\mathrm{dt}.$$

Подсчитаем теперь величину $a(n)$ для всех трех предположений о характере смертности для дробных возрастов.

Равномерное распределение смертей. Ясно, что

$$a(n) = \frac{\int_0^1 [(n+1-n-t)s(n) + (n+t-n)s(n+1) - s(n+1)]\mathrm{dt}}{s(n) - s(n+1)} =$$

$$= \frac{\int_0^1 [(1-t)s(n) + ts(n+1) - s(n+1)]\mathrm{dt} = \int_0^1 (1-t)\mathrm{dt}}{s(n) - s(n+1)} = \frac{1}{2},$$

т.е. $a(n)$ совпадает с серединой одногодичного временного промежутка, что интуитивно мы и ожидали получить.

Постоянная интенсивность смертности. В этом случае

$$a(n) = \frac{1}{s(n) - s(n+1)} \int_0^1 [s(n)p_n^{n+t-n} - s(n+1)]\mathrm{dt} =$$

$$= \frac{s(n)}{s(n)-s(n+1)} \int_0^1 \left[p_n^t - p_n\right] \mathrm{dt} = \frac{1}{q_n} \left[\left. \frac{p_n^t}{\ln p_n} \right|_0^1 - p_n \right] =$$

$$= \frac{1}{q_n} \left(\frac{p_n - 1}{\ln p_n} - p_n \right) = \frac{1}{q_n} \left(-p_n - \frac{q_n}{\ln p_n} \right) = -\frac{1}{\ln p_n} - \frac{p_n}{q_n}.$$

Поскольку $p_n = 1 - q_n$, а величина q_n достаточно мала, то, воспользовавшись представлением

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots - -\frac{x^n}{n} + o(x^n),$$

разложим $\ln p_n$ в ряд по степеням q_n

$$\ln p_n = -q_n - \frac{q_n^2}{2} - \frac{q_n^3}{3} - ...,$$

после чего приходим к следующим оценкам для $a(n)$:

$$a(n) = \frac{1}{2} - \frac{q_n}{12} + O(q_n^2) = \frac{1}{2} - \frac{q_n}{12} + o(q_n). \tag{4.3.1}$$

Докажем равенство (4.3.1). Так как

$$-\frac{p_n}{q_n} - \frac{1}{\ln p_n} = -\frac{1-q_n}{q_n} + \frac{1}{q_n + \frac{q_n^2}{2} + \frac{q_n^3}{3} + ...} =$$

$$= 1 + \frac{1}{q_n + \frac{q_n^2}{2} + \frac{q_n^3}{3} + ...} - \frac{1}{q_n} = 1 - \frac{\frac{q_n^2}{2} + \frac{q_n^3}{3} + ...}{q_n^2 + \frac{q_n^3}{2} + \frac{q_n^4}{3} + ...},$$

то, разложив функцию (отношение) двух переменных в ряд Тейлора в окрестности точки $\left(\frac{q_n^2}{2}, q_n^2\right)$, получаем

$$-\frac{p_n}{q_n} - \frac{1}{\ln p_n} = 1 - \left[\frac{1}{2} + \frac{1}{q_n^2} \left(\frac{q_n^3}{3} + ... \right) - \frac{\frac{q_n^2}{2}}{q_n^4} \left(\frac{q_n^3}{2} + ... \right) \right] =$$

$$= \frac{1}{2} - \frac{q_n}{3} + \frac{q_n}{4} + o(q_n) = \frac{1}{2} - \frac{q_n}{12} + o(q_n),$$

что и требовалось доказать.

Если q_n не слишком мало, то имеет смысл учитывать и слагаемые порядка $O(q_n^2)$:

$$a(n) = 1 - \frac{1}{q_n^2}\left(\frac{q_n^3}{3} + \frac{q_n^4}{4} + ...\right) + \frac{1}{2q_n^2}\left(\frac{q_n^3}{2} + \frac{q_n^4}{3} + ...\right) =$$

$$= \frac{1}{2} - \frac{q_n}{3} + \frac{q_n}{4} - \frac{q_n^2}{4} + \frac{q_n^2}{6} + ... = \frac{1}{2} - \frac{q_n}{12} - \frac{q_n^2}{12} + o(q_n^2).$$

Постулат Балдуччи. Здесь

$$a(n) = \frac{1}{s(n) - s(n+1)} \int_0^1 \left(\frac{s(n+1)}{p_n + tq_n} - s(n+1)\right) \mathrm{dt} =$$

$$= \frac{s(n+1)}{s(n) - s(n+1)} \int_0^1 \left(\frac{1}{p_n + tq_n} - 1\right) \mathrm{dt} = \frac{p_n}{q_n} \int_0^1 \left(\frac{1}{p_n + tq_n} - 1\right) \mathrm{dt} =$$

$$= \frac{p_n}{q_n}\left[\frac{\ln(p_n + tq_n)}{q_n}\bigg|_0^1 - 1\right] = \frac{p_n}{q_n}\left[-\frac{\ln p_n}{q_n} - 1\right] =$$

$$-\frac{p-n}{q_n^2}(q_n + \ln p - n) = \frac{q_n - 1}{q_n^2}\left(q_n - q_n - \frac{q_n^2}{2} - \frac{q_n^3}{3} - ...\right) =$$

$$= (q_n - 1)\left(-\frac{1}{2} - \frac{q_n}{3} - \frac{q_n^2}{4} - ...\right) = \frac{1}{2} - \frac{q_n}{2} + \frac{q-n}{3} - \frac{q_n^2}{3} + \frac{q_n^2}{4} + o(q_n) =$$

$$= \frac{1}{2} - \frac{q_n}{6} - \frac{q_n^2}{12} + o(q_n).$$

Дисперсию будем находить по следующей формуле:

$$b(n) = \mathbf{D}\{\tau | K(0) = n\} =$$

$$= \frac{2}{s(n) - s(n+1)} \int_0^1 [s(n+t) - s(n+1)]\mathrm{dt} - a^2(n).$$

Подсчитаем величину $b(n)$ для всех трех постулатов смертности дробных возрастов.

Равномерное распределение смертей. Для этого постулата

$$b(n) = \frac{2}{s(n) - s(n+1)} \int_0^1 t(1-t)[s(n) - s(n+1)]\mathrm{dt} - \frac{1}{4} =$$

$$= 2\int_0^1 (t - t^2)\mathrm{dt} - \frac{1}{4} = 2\left(\frac{1}{2} - \frac{1}{3}\right) - \frac{1}{4} = 1 - \frac{2}{3} - \frac{1}{4} = \frac{1}{12}.$$

Постоянная интенсивность смертности. Ранее мы получили, что $a(n)=\dfrac{1}{q_n}\displaystyle\int_0^1[p_n^t-p_n]\mathrm{dt}$, поэтому

$$b(n)=\frac{2}{q_n}\int_0^1 t(p_n^t-p_n)\mathrm{dt}-a^2(n)=$$

$$=\frac{2}{q_n}\left[\int_0^1 tp_n^t\mathrm{dt}-\int_0^1 tp_n\mathrm{dt}\right]-\left[\frac{1}{2}-\frac{q_n}{12}+o(q_n)\right]^2=$$

$$=\frac{2}{q_n}\left[\frac{p_n}{\ln p_n}-\frac{p_n-1}{\ln^2 p_n}-\frac{p_n}{2}\right]-\left[\frac{1}{4}-\frac{q_n}{12}+o(q_n)\right]=$$

$$=2\left[\frac{1-q_n}{q_n\ln p_n}+\frac{q_n}{q_n\ln^2 p_n}-\frac{1-q_n}{2q_n}\right]-[...]=$$

$$=2\left[\frac{1}{q_n\ln p_n}-\frac{1}{\ln p_n}+\frac{1}{\ln^2 p_n}-\frac{1}{2q_n}+\frac{1}{2}\right]-[...]=$$

$$=2\left[\frac{2\ln p_n-2q_n\ln p_n+2q_n-\ln^2 p_n}{2q_n\ln^2 p_n}+\frac{1}{2}\right]-[...]=$$

$$=2\left[\frac{-2q_n-q_n^2-\frac{2}{3}q_n^3-\frac{1}{2}q_n^4-...+2q_n^2+q_n^3+\frac{2}{3}q_n^4+...}{2q_n^3+2q_n^4+...}\right]+$$

$$+2\left[\frac{2q_n-q_n^2-q_n^3-\frac{11}{12}q_n^4-...}{2q_n^3+2q_n^4+...}+\frac{1}{2}\right]-[...]=$$

$$=2\left[\frac{-\frac{2}{3}q_n^3-\frac{3}{4}q_n^4-...}{2q_n^3+2q_n^4+...}+\frac{1}{2}\right]-[...].$$

После разложения отношения в ряд Тейлора в окрестности точки $\left(-\dfrac{2}{3}q_n^3,2q_n^3\right)$, имеем

$$b(n)=2\left[\frac{1}{2}-\frac{1}{3}-\frac{1}{2q_n^3}\cdot\frac{3}{4}q_n^4+\frac{\frac{2}{3}q_n^3}{4q_n^6}2q_n^4-...\right]-[...]=$$

$$=2\left[\frac{1}{6}-\frac{3}{8}q_n+\frac{1}{3}q_n-...\right]-\left[\frac{1}{4}-\frac{q_n}{12}+...\right]=\frac{1}{12}+o(q_n).$$

Постулат Балдуччи. Здесь

$$b(n)=\frac{2p_n}{q_n}\int_0^1 t\left(\frac{1}{p_n+tq_n}-1\right)\mathrm{dt}-a^2(n)=$$

$$= \frac{2p_n}{q_n}\left[\int_0^1 \frac{t}{p_n+tq_n}\mathrm{dt} - \int_0^1 t\mathrm{dt}\right] - a^2(n) =$$

$$= \frac{2p_n}{q_n^2}\left[\int_0^1 \frac{p_n+tq_n-p_n}{p_n+tq_n}\mathrm{dt}\right] - \frac{1}{2}\cdot\frac{2p_n}{q_n} - a^2(n) =$$

$$= \frac{2p_n}{q_n^2}\left[1 - p_n\int_0^1 \frac{\mathrm{dt}}{p_n+tq_n}\right] - \frac{p_n}{q_n} - a^2(n) =$$

$$= \frac{2p_n}{q_n^2}\left[1 - \frac{p_n}{q_n}\ln(p_n+tq_n)\Big|_0^1\right] - \frac{p_n}{q_n} - a^2(n) =$$

$$= \frac{2p_n}{q_n^2}\left[1 + p_n\frac{p_n\ln p_n}{q_n}\right] - \frac{p_n}{q_n} - a^2(n) =$$

$$= \frac{2(1-q_n)}{q_n^2} + \frac{2(1-q_n)^2\ln p_n}{q_n^3} - \frac{1-q_n}{q_n} - a^2(n) =$$

$$= \frac{2}{q_n^2} - \frac{2}{q_n} + \frac{2(1-2q+q_n^2)\ln p_n}{q_n^3} - \frac{1}{q_n} + 1 - a^2(n) =$$

$$= \frac{2}{q_n^2} - \frac{2}{q_n} + \frac{2\ln p_n}{q_n^3} - \frac{4\ln p_n}{q_n^2} + \frac{2\ln p_n}{q_n} - \frac{1}{q_n} + 1 - a^2(n) =$$

$$= \frac{2}{q_n^3}q_n^2 - \frac{2}{q_n} + \frac{2}{q_n^3}\left(-q_n - \frac{q_n^2}{2} - \frac{q_n^3}{3} - \frac{q_n^4}{4} - ...\right) +$$

$$+\frac{4}{q_n^2}\left(q_n + \frac{q_n^2}{2} + \frac{q_n^3}{3} + ...\right) + \frac{2}{q_n}\left(-q_n - \frac{q_n^2}{2} - ...\right) - \frac{1}{q_n} + 1 - a^2(n) =$$

$$= \frac{2}{q_n^2} - \frac{2}{q_n} - \frac{2}{q_n^2} - \frac{1}{q_n} - \frac{2}{3} - \frac{1}{2}q_n - ... + \frac{4}{q_n} + 2 + \frac{4}{3}q_n + ... -$$

$$-2 - q_n - ... - \frac{1}{q_n} + 1 - \frac{1}{4} + \frac{q_n}{6} + ... =$$

$$= 1 - \frac{1}{4} - \frac{2}{3} - \frac{q_n}{2} + \frac{4}{3}q_n - q_n + \frac{q_n}{6} + ... = \frac{1}{12} + o(q_n).$$

Обсудим полученные результаты. Для женщин СССР возраста 30 лет $q_{30} = 0,00106$. Поэтому порядок ошибки для средних и дисперсий постулатов постоянной интенсивности смертности и Балдуччи составляет $O(q_{30}^2) \approx 10^{-6}$. Для больших возрастов формулы для $a(n)$ и $b(n)$ следует использовать с учетом слагаемых q_n^2, так как, например, для женщин СССР возраста 82 лет $q_{82} = 0,10155$, и в этом случае $O(q_{82}^2) \approx 10^{-2}$.

Также на практике в некоторых случаях согласно полученным выше результатам, предположив независимость $K(0)$ и τ, можно использовать следующие простые аппроксимации для среднего и дисперсии остаточного времени жизни:

$$\overset{\circ}{e}_x \approx e_x + \frac{1}{2},$$

$$\mathbf{D}T(x) \approx \mathbf{D}K(0) + \frac{1}{12}.$$

4.4 Табличные величины L_x, T_x, их связь между собой и с $a(x)$

Величины L_x и T_x используют в более подробных ТПЖ. В качестве образца в Приложении 3 приведен фрагмент ТПЖ для населения США (1979–81) [1, с.55–58].

Будем обозначать символом L_x среднее суммарное число лет, прожитое между моментами x и $x+1$, x — целое, всеми представителями исходной группы l_0. Понятно, что оно складывается из двух слагаемых:

1) среднее суммарное число лет, прожитое между моментами x и $x+1$ теми представителями исходной группы, которые умерли в возрасте от x до $x+1$ (эта величина равна $d_x a(x)$);

2) среднее суммарное число лет, даваемое живыми представителями исходной группы к моменту $x+1$ (эта величина равна $l_{x+1} \cdot 1$год $= l_{x+1}$).

Итак,

$$L_x = d_x a(x) + l_{x+1} = \frac{d_x}{s(x) - s(x+1)} \int_0^1 [s(x+t) - s(x+1)]\mathrm{dt} + l_{x+1} =$$

$$= \frac{d_x}{l_x - l_{x+1}} \int_0^1 [l_{x+t} - lx + 1]\mathrm{dt} + l_{x+1} =$$

$$= \int_0^1 [l_{x+t} - lx + 1]\mathrm{dt} + l_{x+1} = \int_0^1 l_{x+t}\mathrm{dt}. \qquad (4.4.1)$$

Поскольку в ТПЖ обычно табулируют величины l_x и L_x, то с их помощью можно подсчитать $a(x)$. Действительно, из (4.4.1) следует, что

$$L_x - l_{x+1} = d_x a(x),$$

откуда

$$a(x) = \frac{L_x - l_{x+1}}{l_x - l_{x+1}} = \frac{L_x - l_{x+1}}{d_x}. \qquad (4.4.2)$$

Пример 4.4.1. По данным общей ТПЖ (Приложение 3) найдите $a(20)$, $a(80)$, $a(106)$, $a(107)$, $a(108)$.

Р е ш е н и е. Согласно формуле (4.4.2) имеем

$$a(20) = \frac{L_{20} - l_{21}}{d_20} = \frac{97682 - 97623}{118} = \frac{59}{118} = 0,5;$$

$$a(80) = \frac{L_{80} - l_{81}}{d_80} = \frac{41694 - 40208}{2972} = \frac{1486}{2972} = 0,5;$$

$$a(100) = \frac{983 - 815}{335} = 0,501; \quad a(106) = \frac{99 - 78}{41} = 0,512;$$

$$a(107) = \frac{99 - 78}{41} = 0,512; \quad a(108) = \frac{42 - 33}{18} = 0,5.$$

Так как $a(i) \sim 0,5$, то из этого примера видно, что, по-видимому, имеет место равномерная смертность между двумя днями рождения для индивидуумов любых возрастов.

Будем обозначать символом T_x среднее суммарное число лет, прожитых всеми представителями группы из l_0 новорожденных на интервале (x, ∞). Понятно, что

$$T_x = l_0 \mathbf{E}[(X - x)I(x - x > 0)] = l_0 \int_0^\infty (X - x) d\mathbf{P}\{X - x \le t\} =$$

$$= l_0 \int_0^\infty \mathbf{P}\{X - x > t\}\mathrm{dt} = l_0 \int_0^\infty \mathbf{P}\{X > x + t\}\mathrm{dt} =$$

$$= l_0 \int_0^\infty s(x + t)\mathrm{dt} = l_0 \int_x^\infty s(u)\mathrm{du}.$$

Теперь формуле $\overset{\circ}{e}_x = \mathbf{E}T(x)$ можно придать следующий вид:

$$\overset{\circ}{e}_x = \mathbf{E}T(x) = \frac{1}{s(x)} \int_0^\infty s(u)\mathrm{du} = \frac{1}{l_0 s(x)} l_0 \int_0^\infty s(u)\mathrm{du} = \frac{T_x}{l_x}. \qquad (4.4.3)$$

Пример 4.4.2. По данным общей ТПЖ (Приложение 3) найдите $\overset{\circ}{e}_{20}$, $\overset{\circ}{e}_{80}$.

Р е ш е н и е. Согласно формуле (4.4.3) имеем

$$\overset{\circ}{e}_{20} = \frac{T_{20}}{l_20} = \frac{5420937}{97741} = 55,462; \quad \overset{\circ}{e}_{80} = \frac{T_{80}}{l_80} = \frac{344612}{43180} = 7,98.$$

Также с помощью величин L_n, $n \geq x$, можно подсчитать величину T_x :

$$T_x = l_0 \int_x^{\infty} s(u)\mathrm{du} = l_0 \sum_{k=0}^{\infty} \int_{x+k}^{x+k+1} s(u)\mathrm{du} =$$

$$= l_0 \sum_{k=0}^{\infty} \int_0^1 s(x+k+t)\mathrm{dt} = \sum_{k=0}^{\infty} \int_0^1 l_{x+k+t}\mathrm{dt} = \sum_{k=0}^{\infty} L_{x+k} = \sum_{n=x}^{\infty} L_n.$$

Таким образом,

$$\boxed{T_x = \sum_{n=x}^{\infty} L_n.} \tag{4.4.4}$$

Пример 4.4.3. По данным общей ТПЖ (Приложение 3) с помощью формулы (4.4.4) подсчитайте T_{100} и T_{105}.

Р е ш е н и е. Удобнее сначала найти T_{105} :

$$T_{105} = L_{105} + L_{106} + L_{107} + L_{108} + L_{109} + \sum_{n=110}^{\infty} L_n = L_{105} + L_{106} + L_{107} +$$

$$+ L_{108} + L_{109} + (T_{109} - L_{109}) = 150 + 99 + 64 + 42 + 27 + (73 - 27) = 428.$$

Далее,

$$T_{100} = T_{105} + L_{104} + L_{103} + L_{102} + L_{101} + L_{100} =$$

$$= 428 + 223 + 330 + 481 + 692 + 983 = 3137.$$

Можно применять для интерполяции, вообще говоря, и параболические, и кубические сплайны, но, по-видимому, улучшение точности результатов не оправдывается усложнением вычислительных формул.

4.5 Контрольные задания

Задание 4.5.1

Используя ТПЖ из Приложения 1 подсчитайте вероятности (в процентах), что мужчина (77) умрет в возрасте от $77\frac{5}{12}$ до $78\frac{11}{12}$ (интервал 1,5 года):

1) в предположении о равномерном распределении смертей,
2) в предположении Балдуччи для дробных возрастов,
3) в предположении о постоянной интенсивности смертности.

Проведите сравнительный анализ полученных результатов.

Задание 4.5.2

Вывести формулы для $f(t|\cdot)$, $s(t|\cdot)$, $\mu_{t|\cdot}$ в предположении Балдуччи для дробных возрастов.

Задание 4.5.3

Постройте графики кривых смертей в промежутке (77,79) в предположениях Задания 4.5.1.

Задание 4.5.4

Постройте графики кривых функций интенсивности смертности в промежутке (77,79) в предположениях Задания 4.5.1.

5 КОЛЛЕКТИВНОЕ СТРАХОВАНИЕ

5.1 Страхование жизни нескольких лиц. Статус совместной жизни (joint-life status)

Результаты, представленные в главах 1–4, могут быть обобщены на многомерный случай. Такие обобщения необходимы при расчетах, связанных с пенсионным страхованием, с коллективным страхованием жизни, здоровья и др. Теория излагается в главе 8 фундаментальной монографии [1, с.231–258, с.483-510] и [4, гл. 8].

Кратко изложим особенности страхования жизни нескольких лиц. Рассмотрим случай коллективного страхования жизни, для которого полезной абстракцией является понятие статуса. Пусть m индивидуумов с возрастами $(x_1, x_2, \dots, x_m)$ желают заключить страховой договор. Предстоящее время жизни k-го индивидуума обозначим через $T(x_k) = X - x_k$.

Совокупности m чисел $T(x_1), T(x_2), \dots, T(x_m)$ поставим в соответствие статус (status) U, которому соответствует своя продолжительность жизни $T(U)$. Двумя самыми распространенными статусами являются статус совместной жизни и статус выживания последнего [1].

Статус совместной жизни обозначается $U := x_1 : x_2 : \dots : x_m$ или $(x_1 : x_2 : \dots : x_m)$ и считается разрушенным, если наступила смерть хотя бы одного из индивидуумов, т.е.

$$T(U) = \min\{T(x_1), T(x_2), \dots, T(x_m)\}.$$

Понятно, что

$$\mathbf{P}\{T(U) > t\} = \mathbf{P}\{\min(T(x_1), T(x_2), \dots, T(x_m)) > t\} =$$
$$= \mathbf{P}\{T(x_1) > t, T(x_2) > t, \dots, T(x_m) > t\},$$

и в предположении независимости смертей

$$\mathbf{P}\{T(U) > t\} = \prod_{i=1}^{m} {}_tp_{x_i}. \tag{5.1.1}$$

Смысл вероятностей ${}_tp_{x_i} = \mathbf{P}\{T(x_i) > t\}$ определяется в п.3.2.

Теперь легко вывести другие вероятностные характеристики продолжительности жизни для $T(U)$, например,

$${}_tq_{x_1:x_2:\dots:x_m} = 1 - {}_tp_{x_1:x_2:\dots:x_m} = 1 - \prod_{i=1}^{m}(1 - {}_tq_{x_i}).$$

Для плотности распределения времени разрушения рассматриваемого статуса справедливы следующие соотношения:

$$f_{x_1:x_2:\ldots:x_m}(t) = -\frac{d}{d\,t}\mathbf{P}\{T(U) > t\} =$$

$$= -\frac{d}{d\,t}\prod_{i=1}^{m} {}_tp_{x_i} = -\frac{d}{d\,t}\prod_{i=1}^{m}\frac{s(x_i+t)}{s(x_i)}. \tag{5.1.2}$$

Пример 5.1.1. В предположении независимости смертей найти плотность распределения статуса совместной жизни двух индивидуумов $U := x_1 : x_2$

1) в общем случае;

2) для модели де Муавра.

Р е ш е н и е. Так как $m = 2$, а $T(U) = \min(T(x_1), T(x_2))$, то для плотности распределения статуса в соответствии с (5.1.2) имеем:

$$f_{x_1:x_2}(t) = \frac{d}{dt}\{-\mathbf{P}\{\min(T(x_1), T(x_2)) > t\}\} =$$

$$\left\{-\frac{s(x_1+t)}{s(x_1)}\,\frac{s(x_2+t)}{s(x_2)}\right\}_t' =$$

$$== \frac{f(x_1+t)}{s(x_1)}\,\frac{s(x_2+t)}{s(x_2)} + \frac{f(x_2+t)}{s(x_2)}\,\frac{s(x_1+t)}{s(x_1)} =$$

$$= f_{x_1}(t)\,s_{x_2}(t) + f_{x_2}(t)\,s_{x_1}(t) < f_{x_1}(t) + f_{x_2}(t),$$

где $s_x(t) = \dfrac{s(x+t)}{s(x)}$ — функция выживания случайной величины $T(x)$. Как будет показано позднее, доказанное неравенство позволяет снижать страховой компании размер премии участнику коллективного страхования по сравнению со случаем индивидуального страхования.

Для модели де Муавра $f_x(t)$ задается формулой (3.1.2), функция выживания $s_x(t) = I_t(-\infty, \omega - x) - \dfrac{t\,I_t(0, \omega - x)}{\omega - x}$, поэтому

$$f_{x_1:x_2}(t) = \frac{I_t(0, \omega - x_1)}{\omega - x_1}\left[I_t(-\infty, \omega - x_2) - \frac{t\,I_t(0, \omega - x_2)}{\omega - x_2}\right] +$$

$$+\frac{I_t(0, \omega - x_2)}{\omega - x_2}\left[I_t(-\infty, \omega - x_1) - \frac{t\,I_t(0, \omega - x_1)}{\omega - x_1}\right] =$$

$$= \left[\frac{\omega - x_2 - t}{(\omega - x_1)(\omega - x_2)} + \frac{\omega - x_1 - t}{(\omega - x_1)(\omega - x_2)} \right] I_t(0, \min(\omega - x_1, \omega - x_2)). \tag{5.1.3}$$

Перейдем теперь к функции интенсивности разрушения состояния совместной жизни. Функция интенсивности остаточного времени жизни $T(x) = X - x$ удовлетворяет равенствам

$$\mu_x(t) = \frac{f_x(t)}{s_x(t)} = \frac{f(x+t)}{s(x+t)} = \mu_{x+t} = -\frac{d}{d\,t} \ln s(x+t) =$$

$$= -\frac{d}{d\,t} [\ln s(x+t) - \ln s(x)] = -\frac{d}{d\,t} \ln \frac{s(x+t)}{s(x)} = -\frac{d}{d\,t} \ln {}_t p_x,$$

т.е.

$$\boxed{\mu_x(t) = \mu_{x+t} = -\frac{d}{d\,t} \ln {}_t p_x.} \tag{5.1.4}$$

Принимая во внимание (5.1.4), получаем

$$\mu_{x_1:x_2:\ldots:x_m}(t) = -\frac{d}{d\,t} \ln {}_t p_{x_1:x_2:\ldots:x_m} = -\frac{d}{d\,t} \sum_{i=1}^{m} \ln {}_t p_{x_i} = \sum_{i=1}^{m} \mu_{x_i}(t)$$

и, следовательно,

$$\boxed{\mu_{x_1:x_2:\ldots:x_m}(t) = \sum_{i=1}^{m} \mu_{x_i}(t).} \tag{5.1.5}$$

5.2 Упрощения для моделей Гомпертца и Мэйкхама

Если смертность всех людей рассматриваемой группы распределена по одному и тому же закону Гомпертца (см. п.2.6), то

$$\mu_{x_i}(t) = \mu_{x_i+t} = B\, e^{\alpha(x_i+t)} = B\, r^{x_i+t}, \tag{5.2.1}$$

где $r = e^{\alpha},\ t \geq 0,\ i = 1, \ldots, m$.

Равенство (5.1.5) с учетом (5.2.1) позволяет составить уравнение

$$r^{x_1} + r^{x_2} + \cdots + r^{x_m} = r^{\omega},$$

решив которое относительно ω, получаем

$$\omega = \frac{1}{\alpha} \ln \{ r^{x_1} + r^{x_2} + \cdots + r^{x_m} \} =$$

$$= \frac{1}{\alpha} \ln \{ e^{\alpha x_1} + e^{\alpha x_2} + \dots + e^{\alpha x_m} \} . \qquad (5.2.2)$$

Таким образом, функцию интенсивности разрушения состояния совместной жизни можно представить в виде

$$\boxed{\mu_{x_1:x_2:\ldots:x_m}(t) = \mu_\omega(t), \quad t \geq 0,}$$

откуда видно, что интенсивность разрушения состояния совместной жизни также подчиняется закону Гомпертца некоторого условного индивидуума с начальным возрастом ω, вычисляемым по формуле (5.2.2) и, естественно, выражаемым через начальные возрасты $x_1, x_2, \ldots, x_m$ всех участников коллективного страхования. Это позволяет все расчеты, касающиеся состояния совместной жизни, проводить в терминах одного индивидуума (ω).

Определенные упрощения возникают также, когда смертность всех рассматриваемых индивидуумов подчиняется одному и тому же закону Мэйкхама (см. п.2.6):

$$\mu_{x_i}(t) = \mu_{x_i+t} = A + B\, e^{\alpha(x_i+t)} = A + B\, r^{x_i+t},$$

$t \geq 0,\ i = 1, \ldots, m,\ r = e^\alpha$.

Составив равенство

$$A + B\, r^{x_1+t} + \dots + A + B\, r^{x_m+t} = m\left(A + B\, r^{\omega+t}\right)$$

и рассуждая как и в случае закона Гомпертца, приходим к уравнению

$$r^{x_1} + r^{x_2} + \dots + r^{x_m} = m\, r^\omega,$$

решение которого имеет вид:

$$\omega = \frac{1}{\alpha} \left[\ln\{ r^{x_1} + r^{x_2} + \dots + r^{x_m} \} - \ln m \right] =$$

$$= \frac{1}{\alpha} \left[\ln\{ e^{\alpha x_1} + e^{\alpha x_2} + \dots + e^{\alpha x_m} \} - \ln m \right] . \qquad (5.2.3)$$

Таким образом, для закона Гомпертца

$$\boxed{\mu_{x_1:x_2:\ldots:x_m}(t) = m\, \mu_\omega(t) = \mu_{\omega:\omega:\ldots:\omega}(t),}$$

и, следовательно, m лиц возраста $x_1, x_2, \ldots, x_m$ могут быть заменены m лицами одинакового "начального"возраста ω, который вычисляется по формуле (5.2.3).

5.3 Статус выживания последнего (last-survivor status)

Статус выживания последнего обозначается

$$U := \overline{x_1 : x_2 : \ldots : x_m} \text{или} (\overline{x_1 : x_2 : \ldots : x_m})$$

и считается разрушенным, если все представители коллектива умерли, т.е.

$$T(U) = \max(T(x_1), T(x_2), \ldots, T(x_m)).$$

Состояние совместной жизни соответствует последовательному соединению электролампочек в цепи, состояние выживания последнего — их параллельному соединению. Понятно, что

$$ {}_tq_{\overline{x_1:\ldots:x_m}} = \mathbf{P}\{T(U) \le t\} = \mathbf{P}\{\max(T(x_1), T(x_2), \ldots, T(x_m)) \le t\} =$$
$$= \mathbf{P}\{T(x_1) \le t, T(x_2) \le t, \ldots, T(x_m) \le t\},$$

и в предположении независимости смертей

$${}_tq_{\overline{x_1:x_2:\ldots:x_m}} = \prod_{i=1}^{m} {}_tq_{x_i}, \qquad {}_tp_{\overline{x_1:x_2:\ldots:x_m}} = 1 - \prod_{i=1}^{m} (1 - {}_tp_{x_i}).$$

Плотность распределения времени разрушения статуса выживания последнего равна

$$f_{\overline{x_1:x_2:\ldots:x_m}}(t) = \frac{d}{dt}\mathbf{P}\{T(U) \le t\} = \frac{d}{dt}\prod_{i=1}^{m} (1 - {}_tp_{x_i}).$$

Пример 5.3.1. В предположении независимости смертей найти плотность распределения статуса выживания последнего двух индивидуумов $U := \overline{x_1 : x_2}$

1) в общем случае;

2) для модели де Муавра.

Р е ш е н и е. Так как $m = 2$, а $T(U) = \max(T(x_1), T(x_2))$, то для плотности распределения статуса имеем:

$$f_{\overline{x_1:x_2}}(t) = \frac{d}{dt}\{\mathbf{P}(T(x_1) \le t)\mathbf{P}(T(x_2) \le t)\} = \frac{d}{dt}\{F_{x_1}(t)\, F_{x_2}(t)\} =$$
$$= \frac{d}{dt}\left\{\frac{F(x_1+t)}{s(x_1)}\,\frac{F(x_2+t)}{s(x_2)}\right\} = \frac{f(x_1+t)}{s(x_1)}\,\frac{F(x_2+t)}{s(x_2)} +$$
$$+ \frac{f(x_2+t)}{s(x_2)}\,\frac{F(x_1+t)}{s(x_1)} = f_{x_1}(t)\,F_{x_2}(t) + f_{x_2}(t)\,F_{x_1}(t) < f_{x_1}(t) + f_{x_2}(t).$$

Для модели де Муавра

$$f_{\overline{x_1:x_2}}(t) = \frac{I_t(0,\omega - x_1)}{\omega - x_1}\left[\frac{t\,I_t(0,\omega-x_2)}{\omega-x_2} + I_t(-\infty,\omega-x_2)\right] +$$

$$+\frac{I_t(0,\omega - x_2)}{\omega - x_2}\left[\frac{t\,I_t(0,\omega-x_1)}{\omega-x_1} + I_t(-\infty,\omega-x_1)\right] =$$

$$= \frac{2t\,I_t(0,\min(\omega-x_1,\omega-x_2))}{(\omega-x_1)(\omega-x_2)} +$$

$$+\frac{I_t(\min(\omega-x_1,\omega-x_2),\max(\omega-x_1,\omega-x_2))}{\min(\omega-x_1,\omega-x_2)}. \qquad (5.3.1)$$

Понятно, что

$$\mu_{\overline{x_1:\ldots:x_m}}(t) = \frac{f_{\overline{x_1:\ldots:x_m}}(t)}{s_{\overline{x_1:\ldots:x_m}}(t)} = \frac{d}{d\,t}\prod_{i=1}^{m}(1-{}_tp_{x_i}) \Bigg/ \left(1-\prod_{i=1}^{m}(1-{}_tp_{x_i})\right).$$

5.4 Примеры на оба статуса

Прежде всего проиллюстрируем возможности использования характеристик ТПЖ при вычислении вероятностей, связанных с рассмотренными статусами.

Пример 5.4.1. Предполагая, что $T(70)$ и $T(75)$ независимы, получить выражения вероятностей, что

(i) первая смерть произойдет в промежутке от 5 до 10 лет;

(ii) последняя смерть произойдет в том же промежутке.

(iii) Подсчитать эти вероятности для мужчин и женщин СССР, сравнить с соответствующими индивидуальными вероятностями.

Р е ш е н и е. (i) Для искомой вероятности статуса совместной жизни двух лиц (70 : 75) с учетом (5.1.1) справедлива следующая цепочка равенств:

$$\mathbf{P}\{5 < T(70:75) \le 10\} = \mathbf{P}\{T(70:75) > 5\} - \mathbf{P}\{T(70:75) > 10\} =$$

$$= {}_5p_{70:75} - {}_{10}p_{70:75} = {}_5p_{70}\,{}_5p_{75} - {}_{10}p_{70}\,{}_{10}p_{75} =$$

$$= \frac{l_{75}}{l_{70}}\frac{l_{80}}{l_{75}} - \frac{l_{80}}{l_{70}}\frac{l_{85}}{l_{75}} = \frac{l_{80}}{l_{70}}\left(1 - \frac{l_{85}}{l_{75}}\right).$$

(ii) Здесь для статуса выживания последнего двух лиц $(\overline{70:75})$ получаем:

$$\mathbf{P}\{5 < T(\overline{70:75}) \le 10\} = \mathbf{P}\{T(\overline{70:75}) \le 10\} - \mathbf{P}\{T(\overline{70:75}) \le 5\} =$$

$$= {}_{10}q_{70:75} - {}_{5}q_{70:75} = {}_{10}q_{70}\,{}_{10}q_{75} - {}_{5}q_{70}\,{}_{5}q_{75} =$$

$$= (1 - {}_{10}p_{70})(1 - {}_{10}p_{75}) - (1 - {}_{5}p_{70})(1 - {}_{5}p_{75}) =$$

$$= \left(1 - \frac{l_{80}}{l_{70}}\right)\left(1 - \frac{l_{85}}{l_{75}}\right) - \left(1 - \frac{l_{75}}{l_{70}}\right)\left(1 - \frac{l_{80}}{l_{75}}\right).$$

(iii) Согласно ТПЖ (Приложение 1) имеем для мужчин СССР

$$\mathbf{P}_m\{5 < T(70:75) \le 10\} = \frac{18,787}{43,405}\left(1 - \frac{9,063}{30,857}\right) = 0,3057,$$

$$\mathbf{P}_m\{5 < T(\overline{70:75}) \le 10\} = \left(1 - \frac{18,787}{43,405}\right)\left(1 - \frac{9,063}{30,857}\right) -$$

$$-\left(1 - \frac{30,857}{43,405}\right)\left(1 - \frac{18,7870}{30,857}\right) = 0,2875$$

$$\mathbf{P}_{m1} = \mathbf{P}\{5 < T(70) \le 10\} = \frac{l_{75}}{l_{70}} - \frac{l_{80}}{l_{70}} = \frac{30,857}{43,405} - \frac{18,787}{43,405} = 0,2781,$$

$$\mathbf{P}_{m2} = \mathbf{P}\{5 < T(75) \le 10\} = \frac{l_{80}}{l_{75}} - \frac{l_{85}}{l_{75}} = \frac{18,787}{30,857} - \frac{9,063}{30,857} = 0,3151,$$

а для женщин СССР

$$\mathbf{P}_w\{5 < T(70:75) \le 10\} = \frac{41,674}{70,043}\left(1 - \frac{24,265}{57,679}\right) = 0,3447,$$

$$\mathbf{P}_w\{5 < T(\overline{70:75}) \le 10\} = \left(1 - \frac{41,674}{70,043}\right)\left(1 - \frac{24,265}{57,679}\right) -$$

$$-\left(1 - \frac{57,679}{70,043}\right)\left(1 - \frac{41,674}{57,679}\right) = 0,1856,$$

$$\mathbf{P}_{w1} = \mathbf{P}\{5 < T(70) \le 10\} = \frac{57,679}{70,043} - \frac{41,674}{70,043} = 0,2285,$$

$$\mathbf{P}_{w2} = \mathbf{P}\{5 < T(75) \le 10\} = \frac{41,674}{57,679} - \frac{24,265}{57,679} = 0,30182.$$

Можно найти и различные числовые характеристики для статусов $U := \overline{x_1 : x_2 : \ldots : x_m}$ и $U := x_1 : x_2 : \ldots : x_m$, в частности, согласно п.3.3

$$\overset{\circ}{e}_U = \mathbf{E}T(U) = \int_0^\infty t\, f_U(t)\, dt = \int_0^\infty {}_tp_U\, dt,$$

$$\mathbf{D}T(U) = \int_0^{\infty} t^2\, f_U(t)\, dt - \left(\overset{\circ}{e}_U\right)^2 = 2\int_0^{\infty} t\, {}_tp_U\, dt - \left(\overset{\circ}{e}_U\right)^2.$$

Пример 5.4.2. В предположении независимости смертей найти $\overset{\circ}{e}_{\overline{x_1:x_2}}$ для модели де Муавра.

Р е ш е н и е. Воспользовавшись формулой (5.3.1), имеем

$$\overset{\circ}{e}_{\overline{x_1:x_2}} = \int_0^{\infty} t\, f_{\overline{x_1:x_2}}(t)\, dt = \int_0^{\min(\omega-x_1,\omega-x_2)} t\frac{2t}{(\omega-x_1)(\omega-x_2)}\, dt +$$

$$+\int_{\min(\omega-x_1,\omega-x_2)}^{\max(\omega-x_1,\omega-x_2)} t\frac{dt}{(\omega-x_1)(\omega-x_2)} =$$

$$= \frac{2\min^3(\omega-x_1,\omega-x_2)}{3(\omega-x_1)(\omega-x_2)} +$$

$$+\frac{\max^2(\omega-x_1,\omega-x_2)-\min^2(\omega-x_1,\omega-x_2)}{2\max(\omega-x_1)(\omega-x_2)} =$$

$$= \frac{2\min^2(\omega-x_1,\omega-x_2)}{3\max(\omega-x_1)(\omega-x_2)} +$$

$$+\frac{\max^2(\omega-x_1,\omega-x_2)-\min^2(\omega-x_1,\omega-x_2)}{2\max(\omega-x_1)(\omega-x_2)} =$$

$$= \frac{\min^2(\omega-x_1,\omega-x_2)+3\max^2(\omega-x_1,\omega-x_2)}{6\max(\omega-x_1)(\omega-x_2)}.$$

Пример 5.4.3. В предположении независимости смертей найти $\overset{\circ}{e}_{x:x}$ для модели де Муавра.

Р е ш е н и е. Воспользовавшись формулой (5.1.3), получаем

$$\overset{\circ}{e}_{x:x} = 2\int_0^{\omega-x} t\,\frac{\omega-x-t}{(\omega-x)^2)}\, dt = \frac{\omega-x}{3}.$$

5.5 Статусы k выживших, смешанные статусы (compound statuses)

Статус выживания k последних (k-survivor status) обозначается

$$U := \frac{\qquad k}{x_1 : x_2 : \ldots : x_m} \tag{5.5.1}$$

и существует до тех пор, пока живы по крайней мере k из m индивидуумов $(x_1), (x_2), \dots, (x_m)$, т.е. он считается разрушенным при наступлении $(m-k+1)$-й смерти. Понятно, что

$$\left(\frac{\hfill m}{x_1 : x_2 : \dots : x_m}\right) = (x_1 : x_2 : \dots : x_m),$$

$$\left(\frac{\hfill 1}{x_1 : x_2 : \dots : x_m}\right) = (\overline{x_1 : x_2 : \dots : x_m}),$$

и, следовательно, состояние совместной жизни ($k = m$) и состояние выживания последнего ($k = 1$) являются частными случаями статуса (5.5.2).

Точный статус выживания k последних ($[k]$-deferred survivor status) обозначается

$$U := \frac{\hfill [k]}{x_1 : x_2 : \dots : x_m} \tag{5.5.2}$$

и существует, если живы в точности k из m индивидуумов $(x_1), (x_2), \dots, (x_m)$, т.е. он начинается в момент $(m-k)$-й смерти и прекращается в момент $(m-k+1)$-й смерти. Этот статус находит широкое применение при расчете аннуинитетов (последовательностей платежей с ограниченным сроком длительности) [1, с.484], [4, с.107].

Итак, мы определили статусы для группы индивидуумов через общий статус k выживших. Отметим, что новые статусы можно также комбинировать из рассмотренных в данной главе базовых статусов.

Смешанным статусом назовем состояние, в основе которого лежит комбинация статусов, причем хотя бы один из них задан для более, чем одного индивидуума.

Пример 5.5.1. Опишите следующие смешанные статусы:
(i) $(\overline{x_1 : x_2} : \overline{x_3 : x_4})$; (ii) $\left(\overline{\overline{x_1 : x_2} : (x_3 : x_4)}\right)$; (iii) $(x_1 : x_2 : \overline{x_3 : x_4})$.

Р е ш е н и е. (i) Это состояние сохраняется, если жив по крайней мере один из (x_1) и (x_2) и по крайней мере один из (x_3) и (x_4). Моментом разрушения статуса $(\overline{x_1 : x_2} : \overline{x_3 : x_4})$ является

$$T(U) = \min\{\max\{T(x_1), T(x_2)\}, \max\{T(x_3), T(x_4)\}\}.$$

(ii) Такое состояние сохраняется, если живы по крайней мере двое из четырех, именно, (x_3) и (x_4), или, когда только один жив, и это либо (x_1), либо (x_2). Моментом разрушения статуса $\left(\overline{\overline{x_1 : x_2} : (x_3 : x_4)}\right)$ является

$$T(U) = \max\{\max\{T(x_1), T(x_2)\}, \min\{T(x_3), T(x_4)\}\}.$$

(iii) Состояние сохраняется, если живы (x_1), (x_2) и, когда ещё один жив, и это либо (x_3), либо (x_4). Моментом разрушения статуса ($x_1 : x_2 : \overline{x_3 : x_4}$) является

$$T(U) = \min\{T(x_1), T(x_2), \max\{T(x_3), T(x_4)\}\}.$$

5.6 Контрольные задания

Задание 5.6.1

Убедитесь, что плотность распределения (5.1.2) статуса совместной жизни двух индивидуумов $U := x_1 : x_2$ для модели де Муавра удовлетворяет условию нормировки.

Задание 5.6.2

Убедитесь, что плотность распределения (5.2.1) статуса выживания последнего двух индивидуумов $U := \overline{x_1 : x_2}$ для модели де Муавра удовлетворяет условию нормировки.

Задание 5.6.3

Покажите, что для экспоненциальной модели с плотностью распределения $f(t) = e^{-t}, \quad t \geq 0$, справедливы формулы:

$$f_{x_1:x_2}(t) = 2\,e^{-2t}, \quad f_{\overline{x_1:x_2}}(t) = 2\,e^{-t}(1 - e^{-t}), \quad t \geq 0.$$

Нарисуйте графики этих плотностей и сравните их поведение со случаем модели де Муавра.

Задание 5.6.4

Предполагая, что $T(65)$, $T(70)$ и $T(75)$ независимы, получить выражения вероятностей, что

(i) первая смерть произойдет в промежутке от 5 до 10 лет;

(ii) последняя смерть произойдет в том же промежутке.

(iii) Подсчитать эти вероятности для мужчин и женщин СССР, сравнить с соответствующими индивидуальными вероятностями.

Задание 5.6.5

В предположении независимости смертей для модели де Муавра и экспоненциальной модели Задания 5.6.3 найти $\overset{\circ}{e}_{x_1:x_2}$.

Задание 5.6.6

Убедитесь, что в предположении независимости смертей для модели де Муавра

$$\overset{\circ}{e}_{\overline{x:x}} = \frac{2(\omega - x)}{3}, \quad \mathbf{D}T(x:x) = \mathbf{D}T(\overline{x:x}) = \frac{(\omega - x)^2}{18}.$$

Задание 5.6.7

В предположении независимости смертей для экспоненциальной модели с плотностью распределения $f(t) = \lambda e^{-\lambda t}$, $\lambda > 0$, $t \geq 0$, получите формулы для

$$f_{x_1:x_2:x_3}(t), \quad f_{\overline{x_1:x_2}:x_3}(t), \quad f_{\overline{x_1:x_2:x_3}}(t),$$

$$\overset{\circ}{e}_{x_1:x_2:x_3}, \quad \overset{\circ}{e}_{\overline{x_1:x_2}:x_3}, \quad \overset{\circ}{e}_{\overline{x_1:x_2:x_3}} .$$

Задание 5.6.8

Опишите следующие смешанные статусы:
$\left(\overline{x_1:x_2} : \overline{(x_3:x_4):x_5}\right)$; $\left(x_1 : \overline{(x_2:x_3):(x_4:x_5)} : x_6\right)$.

6 СТАТИСТИЧЕСКИЕ МЕТОДЫ ОЦЕНИВАНИЯ ОСНОВНЫХ ВЕРОЯТНОСТНЫХ ХАРАКТЕРИСТИК АКТУАРНОЙ МАТЕМАТИКИ

В этой главе будут показаны возможности применения методов математической статистики в страховом деле. При этом здесь уместно еще раз напомнить, что идеология предыдущих глав в соответствии с п.2.1 очевидным образом распространяется и на такие области страхования, которые не связаны с личным страхованием, именно:

— оборудование предприятий и фирм (equipment of organizations and firms),

— наземный и воздушный транспорт (machine),

— предпринимательские риски (business ventures), предпринимательские ссуды (business loans), кредиты (credits) и т.д.

Также напомним, что, как и ранее, конец доказательств теорем и лемм будем отмечать указателем ♠.

6.1 Оценивание вероятностей

Примерами вероятностей, подлежащих оцениванию в страховой математике, служат ${}_tq_x$, ${}_tp_x$, q_x, p_x, ${}_{t|u}q_x$, ${}_{t|}q_x$.

Пусть A — случайное событие, $A_1, A_2, \ldots, A_N$ — однородная серия независимых опытов, в результате каждого из которых событие A либо наступает с вероятностью $\mathbf{P}(A)$, либо не наступает с вероятностью $1 - \mathbf{P}(A)$. В этом случае в качестве оценки $\mathbf{P}(A)$ естественно взять

$$\mathbf{P}_N(A) = \frac{1}{N}\sum_{i=1}^{N} I(A_i) = \frac{k}{N},$$

где $I(A)$ — случайная величина, равная индикатору события :

$$I(A) = \begin{cases} 1, & \text{если} \quad \omega \in A, \\ 0, & \text{если} \quad \omega \notin A, \end{cases}$$

ω — элементарное событие (см. [8, с.15]), а отношение $\dfrac{k}{N}$ есть частота события A в N однородных независимых опытах.

Оценка $\mathbf{P}_N(A)$ обладает следующими свойствами:

1) $\mathbf{P}_N(A)$ — несмещенная оценка для $\mathbf{P}(A)$, т.е.

$$\mathbf{E}\,\mathbf{P}_N(A) = \mathbf{P}(A);$$

2) ее дисперсия с ростом числа опытов убывает:

$$\mathbf{D}\,\mathbf{P}_N(A) = \frac{1}{N}\mathbf{P}(A)(1-\mathbf{P}(A)) < \frac{1}{N} \to 0 \quad \text{при} \quad N \to \infty;$$

3) случайная величина $N\,\mathbf{P}_N(A)$ распределена по биномиальному закону:

$$\mathbf{P}\{N\,\mathbf{P}_N(A) = k\} = C_N^k[\mathbf{P}(A)]^k[1-\mathbf{P}(A)]^{N-k};$$

4) $\mathbf{P}_N(A)$ — состоятельная оценка для $\mathbf{P}(A)$, т.е.

$$\lim_{N\to\infty}\mathbf{P}\{|\mathbf{P}_N(A)-\mathbf{P}(A)| < \varepsilon\} = 1 \quad \text{для любого} \quad \varepsilon > 0;$$

5) $\mathbf{P}_N(A)$ — оценка с минимальной дисперсией в классе линейных несмещенных оценок вида $\sum\limits_{i=1}^{N} c_i\, I(A_i), \quad c_i > 0;$

6) $\mathbf{P}_N(A)$ сходится с вероятностью 1 к $\mathbf{P}(A)$, т.е.

$$\mathbf{P}\{\lim_{N\to\infty}|\mathbf{P}_N(A)-\mathbf{P}(A)| = 0\} = 1.$$

Доказательства свойств 1–4 можно найти в большинстве учебников по математической статистике; свойство 5 доказывается методом неопределенных множителей Лагранжа (задача на условный экстремум) [7, с.128], а свойство 6 вытекает из усиленного закона больших чисел [8, с.114], [9, §2], [10].

Замечание 6.1.1. *Смещенная оценка*

$$\widehat{\mathbf{P}}(A) = \mathbf{P}_N(A)\left[1+\frac{\widehat{\mathbf{D}}\,\mathbf{P}_N(A)}{\mathbf{P}_N^2(A)}\right]^{-1} = \mathbf{P}_N(A)\left[1+\frac{1-\mathbf{P}_N(A)}{\mathbf{P}_N(A)\,N}\right]^{-1}$$

имеет среднеквадратическое отклонение (СКО) меньшее, чем дисперсия несмещенной оценки $\mathbf{D}\,\mathbf{P}_N(A)$ (*см.* [11, §1–6]), *т.е.*

$$u^2\left(\widehat{\mathbf{P}}(A)\right) = \mathbf{D}\widehat{\mathbf{P}}(A) + b^2(\widehat{\mathbf{P}}(A)) < \mathbf{D}\,\mathbf{P}_N(A),$$

где $b(\widehat{\mathbf{P}}(A)) = \mathbf{E}\,\widehat{\mathbf{P}}(A) - \mathbf{P}(A)$ — *смещение оценки* $\widehat{\mathbf{P}}(A)$.

При определенных условиях и малых объемах наблюдений СКО оценки $\widehat{\mathbf{P}}(A)$ иногда может оказаться меньше дисперсии оценки $\mathbf{P}_N(A)$ в два-три раза.

О специфических тонкостях, связанных с оцениванием вероятностей в актуарной математике, можно прочитать в главе 11 "Оценочные вероятности смерти"книги Х.Гербера [4].

6.2 Параметрические и непараметрические оценки. Эмпирические функции распределения и выживания

При оценивании функций распределения и выживания, зависящих от конечного набора неизвестных параметров, задача нахождения неизвестных функций сводится к оцениванию неизвестных параметров. Действительно, некоторые распределения достаточно точно описывают процесс смертности индивидуумов (появления отказов тех или иных элементов). Примерами таких распределений актуарной математики могут служить распределения моделей де Муавра, Гомпертца, Мэйкхама, Вейбулла и Эрланга рассмотренные в п.2.6.

В теории надежности, имеющей много точек соприкосновения с актуарной математикой, как отмечалось ранее, широко используется экспоненциальное распределение $F(t,\lambda) = 1 - e^{-\lambda t}$, $t \geq 0$. Оно описывает моменты отказов элементов, у которых остаточное время службы не зависит от длительности предшествующей работы. Заметим, что, если функция распределения отказа — экспоненциальное распределение, то функция интенсивности равна константе λ. Из других распределений можно выделить [12] распределение Вейбулла $F(t,\lambda,\alpha) = 1 - e^{-(\lambda t)^{\alpha}}$, $t \geq 0$, $\lambda, \alpha > 0$, которое "используется для описания усталостных явлений [13], отказов электровакуумных приборов [14], поломок подшипников [15]."

Если функция распределения известна с точностью до конечного набора неизвестных параметров, то имеем случай *параметрической априорной неопределенности*. Классические методы (например: метод максимального правдоподобия, метод моментов, метод наименьших квадратов) позволяют достаточно эффективно оценивать по наблюдениям неизвестные параметры.

В задачах оценивания надежности экспериментальных систем статистическими данными являются моменты отказов исследуемых элементов, которые, как правило, получают в результате проведения большого числа дорогостоящих экспериментов. К тому же исследователи часто не обладают достаточной информацией о самих элементах и о природе возникновения их отказов. Возможен также и случай, когда эта информация может оказаться несоответствующей реальному объекту, что усложняет, а иногда делает невозможным, построение адекватной параметрической модели.

Если априорная информация о природе отказов носит общий характер (например, известно, что производные функции распределения отказа до некоторого порядка существуют, непрерывны и т.д.), задачу оценивания неизвестных функций распределения, надежности, интенсивности и др. естественно рассматривать с точки зрения непараметрической статистики — одного из разделов математической статистики. Изложенные выше проблемы, конечно, имеют место и при решении различных задач оценивания в актуарной математике, в частности, при расчете нетто-премий в новых и нестандартных видах страхования.

Определим термин "непараметрический" согласно монографии Ф.П.Тарасенко [16].

"Непараметрическая задача — это статистическая задача, определенная на таких классах распределений, среди которых хотя бы один не сводится к параметрическому семейству функций."

Главное отличие непараметрических процедур от параметрических состоит в том, что они работоспособны тогда, когда априорная информация о распределениях не позволяет воспользоваться каким-либо параметрическим семейством распределений при определении математической модели объекта.

Введем о б о з н а ч е н и я: $\Longrightarrow$ — символ сходимости по распределению; $\mathcal{N}_s\{\mu,\sigma\}$ — s-мерная случайная величина, распределенная нормально с вектором средних $\mu = (\mu_1,\ldots,\mu_s)$ и ковариационной матрицей $\sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1s} \\ \cdots & \cdots & \cdots \\ \sigma_{s1} & \cdots & \sigma_{ss} \end{bmatrix}$, $0 \le \sigma_{ij} = \sigma_{ij}(x) < \infty$, $i,j = \overline{1,s}$.

Так как $F(x) = \mathbf{P}(X \le x)$, $s(x) = \mathbf{P}(X < x)$, то по выборке объема N независимых одинаково распределенных случайных величин $\{X_i,\ i = \overline{1,N}\}$, являющихся продолжительностями жизней N индивидуумов, в качестве простейших оценок естественно взять:

$$F_N(x) = \frac{1}{N}\sum_{i=1}^{N} I(X_i \le x), \quad s_N(x) = \frac{1}{N}\sum_{i=1}^{N} I(X_i > t). \qquad (6.2.1)$$

Оценка $F_N(x)$ называется *эмпирической функцией распределения*, вообще говоря является непараметрической оценкой для $F(x)$ и обладает всеми свойствами функции распределения. Статистические свойства эмпирической функции распределения $F_N(x)$ хорошо известны [9], [10]. Понятно, что они по сути идентичны свойствам оценки $\mathbf{P}_N(A)$. Приведем здесь лишь те из них, которые понадобятся нам в дальнейшем при проведении сравнительного анализа со свойствами эмпирической функции выживания и гладких эмпирических оценок функций распределения и выживания:

1) $\mathbf{E}F_N(x) = F(x)$;

2) $\mathbf{D}F_N(x) = \dfrac{1}{N}F(x)(1-F(x))$;

3) в силу центральной предельной теоремы

$$F_N(x) = F(x) + \frac{\xi_N(x)}{\sqrt{N}},$$

где распределение случайной величины $\xi_N(x)$ сходится по распределению к нормальному закону с параметрами $\{0, F(x)\,(1-F(x))\}$, т.е.

$$\xi_N(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [I(X_i \le x) - F(x)] \Longrightarrow \mathcal{N}_1 \{0, F(x)\,(1-F(x))\}\,. \quad (6.2.2)$$

4) $F_N(x)$ — оценка с минимальной дисперсией в классе линейных несмещенных оценок вида $\sum\limits_{i=1}^{N} c_i\, I(X_i \le x), \quad c_i > 0.$

Так как

$$s_N(x) = \frac{1}{N}\sum_{i=1}^{N} I(X_i > x) = \frac{1}{N}\sum_{i=1}^{N}[1 - I(X_i \le x)] = 1 - F_N(x),$$

то оценка функции выживания (надежности) $s_N(x)$ обладает свойствами, аналогичными свойствам эмпирической функции распределения $F_N(x)$. В частности,

1) $\mathbf{E}s_N(x) = s(x)$;

2) $\mathbf{D}s_N(x) = \dfrac{1}{N}F(x)(1-F(x)) = \dfrac{1}{N}s(x)(1-s(x))$.

Приведенные выше оценки $F_N(x)$ и $s_N(x)$ имеют два недостатка:

1) они имеют разрывы в точках $X_1, \ldots, X_n$;

2) оценка $F_N(x) = 0$ в области $\Omega_0 = [0, \min{(X_1, \ldots, X_N)}]$, а $s_N(x) = 0$ в области $\Omega_\infty = [\max{(X_1, \ldots, X_N)}, \infty]$.

Так, оценка функции интенсивности, полученная методом подстановки

$$\widehat{\mu_x} = \frac{f_N(x)}{s_N(x)}, \quad (6.2.3)$$

ввиду недостатка 2) в области Ω_∞ неработоспособна для любой оценки плотности $f_N(x)$.

6.3 Гладкая эмпирическая функция выживания, ее асимптотическая несмещенность и порядок сходимости смещения

Определение 6.3.1 . *Борелевская функция $\mathcal{S}(u)$ принадлежит классу функций выживания Suv (в знаках: $\mathcal{S}(u) \in Suv$), если $\mathcal{S}(u)$ — непрерывная, строго монотонно убывающая функция, такая, что $\mathcal{S}(u): \ \mathbf{R}^1 \to \mathbf{R}^1$, $\mathcal{S}(-\infty) = 1$, $\mathcal{S}(\infty) = 0$.*

Указанных в предыдущем пункте недостатков оценки $s_N(x)$ лишена следующая гладкая эмпирическая функция выживания:

$$\widetilde{s_N}(x) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{S}\left(\frac{x - X_i}{a_N}\right), \tag{6.3.1}$$

где $\mathcal{S}(u) \in Suv$, последовательность чисел $a_N \downarrow 0$. Функцию $\mathcal{S}(u)$ назовем *ядром* оценки (6.3.1).

Замечание 6.3.1. *Гладкая эмпирическая функция выживания* $\widetilde{s_N}(x)$ (7.3.1) *при* $a_N = 0$ *совпадает с эмпирической функцией выживания* $s_N(x)$ (7.2.1), т.е. $\widetilde{s_N}(x)\big|_{a_N=0} = s_N(x)$.

Определение 6.3.2 . *Функция* $H(z) : \mathbf{R}^s \to \mathbf{R}^1$ *принадлежит классу* $\mathcal{N}_{\nu,s}(x)$, *если функция* $H(z)$ *и все ее частные производные до* ν*-го пор ядка включительно непрерывны в точке* x; $H(z) \in \mathcal{N}_{\nu,s}(\mathbf{R})$, *если указанные свойства функции* $H(z)$ *выполняются для всех* $x \in \mathbf{R}^s$.

Выпишем условия, при которых оценка (6.3.1) является асимптотически несмещенной для $s(x)$.

Лемма 6.3.1 (асимптотическая несмещенность $\widetilde{s_N}$). *Если функция выживания* $s(z) \in \mathcal{N}_{0,1}(x)$, *т.е.* $s(z)$ *непрерывна в точке* x, $\mathcal{S}(u) \in Suv$, *а последовательность вещественных чисел* $a_N \downarrow 0$, *то*

$$\lim_{N\to\infty} \mathbf{E}\widetilde{s_N}(x) = s(x). \tag{6.3.2}$$

Д о к а з а т е л ь с т в о. По определению математического ожидания, учитывая непрерывность функции выживания $s(\cdot)$ в точке x, имеем

$$\mathbf{E}\widetilde{s_N}(x) = \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} \mathcal{S}\left(\frac{x - X_i}{a_N}\right)\right] = \int\limits_{\mathbf{R}^{1+}} \mathcal{S}\left(\frac{x-y}{a_N}\right) dF(y) =$$

$$= \int\limits_0^x \mathcal{S}\left(\frac{x-y}{a_N}\right) dF(y) + \int\limits_x^\infty \mathcal{S}\left(\frac{x-y}{a_N}\right) dF(y), \tag{6.3.3}$$

где $\mathbf{R}^{1+} = [0, \infty)$.

Так как

$$\lim_{N\to\infty} \mathcal{S}\left(\frac{x-y}{a_N}\right) = \begin{cases} 0, & \text{если} \quad y < x, \\ 1, & \text{если} \quad y > x, \end{cases}$$

то в силу теоремы Лебега о мажорируемой сходимости (см. приложение 4, теорема 1П)

$$\lim_{N\to\infty}\int\limits_{\mathbf{R}^{1+}}\mathcal{S}\left(\frac{x-y}{a_N}\right)dF(y)=\int\limits_x^\infty dF(y)=1-F(x)=s(x).\quad\spadesuit$$

Для построения оценок $\widetilde{s_N}(x)$, у которых главная часть СКО $u^2\left(\widetilde{s_N}(x)\right)=\mathbf{E}\left(\widetilde{s_N}(x)-s(x)\right)^2$ совпадает с дисперсией $\dfrac{s(x)\left(1-s(x)\right)}{N}$ эмпирической функции выживания $s_N(x)$, определенной в (6.2.1), необходимо установить порядок сходимости к нулю смещения

$$b\left(\widetilde{s_N}(x)\right)=\mathbf{E}\widetilde{s_N}(x)-s(x).$$

$\widetilde{s_N}(x)|_{a_N=0}=s_N(x)$.

Так как исследователь может сам выбирать подходящие ядра $\mathcal{S}(u)$ оценки $\widetilde{s_N}(x)$, сначала изучим свойства оценки $\widetilde{s_N}(x)$ с ядром Лапласа $\mathcal{S}_{LAP}(u)=1-\mathcal{F}_{LAP}(u)$, где $\mathcal{F}_{LAP}(u)$ — ядро-распределение Лапласа:

$$\mathcal{F}_{LAP}(u)=\begin{cases}0.5e^{u}, & -\infty<u<0,\\ 1-0.5e^{-u}, & 0\le u<\infty.\end{cases}$$

В этом случае

$$\mathcal{S}_{LAP}(u)=\begin{cases}1-0.5e^{u}, & -\infty<u<0,\\ 0.5e^{-u}, & 0\le u<\infty.\end{cases}\tag{6.3.4}$$

Найдем порядок сходимости смещения оценки

$$\widetilde{s_N}_{LAP}(x)=\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_{LAP}\left(\frac{x-X_i}{a_N}\right).$$

Лемма 6.3.2 (порядок сходимости смещения $\widetilde{s_N}_{LAP}$). *Пусть* $s(z)\in\mathcal{N}_{0,1}(x)$, $\sup\limits_{x\in\mathbf{R}^{1+}}f(x)\le C<\infty$, $a_N\downarrow 0$. *Тогда при* $N\to\infty$

$$\left|b\left(\widetilde{s_N}_{LAP}(x)\right)\right|=O\left(a_N\right).\tag{6.3.5}$$

Д о к а з а т е л ь с т в о. Представим

$$\mathbf{E}\widetilde{s_N}_{LAP}(x)=\int\limits_{\mathbf{R}^{1+}}\mathcal{S}_{LAP}\left(\frac{x-y}{a_N}\right)dF(y)=$$

$$= s(x) + \int\limits_0^x \mathcal{S}_{LAP}\left(\frac{x-y}{a_N}\right) dF(y) + \int\limits_x^\infty \left[\mathcal{S}_{LAP}\left(\frac{x-y}{a_N}\right) - 1\right] dF(y) =$$

$$= s(x) + \int\limits_0^x 0.5e^{-\frac{x-y}{a_N}} f(y)dy + \int\limits_x^\infty \left[1 - 0.5e^{\frac{x-y}{a_N}} - 1\right] f(y)dy.$$

Сделав в интегралах замену переменных $u = \dfrac{x-y}{a_N}$ и учитывая ограниченность кривой смертей, получим

$$|b\left(\widetilde{s_N}_{LAP}(x)\right)| \le C\frac{a_N}{2}\left(\int\limits_0^{x/a_N} e^{-u}du + \int\limits_{-\infty}^0 e^u du\right) =$$

$$= C\frac{a_N}{2}\left(-e^{-x/a_N} + 1 + 1 - e^{-\infty}\right) = O\left(a_N\right). \ \spadesuit$$

Из леммы 6.3.2 следует, что при нахождении порядка сходимости к нулю смещения оценки $\widetilde{s_N}(x)$ возникает необходимость определять порядок сходимости к нулю при $N \to \infty$ интегралов

$$\int\limits_x^\infty \left|\mathcal{S}\left(\frac{x-y}{a_N}\right) - 1\right| dy, \quad \int\limits_0^x \mathcal{S}\left(\frac{x-y}{a_N}\right) dy. \tag{6.3.6}$$

В общем случае, когда ядра $\mathcal{S}(u) \in Suv$, решение такой задачи наталкивается на определенные трудности. Например, выбрав ядро $\mathcal{S}(u) = 1 - \mathcal{F}(u)$, где $\mathcal{F}(u)$ — функция распределения Коши:

$$\mathcal{S}(u) = \frac{1}{2} - \frac{\arctan(u)}{\pi},$$

получим, что первый интеграл в (6.3.6) расходится, а второй сходится к некоторой константе.

В связи с этим определим такой класс ядер, для которого уже можно найти порядок сходимости смещения к нулю.

Определение 6.3.3 . *Функция $\mathcal{S}(u)$ принадлежит классу финитных функций выживания Fin_S ($\mathcal{S}(u) \in Fin_S$), если*

$$\mathcal{S}(u) = \begin{cases} 1, & -\infty < u < C_1, \\ Z(u), & C_1 \le u \le C_2, \\ 0, & C_2 < u < \infty, \end{cases}$$

где $Z(u)$ — непрерывная, строго монотонно убывающая функция, такая, что $Z(C_1) = 1$, $Z(C_2) = 0$, $C_1 < C_2$.

В качестве финитного ядра гладких эмпирических функций выживания можно взять, например, равномерное в $[C_1, C_2]$ ядро, для которого $Z(u) = 1 - \dfrac{u - C_1}{C_2 - C_1}$.

Найдем скорость сходимости к нулю смещения оценки $\widetilde{s_{N\,Fin}}(x)$ с финитным ядром $\mathcal{S}(u) \in Fin_S$.

Лемма 6.3.3 (порядок сходимости смещения $\widetilde{s_{N\,Fin}}$). *Если*

$$s(z) \in \mathcal{N}_{0,1}(x), \ \sup_{x \in \mathbf{R}^{1+}} f(x) \le C < \infty, \ a_N \downarrow 0,$$

то при $N \to \infty$

$$|b\left(\widetilde{s_{N\,Fin}}(x)\right)| = O\left(a_N\right). \tag{6.3.7}$$

Д о к а з а т е л ь с т в о. Представим

$$\mathbf{E}\widetilde{s_{N\,Fin}}(x) = \int\limits_0^\infty \mathcal{S}\left(\frac{x-y}{a_N}\right) dF(y) =$$

$$= s(x) + \int\limits_0^x \mathcal{S}\left(\frac{x-y}{a_N}\right) dF(y) + \int\limits_x^\infty \left[\mathcal{S}\left(\frac{x-y}{a_N}\right) - 1\right] dF(y).$$

В силу $dF(y) = f(y)dy$ и ограниченности кривой смертей

$$\mathbf{E}\widetilde{s_{N\,Fin}}(x) \le s(x) + C\left[\int\limits_0^x \mathcal{S}\left(\frac{x-y}{a_N}\right) dy + \int\limits_x^\infty \left|\mathcal{S}\left(\frac{x-y}{a_N}\right) - 1\right| dy\right].$$

Сделав в интегралах замену переменных $u = \dfrac{x-y}{a_N}$, получим

$$|b\left(\widetilde{s_{N\,Fin}}(x)\right)| \le C a_N \left[\int\limits_0^{x/a_N} \mathcal{S}\left(u\right) du + \int\limits_{-\infty}^0 \left|\mathcal{S}\left(u\right) - 1\right| du\right].$$

Так как при любом N интегралы

$$\int\limits_0^{x/a_N} \mathcal{S}\left(u\right) du \le \int\limits_{C_1}^{C_2} \mathcal{S}\left(u\right) du < \infty,$$

$\displaystyle\int\limits_{-\infty}^0 \left|\mathcal{S}\left(u\right) - 1\right| du \le \int\limits_{C_1}^{C_2} \left|\mathcal{S}\left(u\right) - 1\right| du < \infty$, то $|b\left(\widetilde{s_N}(x)\right)| = O\left(a_N\right)$. ♠

6.4 Предельная дисперсия, скорость сходимости СКО и асимптотическая нормальность гладкой эмпирической функции выживания

Найдем предельные дисперсии оценок $\widetilde{s_N}(x)$ с ядром $\mathcal{S}(u) \in Suv$ (см. определение 6.3.1) и $\widetilde{s}_{NFin}(x)$.

Лемма 6.4.1 (дисперсия $\widetilde{s_N}$). *Если выполнены условия леммы* 6.3.1, *то при* $N \to \infty$

$$\mathbf{D}\widetilde{s_N}(x) = \frac{s(x)\,(1-s(x))}{N} + o\left(\frac{1}{N}\right). \tag{6.4.1}$$

Доказательство. По определению дисперсии, учитывая (6.3.3), имеем:

$$\mathbf{D}\widetilde{s_N}(x) = \frac{1}{N}\mathbf{D}\mathcal{S}\left(\frac{x-X_1}{a_N}\right) =$$

$$= \frac{1}{N}\left\{\int\limits_{\mathbf{R}^{1+}} \mathcal{S}^2\left(\frac{x-y}{a_N}\right) dF(y) - \left[\int\limits_{\mathbf{R}^{1+}} \mathcal{S}\left(\frac{x-y}{a_N}\right) dF(y)\right]^2\right\}. \tag{6.4.2}$$

Применяя к интегралам приемы из леммы 6.3.1, получаем:

$$\mathbf{D}\widetilde{s_N}(x) = \frac{1}{N}\left[s(x) - s^2(x) + o(1)\right] = \frac{s(x)\,(1-s(x))}{N} + o\left(\frac{1}{N}\right). \ \spadesuit$$

Следствие 6.4.1. *Так как ядро* $\mathcal{S}_{LAP}(u) \in Suv$, *то в силу леммы* 6.4.1

$$\mathbf{D}\widetilde{s}_{NLAP}(x) = \frac{s(x)\,(1-s(x))}{N} + o\left(\frac{1}{N}\right).$$

Лемма 6.4.2 (дисперсия $\widetilde{s}_{NFin}$). *Если выполнены условия леммы* 6.3.3, *то при* $N \to \infty$

$$\mathbf{D}\widetilde{s}_{NFin}(x) = \frac{s(x)\,(1-s(x))}{N} + O\left(\frac{a_N}{N}\right) = \frac{s(x)\,(1-s(x))}{N} + o\left(\frac{1}{N}\right). \tag{6.4.3}$$

Доказательство. Применив к интегралам представления (6.4.2) приемы из леммы 6.3.3 приходим к соотношению (6.4.3). $\spadesuit$

Найдем главные части асимптотических СКО оценок $\widetilde{s}_{NFin}$ и $\widetilde{s}_{NLAP}(x)$.

Теорема 6.4.1 (СКО $\widetilde{s_{N\,Fin}}$). *Пусть выполнены условия леммы* 6.3.3. *Тогда при* $N \to \infty$

$$u^2\left(\widetilde{s_{N\,Fin}}\right) = \begin{cases} \dfrac{s(x)\,(1-s(x))}{N} + o\left(\dfrac{1}{N}\right), & a_N = o\left(1/\sqrt{N}\right); \\ O\left(\dfrac{1}{N}\right), & a_N = O\left(1/\sqrt{N}\right). \end{cases}$$

Д о к а з а т е л ь с т в о. Утверждение теоремы следует из представления $u^2\left(\widetilde{s_{N\,Fin}}(x)\right) = \mathbf{D}\widetilde{s_{N\,Fin}}(x) + b^2(\widetilde{s_{N\,Fin}}(x))$ и результатов (6.4.3), (6.3.7). ♠

Теорема 6.4.2 (СКО $\widetilde{s_{N\,LAP}}$). *Пусть выполнены условия леммы* 6.3.2. *Тогда при* $N \to \infty$

$$u^2\left(\widetilde{s_{N\,LAP}}\right) = \begin{cases} \dfrac{s(x)\,(1-s(x))}{N} + o\left(\dfrac{1}{N}\right), & a_N = o\left(1/\sqrt{N}\right); \\ O\left(\dfrac{1}{N}\right), & a_N = O\left(1/\sqrt{N}\right). \end{cases}$$

Д о к а з а т е л ь с т в о. Следствие 6.4.1 и лемма 6.3.2 немедленно приводят к утверждению теоремы. ♠

Определим предельное распределение оценки $\widetilde{s_{N\,Fin}}(x)$.

Введем обозначение: $\{\xi_{j,N}\}_{j=1}^{N}$, $N = 1, 2, \ldots$ — последовательность независимых одинаково распределенных случайных величин в схеме серий (распределение $\xi_{j,N}$ зависит от N).

Теорема 6.4.3 *Если выполнены условия леммы* 6.3.3 *и* $a_N = o\left(N^{-1/2}\right)$ *при* $N \to \infty$, *то*

$$\sqrt{N}\left[\widetilde{s_{N\,Fin}}(x) - s(x)\right] \Longrightarrow \mathcal{N}_1\left\{0, s(x)(1-s(x))\right\}. \tag{6.4.4}$$

Д о к а з а т е л ь с т в о. Представим

$$\sqrt{N}\left[\widetilde{s_{N\,Fin}}(x) - s(x)\right] =$$

$$= \sqrt{N}\left[\widetilde{s_{N\,Fin}}(x) - \mathbf{E}\widetilde{s_{N\,Fin}}(x)\right] + \sqrt{N}\; b\left(\widetilde{s_{N\,Fin}}(x)\right). \tag{6.4.5}$$

Ясно, что второе слагаемое в правой части (6.4.5) согласно (6.3.7) сходится к нулю при $N \to \infty$:

$$\sqrt{N} b\left(\widetilde{s_{N\,Fin}}(x)\right) = \sqrt{N}\left[o\left(N^{-1/2}\right)\right] \longrightarrow 0. \tag{6.4.6}$$

Покажем, что для первого слагаемого в правой части (6.4.5) выполняются все условия центральной предельной теоремы в схеме серий [**?**, с. 435]or1. Пусть

$$\xi_{j,N} = \frac{1}{\sqrt{N}} \left[\mathcal{S}\left(\frac{x - X_j}{a_N}\right) - \mathbf{E}\mathcal{S}\left(\frac{x - X_j}{a_N}\right) \right].$$

Таким образом, $\widetilde{s_N}_{Fin}(x) - \mathbf{E}\widetilde{s_N}_{Fin}(x) = \dfrac{1}{\sqrt{N}} \sum\limits_{j=1}^{N} \xi_{j,N}$. Очевидно, что $\mathbf{E}\xi_{j,N} = 0$ и, учитывая результат леммы 6.4.1,

$$\mathbf{E}\xi_{j,N}^2 = \frac{1}{N} \mathbf{D}\mathcal{S}\left(\frac{x - X_j}{a_N}\right) < \infty.$$

Также, согласно лемме 6.4.1 $\lim\limits_{N \to \infty} n\mathbf{E}\xi_{1,N}^2 = s(x)\left(1 - s(x)\right)$.

Проверим выполнение условия Линдеберга, которое является достаточным условием применимости центральной предельной теоремы. Учитывая, что $\sup\limits_{u \in \mathbf{R}^1} \mathcal{S}(u) \leq 1$, для любого $\tau > 0$ имеем

$$\beta_N = n\mathbf{E}\left(\left|\xi_{1,N}\right|^2, \left|\xi_{1,N}\right| > \tau\right) < \frac{N}{\tau} \mathbf{E}\left|\xi_{1,N}\right|^3 \leq$$

$$\leq \frac{C}{\sqrt{N}} \left[\mathbf{E}\left|\mathcal{S}\left(\frac{x - X_1}{a_N}\right)\right|^3 + \left|\mathbf{E}\mathcal{S}\left(\frac{x - X_1}{a_N}\right)\right|^3 \right] < \frac{2C}{\sqrt{N}}$$

здесь $C < \infty$ — некоторая положительная константа. Следовательно, $\beta_N = O\left(N^{-1/2}\right) \longrightarrow 0$ при $N \to \infty$, т.е. условие Линдеберга выполняется. Применяя центральную предельную теорему в схеме серий, получаем, что

$$\sqrt{N}\, \xi_N = \sqrt{N}\, \sum_{j=1}^{N} \xi_{j,N} \Longrightarrow \mathcal{N}_1 \left\{0, s(x)(1 - s(x))\right\}.$$

Учитывая (6.4.6), получаем требуемое утверждение (6.4.4). ♠

6.5 Гладкая эмпирическая функция распределения

Определение 6.5.1 . *Борелевская функция $\mathcal{F}(u)$ принадлежит классу функций распределения Dis ($\mathcal{F}(u) \in Dis$), если $\mathcal{F}(u)$ — непрерывная, строго монотонно возрастающая функция, такая, что* $\mathcal{F}(\cdot): \ \mathbf{R}^1 \to \mathbf{R}^1$, $\mathcal{F}(-\infty) = 0$, $\mathcal{F}(\infty) = 1$.

По аналогии с гладкой эмпирической функцией выживания $\widetilde{s_N}(x)$ строится гладкая эмпирическая функция распределения

$$\widetilde{F_N}(x) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{F}\left(\frac{x - X_i}{a_N}\right), \tag{6.5.1}$$

где $\mathcal{F}(u) \in Dis$, последовательность чисел $a_N \downarrow 0$.

Впервые оценка типа (6.5.1) была предложена Э.А. Надарая в 1964 г. [17]. Функция $\mathcal{F}(u)$ называется *ядром-распределением* оценки (6.5.1). Если взять ядро-распределение $\mathcal{F}(u) = 1 - \mathcal{S}(u)$, где $\mathcal{S}(u) \in Suv$ или $\mathcal{S}(u) \in Fin_S$, то

$$\widetilde{F_N}(x) = 1 - \frac{1}{N}\sum_{i=1}^{N} \mathcal{S}\left(\frac{x - X_i}{a_N}\right) = 1 - \widetilde{s_N}(x). \tag{6.5.2}$$

Свойства оценки $\widetilde{F_N}(x)$ отражены в леммах и теореме, доказательство которых предоставляется читателю.

Лемма 6.5.1 (асимптотическая несмещенность и дисперсия $\widetilde{F_N}$). *Если функция распределения $F(z) \in \mathcal{N}_{0,1}(x)$, $\mathcal{F}(u) \in Dis$, а последовательность вещественных чисел $a_N \downarrow 0$, то*

$$\mathbf{E}\widetilde{F_N}(x) = F(x) + o(1), \quad \mathbf{D}\widetilde{F_N}(x) = \frac{1}{N}F(x)\left(1 - F(x)\right) + o\left(\frac{1}{N}\right).$$

По аналогии с классом финитных ядер Fin_S введем класс финитных ядер-распределений Fin_F.

Определение 6.5.2 . *Функция $\mathcal{F}(u)$ принадлежит классу финитных функций распределения Fin_F ($\mathcal{F}(u) \in Fin_F$), если*

$$\mathcal{F}(u) = \begin{cases} 0, & -\infty < u < C_1, \\ Y(u), & C_1 \le u \le C_2, \\ 1, & C_2 < u < \infty, \end{cases}$$

где $Y(u)$ — непрерывная, строго монотонно возрастающая функция, такая, что $Y(C_1) = 0$, $Y(C_2) = 1$, $C_1 < C_2$.

Ядро $\mathcal{F}(u) \in Fin_F$ будем называть финитным ядром-распределением. Очевидно, что, если $\mathcal{F}(u) \in Fin_F$, то $\mathcal{F} \in Dis$.

Определим порядок сходимости к нулю смещения, главную часть асимптотического СКО и предельное распределение оценки $\widetilde{F_{N\,Fin}}(x)$ с финитным ядром-распределением $\mathcal{F}(u) \in Fin_F$.

Лемма 6.5.2 (порядок сходимости смещения $\widetilde{F}_{N\,Fin}$). *Если* $F(z) \in \mathcal{N}_{0,1}(x)$, $\sup\limits_{t\in\mathbf{R}^{1+}} f(x) < \infty$, $a_N \downarrow 0$, *то при* $N \to \infty$

$$\left| b\left(\widetilde{F}_{N\,Fin}(x)\right)\right| = O\left(a_N\right).$$

Теорема 6.5.1 (СКО $\widetilde{F}_N$). *Пусть выполнены условия леммы* 6.5.2. *Тогда при* $N \to \infty$

$$u^2\left(\widetilde{F}_{N\,Fin}\right) = \begin{cases} \dfrac{F(x)\,(1-F(x))}{N} + o\left(\dfrac{1}{N}\right), & a_N = o\left(1/\sqrt{N}\right), \\ O\left(\dfrac{1}{N}\right), & a_N = O\left(1/\sqrt{N}\right). \end{cases}$$

Теорема 6.5.2 *Если выполнены условия леммы* 6.5.2 *и* $a_N = o\left(N^{-1/2}\right)$, *то*

$$\sqrt{N}\left[\widetilde{F}_N(x) - F(x)\right] \Longrightarrow \mathcal{N}_1\left\{0, F(x)(1-F(x))\right\}.$$

6.6 Непараметрическое оценивание кривой смертей

Предполагая абсолютную непрерывность ядра-распределения $\mathcal{F}(u)$, в качестве непараметрической оценки кривой смертей $f(x) = F'(x)$ естественно взять оценку вида:

$$f_N(x) = \frac{d}{dx}\widetilde{F}_N(x) = \frac{1}{Na_N}\sum_{i=1}^{N} K\left(\frac{x-X_i}{a_N}\right), \tag{6.6.1}$$

где $K(u) = \mathcal{F}'(u)$ почти всюду. Так как на последовательность чисел $a_N \downarrow 0$ для оценки кривой смертей (6.6.1) по сравнению с оценкой функции выживания (6.3.1) будут накладываться дополнительные условия, то в дальнейшем оценку (6.6.1) будем записывать в виде

$$f_N(x) = \frac{1}{Nh_N}\sum_{i=1}^{N} K\left(\frac{x-X_i}{h_N}\right), \tag{6.6.2}$$

где последовательность чисел $h_N \downarrow 0$, $K(u)$ — ядро, вообще говоря, не обязательно обладающее свойствами плотности распределения.

Оценку (6.6.2) обычно называют ядерной или парзеновской, или оценкой типа Розенблатта—Парзена. Параметр h_N, как видно из структуры статистики (6.6.2), играет роль параметра масштаба ядра $K(\cdot)$ и поэтому называется *параметром размытости* ядерной оценки кривой смертей.

Впервые класс оценок (6.6.2) был предложен М. Розенблаттом в 1956 году [18]. В этой работе была доказана асимптотическая несмещенность и состоятельность ядерных оценок. Позднее, в 1962 году, Е. Парзен доказал асимптотическую нормальность этих оценок [19].

Для непараметрических оценок плотности известен следующий интересный результат [18].

Лемма 6.6.1 . *В условиях непараметрической априорной неопределенности не существует несмещенных оценок неизвестной плотности распределения* $f(x)$.

Доказательство леммы 6.6.1 имеется в [18], [16].

Определение 6.6.1 . *Борелевская функция* $K(u)$ принадлежит классу $\mathcal{A}$, *если*

$$\sup_{u\in\mathbf{R}^1}|K(u)|<\infty,\ \int\limits_{-\infty}^{\infty}|K(u)|\,du<\infty,\ \int\limits_{-\infty}^{\infty}K(u)du=1;\ K(u)\in\mathcal{A}_\nu,$$

если $K(u)\in\mathcal{A}$ *и* $K(u)$ *удовлетворяет дополнительным условиям*

$$\int\limits_{-\infty}^{\infty}|u^\nu K(u)|\,du<\infty,\ \ \mathcal{T}_j=\int\limits_{-\infty}^{\infty}u^jK(u)du=0,\ j=1,\ldots,\nu-1.$$

Выясним, при выполнении каких условий оценка (6.6.2) является асимптотически несмещенной для кривой смертей $f(x)$, т.е. для плотности распределения неотрицательной случайной величины X. Отметим, что в этом случае усложняются доказательства известных классических результатов, в чем можно убедиться на примере всех лемм и теорем данного раздела.

Лемма 6.6.2 (асимптотическая несмещенность f_N).
Если кривая смертей $f(x)\in\mathcal{N}_{0,1}(\mathbf{R})$, *т.е.* $f(x)$ *непрерывна на* $\mathbf{R}^1=(-\infty,\infty)$, $\sup_{x\in\mathbf{R}^{1+}}f(x)\le C_1<\infty$, $\int\limits_{-\infty}^{\infty}|K(u)|\,du<\infty$, $\int\limits_{-\infty}^{\infty}K(u)du=1$, *последовательность вещественных чисел* $h_N\downarrow 0$, *то*

$$\lim_{N\to\infty}\mathbf{E}f_N(x)=f(x). \qquad (6.6.3)$$

Д о к а з а т е л ь с т в о. По определению математического ожидания имеем

$$\mathbf{E}f_N(x) = \frac{1}{h_N}\int\limits_0^\infty K\left(\frac{x-y}{h_N}\right) f(y)dy. \qquad (6.6.4)$$

Сделав в интеграле (6.6.4) замену переменных $u = \dfrac{x-y}{h_N}$, получаем

$$\mathbf{E}f_N(x) = \frac{1}{h_N}\int\limits_{x/h_N}^{-\infty} (-h_N)K\left(u\right) f(x-uh_N)du =$$

$$= \int\limits_{-\infty}^{\infty} K\left(u\right) f(x-uh_N)du - \int\limits_{x/h_N}^{\infty} K\left(u\right) f(x-uh_N)du. \qquad (6.6.5)$$

В силу теоремы Лебега о мажорируемой сходимости (см. теорему 1П приложения 4) в (6.6.5) получаем: для первого интеграла

$$\lim_{N\to\infty}\int\limits_{-\infty}^{\infty} K\left(u\right) f(x-uh_N)du = f(x)\int\limits_{-\infty}^{\infty} K\left(u\right) du = f(x), \qquad (6.6.6)$$

а для второго интеграла

$$\lim_{N\to\infty} - \int\limits_{x/h_N}^{\infty} K\left(u\right) f(x-uh_N)du \le C_1 \lim_{N\to\infty}\int\limits_{x/h_N}^{\infty} K\left(u\right) du = 0. \qquad (6.6.7)$$

Теперь из (6.6.6) и (6.6.7) сразу следует утверждение (6.6.3). ♠

Найдем дисперсию оценки (6.6.2).

Определение 6.6.2 . *Для последовательности вещественных чисел* $h_N > 0$ *условия* $\lim\limits_{N\to\infty} h_N = 0$, $\lim\limits_{N\to\infty}\left(h_N + \dfrac{1}{Nh_N}\right) = 0$, *обозначим соответственно* $\mathcal{H}_1$, $\mathcal{H}_2$.

Лемма 6.6.3 (дисперсия f_N). *Если выполнены условия леммы* 6.6.2, $\int\limits_{-\infty}^{\infty} K^2(u)du < \infty$ *и* $h_N \in \mathcal{H}_2$, *то при* $N \to \infty$

$$\mathbf{D}f_N(x) = \frac{f(x)}{Nh_N}\int\limits_{-\infty}^{\infty} K^2(u)du + o\left(\frac{1}{Nh_N}\right). \qquad (6.6.8)$$

Д о к а з а т е л ь с т в о. По определению дисперсии

$$\mathbf{D}f_N(x) = \frac{1}{Nh_N^2}\mathbf{D}K\left(\frac{x-X_1}{h_N}\right) = \frac{1}{Nh_N^2}\left[\int\limits_0^\infty K^2\left(\frac{x-y}{h_N}\right)f(y)dy - \right.$$

$$\left. - \left(\int\limits_0^\infty K\left(\frac{x-y}{h_N}\right)f(y)dy\right)^2\right]. \qquad (6.6.9)$$

После замены переменных $u = \dfrac{x-y}{h_N}$ в интегралах (6.6.9) получаем

$$\mathbf{D}f_N(x) =$$

$$= \frac{1}{Nh_N}\left[\int\limits_{-\infty}^{\infty} K^2(u)f(x-uh_N)du - h_N\left(\int\limits_{-\infty}^{\infty} K(u)f(x-uh_N)du - \right.\right.$$

$$\left.\left. - \int\limits_{x/h_N}^{\infty} K(u)f(x-uh_N)du\right)^2 - \int\limits_{x/h_N}^{\infty} K^2(u)f(x-uh_N)du\right].$$

Далее, рассуждая также, как и в лемме 6.6.2, приходим к доказываемому утверждению:

$$\mathbf{D}f_N(x) = \frac{1}{Nh_N}\left[f(x)\int\limits_{-\infty}^{\infty} K^2(u)du + h_N\left(f(x)+o(1)\right)^2 + o(1)\right] =$$

$$= \frac{f(x)}{Nh_N}\int\limits_{-\infty}^{\infty} K^2(u)du + o\left(\frac{1}{Nh_N}\right). \ \spadesuit$$

Для построения оценок $f_N(x)$, оптимальных по скорости сходимости СКО, необходимо определить порядок сходимости к нулю смещения $b\left(f_N(x)\right)$.

Обозначим $w_\nu(x) = \dfrac{f^{(\nu)}(x)}{\nu!}\mathcal{T}_\nu$, где

$$f^{(\nu)}(x) = \frac{\partial^\nu f(x)}{\partial x^\nu}, \ f^{(0)}(x) = f(x).$$

Лемма 6.6.4 (скорость сходимости смещения $b(f_N)$).

Пусть

1) *кривая смертей* $f(z) \in \mathcal{N}_{\nu,1}(\mathbf{R})$;

2) $\sup\limits_{x\in\mathbf{R}^1} \left| f^{(m)}(x) \right| < \infty, \quad m = 0, \nu;$

3) *ядро* $K(u) \in \mathcal{A}_\nu$ *без условия* $\sup\limits_{u\in\mathbf{R}^1} |K(u)| < \infty$;

4) $1 - \mathcal{K}(x) = o\left(x^{-\nu}\right)$ *при* $x \to \infty$, *где* $\mathcal{K}(x) = \int\limits_{-\infty}^{x} K(u)du$;

5) $h_N \downarrow 0$.

Тогда при $N \to \infty$ *смещение* $b(f_N)$ *удовлетворяет соотношению*

$$|b(f_N(x)) - w_\nu(x)h_N^\nu| = o\left(h_N^\nu\right). \tag{6.6.10}$$

Д о к а з а т е л ь с т в о. Ранее, в (6.6.5), было показано, что

$$\mathbf{E}f_N(x) = \int\limits_{-\infty}^{\infty} f(x - uh_N)K(u)du - \int\limits_{x/h_N}^{\infty} f(x - uh_N)K(u)du. \tag{6.6.11}$$

Для второго интеграла представления (6.6.11), учитывая ограниченность $f(x)$ и условие 4) доказываемой леммы, имеем при $N \to \infty$

$$\int\limits_{x/h_N}^{\infty} f(x - uh_N)K(u)du \leq C \int\limits_{x/h_N}^{\infty} K(u)du =$$

$$= C\left[1 - \mathcal{K}\left(\frac{x}{h_N}\right)\right] = o(h_N^\nu). \tag{6.6.12}$$

В первом интеграле (6.6.11), разложив функцию $f(x - uh_N)$ по формуле Тейлора с остаточным членом в форме Лагранжа (см. утверждение 3П приложения 4), получим

$$\mathbf{E}f_N(x) = f(x) + \sum_{i=1}^{\nu-1}(-1)^i f^{(i)}(x)\frac{h_N^i}{i!}\mathcal{T}_i +$$

$$+(-1)^\nu f^{(\nu)}(x)\frac{h_N^\nu}{\nu!}\mathcal{T}_\nu + \gamma_N - \int\limits_{x/h_N}^{\infty} f(x - uh_N)K(u)du, \tag{6.6.13}$$

где

$$\gamma_N = \frac{(-1)^\nu h_N^\nu}{\nu!} \int\limits_{-\infty}^{\infty} \left[f^{(\nu)}\left(x + (-1)^\nu u h_N \theta\right) - f^{(\nu)}(x) \right] u^\nu K(u) du,$$

$$0 < \theta < 1.$$

Так как $\int\limits_{-\infty}^{\infty} |u^\nu K(u)| \, du < \infty$, а для каждого $x \in \mathbf{R}^1$ при $N \to \infty$ $f^{(\nu)}\left(x + (-1)^\nu u h_N \theta\right) \to f^{(\nu)}(x)$, то условия теоремы Лебега о мажорируемой сходимости выполнены, и, следовательно, $|\gamma_N| = o\left(h_N^\nu\right)$. С учетом того, что сумма в (6.6.13) равна нулю (из условия 3): $\mathcal{T}_i = 0,\ i = 1, \ldots, \nu - 1$), теперь следует справедливость утверждения (6.6.10). ♠

Лемма 6.6.4 указывает путь к нахождению непараметрических оценок плотности распределения со сколь угодно высокой скоростью сходимости их смещений $b\left(f_N(x)\right)$. Для этого необходимо использовать ядра $K(u)$ такие, что $K(u) \in \mathcal{A}_\nu$, $\nu \geq 4$.

Пример 6.6.1. Ядро-плотность $K(u) \in \mathcal{A}_2$, если $\sup\limits_u K(u) < \infty$, $K(u) = K(-u)$ и $\int u^2 K(u) du < \infty$.

Пример 6.6.2. Ядро $K(u) \in \mathcal{A}_4$, если

$$K(u) = \begin{cases} 15(3 - 10u^2 + 7u^4)/2^5, & \text{при} \quad |u| \leq 1, \\ 0, & \text{при} \quad |u| > 1. \end{cases}$$

Ядро $K(u) \in \mathcal{A}_6$, если

$$K(u) = \begin{cases} 105(5 - 35u^2 + 63u^4 - 33u^6)/2^8, & \text{при} \quad |u| \leq 1, \\ 0, & \text{при} \quad |u| > 1. \end{cases}$$

Указанные ядра получаются с помощью рекуррентной процедуры, использующей полиномы Якоби, ортонормированные с весовой функцией $\rho(u) = \{1 - u^2, |u| \leq 1;\ 0, |u| > 1\}$: $K(u) \in \mathcal{A}_\nu$, если

$$K(u) = \rho(u) \sum_{j=0}^{\nu-2} p_j(0) p_j(u) (2j+3)(j+2)/8(j+1),$$

$$p_{j+2}(u) = \frac{j+3}{j+4} \left[\frac{2j+5}{j+2} u p_{j+1}(u) - p_j(u) \right],$$

$$p_0(u) = p_0(0) = 1, \quad p_1(u) = 2u, \quad p_1(0) = 0.$$

Замечание 6.6.1. Ядра $K(u) \in \mathcal{A}_\nu$, $\nu \geq 4$, не обладают свойством плотности $K(u) \geq 0$ и могут принимать отрицательные значения.

Замечание 6.6.2. Ядра $K(u) \in \mathcal{A}_\nu$, $\nu \geq 4$, примера 4 являются оптимальными в классе полиномов при некоторых ограничениях на $K(u)$ [20], [21].

Определение 6.6.3 . *Будем говорить, что оценка t_N для t имеет скорость сходимости $O\left(N^{-\alpha}\right)$, $\alpha > 0$ (в знаках: $t_N \in \mathcal{V}\left(N^{-\alpha}\right)$,), если для СКО $u^2(t_N)$ справедливо предельное соотношение*

$$\lim_{N\to\infty} N^\alpha u^2\left(t_N\right) = C, \quad 0 < C < \infty.$$

Определение 6.6.4 . *Последовательности α_N и β_N называются эквивалентными (в знаках: $\alpha_N \sim \beta_N$), если $\lim\limits_{N\to\infty} |\alpha_N/\beta_N| = 1$.*

Теперь покажем, как возможность улучшения скорости сходимости смещения $b\left(f_N(x)\right)$ позволяет повышать скорость сходимости оценки кривой смертей в смысле определения 6.6.3.

Теорема 6.6.1 (оптимальное СКО $u^2(f_N)$). *Пусть выполнены условия леммы 6.6.4 и $h_N \in \mathcal{H}_2$. Тогда при $N \to \infty$ найдется такая оптимальная последовательность*

$$h_{N,o} = \mathrm{argmin}_{h_N > 0} u^2\left(f_N(x)\right) \sim \left(\frac{A}{2\nu B^2 N}\right)^{\frac{1}{2\nu+1}}, \tag{6.6.14}$$

где

$$A = f(x)\int\limits_{-\infty}^{\infty} K^2(u)du, \quad B = \frac{f^{(\nu)}\mathcal{T}^\nu}{\nu!},$$

для которой оптимальное СКО

$$u^2\left(f_N(x)|_{h_N = h_{N,o}}\right) = u^2\left(f_{N,o}(x)\right) \sim O\left(n^{-\frac{2\nu}{2\nu+1}}\right). \tag{6.6.15}$$

Д о к а з а т е л ь с т в о. В силу лемм 6.6.4 и 6.6.3

$$u^2\left(f_N(x)\right) = \frac{A}{Nh_N} + B^2 h_N^{2\nu} + o\left(\frac{1}{Nh_N} + h_N^{2\nu}\right). \tag{6.6.16}$$

Продифференцировав главную часть СКО (6.6.16) по h_N и приравняв полученное выражение к нулю, находим

$$h_{N,o} \sim \left(\frac{A}{2\nu B^2 N}\right)^{\frac{1}{2\nu+1}} = O\left(N^{-\frac{1}{2\nu+1}}\right). \tag{6.6.17}$$

Подставляя (6.6.17) в (6.6.16), имеем:

$$u^2\left(f_{N,o}(x)\right) \sim \frac{A}{N}\left(\frac{2\nu B^2 N}{A}\right)^{\frac{1}{2\nu+1}} + B^2\left(\frac{A}{2\nu B^2 N}\right)^{\frac{2\nu}{2\nu+1}} =$$

$$= \left(\frac{A}{N}\right)^{\frac{2\nu}{2\nu+1}} B^{\frac{2}{2\nu+1}}\left[(2\nu)^{\frac{1}{2\nu+1}} + \left(\frac{1}{2\nu}\right)^{\frac{2\nu}{2\nu+1}}\right] = O\left(N^{-\frac{2\nu}{2\nu+1}}\right). \ \spadesuit$$

Таким образом, из теоремы 6.6.1 следует, что при $\nu \to \infty$

$$f_{N,o}(x) \in \mathcal{V}\left(N^{-\frac{2\nu}{2\nu+1}}\right) \to \mathcal{V}\left(N^{-1}\right). \tag{6.6.18}$$

Из (6.6.18) видно, что в смысле определения 6.6.3 скорость сходимости оптимальной оценки кривой смертей $f_{N,o}(x)$ может достичь скорости сходимости параметрических оценок, а также оценок функций распределения $F_N(x)$, $\widetilde{F_N}(x)$ и функций надежности $s_N(x)$, $\widetilde{s_N}(x)$.

6.7 Нахождение оптимальных параметров размытости в ядерных оценках кривой смертей

Одной из основных проблем при непараметрическом оценивании плотностей — это нахождение оптимального значения $h_{N,o}\left(X_1,\ldots,X_N\right)$ для определенного набора $X_1,\ldots,X_N$. Теорема 6.6.1 указывает путь к нахождению оптимального значения параметра $h_{N,o}$ адаптивной ядерной оценки $f_{N,o}(x) = f_N(x)|_{h_N=h_{N,o}}$ плотностей из класса, определенного условиями 1), 2) леммы 6.6.4. Отметим также, что путем выбора ядра $K(u)$ из класса ядер, удовлетворяющих условию 3) леммы 6.6.4, можно улучшить скорость сходимости СКО оценки $f_{N,o}(x)$ (см. (6.6.18)). Из (6.6.14) видно, что последовательность $h_{N,o}$, минимизирующую главные части асимптотического СКО, трудно выписать в явном виде, так как она выражается через неизвестную плотность $f(x)$ и ее производную ν-го порядка $f^{(\nu)}(x)$.

В этом разделе рассмотрены методы адаптивного ядерного оценивания плотностей, которые делятся на два основных типа.

К первому типу относятся методы, связанные с оцениванием по выборке неизвестных параметров главной части асимптотического разложения некоторого критерия качества (например, нахождение $h_{N,o}$ путем оценивания неизвестных констант A и B в (6.6.14)).

Методы нахождения оптимального параметра размытости путем непосредственной минимизации некоторого критерия (например, видоизмененного критерия максимума правдоподобия, который рассмотрим ниже) относятся ко второму типу.

Первому типу принадлежит параметрический метод, исследованный в работах [22], [23] для одномерных плотностей и критерия $J_N = \mathbf{E}\left(\int\limits_{\mathbf{R}^1} \left(f_N(x) - f(x)\right)^2 dt\right)$. Согласно [18], [24], если $K(u)$ — ограниченное симметричное ядро-плотность, $K(u) \in \mathcal{A}_2$, $f(x)$ — ограниченная и дважды непрерывно-дифференцируемая плотность и $\int\limits_{\mathbf{R}^1} \left(f''(x)\right)^2 dx < \infty$, то при $h_N \in \mathcal{H}_2$:

$$J_N \sim (Nh_N)^{-1} \int_{\mathbf{R}^1} K^2(u)du + \frac{1}{4} h_N^4 \left(\int_{\mathbf{R}^1} u^2 K(u)du\right)^2 \int_{\mathbf{R}^1} \left(f''(x)\right)^2 dt. \tag{6.7.1}$$

Из (6.7.1) следует, что

$$h_{N,o} = \left[C \Big/ N \int_{\mathbf{R}^1} \left(f''(x)\right)^2 dx\right]^{1/5}, \tag{6.7.2}$$

где коэффициент $C = \int_{\mathbf{R}^1} K^2(u)du \Big/ \left(\int_{\mathbf{R}^1} u^2 K(u)du\right)^2$ зависит только от выбранного ядра $K(u)$. Таким образом, неизвестной величиной в (6.7.2) является интеграл $\int_{\mathbf{R}^1} \left(f''(x)\right)^2 dx$, который, например, для нормальной плотности равен $3/\left(8\sqrt{\pi}\sigma^5\right)$.

В работах [25], [23], [30], [31], [53] интеграл $\int_{\mathbf{R}^1} \left(f''(x)\right)^2 dx$ оценивается с помощью непараметрических методов. В данном случае процедура построения адаптивной ядерной оценки состоит из двух этапов: сначала строятся оценки интеграла $\int_{\mathbf{R}^1} \left(f''(x)\right)^2 dx$, затем, используя эти оценки в формуле (6.7.2), оценивается неизвестная $f(x)$.

При параметрическом методе адаптивного ядерного оценивания неизбежно возникает необходимость вынесения предположения о принадлежности $f(x)$ к некоторому параметрическому семейству. Недостатком данного метода является то, что, если предположение неверно, то мы получим неоптимальную по скорости сходимости (в смысле выбранного критерия) оценку.

При построении адаптивных ядерных оценок с привлечением непараметрических методов снова возникает задача нахождения параметров непараметрических оценок, которая является аналогичной по сложности той, которую решаем.

Указанных выше недостатков удается избежать, если выбирать параметр h_N так, чтобы максимизировать (минимизировать) некоторый эмпирический критерий. Данная процедура относится ко второму типу методов адаптивного ядерного оценивания. Отметим критерий, впервые примененный для ядерной оценки плотности и описанный в работах [29], [52], который основывается на принципе максимума правдоподобия (см. также обзор [27]). Согласно этим работам параметр $h_{N,o}$ выбирается из условия максимума эмпирической функции правдоподобия:

$$h_{N,o} = h_{N,o}\left(X_1, \ldots, X_N\right) = \operatorname{argmax}_{h>0}\left[\prod_{i=1}^{N} f_{N-1}(X_i)\right], \tag{6.7.3}$$

$$f_{N-1}(X_i) = \frac{1}{(N-1)h} \sum_{j=1, j\neq i}^{N-1} K\left(\frac{X_i - X_j}{h}\right), \tag{6.7.4}$$

и полученное значение $h_{N,o}$ используется при построении адаптивной ядерной оценке $f_{N,o}(x) = f_N(x)|_{h_N=h_{N,o}}$. Оценки, получаемые данным образом, называются кросс-проверочными ядерными оценками плотности.

В монографии [28, гл. 6.4] изучены проблемы состоятельности кросс-проверочных оценок, сделан обзор работ, посвященных методам кросс-проверки.

6.8 Оценивание средних $\overset{\circ}{e}_\circ$, $\overset{\circ}{e}_x$, $\overset{\circ}{e}_{x:n\rceil}$, $\overset{\circ}{e}_{x_1:x_2}$ и дисперсий $\mathbf{D}\,X$, $\mathbf{D}\,T(x)$

Сначала рассмотрим идеологию статистического оценивания на примере простейшей актуарной вероятностной характеристики — среднего времени жизни $\overset{\circ}{e}_\circ$.

Так как $\overset{\circ}{e}_\circ = \mathbf{E}\,X = \displaystyle\int_0^\infty x\,dF(x) = \int_0^\infty s(x)\,dx$, то в качестве оценок подстановок естественно взять

$\widehat{\overset{\circ}{e}}_\circ = \displaystyle\int_0^\infty x\,dF_N(x)$ или $\widehat{\widehat{\overset{\circ}{e}}}_\circ = \displaystyle\int_0^\infty s_N(x)\,dx$.

Легко убедиться, что эти оценки совпадают между собой. Действительно,

$$\widehat{\overset{\circ}{e}}_\circ = \int_0^\infty x\,\frac{1}{N}\sum_{i=1}^{N}\delta(x - X_i)\,dx = \frac{1}{N}\sum_{i=1}^{N} X_i = \overline{x},$$

$$\widehat{\overset{\circ}{e}}_\circ = \int_0^\infty \frac{1}{N} \sum_{i=1}^N I(X_i > x) = \frac{1}{N} \sum_{i=1}^N \int^{X_i} dx = \overline{x}.$$

Теперь можно синтезировать оценки для более сложных актуарных характеристик. Например, из представлений дисперсий

$$\mathbf{D}\, X = \mathbf{E}(X - \overset{\circ}{e}_x)^2 = \int_0^\infty (x - \overset{\circ}{e}_\circ)^2 \, dF(x),$$

$$\mathbf{D}\, X = 2 \int_0^\infty x\, s(x)\, dx - (\overset{\circ}{e}_\circ)^2,$$

следует, что соответственно

$$\widehat{\mathbf{D}X} = \frac{1}{N} \sum_{i=1}^N (X_i - \overline{x})^2,$$

$$\widehat{\widehat{\mathbf{D}X}} = 2 \int_0^\infty x \frac{1}{N} \sum_{i=1}^N I(X_i > x)\, dx - (\overline{x})^2 =$$

$$= \frac{1}{N} \sum_{i=1}^N X_i^2 - (\overline{x})^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \overline{x})^2.$$

Из формулы $\overset{\circ}{e}_x = \dfrac{1}{s(x)} \displaystyle\int_0^\infty s(x+t)\, dt$ получаем, что

$$\widehat{\overset{\circ}{e}}_x = \frac{1}{s_N(x)} \int_0^\infty \frac{1}{N} \sum_{i=1}^N I(X_i > x + t)\, dt =$$

$$= \frac{1}{s_N(x)\, N} \sum_{i=1}^N \int_0^{(X_i - x) I(X_i - x > 0)} dt =$$

$$= \sum_{i=1}^N (X_i - x) I(X_i - x > 0) \Bigg/ \sum_{i=1}^N I(X_i - x > 0). \tag{6.8.1}$$

В области $[X_{(N)}, \infty)$ оценка $\widehat{\overset{\circ}{e}}_x$ не определена (ее знаменатель принимает нулевое значение), поэтому в ряде случаев имеет смысл применять вместо (6.8.1) следующую усложненную модификацию:

$$\tilde{\overset{\circ}{e}}_x = \sum_{i=1}^N (X_i - x) I(X_i - x > 0) \Bigg/ \sum_{i=1}^N \mathcal{S}\left(\frac{x - X_i}{a_N}\right).$$

Рассуждая аналогично, для $\overset{\circ}{e}_{x:n\rceil} = \dfrac{1}{s(x)} \displaystyle\int_0^n s(x+t)\,dt$, $\mathbf{D}\,T(x)$ и $\mathbf{D}\{\min(T(x), n)\}$ имеем:

$$\widehat{\overset{\circ}{e}_{x:n\rceil}} = \frac{1}{s_N(x)} \int_0^n \frac{1}{N} \sum_{i=1}^{N} I(X_i > x+t)\,dt =$$

$$= \frac{1}{s_N(x)\,N} \sum_{i=1}^{N} \int_0^{\min(n, X_i - x) I(X_i - x > 0)} dt =$$

$$= \sum_{i=1}^{N} \min(n, X_i - x) I(X_i - x > 0) \Bigg/ \sum_{i=1}^{N} I(X_i - x > 0),$$

$$\mathbf{D}\,\widehat{\widehat{T}}(x) = \frac{\sum\limits_{i=1}^{N} (X_i - x)^2 I(X_i - x > 0)}{\sum\limits_{i=1}^{N} \mathcal{S}\left(\dfrac{x - X_i}{a_N}\right)} - (\hat{\overset{\circ}{e}}_x)^2,$$

$$\mathbf{D}\,\widehat{\widehat{\min(T}}(x), n) = \frac{\sum\limits_{i=1}^{N} (\min(n, X_i - x))^2 I(X_i - x > 0)}{\sum\limits_{i=1}^{N} \mathcal{S}\left(\dfrac{x - X_i}{a_N}\right)} - (\hat{\overset{\circ}{e}}_x)^2.$$

В заключение рассмотрим особенности применения рассмотренных выше приемов оценивания в коллективном страховании. Например, в качестве оценки $\overset{\circ}{e}_{x_1:x_2}$ можно взять

$$\widehat{\overset{\circ}{e}_{x_1:x_2}} = \frac{1}{s_N(x_1, x_2)} \int_0^{\infty} \hat{\mathbf{P}}\{\min(X - x_1, Y - x_2) > t\}\,dt =$$

$$= \frac{1}{s_N(x_1, x_2)} \int_0^{\infty} \frac{1}{N} \sum_{i=1}^{N} I\{\min(X_i - x_1, Y_i - x_2) > t\}\,dt =$$

$$= \frac{1}{s_N(x_1, x_2)\,N} \sum_{i=1}^{N} \int_0^{\min(X_i - x_1, Y_i - x_2)} dt =$$

$$= \frac{1}{s_N(x_1, x_2)\,N} \sum_{i=1}^{N} \min(X_i - x_1, Y_i - x_2) I\{\min(X_i - x_1, Y_i - x_2) > 0\},$$

где $s_N(x_1,x_2) = \dfrac{1}{N}\sum\limits_{i=1}^{N} I\{X_i - x_1 > 0\}\, I\{Y_i - x_2 > 0\}$). Заметим, что приведенная формула имеет смысл и в случае, когда случайные величины X и Y статистически связаны между собой. Очевидно, что такую зависимость необходимо учитывать, например, при страховании родственников, работников с одного предприятия и т.п.

6.9 Асимптотическая нормальность оценки функции интенсивности

Так как параметры предельных распределений оценок функций интенсивности выражаются через ковариации оценок функций надежности и плотности распределения, то в этом пункте найдем предельные ковариационные матрицы этих оценок.

Обозначим: $b_N = (b_{1N}, \ldots, b_{sN})$ — вектор оценок,

$$\mathcal{C}(b_n) = \begin{bmatrix} \operatorname{cov}(b_{1N}, b_{1N}) & \ldots & \operatorname{cov}(b_{1N}, b_{sN}) \\ \ldots & \ldots & \ldots \\ \operatorname{cov}(b_{sN}, b_{1N}) & \ldots & \operatorname{cov}(b_{sN}, b_{sN}) \end{bmatrix}.$$

Найдем предельную ковариационную матрицу оценок плотности и функции надежности.

Лемма 6.9.1 (ковариационная матрица оценок $f_N, \widetilde{s_N}$). *Если* $f(x) \in \mathcal{N}_{0,1}(\mathbf{R})$, $\sup\limits_{x\in\mathbf{R}^{1+}} f(x) < \infty$, *ядро* $K(u) \in \mathcal{A}$, *выполнены условия леммы* 6.3.1 *и* $h_N \in \mathcal{H}_2$, *то*

$$\lim_{N\to\infty} N h_N \mathcal{C}\left(f_N(x), \widetilde{s_N}(x)\right) = f(x) \begin{bmatrix} \int\limits_{\mathbf{R}^1} K^2(u)du & 0 \\ 0 & 0 \end{bmatrix} = \bar{\sigma}_\lambda. \qquad (6.9.1)$$

Д о к а з а т е л ь с т в о. Согласно (6.4.1) и (6.6.8)

$$\operatorname{cov}\left(\widetilde{s_N}(x), \widetilde{s_N}(x)\right) = \mathbf{D}\widetilde{s_N}(x) = \frac{1}{N} s(x)\left(1 - s(x)\right) + o\left(\frac{1}{N}\right),$$

$$\operatorname{cov}\left(f_N(x), f_N(x)\right) = \mathbf{D} f_N(x) = \frac{f(x)}{N h_N} \int\limits_{-\infty}^{\infty} K^2(u)du + o\left(\frac{1}{N h_N}\right),$$

следовательно

$$\lim_{N\to\infty} n h_N \operatorname{cov}\left(f_N(x), f_N(x)\right) = f(x) \int\limits_{-\infty}^{\infty} K^2(u)du,$$

$$\lim_{N\to\infty} Nh_N \operatorname{cov}\left(\widetilde{s_N}(x), \widetilde{s_N}(x)\right) = \lim_{N\to\infty} h_N s(x)\left(1 - s(x)\right) = 0.$$

Далее,

$$\operatorname{cov}\left(f_N(x), \widetilde{s_N}(x)\right) = \frac{1}{Nh_N}\left[\int\limits_0^\infty K\left(\frac{x-y}{h_N}\right)\mathcal{S}\left(\frac{x-y}{a_N}\right) f(y)dy - \right.$$

$$\left. - \int\limits_0^\infty K\left(\frac{x-y}{h_N}\right) f(y)dy \int\limits_0^\infty \mathcal{S}\left(\frac{x-y}{a_N}\right) f(y)dy\right].$$

Так как $\mathcal{S}(\cdot)$ — ограниченная функция, то $\sup\limits_{x\in\mathbf{R}^{1+}} \mathcal{S}(x) \le C < \infty$ и, таким образом,

$$I_1 = \frac{1}{Nh_N}\int\limits_0^\infty K\left(\frac{x-y}{h_N}\right)\mathcal{S}\left(\frac{x-y}{a_N}\right) f(y)dy \le$$

$$\le C\frac{1}{Nh_N}\int\limits_0^\infty K\left(\frac{x-y}{h_N}\right) f(y)dy.$$

Согласно (6.6.3), имеем при $N\to\infty$: $I_1 = O\left(N^{-1}\right)$. Далее, учитывая (6.6.3) и (6.3.2), получаем при $N\to\infty$:

$$\frac{1}{Nh_N}\int\limits_0^\infty K\left(\frac{x-y}{h_N}\right) f(y)dy \int\limits_0^\infty \mathcal{S}\left(\frac{x-y}{a_N}\right) f(y)dy = O(N^{-1}),$$

$$\operatorname{cov}\left(f_N(x), \widetilde{s_N}(x)\right) = \operatorname{cov}\left(\widetilde{s_N}(x), f_N(x)\right) = O\left(N^{-1}\right). \qquad (6.9.2)$$

Лемма 6.9.1 доказана.

Выбирая в качестве оценки функции надежности оценку $\widetilde{s_N}(x)$ (6.3.1), а в качестве оценки плотности — ядерную оценку $f_N(x)$ (6.6.2), построим оценку подстановки функции интенсивности

$$\lambda_N(x) = \frac{f_N(x)}{\widetilde{s_N}(x)}. \qquad (6.9.3)$$

Оценка (6.9.3) представляет собой случайную дробь. Очевидно, что исследовать свойства оценки $\lambda_N(x)$ труднее, чем оценок числителя и знаменателя. В связи с этим, при нахождении предельного распределения оценки $\lambda_N(x)$ будем использовать теорему 2.1.1 из [32, с. 20], которая приведена в приложении 4 как теорема 3П. Чтобы применить теорему 3П, найдем предельное распределение двумерного случайного вектора

$$\sqrt{Nh_N}\left(f_N(x)-f(x),\widetilde{s_N}(x)-s(x)\right).$$

Для нахождения предельного распределения двумерного случайного вектора нам потребуется центральная предельная теорема в схеме серий, которая сформулирована для двумерного случая и приведена в приложении 4.

Введем обозначения: $\{\xi_{j,N},\eta_{j,N}\}_{j=1}^{N}$, $N=1,2,\ldots$ — последовательность независимых одинаково распределенных двумерных векторов в схеме серий (распределение $(\xi_{j,N},\eta_{j,N})$ зависит от N); $\sigma_N=N\mathbf{E}\left(\xi_{1,N},\eta_{1,N}\right)^T\left(\xi_{1,N},\eta_{1,N}\right)$, здесь ^{T} — знак транспонирования; $\|(\xi,\eta)\|=\sqrt{\xi^2+\eta^2}$.

Теорема 6.9.1 . *Пусть* 1) *плотность* $f(z)\in\mathcal{N}_{\nu,1}\left(\mathbf{R}\right)$; 2) $\sup\limits_{x\in\mathbf{R}^1}\left|f^{(m)}(x)\right|<\infty$, $m=\overline{0,\nu}$; 3) *ядро* $K(u)\in\mathcal{A}_\nu$; 4) $1-\mathcal{K}(x)=o\left(x^{-\nu}\right)$ *при* $x\to\infty$; 5) *выполнены условия леммы* 6.3.3; 6) $h_N\in\mathcal{H}_2$, $\lim\limits_{N\to\infty}\sqrt{Nh_N}\left(a_N+h_N^\nu\right)=0$. *Тогда вектор* $\sqrt{Nh_N}\left(f_N(x)-f(x),\widetilde{s_N}(x)-s(x)\right)$ *имеет двумерное предельное нормальное распределение* $\mathcal{N}_2\left\{(0,0),\bar\sigma_\lambda\right\}$, *где* $\bar\sigma_\lambda$ — *предельная ковариационная матрица, определенная в* (6.9.1).

Д о к а з а т е л ь с т в о. Представим

$$\sqrt{Nh_N}\left(f_N(x)-f(x),\widetilde{s_N}(x)-s(x)\right)=$$

$$=\sqrt{Nh_N}\left(f_N(x)-\mathbf{E}f_N(x),\widetilde{s_N}(x)-\mathbf{E}\widetilde{s_N}(x)\right)+$$

$$+\sqrt{Nh_N}\left(b\left(f_N(x)\right),b\left(\widetilde{s_N}(x)\right)\right). \tag{6.9.4}$$

Второе слагаемое в правой части (6.9.4) сходится при $N\to\infty$ к нулевому вектору:

$$\sqrt{Nh_N}\left(b\left(f_N(x)\right),b\left(\widetilde{s_N}(x)\right)\right)=\sqrt{Nh_N}\left(O\left(h_N^\nu\right),O\left(a_N\right)\right)=$$

$$=\left(O\left(\sqrt{Nh_N^{2\nu+1}}\right),O\left(\sqrt{Nh_N}a_N\right)\right)\longrightarrow(0,0).$$

Покажем, что для первого слагаемого в правой части (6.9.4) выполняются все условия двумерной центральной предельной теоремы в схеме серий (см. приложение 4, теорема 6П). Пусть

$$\xi_{j,N} = \frac{1}{\sqrt{Nh_N}} \left[K\left(\frac{x - X_j}{h_N}\right) - h_N \mathbf{E} f_N(x) \right],$$

$$\eta_{j,N} = \frac{\sqrt{h_N}}{\sqrt{N}} \left[\mathcal{S}\left(\frac{x - X_j}{a_N}\right) - \mathbf{E}\mathcal{S}\left(\frac{x - X_j}{a_N}\right) \right].$$

Таким образом

$$f_N(x) - \mathbf{E} f_N(x) = \frac{1}{\sqrt{Nh_N}} \sum_{j=1}^{N} \xi_{j,N}, \quad \widetilde{s_N}(x) - \mathbf{E}\widetilde{s_N}(x) = \frac{1}{\sqrt{Nh_N}} \sum_{j=1}^{N} \eta_{j,N}.$$

Очевидно, что $\mathbf{E}\left(\xi_{j,N}, \eta_{j,N}\right) = (0,0)$. Проверим выполнение условия 2) теоремы 6П. Для этого необходимо показать, что

$$\mathbf{E}\left\|\xi_{j,N}, \eta_{j,N}\right\|^2 = \frac{1}{Nh_N} \mathbf{D} K\left(\frac{x - X_j}{h_N}\right) + \frac{h_N}{N} \mathbf{D}\mathcal{S}\left(\frac{x - X_j}{a_N}\right) < \infty.$$

Учитывая результаты лемм о дисперсиях оценок $\widetilde{s_N}(x)$ и $f_N(x)$ — (6.4.1), (6.6.8), получаем:

$$\mathbf{E}\left\|\xi_{j,N}, \eta_{j,N}\right\|^2 \le \frac{1}{N} C_1 \int\limits_{-\infty}^{\infty} K^2(u) du + \frac{h_N}{N} C_2 < \infty,$$

здесь $0 \le C_1, C_2 < \infty$ — некоторые константы.

По лемме 6.9.1 $\lim\limits_{N\to\infty} N\mathbf{E}\left(\xi_{1,N}, \eta_{1,N}\right)^T \left(\xi_{1,N}, \eta_{1,N}\right) = \bar{\sigma}_\lambda$.

Проверим выполнение условия Линдеберга, которое является достаточным условием применимости центральной предельной теоремы в схеме серий. Для любого $\tau > 0$ имеем

$$\beta_N = N\mathbf{E}\left(\left\|\xi_{1,N}, \eta_{1,N}\right\|^2; \left\|\xi_{1,N}, \eta_{1,N}\right\| > \tau\right) < \frac{N}{\tau} \mathbf{E}\left\|\xi_{1,N}, \eta_{1,N}\right\|^3 =$$

$$= \frac{N}{\tau} \mathbf{E}\left(\xi_{1,N}^2 + \eta_{1,N}^2\right)^{3/2} \le \frac{CN}{\tau} \left[\mathbf{E}\left|\xi_{1,N}\right|^3 + \mathbf{E}\left|\eta_{1,N}\right|^3 \right] =$$

$$= \frac{CN}{\left(Nh_N\right)^{3/2}} \left[\mathbf{E}\left| K\left(\frac{x - X_1}{h_N}\right) - \mathbf{E}K\left(\frac{x - X_1}{h_N}\right) \right|^3 + \right.$$

$$+h_N^3 \mathbf{E}\left|\mathcal{S}\left(\frac{x-X_1}{h_N}\right)-\mathbf{E}\mathcal{S}\left(\frac{x-X_1}{h_N}\right)\right|^3\Bigg] \le$$

$$\le \frac{C}{(Nh_N)^{1/2}}\left[\frac{1}{h_N}\left\{\mathbf{E}\left|K\left(\frac{x-X_1}{h_N}\right)\right|^3+h_N^3\mathbf{E}^3 f_N(x)\right\}+\right.$$

$$\left.+h_N^2\left\{\mathbf{E}\left|\mathcal{S}\left(\frac{x-X_1}{h_N}\right)\right|^3+\mathbf{E}^3\mathcal{S}\left(\frac{x-X_1}{h_N}\right)\right\}\right].$$

Далее, так как ядра $K(\cdot), \mathcal{S}(\cdot)$ ограничены на $\mathbf{R}^1$, то

$$\beta_N = O\left(\frac{1}{\sqrt{Nh_N}}\right) \longrightarrow 0$$

при $N \to \infty$, т.е. условие Линдеберга выполняется, следовательно, по теореме 6П получаем требуемый результат.

Теорема 6.9.1 доказана.

Введем обозначения:

$x_N = (x_{N1}, \ldots, x_{Ns})$ — векторная статистика с компонентами $x_{Nj} = x_{Nj}(x) = x_{Nj}(t; X_1, \ldots, X_N)$, $\quad j = \overline{1,s}$;

функция $H(z) : \mathbf{R}^s \to \mathbf{R}^1$;

$H^{(1)}(z) = \nabla H(z) = (H_1(z), \ldots, H_s(z))\,,\ H_j(z) = \partial H(z)/\partial z_j$;

d_N — положительная числовая последовательность, неограниченно возрастающая с ростом N;

$\phi = \phi(x) = (\phi_1(x), \ldots, \phi_s(x))$ — ограниченная вектор-функция, которая, например, может быть математическим ожиданием статистики x_N.

Теперь, согласно теоремам 6.9.1 и 3П можно найти предельное распределение оценки функции интенсивности.

Теорема 6.9.2 (асимптотическая нормальность λ_N). *Если выполнены условия теоремы* 6.9.1 *и* $\lambda(x), s(x) \neq 0$, *то*

$$\sqrt{Nh_N}\,(\lambda_N(x)-\lambda(x)) \Longrightarrow \mathcal{N}_1\left\{0, \frac{\lambda(x)\int\limits_{\mathbf{R}^1} K^2(u)du}{s(x)}\right\}. \qquad (6.9.5)$$

Д о к а з а т е л ь с т в о. В обозначениях теоремы 3П (см. приложение 4) имеем:

$d_N = \sqrt{Nh_N}$, $\lambda_N = H(x_N) = H\left(x_{N1}, x_{N2}\right) = \dfrac{x_{N1}}{x_{N2}} = \dfrac{f_N}{\widetilde{s_N}}$,

$$\lambda = H(\phi) = H(\phi_1, \phi_2) = \frac{\phi_1}{\phi_2} = \frac{f}{s},$$
$$H^{(1)}(\phi) = \left(\frac{1}{\phi_2},\ -\frac{\phi_1}{\phi_2^2}\right) = \left(\frac{1}{s},\ -\frac{f}{s^2}\right),\ \mu_1 = \mu_2 = 0,\ \sigma = \bar{\sigma}_\lambda$$
(матрица $\bar{\sigma}_\lambda$ определена в (6.9.1)).

После подстановки найденных величин $H_j^{(1)}(\phi)$, μ_j и σ_{jp} в (П.1) (см. приложение 4) приходим к утверждению теоремы.

Теорема 6.9.2 доказана.

Замечание 6.9.1 *По аналогии с теоремой* 6.9.2 *в работе* [33] *показана асимптотическая нормальность непараметрических оценок функции интенсивности вида:* $\lambda_{N,2}(x) = f_N(x)/(1 - F_{N,2}(x))$, *где* $f_N(x) = f_N(x, h_N)$ *(оценка плотности* $f_N(x)$ *определена согласно* (6.6.2) *данной работы),* $F_{N,2}(x) = \int\limits_0^x f_N(t, a_N)dt$.

6.10 Интервальное оценивание функции интенсивности

Результат теоремы 6.9.2 позволяет найти преобразование оценки функции интенсивности $\lambda_N(x)$, которое имеет предельное стандартное нормальное распределение. Нам потребуется следующее простое обобщение результата из [34], которое приведено в приложении 4 в виде теоремы 4П.

В теореме 4П показано, что, если $\{\mathcal{T}_N,\ \ N = 1, 2, \ldots\}$ — последовательность статистик такая, что $\sqrt{q_N}\,[\mathcal{T}_N - \theta] \Longrightarrow \mathcal{N}_1\left\{0, \sigma^2(\theta)\right\}$, числовая последовательность $q_N \uparrow \infty$, то

$$\sqrt{q_N}\,[g(\mathcal{T}_N) - g(\theta)] \Longrightarrow \mathcal{N}_1\left\{0, [g'(\theta)\sigma(\theta)]^2\right\}, \tag{6.10.1}$$

где g — функция, имеющая первую производную и $g'(\theta) \neq 0$.

Согласно [34, стр. 377], функцию g следует выбирать так, чтобы

$$g'(\theta)\sigma(\theta) = c, \tag{6.10.2}$$

где c не зависит от θ. В этом случае асимптотическая дисперсия статистики $g(\mathcal{T}_N)$ не будет зависеть от θ.

Функцию g найдем из уравнения $g = \int \frac{cd\theta}{\sigma(\theta)}$. В нашем случае, учитывая (6.9.5), имеем

$$g=\int\frac{cd\theta}{\sigma(\theta)}=\int\frac{\sqrt{s(x)}d\lambda(x)}{\sqrt{\lambda(x)}\sqrt{\int_{\mathbf{R}^1}K^2(u)du}}=\int\frac{cd\lambda(x)}{\sqrt{\lambda(x)}}. \qquad (6.10.3)$$

Из (6.10.3) следует, что $g(x)=\sqrt{x}$.

Найдем преобразование оценки функции интенсивности, имеющее предельное стандартное нормальное распределение.

Теорема 6.10.1 . *В условиях теоремы* 6.9.2

$$\frac{2\sqrt{Nh_N}\sqrt{s(x)}}{\sqrt{\int\limits_{-\infty}^{\infty}K^2(u)du}}\left[\sqrt{\lambda_N(x)}-\sqrt{\lambda(x)}\right]\Longrightarrow\mathcal{N}_1\left\{0,1\right\}. \qquad (6.10.4)$$

Д о к а з а т е л ь с т в о. В обозначениях теоремы 4П, полагая:

$$q_N=Nh_N,\ \mathcal{T}_N=\lambda_N,\ \theta=\lambda\,g(\lambda)=\sqrt{\lambda}$$

$$g'(\lambda_N)\sigma(\lambda_N)=\frac{\sqrt{\int\limits_{-\infty}^{\infty}K^2(u)du}}{2\sqrt{s(x)}},$$

получаем утверждение 6.10.4.

Теорема 6.10.1 доказана.

Учитывая, что $|b\left(\widetilde{s_N}(x)\right)|=O(a_N)$ (см. лемму 6.3.3) и $\lim\limits_{N\to\infty}a_N\sqrt{Nh_N}=0$ (условие 6) теоремы 6.9.1), взяв $a_N=o\left(n^{-1/2}\right)$, в формуле (6.10.4) правомерно заменить $\sqrt{s(x)}$ на $\sqrt{\widetilde{s_N}(x)}$. Таким образом, при выполнении условий теоремы 6.9.2

$$\frac{2\sqrt{Nh_N}\sqrt{\widetilde{s_N}(x)}}{\sqrt{\int\limits_{-\infty}^{\infty}K^2(u)du}}\left[\sqrt{\lambda_N(x)}-\sqrt{\lambda(x)}\right]\Longrightarrow\mathcal{N}_1\left\{0,1\right\}. \qquad (6.10.5)$$

На основе утверждения (6.10.5) можно построить интервальную оценку заданной надежности $(1-\alpha)$ для функции интенсивности $\lambda(x)$:

$$\frac{2\sqrt{Nh_N}\sqrt{\widetilde{s_N}(x)}}{\sqrt{\int\limits_{-\infty}^{\infty}K^2(u)du}}\left|\sqrt{\lambda_N(x)}-\sqrt{\lambda(x)}\right|<U_{1-\alpha/2}, \qquad (6.10.6)$$

где $U_{1-\alpha/2}$ — квантиль уровня $1-\frac{\alpha}{2}$ стандартного нормального распределения. Таким образом

$$\left(\sqrt{\lambda_N(x)}-\frac{\sqrt{\int\limits_{-\infty}^{\infty}K^2(u)du}}{2\sqrt{Nh_N}\sqrt{\widetilde{s_N}(x)}}U_{1-\alpha/2}\right)^2<\lambda(x)<$$

$$<\left(\sqrt{\lambda_N(x)}+\frac{\sqrt{\int\limits_{-\infty}^{\infty}K^2(u)du}}{2\sqrt{Nh_N}\sqrt{\widetilde{s_N}(x)}}U_{1-\alpha/2}\right)^2. \tag{6.10.7}$$

Достоинством интервальной оценки (6.10.7) является то, что она определяется через известные функции (сравните с дисперсией предельного распределения оценки $\lambda_N(x)$).

Интервальные непараметрические оценки (6.10.7) рассмотрены в работах [35]—[38], [39], [40].

6.11 Сходимость в среднеквадратическом оценок функции интенсивности

Одной из основных точностных характеристик оценки является ее среднеквадратическое отклонение (СКО) от истинного значения. При нахождении СКО оценки подстановки $\lambda_N(x)$ возникают трудности, обусловленные их возможной неограниченностью, например, когда оценки знаменателей принимают значения, равные нулю. Решение этой проблемы состоит либо в использовании усеченных модификаций оценок $\lambda_N(x)$ [32], либо их кусочно-гладких аппроксимаций [42], [43]:

$$\Psi\left(\lambda_N,\delta\right)=\widetilde{\Psi}\left(x,\delta\right)=\frac{\lambda_N(x)}{\left[1+\delta\left|\lambda_N(x)\right|^{\tau}\right]^{\rho}}, \tag{6.11.1}$$

где $\tau>0,\rho>0,\rho\tau\geq 1,\delta>0$.

В данной работе рассмотрим кусочно-гладкие аппроксимации (6.11.1)

Введем обозначения: $M_\nu\left\|y_N\right\|=\mathbf{E}\left\|y_N-\phi\right\|^\nu$ — момент порядка ν нормы отклонения оценки $y_N(x)$ от функции $\phi(x)$ в точке x. Введем для тройки (τ,k,m), где k,m — натуральные числа, множество
$T(m)=\{(\tau,k):\tau\geq\tau(m)=2k/(m-k-1)>0,$
$m\geq m_0=[3,k=1;2k,k\geq 2]\}$.

Для нахождения СКО оценок $\Psi\left(\lambda_N^{(r)}, \delta_r\right)$ потребуется следствие 4 из [43], которое приведено в приложении 4 в виде теоремы 5П.

Чтобы использовать формулу (П.2) приложения 4 для определения главных частей СКО кусочно-гладких аппроксимаций $\widetilde{\Psi}(x, \delta)$, необходимо, согласно условию 2) теоремы 5П, знать порядок сходимости к нулю четвертого момента отклонения оценок $\widetilde{s_N}(x)$, $f_N(x)$. Эти результаты приводятся ниже в виде лемм 6.11.1 и 6.11.2.

Лемма 6.11.1 . *Если выполнены условия леммы* 6.3.3 *и*

$$a_N = o\left(N^{-1/2}\right),$$

то при $N \to \infty$

$$M_4 \left\|\widetilde{s_N}(x)\right\| = \mathbf{E}\left(\widetilde{s_N}(x) - s(x)\right)^4 = O\left(N^{-2}\right). \tag{6.11.2}$$

Д о к а з а т е л ь с т в о. Представим

$$\mathbf{E}\left(\widetilde{s_N}(x) - s(x)\right)^4 = \mathbf{E}\left[\left(\widetilde{s_N}(x) - \mathbf{E}\widetilde{s_N}(x)\right) + b\left(\widetilde{s_N}(x)\right)\right]^4.$$

Привлекая при $p = 4$ и $m = 2$ неравенство

$$\left(\sum_{i=1}^{m} |a_i|\right)^p \le m^{p-1} \sum_{i=1}^{m} |a_i|^p, \quad p > 1,$$

имеем

$$\mathbf{E}\left(\widetilde{s_N}(x) - s(x)\right)^4 \le 2^3 \left[\mathbf{E}\left(\widetilde{s_N}(x) - \mathbf{E}\widetilde{s_N}(x)\right)^4 + b^4\left(\widetilde{s_N}(x)\right)\right]^4.$$

Согласно лемме 6.3.3 $b\left(\widetilde{s_N}(x)\right) = O\left(a_N\right)$, значит $b^4\left(\widetilde{s_N}(x)\right) = o\left(N^{-2}\right)$. Согласно [44] $\mathbf{E}\left(\widetilde{F_N}(x) - \mathbf{E}\widetilde{F_N}(x)\right)^4 = O\left(N^{-2}\right)$, следовательно $\mathbf{E}\left(\widetilde{s_N}(x) - \mathbf{E}\widetilde{s_N}(x)\right)^4 = O\left(N^{-2}\right)$, отсюда немедленно следует справедливость соотношения (6.11.2).

Лемма 6.11.1 доказана.

Найдем порядок сходимости четвертого момента отклонения оценок $f_N(x)$ от истинной кривой смертей $f(x)$.

Лемма 6.11.2 . *Пусть выполняются условия*:
1) $f(x) \in \mathcal{N}_{2,1}(\mathbf{R})$; 2) $\sup\limits_{x \in \mathbf{R}^1} \left|f^{(m)}(x)\right| < \infty$, $m = \overline{0, 2}$;
3) *ядро* $K(u) \in \mathcal{A}_2$; 4) *при* $x \to \infty$: $1 - \mathcal{K}(x) = o\left(x^{-2}\right)$;

5) $\lim\limits_{N\to\infty}\left(h_N+\dfrac{1}{Nh_N}\right)=0$;

Тогда при $N\to\infty$

$$M_4\left\|f_N(x)\right\|=\mathbf{E}\left(f_N(x)-f(x)\right)^4=O\left(\left[\frac{1}{Nh_N}+h_N^4\right]^2\right). \qquad (6.11.3)$$

Д о к а з а т е л ь с т в о. Рассуждая, как при доказательстве леммы 6.11.1, и учитывая, что в силу леммы 2.3.3 [32] $\mathbf{E}\left(f_N(x)-f(x)\right)^4=O\left(\left(Nh_N\right)^{-2}\right)$, согласно лемме 6.6.4 при $\nu=2:$
$b\left(f_N(x)\right)=O\left(h_N^2\right)$, получаем выражение (6.11.3).

Лемма 6.11.2 доказана.

Теперь, используя теорему 5П приложения 4, найдем СКО кусочно-гладких аппроксимаций $\Psi\left(\lambda_N,\delta\right)$.

Теорема 6.11.1 . *Пусть выполнены следующие условия:* 1) $\lambda(x)\neq 0$, $s(x)\neq 0$; 2) *выполнены условия лемм* 6.11.1 *и* 6.11.2; 3) $\delta=\left(\dfrac{1}{Nh_N}+h_N^4\right)$.

Тогда для любых $(\tau,2)\in T(m)$ *при* $N\to\infty$*:*

$$\left|\mathbf{E}\left[\Psi\left(\lambda_N,\delta\right)-\lambda(x)\right]^2-\frac{\lambda(x)}{s(x)Nh_N}\int\limits_{-\infty}^{\infty}\left(K(u)\right)^2du-\right.$$

$$\left.-\frac{\mathcal{T}_2^2h_N^4}{4}\left(\frac{\left(f^{(2)}(x)\right)^2}{s^2(x)}\right)\right|=O\left(\left[\frac{1}{Nh_N}+h_N^4\right]^{3/2}\right). \qquad (6.11.4)$$

Д о к а з а т е л ь с т в о. Покажем справедливость соотношения (6.11.4). В обозначениях теоремы 5П (см. приложение 4) имеем: $s=2$, $z=\left(z_1,z_2\right)$, $H(z)=z_1/z_2$, $y_N=\left(y_{1N},y_{2N}\right)=\left(f_N,\widetilde{s_N}\right)$, $\phi=\left(\phi_1,\phi_2\right)=\left(f(x),s(x)\right)$, $d_N=O\left(Nh_N+h_N^{-4}\right)$, $k=2$. Возьмем $m=m_0=4$ и покажем, что $M_4\left\|f_N,\widetilde{s_N}\right\|=O\left(d_N^{-2}\right)$. Это сразу следует из соотношения (6.11.2), соотношения (6.11.3) и неравенства $M_4\left\|f_N,\widetilde{s_N}\right\|\leq 2\left[M_4\left\|f_N\right\|+M_4\left\|\widetilde{s_N}\right\|\right]$.

Далее, так как $s(x)\neq 0$, а при $z_2\neq 0$ функция

$$z_1/z_2\in\mathcal{N}_{2,2}\left(f(x),s(x)\right),$$

то все условия теоремы 5П выполнены для $\Psi\left(\lambda_N,\delta\right)=\Psi\left(\lambda_N,\delta\right)$ при $\tau\geq\tau_0=4$. Теорема 6.11.1 доказана.

6.12 Контрольные задания

Задание 6.11.1

Докажите свойства 1)–6) из п.6.1 для оценки вероятности $\mathbf{P}_N(A)$.

Задание 6.11.2

Сформулируйте гипотезу о методе получения аппроксимации вероятности q_x в таблице смертности из Приложения 1. Дайте обоснование Ваших предположений. Приведите пример смещенной оценки вероятности с меньшей среднеквадратической ошибкой, чем дисперсия несмещенной оценки.

Задание 6.11.3

Опишите организацию однородной серии опытов для получения оценок вероятностей ${}_tq_x$, ${}_tp_x$,p_x, ${}_{t|u}q_x$, ${}_{t|}q_x$.

Задание 6.11.4

По 5 числам, взятым из таблицы равномерно распределенных случайных чисел [41, с. 426], постройте эмпирические функцию распределения и функцию выживания. Сравните с теоретическими кривыми. В качестве 5 чисел можете взять 10, 9, 73, 25, 33.

Задание 6.11.5

В условиях Задания 6.11.4 постройте гладкие эмпирические функцию распределения и функцию выживания. В качестве базовых функций распределения $\mathcal{F}(u)$ (ядер) можно взять:

равномерная функция распределения

$$\mathcal{F}(u) = \begin{cases} 0, & u \le -\frac{1}{2}, \\ \frac{1}{2} + u, & -\frac{1}{2} < u \le \frac{1}{2}, \\ 1, & u > \frac{1}{2}; \end{cases}$$

функция распределения Коши

$$\mathcal{F}(u) = \frac{1}{2} + \frac{1}{\pi}\mathrm{arctg}u, \quad -\infty < u < \infty;$$

логистическая функция распределения

$$\mathcal{F}(u) = \frac{1}{1+e^{-u}}, \quad -\infty < u < \infty;$$

гиперболический косинус

$$\mathcal{F}(u) = 1 - \frac{2}{\pi}\mathrm{arctg}e^{-u}, \quad -\infty < u < \infty;$$

двойной показательный закон

$$\mathcal{F}(u) = e^{-e^{-u}}, \quad -\infty < u < \infty;$$

Задание 6.11.6

Синтезируйте наилучшие смещенные оценки функций распределения и выживания в точке x.

Задание 6.11.7

На основе гладкой эмпирической функции распределения с ядрами Задания 6.11.5 по выборке Задания 6.11.4 постройте ядерные оценки плотности распределения. Вычислите значения оценки плотности и ее дисперсию в точках 0, 50, 100.

Задание 6.11.8

В условиях Задания 6.11.7 постройте алгоритм нахождения эмпирического параметра размытости методом максимума правдоподобия.

Литература

[1] Bowers N.L., Gerber H.U., Hickman J.C., Jones D.A., Nesbitt C.J. Actuarial Mathematics. Itasca, Illinois: The Society of Actuaries, 1986, 624 p.

[2] Фалин Г.И., Фалин А.И. Введение в актуарную математику. М.: Изд-во МГУ, 1994, 86 с.

[3] Фалин Г.И. Математический анализ рисков в страховании. М.: Российский юридический издательский дом, 1994, 130 с.

[4] Гербер Х. Математика страхования жизни. М., Мир, 1995, 154 с.

[5] Четыркин Е.М. Пенсионные фонды. Зарубежный опыт для отечественных предприятий, актуарные расчеты. М.: АО "АРГО", 1993, 100 с.

[6] Штрауб Э. Актуарная математика имущественного страхования. М., Мир, 1988, 146 с.

[7] Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. М.: Инфра-М, 1997, 302 с.

[8] Боровков А.А. Теория вероятностей. М.: Наука, 1986, 432 с.

[9] Боровков А.А. Математическая статистика. Оценка параметров. Проверка гипотез. М.: Наука, 1984, 472 с.

[10] Боровков А.А. Математическая статистика. Новосибирск: Наука; Изд-во Института математики. 1997, 772 с.

[11] Виленкин С.Я. Статистическая обработка результатов исследования случайных функций. М.: Энергия, 1979. – 320 с.

[12] Барлоу Р., Прошан Ф. Статистическая теория надежности и испытания на безотказность. — М.: Наука. Гл. ред. физ.-мат. лит. 1984. — 328 с.

[13] Solovyev A.D. Theory of aging elements// Proc. of the 6 Berkeley Symp. on Math. Stat. and Probability. — 1965. — V.3.

[14] Полляк Ю.Г. О погрешностях прогноза надежности, обусловленных статистической зависимостью между отказами элементов. — Электросвязь. — 1963. — №4.

[15] Смолицкий Х.Л., Чукреев П.А. К вопросу об оптимальном резервировании аппаратуры. Изв. АН СССР, Сер. "Тех. кибернетика". 1965. №4.

[16] Тарасенко Ф.П. Непараметрическая статистика // Томск: Изд-во Том. ун-та, 1976. – 292 с.

[17] Надарая Э.А. Некоторые новые оценки функции распределения// Теория вероятностей и ее применения. — 1964. — Т. IX, вып. 3. — С. 550-554.

[18] Rosenblatt M. Remarks on some nonparametric estimates of a density function // Ann. Math. Statist. 1956. V.27. т 3. P.832–835.

[19] Parzen E. On estimation of a probability density function and mode // Ann. Math. Statist. 1962. V.33. т 3. P.1065–1076.

[20] Gasser T., Müller H.-G. Kernel estimation of regression functions // Lect. Notes Math. 1979. V.757. P.23–68.

[21] Müller H.G., Gasser T. Optimal convergence properties of kernel estimates of derivatives of a density function// Lect. Notes Math. — 1979. — V. 757. — P. 144-154.

[22] Deheuvels P. Estimation non parametrique de la densite par histogrammes generalises// Revue de Statistique Appliguee. — 1977. — V. 25. — P. 5-42.

[23] Deheuvels P., Hominal P. Estimation automatique de la densite// Revue de Statistique Appliguee. — 1980. — V. 28. — P. 25-55.

[24] Rosenblatt M. Curve estimates// Ann. Math. Statist. — 1971. — V. 42. — P.1815-1842.

[25] Надарая Э.А. Об интегральной среднеквадратичной ошибке некоторых непараметрических оценок плотности вероятностей// Теория вероятностей и ее применения. — 1974. — Т. 19, № 1. — С. 131-140.

[26] Habbema J.D.F., Hermans J., Vandenbroek K. A stepwise disriminant analysis program using density estimation in COMPSTAT 1974. — Wien. Physica Verlag. 1974. — P. 101-110.

[27] Rudemo M. Empirical choice of histogram and kernel density estimators// Scandinavian Journal of Statistics. — 1982. — V. 9. — P. 65-78.

[28] Деврой Л., Дьёрфи З. Непараметрическое оценивание плотности. L_1-подход. — М.: Мир. 1988. — 408 с.

[29] Duin R.P.W. On the choice of smoothing parameters for Parzen estimators of propability density functions. IEEE Transactions on Computers. 1976. C–25. — P. 1175-1179.

[30] Scott D.W., Factor L.E. Monte-Carlo study of three data-based nonparametric probability density estimators. Journal of the American Statistical Association. 1981. V. 76. P. 9-15.

[31] Scott D.W., Tapia R.A., Thompson J.R. Kernel density estimation revisited. Journal of Nonlinear Analysis, Theory, Methods and Applications. 1977. V. 1. — P. 339-372.

[32] Добровидов А.В., Кошкин Г.М. Непараметрическое оценивание сигналов. М.: Наука. Физматлит. 1997. – 336 с.

[33] Вааль В.А., Китаева А.В., Кошкин Г.М. Доверительное непараметрическое оценивание функции интенсивности// Межд. конф. "Всесибирские чтения по математике и механике". Тезисы докл. — Томск: Изд-во ТГУ. — 1997. — Т.1. — С.120.

[34] Рао С. Линейные статистические методы и их применения. — М.: Наука. 1968. — 548 с.

[35] Вааль В.А., Кошкин Г.М. Интервальные непараметрические оценки функции интенсивности// Математическое моделирование и теория вероятностей. Сб. научных трудов Томского университета / Под ред. И.А. Александрова, В.Н.Берцуна, А.М.Бубенчикова и Ю.К.Устинова. — Томск: Изд-во ТГУ. — 1998. — С.147–149.

[36] Вааль В.А., Кошкин Г.М. Непараметрическое оценивание функции интенсивности отказов и ее производной// Изв. вузов. Физика. — 1999. — N 3. — С.141-146.

[37] Вааль В.А., Кошкин Г.М. Оценивание функции интенсивности отказов и ее производной в условиях неопределенности// Межд. конф. по проблемам управления (29 июня - 2 июля 1999 г.). Тезисы докл. в трех томах. — М.: Фонд "Проблемы управления". 1999. — Т.2. — С.138-140.

[38] Вааль В.А., Кошкин Г.М. Оценивание функции интенсивности отказов и ее первых двух производных в условиях неопределенности// Межд. конф. по проблемам управления (29 июня - 2 июля 1999 г.). Избранные труды в двух томах. — М.: СИНТЕГ. — 1999. — Т.1. — С.120-126.

[39] Koshkin G.M., Vaal V.A. On Estimation of Hazard Rate Function and its Derivatives// The Third Russian-Korean International Symposium on Science and Technology (June 22-25, 1999, Novosibirsk). Abstracts. — Novosibirsk: Novosibirsk State Technical University. — 1999. — Vol.2. — P.544.

[40] Vaal V.A., Koshkin G.M. Kernel Nonparametric Estimation of the Hazard Rate Function and its Derivatives// The Third Russian-Korean International Symposium on Science and Technology (June 22-25, 1999, Novosibirsk). Proceedings. — Novosibirsk: Novosibirsk State Technical University. — 1999. — Vol.2. — P.496-500.

[41] Сборник задач по математике для втузов. Ч. 3. Теория вероятностей и математическая статистика: Учеб. пособие для втузов / Под ред. А.В.Ефимова. — 2-е изд., перераб. и доп. — М.: Наука, 1990, 428 с.

[42] Пенская М.Я. Статистические методы оценивания и проверки гипотез// Межвузовский сб. науч. трудов. — Пермь: Пермский ун-т. 1990. — С. 44-55.

[43] Кошкин Г.М. Моменты отклонений оценки подстановки и ее кусочно-гладких аппроксимаций // Сибирский математический журнал. 1999. Т.40. т 3. С.605–618.

[44] Крамер Г. Математические методы статистики. — М.: Мир. 1975. — 648 с.

[45] Koshkin G.M. Nonparametric Estimation in Complex Models of Life Insurance // Математическое моделирование и теория вероятностей. Сб. научных трудов Томского университета / Под ред.

В.Н.Берцун, А.М.Бубенчикова и Ю.К.Устинова. Томск: Изд-во ТГУ. 1998. С.207–214.

[46] Кошкин Г.М. Непараметрическое оценивание функционалов в моделях страхования жизни // Обозрение прикладной и промышленной математики. 1997. Т.4. Вып.3. С.362-363.

[47] Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. — М.: Наука. 1968. — 720 с.

[48] Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М.: Наука. 1972. – 496 с.

[49] Китаева А.В., Кошкин Г.М. Устойчивое с улучшенной скоростью сходимости непараметрическое оценивание многомерной функции интенсивности // Автоматика и телемеханика. 1997. т 5. С. 202–214.

[50] Кошкин Г.М. Улучшенная неотрицательная ядерная оценка плотности // Теория вероят. и ее примен. 1988. Т.33. Вып. 4. С.817–822.

[51] Кошкин Г.М. О двух типах непараметрических оценок плотности распределения // Изв. вузов. Физика. 1995. т 9. С.31–38.

[52] Habbema J.D.F., Hermans J., Vandenbroek K. A stepwise disriminant analysis program using density estimation in COMPSTAT 1974, G.Bruckmann (Ed.). Wien. Physica Verlag. 1974. — P. 101-110.

[53] Woodroofe M. On choosing a delta-sequence. Ann. of Math. Stat. 1970. V. 41. — P. 1665-1671.

Г.М.Кошкин

ОСНОВЫ СТРАХОВОЙ МАТЕМАТИКИ

Учебное пособие
Томский государственный университет
634050, г. Томск, пр. Ленина, 36

7 ПРИЛОЖЕНИЯ

Таблица продолжительности жизни (СССР, 1984–1985 г.) [3, с.86]

возраст	мужчины			женщины		
x	l_x	q_x	d_x	l_x	q_x	d_x
14	95,438	0,00068	65	96,407	0,00037	36
15	95,373	0,00082	78	96,371	0,00041	40
16	95,295	0,00101	97	96,331	0,00047	45
17	95,198	0,00124	118	96,286	0,00053	51
18	95,080	0,00149	142	96,235	0,00059	57
19	94,938	0,00173	164	96,178	0,00065	62
20	94,774	0,00196	186	96,116	0,00069	66
21	94,588	0,00216	205	96,050	0,00072	69
22	94,383	0,00234	221	95,981	0,00074	71
23	94,162	0,00249	235	95,910	0,00076	73
24	93,927	0,00263	247	95,837	0,00078	75
25	93,680	0,00277	260	95,762	0,00081	77
26	93,420	0,00293	274	95,685	0,00084	80
27	93,146	0,00312	290	95,605	0,00088	84
28	92,856	0,00333	310	95,521	0,00093	89
29	92,546	0,00356	330	95,432	0,00099	95
30	92,216	0,00381	352	95,337	0,00106	101
31	91,864	0,00405	372	95,236	0,00113	108
32	91,492	0,00425	389	95,128	0,00121	116
33	91,103	0,00445	406	95,012	0,00131	125
34	90,697	0,00465	422	94,887	0,00142	135
35	90,275	0,00487	440	94,752	0,00155	147
36	89,835	0,00514	462	94,605	0,00168	159
37	89,373	0,00550	492	94,446	0,00182	172
38	88,881	0,00595	529	94,274	0,00196	185
39	88,352	0,00649	573	94,089	0,00212	199
40	87,779	0,00708	622	93,890	0,00228	214
41	87,157	0,00770	671	93,676	0,00247	231
42	86,486	0,00831	719	93,445	0,00267	249

возраст мужчины женщины

x	l_x	q_x	d_x	l_x	q_x	d_x
43	85,767	0,00888	762	93,196	0,00289	270
44	85,005	0,00943	801	92,926	0,00314	292
45	84,204	0,00997	840	92,634	0,00341	316
46	83,364	0,01057	881	92,318	0,00369	341
47	82,483	0,01126	929	91,977	0,00399	367
48	81,554	0,01208	985	91,610	0,00430	394
49	80,569	0,01303	1,050	91,216	0,00465	424
50	79,519	0,01409	1,121	90,792	0,00506	459
51	78,398	0,01522	1,193	90,333	0,00554	500
52	77,205	0,01637	1,264	89,833	0,00610	548
53	75,941	0,01754	1,332	89,285	0,00673	601
54	74,609	0,01872	1,397	88,684	0,00740	656
55	73,212	0,01997	1,462	88,028	0,00806	709
56	71,750	0,02136	1,532	87,319	0,00866	756
57	70,218	0,02293	1,610	86,563	0,00919	795
58	68,608	0,02470	1,695	85,768	0,00969	831
59	66,913	0,02665	1,783	84,937	0,01023	869
60	65,130	0,02871	1,870	84,068	0,01094	919
61	63,260	0,03080	1,949	83,149	0,01193	992
62	61,311	0,03296	2,021	82,157	0,01318	1,083
63	59,290	0,03523	2,089	81,074	0,01467	1,189
64	57,201	0,03765	2,153	79,885	0,01634	1,305
65	55,048	0,04027	2,217	78,580	0,01819	1,430
66	52,831	0,04310	2,277	77,150	0,02024	1,561
67	50,554	0,04616	2,333	75,589	0,02249	1,700
68	48,221	0,04947	2,385	73,889	0,02497	1,845
69	45,836	0,05304	2,431	72,044	0,02771	1,997
70	43,405	0,05691	2,470	70,043	0,03073	2,153
71	40,935	0,06108	2,500	67,894	0,03406	2,212
72	38,435	0,06558	2,521	65,582	0,03772	2,474
73	35,914	0,07044	2,530	63,108	0,04176	2,635

возраст мужчины женщины

x	l_x	q_x	d_x	l_x	q_x	d_x
74	33,384	0,07568	2,527	60,473	0,04620	2,794
75	30,857	0,08129	2,508	57,679	0,05106	2,945
76	28,349	0,08738	2,477	54,736	0,05642	3,088
77	25,872	0,09393	2,430	51,646	0,06232	3,218
78	23,442	0,10098	2,367	48,428	0,06879	3,331
79	21,075	0,10857	2,288	45,097	0,07589	3,423
80	18,787	0,11672	2,193	41,674	0,08368	3,487
81	16,594	0,12548	2,082	38,187	0,09221	3,521
82	14,512	0,13489	1,957	34,666	0,10155	3,520
83	12,555	0,14497	1,820	31,146	0,11176	3,481
84	10,735	0,15577	1,672	27,665	0,12291	3,400
85	9,063	0,16733	1,517	24,265	0,13507	3,277
86	7,546	0,20000	1,509	20,988	0,20000	4,197
87	6,037	0,40000	2,414	16,791	0,40000	6,716
88	3,623	0,60000	2,174	10,075	0,60000	6,045
89	1,449	0,80000	1,159	4,030	0,80000	3,224
90	290	1,00000	290	806	1,00000	806

]

Квантили x_p уровня p стандартного нормального распределения [, с.407]

p	0,90	0,95	0,975	0,99	0,995	0,999	0,9999
x_p	1,282	1,645	1,960	2,326	2,576	3,090	3,291

]

Фрагмент таблицы продолжительности жизни (США, 1979–1981 г.) [1, с.55–58]

x	l_x	q_x	d_x	L_x	T_x	$\mathring{e}_x$
0	100 000	0,01260	1 260	98 973	7 387 758	73,88
1	98 740	0,00093	92	98 694	7288785	73,82
2	98 648	0,00065	64	98 617	7190091	72,89
3	98 584	0,00050	49	98 560	7091474	71,93
⋮	⋮	⋮	⋮	⋮	⋮	⋮
19	97 851	0,00112	110	97796	5518733	56,40
20	97 741	0,00120	118	97682	5420937	55,46
21	97 623	0,00127	124	97561	5323255	54,53
22	97 499	0,00132	129	97435	5225694	53,60
⋮	⋮	⋮	⋮	⋮	⋮	⋮
40	94 926	0,00232	220	94817	3492100	36,79
41	94 706	0,00254	241	94585	3397283	35,87
42	94 465	0,00279	264	94334	3302698	34,96
43	94 201	0,00306	288	94057	3208364	34,06
⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	83 726	0,01368	1145	83153	1676326	20,02
61	82 581	0,01493	1233	81965	1593173	19,29
62	81 348	0,01628	1324	80686	1511208	18,58
63	80 024	0,01767	1415	79316	1430522	17,88
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	1 150	0,29120	335	983	3137	2,73
101	815	0,30139	245	692	2154	2,64
102	570	0,31089	177	481	1462	2,57
103	393	0,31970	126	330	981	2,50
104	267	0,32786	88	223	651	2,44
105	179	0,33539	60	150	428	2,38
106	119	0,34233	41	99	278	2,33
107	78	0,34870	27	64	179	2,29

]
Утверждение 1П. Если функция $f(x)$ имеет в точке $x = x_0 \in X$ производную $f^{(n)}(x_0)$, то она в окрестности точки $x = x_0$ непрерывна и имеет производные $f^{'}(x)$, $f^{''}(x), \ldots, f^{(n-2)}(x)$, непрерывные в окрестности точки $x = x_0$, и производную $f^{(n-1)}(x)$, непрерывную в точке $x = x_0$.

Утверждение 2П (формула Тейлора с остаточным членом в форме Пеано). Если функция $f(x)$ n раз дифференцируема в точке $x = x_0$, то в некоторой окрестности этой точки

$$f(x) = f(x_0) + \frac{f^{'}(x_0)}{1!}(x - x_0) + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + o((x - x_0)^n),$$
$$x \to x_0.$$

Утверждение 3П (формула Тейлора с остаточным членом в форме Лагранжа). Если функция $f(x)$ имеет в некоторой окрестности точки x_0 производную порядка $n + 1$, то для любой точки x из указанной окрестности

$$f(x) = f(x_0) + \frac{f^{'}(x_0)}{1!}(x - x_0) + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n +$$
$$+ \frac{f^{(n+1)}(x_0 + \theta(x - x_0))}{(n+1)!}(x - x_0)^n, \quad 0 < \theta < 1.$$

Теорема 1П [48, с. 284] (теорема Лебега о предельном переходе под знаком интеграла или о мажорируемой сходимости). Пусть последовательность $f_n(x)$ и функция $\varphi(x)$ интегрируемы на измеримом множестве A, при всех n и при почти всех $x \in A$

$$|f_n(x)| \leq \varphi(x)$$

и $\lim\limits_{n\to\infty} f_n(x) = f(x)$ для почти всех $x \in A$. Тогда $\int_A |f(x)|dx < \infty$,

$$\lim_{n\to\infty} \int_A f_n(x)dx = \int_A f(x)dx,$$

$$\lim_{n\to\infty} \int_A |f_n(x) - f(x)|dx = 0.$$

Теорема 2П (теорема Лебега о мажорируемой сходимости в терминах математического ожидания). Пусть $\eta, \xi_1, \xi_2, \ldots$ — случайные величины такие, что $|\xi_n| \leq \eta$, $\mathbf{E}\eta < \infty$ и $\xi_n \to \xi$ с вероятностью 1. Тогда $\mathbf{E}\xi < \infty$,

$$\mathbf{E}\xi_n \to \mathbf{E}\xi,$$

$$\mathbf{E}|\xi_n - \xi| \to 0$$

при $n \to \infty$.

Теорема 3П (теорема 2.1.1 из [32]). *Если* $d_n(x_n - \phi) \Longrightarrow \mathcal{N}_s\left\{\mu, \sigma\right\}$, $H(z) \in \mathcal{N}_{1,s}(\phi)$, $H^{(1)}(\phi) \neq 0$, *то*

$$d_n\left(H(x_n) - H(\phi)\right) \Longrightarrow \mathcal{N}_1\left\{\sum_{j=1}^{s} H_j(\phi)\mu_j, \sum_{j,p=1}^{s} H_j(\phi)H_p(\phi)\sigma_{jp}\right\}. \quad (\text{П.1})$$

Теорема 4П [34]. *Пусть* $\{\mathcal{T}_n, \quad n = 1, 2, \ldots\}$ *— последовательность статистик такая, что* $\sqrt{d_n}\left[\mathcal{T}_n - \theta\right] \Longrightarrow \mathcal{N}_1\left\{0, \sigma^2(\theta)\right\}$, *числовая последовательность* $d_n \uparrow \infty$. *Пусть* g *— функция одной переменной, имеющая первую производную* g'. *Тогда, если* $g'(\theta) \neq 0$, *то*

$$\sqrt{d_n}\left[g\left(\mathcal{T}_n\right) - g(\theta)\right] \Longrightarrow \mathcal{N}_1\left\{0, \left[g'(\theta)\sigma(\theta)\right]^2\right\}.$$

Если, кроме того, g' *непрерывна, то*

$$\frac{\sqrt{d_n}\left[g\left(\mathcal{T}_n\right) - g(\theta)\right]}{g'\left(\mathcal{T}_n\right)} \Longrightarrow \mathcal{N}_1\left\{0, \sigma^2(\theta)\right\},$$

а если и $\sigma(\theta)$ *непрерывна, то*

$$\frac{\sqrt{d_n}\left[g\left(\mathcal{T}_n\right) - g(\theta)\right]}{g'\left(\mathcal{T}_n\right)\sigma\left(\mathcal{T}_n\right)} \Longrightarrow \mathcal{N}_1\left\{0, 1\right\}.$$

Замечание 1П. *В* [34, с. 335], *приводится аналогичный результат при* $d_n = n$.

Теорема 5П [1]. *Пусть выполнены следующие условия:*
1) $H(z) \in \mathcal{N}_{2,s}(\phi)$,
2) $M_m\left\|x_n\right\| = O\left(d_n^{-m/2}\right)$ *для некоторого* $m \geq 3$, $m \in N$,
3) $\delta = \delta_n = Cd_n^{-1}$,
4) $H(\phi) \neq 0$ *или* $\tau \in N^+$. *Тогда для любых* $(\tau, k) \in T(m)$

$$\left|\mathbf{E}\left[\widetilde{\Phi}(x_n, \delta) - H(\phi)\right]^k - \mathbf{E}\left[\nabla H(\phi)(x_n - \phi)^T\right]^k\right| = O\left(d_n^{-(k+1)/2}\right), \quad (\text{П.2})$$

где $\widetilde{\Phi}(x_n, \delta) = H(x_n)/\left(1 + \delta\left|H(x_n)\right|^\tau\right)^\rho$, $\tau > 0$, $\rho > 0$, $\rho\tau \geq 1$, $\delta > 0$.

[1] Согласно обозначений на стр. 92.

Сформулируем двумерную центральную предельную теорему в схеме серий в следующих обозначениях: $\{\xi_{j,n}, \eta_{j,n}\}_{j=1}^{n}$, $n = 1, 2, \ldots$ — последовательность независимых одинаково распределенных двумерных векторов в схеме серий (распределение $(\xi_{j,n}, \eta_{j,n})$ зависит от n); $\sigma_n = n\mathbf{E}\left(\xi_{1,n}, \eta_{1,n}\right)^T \left(\xi_{1,n}, \eta_{1,n}\right)$, здесь ^{T} — знак транспонирования; $\|(\xi, \eta)\| = \sqrt{\xi^2 + \eta^2}$.

Теорема 6П [9] (двумерная центральная предельная теорема в схеме серий).

Пусть: 1) $\mathbf{E}\left(\xi_{j,n}, \eta_{j,n}\right) = (0, 0)$, 2) $\mathbf{E}\left\|\xi_{j,n}, \eta_{j,n}\right\|^2 < \infty$, 3) *выполнено условия Линдеберга: для любого* τ *при* $n \to \infty$*:*

$$n\mathbf{E}\left(\left\|\xi_{j,n}, \eta_{j,n}\right\|^2 ; \left\|\xi_{j,n}, \eta_{j,n}\right\| > \tau\right) \longrightarrow 0,$$

4) $\sigma_n \longrightarrow \sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$. *Тогда вектор* $\sum\limits_{j=1}^{n} \left(\xi_{j,n}, \eta_{j,n}\right)$ *имеет двумерное предельное нормальное распределение* $\mathcal{N}_2\left\{(0, 0), \sigma\right\}$.