

С.В. Осовец**ТЕОРЕТИЧЕСКАЯ ОЦЕНКА ФУНКЦИИ РИСКА И КОЭФФИЦИЕНТА ОБЩЕЙ СМЕРТНОСТИ НА ОСНОВАНИИ РАСПРЕДЕЛЕНИЯ ВЕЙБУЛЛА**Южно-Уральский институт биофизики, Озерск, Челябинская обл., Россия. E-mail: clinic@subi.su
С.В. Осовец – вед.н.с., к.т.н.**Реферат**Цель: Теоретическая оценка параметров функции риска общей смертности на основе распределения Вейбулла и вычисление среднего коэффициента смертности.Результаты: В общем виде функция распределения длительности жизни определена следующим образом:

$$R(t) = 1 - \exp \left[- \int_0^t \lambda(t) dt \right],$$

где $\lambda(t)$ – функция интенсивности смертности.

Для распределения Вейбулла функция интенсивности смертности (функция риска) имеет степенной вид:

$$\lambda(t) = \lambda_0 \cdot \alpha \cdot t^{\alpha-1}$$

где λ_0 и α – параметры степенной модели.Построен функционал для нахождения методом максимального правдоподобия параметров λ_0 и α применительно к конкретной схеме наблюдений (m смертей из N возможных событий на интервале $[0; T]$), который имеет следующий вид:

$$L(\lambda_0, \alpha) = \prod_{i=1}^m (\lambda_0 t_i^{\alpha} e^{-\lambda_0 t_i^{\alpha}}) \cdot \prod_{i=m+1}^N e^{-\lambda_0 t_i^{\alpha}}.$$

В результате минимизации функционала были получены обобщенные аналитические оценки параметров λ_0 и α для функции риска на заданном периоде наблюдения. Предложен новый метод степенных человеко-лет, который может быть в перспективе адаптирован к задачам радиационной эпидемиологии. На выбранном временном отрезке наблюдения $[t_1, t_2]$ (при условии $t_2 > t_1$ и $[t_1, t_2] \subset [0; T]$) получена оценка среднего коэффициента смертности $\bar{\lambda}(t)$:

$$\bar{\lambda}(t) = \lambda_0 \frac{(t_2^{\alpha} - t_1^{\alpha})}{(t_2 - t_1)}.$$

Из этой формулы следует, что величина $\bar{\lambda}(t)$ зависит не только от параметров λ_0 и α , но и от величины временного отрезка (t_1, t_2), на котором происходит процесс усреднения. В частном случае (при $\alpha = 1$), значение среднего коэффициента смертности $\bar{\lambda}(t) = \lambda_0$, которое соответствует экспоненциальному распределению.Выводы: Разработан обобщенный метод оценки коэффициента и функции риска общей смертности на основе распределения Вейбулла. Полученные теоретические результаты в дальнейшем могут быть использованы в области радиационной эпидемиологии.**Ключевые слова:** функция риска, общая смертность, распределение Вейбулла, степенные человеко-годы, методы оценки

Поступила: 16.08.2016. Принята к публикации: 20.02.2017

Введение

Наиболее важными характеристиками при изучении процесса общей смертности в различных когортах и больших группах населения являются коэффициент и функция интенсивности (функция риска) смертности [1, 2]. С количественной стороны коэффициент смертности λ_0 определяют как отношение числа случаев смерти к накопленной сумме человеко-лет наблюдения в данной группе населения; с качественной (содержательной) стороны смысл коэффициента смертности – это средняя доля умерших в год за рассматриваемый отрезок времени наблюдения. При более общем подходе коэффициент смертности является функцией времени $\lambda(t)$, и в этом случае он задается так называемой функцией интенсивности смертности. В качестве временной характеристики здесь может выступать наблюдаемое время наступления неблагоприятного события (смерти) или возраст.

Задавая различные формы функции интенсивности смертности, можно получить те или иные статистические модели распределения продолжительности жизни. Например, если функция интенсивности смертности будет содержать два неизвестных параметра, то статистическая модель продолжительности жизни может быть описана с помощью распределения Гомперца, Вейбулла или биномиального закона [3]. В простейшем случае, когда функция интенсивности смертности является постоянной величиной, не зависящей от времени $\bar{\lambda}(t) = \lambda_0 \sim const$, закон распределения продолжительности жизни можно приближенно опи-

сать и однопараметрическими распределениями, такими как экспоненциальное или пуассоновское.

В недавно опубликованной теоретической работе, посвященной методологии получения оценок среднего риска общей смертности в эпидемиологических исследованиях [4], авторы, используя специальный вариант метода максимального правдоподобия, убедительно показали эквивалентность экспоненциальной и пуассоновской моделей при статистической оценке среднего коэффициента общей смертности. Помимо этого, в цитируемой работе были представлены дополнительные теоретические наработки, которые оказались полезны в дальнейшем не только в аналитической, но и радиационной эпидемиологии [5]. С другой стороны, основываясь на разработанной методологии, в нашем кратком сообщении [6] был предложен обобщенный метод степенных человеко-лет для оценки коэффициента и функции интенсивности смертности на основе двухпараметрического распределения Вейбулла.

В связи с этим целью настоящего исследования является развитие указанного подхода и получение обобщенных оценок параметров функции риска и среднего коэффициента общей смертности.

Результаты и обсуждение

Известно, что в общем виде функцию распределения длительности жизни для исследуемой когорты или большой группы населения можно определить следующим образом:

$$R(t) = 1 - \exp \left[- \int_0^t \lambda(t) dt \right], \quad (1)$$

где $\lambda(t)$ – функция интенсивности смертности.

Таким образом, конкретный вид функции распределения $R(t)$ полностью определяется видом функции интенсивности смертности [3, 7].

Допустим, что функция интенсивности смертности является степенной функцией от t следующего вида:

$$\lambda(t) = \lambda_0 \cdot \alpha \cdot t^{\alpha-1} \quad (2)$$

где λ_0 и α – параметры этой степенной модели.

Подставляя $\lambda(t)$ из формулы (2) в обобщенное выражение (1), после интегрирования получаем конкретный вид функции распределения $R(t)$:

$$R(t) = 1 - \exp(-\lambda_0 t^\alpha) \quad (3)$$

Такой закон распределения времени жизни называется вейбулловским [7, 8], и он полезен не только для решения задач эпидемиологии и демографии, но и широко применяется в теории надежности сложных технических систем [9]. Следует обратить внимание и на тот факт, что в частном случае при равенстве безразмерного параметра формы единице ($\alpha = 1$) функция интенсивности смертности $\lambda(t)$, как это видно из (2), становится постоянной величиной, равной $\lambda_0 = \text{const}$. При этом функция распределения (3) переходит к своему частному виду – экспоненциальному распределению с единственным параметром λ_0 .

Преимущество использования вейбулловского закона распределения по сравнению с экспоненциальным распределением для количественного описания функции интенсивности смертности $\lambda(t)$ можно показать с помощью следующего условного графика (рис.).

Из рис. видно, что если под временным фактором «годы» подразумевается возраст, то всю кривую можно условно разделить на три области: I – детская смертность, II – смертность в зрелом возрасте, III – смертность в пожилом возрасте. В каждом возрастном интервале функция интенсивности смертности (2) будет иметь различные значения параметра α ($\alpha < 1$; $\alpha \approx 1$; $\alpha > 1$). Если под временным фактором «годы» понимать длительное время наблюдения за стареющей и постепенно вымирающей когортой, начиная со зрелого возраста, то зависимость коэффициента смертности от времени будет аналогична поведению кривой в областях II и III. В целом можно констатировать, что и в том и другом случае зависимость $\lambda(t)$ будет характеризоваться нелинейной зависимостью от времени.

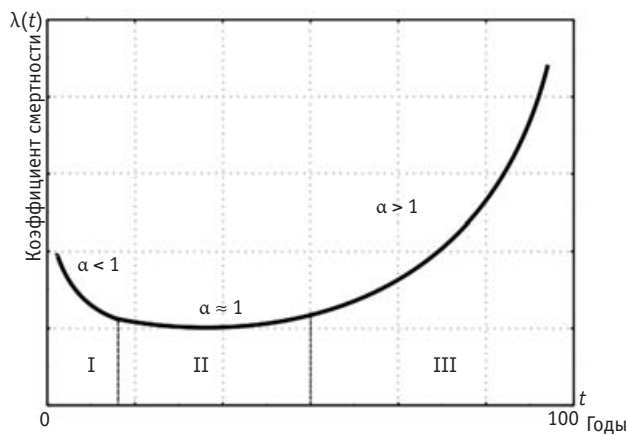


Рис. Зависимость коэффициента смертности от временного фактора (годы)

Следовательно, аппроксимация такой зависимости с помощью единственного параметра λ_0 (на базе экспоненциального распределения) является очень приближенной оценкой и может быть справедливой только для небольших временных отрезков наблюдения. С другой стороны, построение зависимости $\lambda(t)$ на базе двухпараметрического распределения Вейбулла с последующей оценкой параметров λ_0 и α , вероятнее всего, будет более точно описывать коэффициент и функцию интенсивности смертности (необходимо также отметить, что при возрасте более 85 лет зависимость $\lambda(t)$ лучше описывается функцией Гомперца [3] по сравнению с распределением Вейбулла).

Для построения функционала на основе распределения Вейбулла, с помощью которого можно получить формулы для оценки параметров λ_0 и α , воспользуемся методологией, разработанной ранее для экспоненциального распределения [4].

Пусть $t \in [0; T]$, где T – одновременно и продолжительность интервала наблюдения, и время окончания наблюдения $t_k = T$ (при этом $t_0 = 0$). Рассмотрим группу из N индивидуумов на интервале $[0; T]$, в котором произошло m случаев нежелательных событий (смертей), причем каждый i -тый случай произошел в свой i -тый момент времени – t_i , $i = \overline{1, m}$. Индикатором состояния i -того индивидуума Ind_i ($i = \overline{1, m}$) является дискретная величина:

$$Ind_i = \begin{cases} 1, & \text{с вероятностью } R(t_i) = 1 - \exp(-\lambda_0 t_i^\alpha) \\ 0, & \text{с вероятностью } Q(T) = \exp(-\lambda_0 T^\alpha) \end{cases} \quad (4)$$

Единица указывает на то, что событие с вероятностью $R(t_i)$ произошло в интервале $[0, t_i) \subset [0; T]$, а ноль – на то, что с вероятностью $Q(T)$ в интервале $[0; T]$ событие не произошло.

Разложим функцию $R(t)$ в ряд Маклорена:

$$R(t) = 1 - e^{-\lambda_0 t^\alpha} = e^{-\lambda_0 t^\alpha} (e^{\lambda_0 t^\alpha} - 1) \approx \lambda_0 t^\alpha e^{-\lambda_0 t^\alpha}. \quad (5)$$

При таком разложении приближенное равенство (5) будет выполняться при дополнительном условии:

$$|\lambda_0 t^\alpha| < 1. \quad (6)$$

Функционал для нахождения параметров λ_0 и α применительно к данной схеме наблюдений (m событий из N возможных событий на интервале $[0; T]$) будет иметь следующий вид:

$$L(\lambda_0, \alpha) = \prod_{i=1}^m R(t_i) \cdot \prod_{i=m+1}^N Q(T) = \prod_{i=1}^m (\lambda_0 t_i^\alpha e^{-\lambda_0 t_i^\alpha}) \cdot \prod_{i=m+1}^N e^{-\lambda_0 T^\alpha}. \quad (7)$$

Логарифмируя обе части равенства (7), получаем:

$$S = \ln L(\lambda_0, \alpha) = m \ln \lambda_0 - \lambda_0 \left[\sum_{i=1}^m t_i^\alpha + (N - m) T^\alpha \right] + \alpha \sum_{i=1}^m \ln t_i. \quad (8)$$

Дифференцируя функционал (8) по параметрам λ_0 и α , получаем следующую систему уравнений:

$$\begin{cases} \frac{\partial S}{\partial \lambda_0} = \frac{m}{\lambda_0} - \left[\sum_{i=1}^m t_i^\alpha + (N - m) T^\alpha \right] = 0, \\ \frac{\partial S}{\partial \alpha} = \sum_{i=1}^m \ln t_i - \lambda_0 \left[\sum_{i=1}^m (t_i^\alpha \ln t_i) + (N - m) T^\alpha \ln T \right] = 0. \end{cases} \quad (9)$$

Из первого уравнения системы (9) следует:

$$\lambda_0 = m / \left[\sum_{i=1}^m t_i^\alpha + (N - m) T^\alpha \right]. \quad (10)$$

Подставляя λ_0 из (10) во второе уравнение системы (9), получаем нелинейное уравнение относительно параметра α :

$$m \left[\frac{\sum_{i=1}^m (t_i^\alpha \ln t_i) + (N-m)T^\alpha \ln T}{\sum_{i=1}^m t_i^\alpha + (N-m)T^\alpha} \right] = \sum_{i=1}^m \ln t_i. \quad (11)$$

Решая данное нелинейное уравнение относительно параметра α любым итерационным методом (к примеру, методом Ньютона [10]), можно с заданной точностью ε вычислить значение параметра α . После этого, подставляя найденное значение α в формулу (10), можно получить численную оценку параметра λ_0 . В итоге функция интенсивности смертности $\lambda(t)$ будет полностью определена в соответствии с формулой (2).

В формуле (10) стоящее в знаменателе выражение в квадратных скобках представляет суммарное количество человеко-лет наблюдения в степенной форме или, короче, степенные человеко-годы [6].

В частности, если параметр $\alpha = 1$, то коэффициент смертности λ_0 из выражения (10) определяется классически, – как отношение числа случаев смерти к стандартной сумме человеко-лет наблюдения. Единственным ограничением при оценке параметров λ_0 и α с помощью функционала (7) является неравенство (6), гарантирующее точность и качество аппроксимации $R(t)$ при разложении в ряд Маклорена (см. равенство (5)).

Более общим подходом к оценке параметров вейбулловского распределения λ_0 и α , на который не распространяется указанное выше ограничение, является минимизация следующего функционала:

$$L(\lambda_0, \alpha) = \prod_{i=1}^m R(t_i) \cdot \prod_{i=m+1}^N Q(T) = \prod_{i=1}^m [1 - \exp(-\lambda_0 t_i^\alpha)] \cdot \prod_{i=m+1}^N e^{-\lambda_0 T^\alpha} \quad (12)$$

Дифференцируя последовательно по параметрам λ_0 и α данный функционал, можно получить систему двух нелинейных уравнений и, затем, применяя итерационный метод Ньютона, численно оценить параметры λ_0 и α . Однако при таком общем подходе к оценке параметров λ_0 и α вряд ли возможно в явном виде увидеть и выделить из сложных выражений метод степенных человеко-лет, как это было показано в предыдущем алгоритме расчетов с условным ограничением (6). Более того, при минимизации функционала (12) по параметрам λ_0 и α получается довольно громоздкая система нелинейных уравнений, которую, вероятнее всего, нецелесообразно приводить в данном исследовании.

После нахождения параметров и построения функции интенсивности смертности заключительным этапом теоретического исследования является оценка на выбранном временном отрезке наблюдения $[t_1; t_2]$ средней величины коэффициента смертности $\bar{\lambda}(t)$.

Для этого воспользуемся известной теоремой о среднем [11]. Итак, пусть задан временной отрезок $[t_1; t_2]$ для функции $\lambda(t)$ при условии $t_2 > t_1$ и $[t_1; t_2] \subset [0; T]$, тогда средняя величина коэффициента смертности на этом отрезке будет равной

$$\bar{\lambda}(t) = \int_{t_1}^{t_2} \lambda(t) dt / (t_2 - t_1) = \int_{t_1}^{t_2} \lambda_0 \alpha t^{\alpha-1} dt / (t_2 - t_1)$$

или, окончательно,

$$\bar{\lambda}(t) = \lambda_0 \frac{(t_2^\alpha - t_1^\alpha)}{(t_2 - t_1)}. \quad (13)$$

Из этого равенства видно, что величина $\bar{\lambda}(t)$ зависит не только от параметров λ_0 и α , но и от величины временного отрезка (t_1, t_2) , на котором происходит процесс усреднения. Более четко проиллюстрировать смысл формулы (13) можно в частном случае: пусть $t_1 = 0$ и $t_2 = t$, при условии $[0; t] \subset [0; T]$. Тогда равенство (13) трансформируется к следующему виду:

$$\bar{\lambda}(t) = \lambda_0 t^{\alpha-1}. \quad (14)$$

Из этой формулы видно, что величина $\bar{\lambda}(t)$ является функцией времени и отличается от функции интенсивности смертности $\lambda(t)$ (см. формулу (2)) только на постоянный коэффициент α . Очевидно, что при $\alpha = 1$ или $t = 1$ равенство (14) преобразуется в более простое, ранее полученное [4] для однопараметрического экспоненциального распределения: $\bar{\lambda}(t) = \lambda_0$. Это вполне ожидаемый результат исследования, который представляет собой частное следствие предлагаемого подхода к оценке среднего значения коэффициента смертности.

Заключение

Таким образом, основываясь на двухпараметрическом распределении Вейбулла, были получены обобщенные теоретические оценки параметров для функции интенсивности смертности и среднего коэффициента смертности на заданном периоде наблюдения. Предложен новый метод степенных человеко-лет, который может быть в перспективе адаптирован к задачам радиационной эпидемиологии.

Необходимо отметить, что формулы для оценки среднего коэффициента и функции интенсивности общей смертности, а также используемые методы оценки параметров этих величин имеют на настоящий момент в определенной степени эвристический (поисковой) характер. Поэтому в дальнейшем требуется более тщательное теоретическое обоснование и всесторонняя практическая проверка полученных результатов.

СПИСОК ЛИТЕРАТУРЫ

1. Альбом А., Норелл С. Введение в современную эпидемиологию. – Таллин: Институт экспериментальной и клинической медицины. 1996. 122 с.
2. Boyle P., Parkin D. Statistical methods for registries // In: Cancer Registration (Principles and Methods). IARC Publication. – Lyon. 1991. № 95. P. 126–158.
3. Гаврилов Л.А., Гаврилова Н.С. Биология продолжительности жизни: количественные аспекты. – М.: Наука. 1986. 169 с.
4. Белых Л.Н., Бирюков А.П., Васильев Е.В., Невзоров В.П. О теоретических оценках среднего риска общей смертности и правомерности применения различных законов распределения вероятностей в эпидемиологических исследованиях // Мед. радиол. и радиац. безопасность. 2015. Т. 60. № 5. С. 40–45.
5. Белых Л.Н., Бирюков А.П., Васильев Е.В., Невзоров В.П. Оценки пожизненного радиогенного риска онкологической смертности и заболеваемости // Мед. радиол. и радиац. безопасность. 2015. Т. 60. № 6. С. 20–26.
6. Осовец С.В. Метод степенных человеко-лет для оценки коэффициента и функции общей смертности // XVI Все-

- российская научно-практическая конференция «Дни науки – 2016». Материалы конференции. Озерск. 20–23 апреля 2016 г. – Озерск: ОТИ НИЯУ МИФИ. 2016. С. 194–195.
7. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное издание. – М.: Финансы и статистика. 1983. 471 с.
 8. Кобзарь А.Н. Прикладная математическая статистика для инженеров и научных сотрудников. – М.: Физматлит. 2012. 816 с.
 9. Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д. Математические методы в теории надежности: основные характеристики надежности и их статистический анализ. – М.: КД «ЛИБРОКОМ». 2013. 584 с.
 10. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. – М.: Физматлит. 2001. 632 с.
 11. Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. – М.: Физматлит. 2006. Т. 1. 680 с.

DOI 10.12737/article_58f0b9573730e4.55456538

Theoretical Estimation of Risk Function and Total Mortality Rate Based on Weibull Distribution

S.V. Osovets

Southern Urals Biophysics Institute, Ozyorsk, Chelyabinsk region, Russia. E-mail: clinic@subi.su

S.V. Osovets – senior researcher, PhD Sc. Tech.

Abstract

Purpose: To estimate theoretically total mortality risk function parameters based on Weibull distribution and to calculate the average mortality rate.

Results: Lifespan distribution function takes the general form of:

$$R(t) = 1 - \exp \left[- \int_0^t \lambda(t) dt \right],$$

where $\lambda(t)$ is mortality rate function.

To obtain Weibull's distribution mortality rate function (risk function) is represented as power equation:

$$\lambda(t) = \lambda_0 \cdot \alpha \cdot t^{\alpha-1}$$

where λ_0 and α are power model parameters.

We provided a function for estimation of λ_0 and α parameters to be applied to a certain follow-up pattern using maximum likelihood technique (m deaths among N of potential outcomes during [0; T] period of time) which looks as:

$$L(\lambda_0, \alpha) = \prod_{i=1}^m (\lambda_0 t_i^\alpha e^{-\lambda_0 t_i^\alpha}) \cdot \prod_{i=m+1}^N e^{-\lambda_0 T^\alpha}.$$

As a result of minimization of the function we obtained generalized analytical λ_0 and α parameter estimates for risk association during a specified period of follow-up. We developed a new technique using power man-years which may be adapted for radiation epidemiology in future. For a set period of follow-up $[t_1, t_2]$ (with assumptions $t_2 > t_1$ и $[t_1, t_2] \subset [0; T]$) the mean mortality rate $\bar{\lambda}(t)$ was estimated to be:

$$\bar{\lambda}(t) = \lambda_0 \frac{(t_2^\alpha - t_1^\alpha)}{(t_2 - t_1)}.$$

This equation implies that $\bar{\lambda}(t)$ depends not only on λ_0 and α parameters but also on the duration of the time line (t_1, t_2) during which the averaging is carried out. As a special case (with $\alpha = 1$) the mean mortality rate is $\bar{\lambda}(t) = \lambda_0$, and consistent with exponential distribution.

Conclusion: The generalized technique for estimation of the coefficient and function of total mortality risk was developed on the basis of Weibull distribution. The obtained theoretical results may be used in radiation epidemiology in future.

Key words: risk function, total mortality, Weibull distribution, power man-years, estimation technique

REFERENCES

1. Al'bom A., Norell S. Vvedenie v sovremennuyu ehpidemiologiyu. – Tallin: Institut ehksperimental'noj i klinicheskoy mediciny. 1996. 122 s.
2. Boyle P., Parkin D. Statistical methods for registries // In: Cancer Registration (Principles and Methods). IARC Publication. Lyon. 1991. № 95. P. 126–158.
3. Gavrilov L.A., Gavrilova N.S. Biologiya prodolzhitel'nosti zhizni: kolichestvennye aspekty. – М.: Nauka. 1986. 169 s.
4. Belyh L.N., Biryukov A.P., Vasil'ev E.V., Nevzorov V.P. O teoreticheskikh ocnenках srednego riska obshchej smertnosti i pravomernosti primeneniya razlichnyh zakonov raspredeleniya veroyatnostej v ehpidemiologicheskikh issledovaniyah // Med. radiol. i radiac. bezopasnost'. 2015. T. 60. № 5. P. 40–45.
5. Belyh L.N., Biryukov A.P., Vasil'ev E.V., Nevzorov V.P. Ocnki pozhiznennogo radiogennogo riska onkologicheskoy smertnosti i zabolevaemosti // Med. radiol. i radiac. bezopasnost'. 2015. T. 60. № 6. P. 20–26.
6. Osovets S.V. Metod stepennyh cheloveko-let dlya ocenki koehfficienta i funkcii obshchej smertnosti // XVI Vserossiyskaya nauchno-prakticheskaya konferenciya «Dni nauki – 2016». Materialy konferencii. Ozersk. 20–23 aprelya 2016 g. – Ozersk: OТИ НИЯУ МИФИ. 2016. P. 194–195.
7. Ajvazyan S.A., Enyukov I.S., Meshalkin L.D. Prikladnaya statistika: Osnovy modelirovaniya i pervichnaya obrabotka dannyh. Spravochnoe izdanie. – М.: Finansy i statistika. 1983. 471 s.
8. Kobzar' A.N. Prikladnaya matematicheskaya statistika dlya inzhenerov i nauchnyh sotrudnikov. – М.: Fizmatlit. 2012. 816 s.
9. Gnedenko B.V., Belyaev Yu.K., Solov'ev A.D. Matematicheskie metody v teorii nadezhnosti: osnovnye harakteristiki nadezhnosti i ih statisticheskij analiz. – М.: КД «LIBROKOM». 2013. 584 s.
10. Bahvalov N.S., Zhidkov N.P., Kobel'kov G.M. Chislennye metody. – М.: Fizmatlit. 2001. 632 s.
11. Fihtengol'ts G.M. Kurs differencial'nogo i integral'nogo ischisleniya. – М.: Fizmatlit. 2006. T. 1. 680 s.