ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ ВЕЙБУЛЛА ПРИ МАЛОМ ОБЪЕМЕ ВЫБОРКИ В ЗАДАЧАХ ОПТИМИЗАЦИИ ПЛАНИРОВАНИЯ РЕМОНТНЫХ РАБОТ НА ИНЖЕНЕРНЫХ СЕТЯХ

Крыгин А. А.[1]
(*ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва*)

В сконструированной ранее экономико-статистической методике оценки состояния участка инженерной сети и информационной поддержки при принятии решения о продлении его срока службы предъявляются серьезные требования к исходным данным. Одним из условий правильной работы методики является наличие полной статистики повреждений и проведенных ремонтов, начиная с момента ввода участка в эксплуатацию. Однако для многих участков такая статистика имеется только за последние несколько лет. Чтобы расширить область применения на случаи неполной статистики повреждений, было выполнено отдельное исследование по оценке параметров распределения Вейбулла при малом объеме выборки на примере участков сетей теплоснабжения. Данная работа посвящена результатам проведенного исследования. Были определены три класса участков: с достаточно полной статистикой повреждений, с малым объемом статистики и без статистики. Для каждого класса был предложен алгоритм оценки необходимых параметров распределения Вейбулла, далее использующихся в методике оценки состояния участка и определен критерий, по которому участок можно отнести к тому или иному классу. Для участков с достаточно полной статистикой повреждений использовался алгоритм, предложенный в исходной экономико-статистической методике. Для участков с малым объемом статистики к данным по повреждениям добавлялись данные из других, специальным образом отобранных участков. Для участков без статистики получены усредненные значения параметров.

Ключевые слова: оценка параметров распределения Вейбулла при малом объеме выборки, оценка усредненных значений параметров, оптимизация ремонтов участков тепловых сетей.

1. Введение

Анализ этапов жизненного цикла участка инженерной сети в [5] показал, что необходимые при информационной поддерж-

[1] *Андрей Александрович Крыгин, к.т.н., с.н.с.* (*andreyakr@yandex.ru*).

ке управленческих решений методики должны содержать алгоритмы определения комплексного технико-экономического показателя состояния участка и оптимального времени его замены. Одно из требований методики к исходным данным заключается в наличии полной статистики повреждений и ремонтов на участке, начиная с момента его ввода в эксплуатацию. Целью данной работы является расширение области применения методики, предложенной в [5] на участки с неполной статистикой повреждений.

Основная идея методики, заключается в поиске момента времени T_0, в который функция $S_y(t)$ – зависимость удельных суммарных затрат на эксплуатацию участка – достигает своего минимума.

Было установлено, что для определения $S_y(t)$ необходимо спрогнозировать зависимости от времени для следующих величин: общее количество повреждений, общее количество локально-вставочных ремонтов (ЛВР), общая длина ЛВР. Не нарушая общности, в дальнейшем будем рассматривать только задачу определения функции $N(t)$ – зависимости общего количества повреждений от времени эксплуатации участка.

Идея алгоритма определения $N(t)$ в [5] базируется на известном методе регрессии медиан рангов [2]. Время до возникновения повреждения на участке можно рассматривать как случайную величину, описываемую с помощью распределения Вейбулла: $F(t) = 1 - e^{-\left(\frac{t}{b}\right)^c}$ с двумя параметрами b и c. Тогда функцию $N(t)$ можно определить как

$$N(t) = -\ln(1 - F(t)) = \left(\frac{t}{b}\right)^c. \tag{1}$$

Оценки параметров можно получить следующим образом: методом наименьших квадратов определяются коэффициенты прямой линии $y = m \times x + n$ по точкам (x, y), где x соответствует логарифму числа прошедших лет с момента эксплуатации участка, а y – логарифму общего числа повреждений. Тогда параметры b и c распределения Вейбулла определяются по следующим формулам:

(2) $$\begin{cases} b = e^{-\frac{n}{m}}, \\ c = m. \end{cases}$$

При этом было установлено, что лучшие результаты дает предварительная фильтрация точек перед применением метода наименьших квадратов.

В этой предметной области можно выделить четыре основных особенности, отличающих ее от других исследований, посвященных оценкам параметров вероятностного распределения:

– малое количество точек (обусловленное тем, что статистика велась достаточно непродолжительное для этого процесса время);

– нежесткие требования к точности полученных оценок;

– наличие статистики по достаточно большому количеству участков;

– сам факт того, что оценки параметров необходимы для определения комплексного технико-экономического показателя состояния участка и оптимального времени его замены.

Анализ и использование этих особенностей рассмотрено ниже.

Данное исследование методов оценки параметров распределения Вейбулла для участка с неполной статистикой состоит из трех частей: во втором разделе проводится обзор и оценка точности некоторых алгоритмов. В третьем разделе обосновывается необходимость дополнительной классификации участков по объему имеющейся статистики и предлагается ряд алгоритмов по каждому подмножеству. В четвертом разделе строится общий алгоритм оценки параметров.

2. Обзор алгоритмов оценки параметров распределения Вейбулла и оценка их точности

В ГОСТ [2] рассматриваются следующие классы методов оценки: графический, аналитический и анализ Вейбулла – Байеса. Анализ Вейбулла – Байеса не применим к данной задаче, так как этот метод предполагает, что параметр c подбирается каким-либо образом (используя данные предыдущих экспериментов, данные о физике процесса), а параметр b находится методом максимального правдоподобия. Диапазон изменений этого параметра в рассматриваемой области достаточно широк (1,5 – 4,5) и возможности подобрать его нет. Метод оценки, предложенный в [5], относится к аналитическим методам, его можно определить как метод регрессии медиан рангов с добавлением блока фильтрации. Из аналитических методов в ГОСТ также рассмотрен метод максимального правдоподобия и метод доверительных интервалов.

Актуальность задачи оценки параметров распределения Вейбулла подтверждается многочисленными современными работами в различных областях. Помимо методов, рассмотренных в [2], можно отметить методы, предложенные в [4] и [6–20]. В [10] авторы предлагают метод «кросс-энтропии» для оценки максимума функции правдоподобия. Вместо алгоритма наименьших квадратов для метода регрессии медиан рангов в [13] предлагается использовать алгоритм, в котором находится минимум взвешенной суммы квадратов. В [17] на примере определения скорости ветра сравниваются 4 метода оценки параметров: метод максимального правдоподобия, метод регрессии медиан рангов, метод, использующий значения среднего и стандартного отклонения и метод, использующий фактор энергетического рисунка (Energy Pattern Factor Method). Сравнение показало, что нельзя отдать однозначное предпочтение тому или иному методу, при этом метод регрессии медиан рангов был признан одним из лучших.

Во всех работах в качестве примеров рассматривались выборки из 50–500 точек. В данной задаче статистика имеется только за 15 лет, соответственно, в редких случаях количество точек превышает 10. Поэтому логично проверить эти методы для поставленной задачи и найти для них допустимую область применения.

Дальнейшие выкладки для наглядности будут приводиться на примере одного участка; те же операции были проведены на всех участках, по которым имеется достаточно полная статистика. Выводы и результаты этого раздела совпадают с примером практически во всех случаях.

Итак, имеется участок, исходные данные по повреждениям у которого следующие: год ввода в эксплуатацию – 1983, полная статистика повреждений представлена в таблице 1.

Таблица 1. Полная статистика повреждений на участке

Год	Количество повреждений
1995	1
1996	1
2000	1
2001	3
2002	2
2004	1
2005	3
2006	3
2007	2

С помощью предложенных в [5] алгоритмов для этого участка были получены следующие значения параметров распределения Вейбулла: b = 11,82, c = 4,05. На рис. 1 в логарифмических шкалах представлен график всех повреждений и его аппроксимация прямой линией.

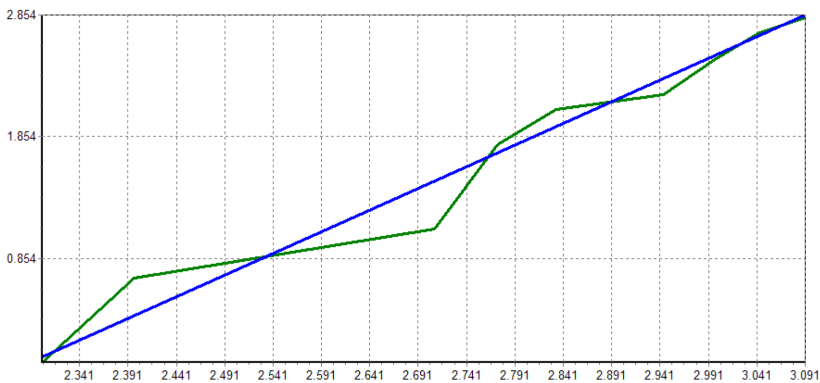

Рис. 1. График повреждений и его аппроксимация

Необходимо отметить, что вначале выполняется фильтрация точек и только после этого по оставшимся точкам определяются коэффициенты прямой, за счет этого достигается более высокая точность прогнозирования. Это непосредственно видно из рисунка: ближе к его концу прямая линия почти совпадает с графиком повреждений. Существенные отклонения графика от прямой линии (на отметках около 2,391 и 2,691) вызваны протяженными локально-вставочными ремонтами (ЛВР) в 1995–96 и 2001 годах, в результате которых была проведена замена примерно 11% наиболее поврежденной части участка. Подобная ситуация встречается практически всегда в случаях существенных отклонений графика от прямой линии. Очевидно, что спрогнозировать такие протяженные ЛВР не представляется возможным, однако долгосрочный прогноз (т.е. усредненное поведение графика) оказывается вполне убедительным.

В данном исследовании рассматриваются три алгоритма, для краткости обозначим их *А*, *Б* и *В*. Также были опробованы и другие алгоритмы, в частности, несколько вариаций метода, предложенного в [10], которые, к сожалению, не дали ощутимых результатов.

АЛГОРИТМ А.

Так как неполная статистика по участку является цензурированной выборкой, то логично применить уже разработанные методы максимального правдоподобия для оценки параметров по цензурированным выборкам [1]. Система уравнений для нахождения параметров распределения Вейбулла в этом случае выглядит следующим образом:

$$(3)\quad \begin{cases} \left(r/c + \sum\limits_{i=1}^{r} \ln t_i \right) \cdot \left(\sum\limits_{i=1}^{r} t_i^c + (K-r)\cdot T^c \right) = \\ = r \cdot \left(\sum\limits_{i=1}^{r} t_i^c \ln t_i + (K-r)\cdot T^c \cdot \ln T \right), \\ b = \left[\sum\limits_{i=1}^{r} t_i^c + (K-r)\cdot T^c \right]^{1/c}. \end{cases}$$

Здесь r – количество повреждений; t_i – время возникновения повреждений; T – период времени от начала эксплуатации участка до окончания периода наблюдения; K – в литературе описывается как общее количество объектов, применительно к участку инженерной сети эту величину можно рассматривать как отношение общей длины участка к длине повреждения. В данном случае участок расположен в непроходном канале и при возникновении повреждения выполняется шурф, диаметр которого обычно составляет 1 метр, поэтому логично принять величину K равной длине участка в метрах.

АЛГОРИТМ Б.

Второй алгоритм, предложенный в [4], работает по следующему принципу: пусть интервал $[t_{begin}; t_{end}]$ соответствует периоду наблюдения за участком. Зададим на этом интервале параметр τ и будем рассматривать два интервала $[t_{begin}; \tau]$ и $[\tau; t_{end}]$. Пусть количество повреждений, произошедших на этих интервалах равно соответственно n_1 и n_2. Тогда, оценку параметров b и c можно получить из следующей системы уравнений:

$$(4)\quad \begin{cases} \left(\dfrac{\tau}{b}\right)^c - \left(\dfrac{t_{begin}}{b}\right)^c = n_1 , \\ \left(\dfrac{t_{end}}{b}\right)^c - \left(\dfrac{\tau}{b}\right)^c = n_2 . \end{cases}$$

Параметр τ обычно выбирается так, чтобы два интервала были примерно одинаковы. Это обусловлено тем, что если интервалы будут перекрываться, то теряется независимость уравнений, а если один из интервалов будет существенно больше другого, то теряется «значимость» одного из уравнений.

АЛГОРИТМ В.

В качестве третьего алгоритма будет использоваться исходный алгоритм оценки параметров по полной статистике, описанный в [5], для краткости в дальнейшем будем называть его «исходный алгоритм».

Допустимую область применения алгоритмов оценки параметров будем определять следующим образом:

– параметры распределения Вейбулла, полученные по имеющейся статистике рассматривать в качестве эталонных;

– определить критерий, по которому сравниваются оценки параметров, и допустимое отклонение значения критерия от эталонного;

– создать из имеющейся статистики несколько статистик повреждений, последовательно исключая из нее точки;

– определять допустимую область применения алгоритмов как область (т.е. количество точек), в которой значения критерия, полученного по ограниченной статистике, имеют допустимое отклонение от эталонного значения.

В качестве критерия выберем величину оптимального времени замены участка T_0. Для этой величины точность в 2 года является допустимой, так как при замене участка в большинстве случаев время между принятием решения о замене и вводом в эксплуатацию замененного участка не превышает трех-четырех лет. По методике, предложенной в [5], функция $S_y(t)$ для рассматриваемого примера после приведения слагаемых имеет следующий вид:

$$S_y(t) = \frac{77484070 + 32625 \cdot t + 73.86 \cdot t^{4,05} + 11704 \cdot t^{2,24} - 14 \cdot t^{3,08} - 13.11 \cdot t^{3,75}}{t}. \quad (5)$$

С точностью до года минимум $S_y(t)$ достигается при $T_0 = 35$.

Оказалось, что уже при сравнении параметров b и c по ограниченным статистикам с эталонными ($b = 11{,}82$, $c = 4{,}05$), можно сделать ряд выводов. Рассмотрим и проанализируем полученные значения параметров (таблица 2) для общего числа повреждений по каждому алгоритму.

Таблица 2. Значения параметров b и c для общего числа повреждений по ограниченным статистикам

Количество точек	Алгоритм *А*		Алгоритм *Б*		Алгоритм *В*	
	b	*c*	*b*	*c*	*b*	*c*
8	74,6	5,25	9,65	3,06	11,6	3,87
7	73,1	5,15	14,3	5,30	11,5	3,81
6	78,3	4,7	14,1	5,05	11,5	3,77
5	44,9	7,2	7,99	3,01	11,6	3,88
4	40,6	7,6	7,42	2,76	11,4	3,52
3	60,9	4,85	3,25	1,11	11,1	2,73
2	24,2	9,96	3,15	1,51	11,9	8,65

Алгоритм *А*, использующий систему уравнений по цензурированным выборкам (4), показал наихудшие результаты: параметр b очень сильно (в 4–7 раз) отличается от эталонного; параметр c также недопустимо сильно отличается от эталонного, учитывая, что в выражении $S_y(t)$ он присутствует в виде t^c. Было сделано предположение, что значение коэффициента K (в качестве которого использовалась длина участка) было выбрано неверно. Это предположение не подтвердилось; были проведены аналогичные расчеты с другими значениями коэффициента K, которые дали такие же неудовлетворительные результаты. Такое большое расхождение с эталоном вызвано слишком малым для этого метода количеством точек, в [7] отмечается, что при количестве точек меньше 20 относительное отклонение оценок от истинных значений может достигать 100 %.

Алгоритм *Б* показал гораздо лучшие, но все равно недостаточно точные результаты: относительное отклонение параметра b составляет 9–45%, параметра c 14–49%.

Алгоритм *В* показал наилучшие результаты: относительное отклонение параметра b не превышает 6%, а параметра c – 13% и 7%, если не рассматривать количество точек меньше 5. Также, в отличие от алгоритмов *А* и *Б*, в этом алгоритме прослеживается интуитивно ожидаемая закономерность: при уменьшении количества точек происходит увеличение относительного отклонения.

Поэтому уже на этом этапе можно сделать вывод, что при оценке параметров распределения Вейбулла для нахождения зависимости общего количества повреждений на участке наиболее точные результаты дает алгоритм *В*.

Определим допустимую область применения алгоритма *В*: это минимальное количество точек, для которых вычисленное с использованием алгоритма *В* значение величины T_0 отличается от эталонного не более чем на 2 года. В таблице 2 показаны результаты, когда статистика ограничивалась сверху, т.е. исключались данные за 2007-й, 2006-й и т.д. годы. В таблице 3 представлены результаты работы алгоритма *В* при ограничении статистики сверху (от 8 до 2 точек) и снизу (когда исключались данные за 1995-й, 1996-й и т.д. годы) и величина T_0 в обоих случаях.

Таблица 3. Значения параметров b и c и оптимального времени замены по ограниченным сверху и снизу статистикам

Количество точек	Ограничение сверху			Ограничение снизу		
	b	c	T_0	b	c	T_0
8	11,6	3,87	35	11,8	4,04	35
7	11,5	3,81	36	12,7	4,46	33
6	11,5	3,77	39	12,9	4,4	44
5	11,6	3,88	41	14,8	5,52	44
4	11,4	3,52	43	16,9	6,98	23
3	11,1	2,73	43	18,8	9,34	23
2	11,9	8,65	35	19,8	11,3	21

Сравнивая полученные значения (а также результаты аналогичных расчетов по другим участкам) с эталонным значением оптимального времени замены (35 лет) и учитывая, что в подавляющем большинстве случаев статистика ограничена снизу, будем считать, что для достаточно точной оценки необходимо 7 точек. Количество точек соответствует суммарному числу лет, в которые были зафиксированы повреждения; например, в рассматриваемом примере наблюдение велось 13 лет, количество точек равно 9.

Таким образом, алгоритм, предложенный в [5], дает (с необходимой точностью) решение задачи оценки параметров распределения Вейбулла, если в имеющейся по участку статистике повреждений больше семи точек, и является наиболее точным по сравнению с другими алгоритмами. В следующем разделе рассматривается решение этой задачи при меньшем количестве точек.

3. Определение параметров распределения Вейбулла при малом количестве статистики

В качестве примера рассмотрим участок, который наблюдался с 15-го по 17-й год его срока службы и за это время на нем зафиксировано одно повреждение. Очевидно, что можно найти два участка с достаточно полной статистикой, удовлетворяющих описанному условию, оптимальное время замены у которых различается на десятки лет. Поэтому логично разбить все участки с малым объемом статистики еще на 2 подмножества: к первому отнести те участки, для которых объем имеющейся статистики повреждений настолько мал, что не дает никакой дополнительной информации для определения технико-экономического показателя состояния участка; будем называть это подмножество «участки без статистики повреждений». Ко второму подмножеству отнести все остальные рассматриваемые участки, назовем его «участки с малым объемом статистики».

3.1. ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ ВЕЙБУЛЛА ДЛЯ УЧАСТКОВ БЕЗ СТАТИСТИКИ ПОВРЕЖДЕНИЙ

Единственным методом оценки для таких участков является использование усредненных по другим участкам (с достаточно полной статистикой) параметров. Расчет усредненных параметров для общего количества повреждений выполнялся следующим образом:

1. Множество участков с достаточно полной статистикой разбивалось на 44 группы: 11 по диаметру (100, 200, …, 1000, 1200, 1400) и 4 по типу канала (непроходной, полупроходной, проходной канал и наземная прокладка) и дальнейшие операции выполнялись по каждой из групп.

2. Было введено понятие «аномального участка»: участок, у которого по тем или иным причинам (включая ошибки при вводе данных по ремонтам) слишком много повреждений и локально-вставочных ремонтов. Такие участки встречаются редко, но сильно искажают значения усредненных параметров. Поэтому логично при определении параметров «аномальные участки» не учитывать.

3. В качестве критерия, по которому участок считался аномальным, было выбрано значение удельной (по длине участка) суммарной (за один год) длины локально-вставочных ремонтов. Была составлена матрица (30*20), в которой по горизонтали откладывалось количество лет эксплуатации (30), а по вертикали с шагом 0,1 – значение удельной суммарной длины ЛВР (от 0

до 2). Значение в ячейке $[i, j]$ равнялось количеству участков, у которых на j-м году эксплуатации значение суммарной удельной длины ЛВР за этот год находится в интервале $((i*0{,}1); ((i + 1)*0{,}1))$.

Таблица 4. Фрагмент матрицы суммарных удельных ЛВР

*i**j*	11	12	13	14	15	16	17	18	19	20
0,1	17	16	13	10	8	7	6	3	2	2
…										
0,4	1	1	1	1	1	0	0	0	1	1
0,5	0	0	0	0	0	1	1	1	0	0
0,6	0	0	0	0	0	0	0	0	0	0
…										
1,0	0	0	0	0	0	0	0	0	0	0
1,1	1	1	1	1	1	0	0	0	0	0
…										

Из приведенной таблицы видно, что для подавляющего большинства участков величина удельной суммарной длины ЛВР находится в диапазоне [0; 0,1], в диапазоне [0,1; 0,5] в зависимости от срока службы находятся 1–2 участка, а в диапазоне [0,6; 1] нет ни одного участка. Далее, в диапазоне [1; 2] появляется один участок, именно его логично считать аномальным. Тогда можно определить величину порога для аномальных участков как значение удельной суммарной длины ЛВР в первой сверху строке, состоящей из одних нулей. Для приведенного примера она соответствует 0,6.

По этому принципу для каждой группы было установлено значение порога для аномальных участков. Если за какой-либо год эксплуатации участка наблюдалось превышение удельной суммарной длины ЛВР величины установленного порога, то такой участок считался аномальным и в дальнейших расчетах не учитывался.

4. Для каждой группы с исключенными аномальными участками заполнялся массив из 35 элементов, элемент состоял из двух ячеек. Номер элемента соответствовал году эксплуатации и содержал в одной ячейке суммарное (по всем участкам из груп-

пы) количество повреждений за этот год, а в другой – суммарную длину участков по которым в этот год велось наблюдение. Тогда отношение суммарного количества повреждений к суммарной длине в i-м элементе массива равняется усредненному числу повреждений для участка длиной 1 метр.

5. Далее по алгоритму аппроксимации общего числа повреждений из [5] определялись параметры распределения Вейбулла. На рис. 2 приведен пример графика зависимости (в логарифмических шкалах) общего числа повреждений от времени для участков $d = 500$ мм в проходных каналах и его аппроксимация.

Для краткости будем называть предложенный метод «метод усредненных параметров».

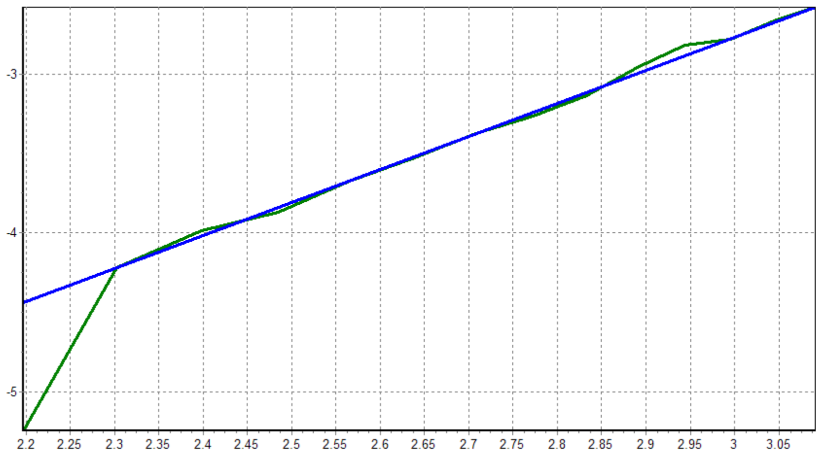

Рис. 2. Пример графика зависимости общего числа повреждений от времени и его аппроксимация прямой линией

3.2. ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ ВЕЙБУЛЛА ДЛЯ УЧАСТКОВ С МАЛЫМ ОБЪЕМОМ СТАТИСТИКИ

Ход рассуждений при построении алгоритма для участков с малым объемом статистики был следующим. Как и в случае с участками без статистики повреждений, имеющийся небольшой объем данных для этого подмножества участков не позволяет использовать алгоритм B из-за невысокой точности конечных результатов. Следовательно, если существует более точный алгоритм оценки параметров, то в нем должны использоваться данные по некоторому множеству других участков, чтобы компенсировать недостаток информации. Очевидно, что это множество должно состоять из участков той же группы (по диаметру и способу прокладки), что и рассматриваемый участок. Но, как будет видно ниже, значения параметров для участков с достаточно полной статистикой из одной группы могут сильно различаться. Поэтому логично поступить так:

– выбрать из этой группы участки, «похожие» на рассматриваемый участок;

– добавить к множеству «похожих» участков рассматриваемый участок и по получившемуся множеству найти усредненные параметры распределения Вейбулла, алгоритмом, описанным в 3.1.

Определим критерий, по которому участок той же группы считается «похожим» на рассматриваемый участок. Пусть рассматриваемый участок наблюдался с m-го по n-й год эксплуатации, при этом за этот период на нем произошло r повреждений на единицу его длины. Будем считать, что участок «похож» на рассматриваемый участок, если с m-го по n-й год эксплуатации количество повреждений на единицу длины лежит в интервале $[r*(1 - \Delta); r*(1 + \Delta)]$, где Δ – параметр в интервале (0; 1).

Вышеизложенное рассуждение хорошо иллюстрируется следующим рисунком (рис. 3): рассмотрим плоскость $b - c$, соответствующую удельным (для 1-го метра) параметрам распределения Вейбулла для различных участков. Определим численными методами такую область (она показана черным цветом), что любая ее точка с координатами (b, c) удовлетворяет сформулированному критерию, т.е. участок с такими параметрами «похож» на рассматриваемый. Также на этом графике поставим точки, координаты которых (b, c) соответствуют параметрам участков с достаточно полной статистикой из той же группы (они показаны серым цветом).

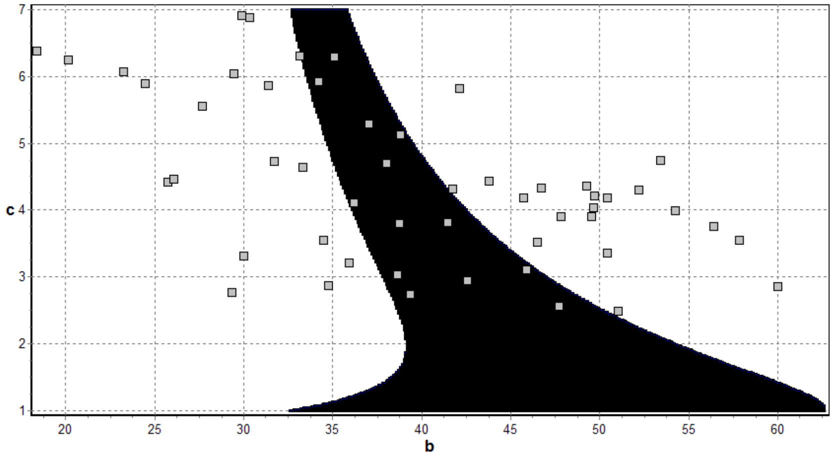

Рис. 3. Пример области «похожих» участков и параметров (b, c) *участков с достаточно полной статистикой*

Из рисунка сразу видно множество «похожих» участков – участки, координаты которых попали в черную область.

Значение параметра Δ было выбрано равным 0,4. При этом значении у наибольшего количества участков с достаточно полной статистикой наблюдалось минимальное отклонение T_0, вычисленного указанным способом при ограничении статистики.

В результате был построен следующий алгоритм, будем называть его «метод похожих участков».

1. Среди всех участков, по которым имеется статистика, выбираются те, для которых выполняются все нижеперечисленные условия:

   – диаметр и способ прокладки совпадают с рассматриваемым участком;

   – статистика повреждений содержит больше семи точек;

   – участок «похож» на рассматриваемый.

2. Аналогично пунктам 4 и 5 алгоритма из раздела 3.1 создается массив из 35 элементов, который заполняется, как показано на рис. 4.

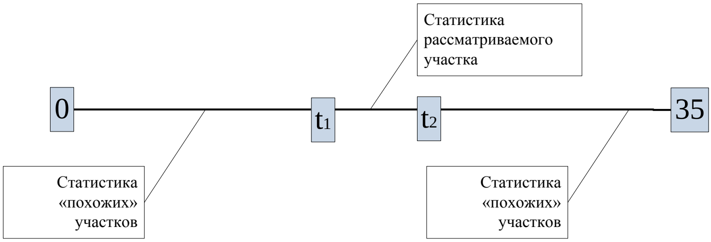

Рис. 4. Алгоритм заполнения массива.

На этом рисунке t_1, t_2 – срок наблюдения за рассматриваемым участком. И далее по этому массиву определяются параметры распределения Вейбулла.

4. Общий алгоритм оценки параметров распределения Вейбулла

В предыдущих разделах рассматривались 3 алгоритма оценки параметров распределения: исходный алгоритм, работающий по достаточно полной статистике, алгоритм для участков без статистики повреждений (метод усредненных параметров) и алгоритм для участков с малым объемом статистики (метод похожих участков). Также было показано, что исходный алгоритм дает оценки параметров с достаточной точностью, если количество точек больше семи.

Чтобы построить общий алгоритм, определим границу между методом усредненных параметров и методом похожих участков. Найдем некоторое число U, такое, что если количество имеющихся точек (n) в статистике повреждений рассматриваемого участка меньше U, то для оценки параметров метод усредненных параметров дает более точные результаты, а если $n > U$, то более точные результаты даст метод похожих участков. Построим для этого таблицу значений T_0, полученных разными алгоритмами, в скобках указано, на сколько значения T_0 отличаются от эталонного (таблица 5).

Таблица 5. Оценки оптимального времени замены участка (T_0), полученные разными алгоритмами.

Количество точек	Исходный алгоритм	Метод похожих участков	Метод усредненных параметров
8	35	–	–
7	33 (2)	–	–
6	44 (9)	41 (6)	47 (12)
5	44 (9)	28 (7)	47 (12)
4	23 (12)	24 (11)	47 (12)
3	23 (12)	89 (54)	47 (12)
2	21 (14)	89 (54)	47 (12)
1	–	89 (54)	47 (12)
0	–	–	47 (12)

Из таблицы видно, что метод похожих участков дает лучшие результаты при количестве точек > 3; при количестве точек ≤ 3 этот метод не применим. Для других участков в некоторых случаях метод усредненных параметров давал лучшие значения при количестве точек ≤ 4, поэтому определим границу $U = 4$.

Суммируя полученные результаты, можно определить общий алгоритм оценки параметров распределения Вейбулла так:

– если количество точек $n \geq 7$, то следует использовать исходный алгоритм, при этом в большинстве случаев будет достигнута необходимая точность полученных оценок;

– если $4 \leq n \leq 6$, то следует использовать метод похожих участков, при этом необходимая точность полученных оценок не достигается, отклонение может составлять до 30%;

– если $n < 4$, то следует использовать метод усредненных параметров, при этом необходимая точность полученных оценок не достигается.

5. Выводы

При конструировании методов оптимизации планирования ремонтных работ на инженерных сетях возникла необходимость в решении известной задачи оценки параметров распределения Вейбулла при малом объеме выборки применительно к участку сети. В данной работе решение этой задачи опирается на использование специфики предметной области и анализ статистики повреждений на совокупности участков сети. Основные результаты следующие:

– сформулирован критерий сравнения алгоритмов оценки параметров, с помощью которого также можно определить допустимую область применения алгоритма;

– по этому критерию было установлено, что алгоритмы оценки параметров, не использующие информации о предметной области, дают неудовлетворительные результаты при количестве точек меньше семи;

– дальнейшее исследование проводилось для множества участков, по которым имеющаяся в наличии статистика повреждений насчитывает меньше семи точек; было показано, что это множество логично разделить на два; для каждого из подмножеств был сконструирован свой алгоритм: метод усредненных параметров и метод похожих участков;

– был предложен общий алгоритм оценки параметров и определены границы применения для каждого из трех предложенных алгоритмов, в пределах которых алгоритм дает наилучшие результаты.

Литература

1. *ГОСТ Р 50779.27-2007 Статистические методы. Критерий согласия и доверительные интервалы для распределения Вейбулла.*
2. *ГОСТ 50779.27-2017 Статистические методы распределения Вейбулла. Анализ данных*.
3. КРЫГИН А.А. *Оптимизация графиков плановых ремонтов совокупности участков инженерных сетей* // Автоматика и телемеханика. – 2010. – №9. – С. 83–102.
4. КРЫГИН А.А. *Оптимизация графиков плановых ремонтов совместно расположенных участков теплосети с учетом практических особенностей* // Автоматизация в промышленности. – 2016. – № 5. – С. 53–56.

5. КРЫГИН А.А. *Расчетно-статистические методы управления обслуживанием протяженных инженерных сетей* // Управление большими системами. – 2019. – Вып. 78. – 2019. – С. 235–249. – DOI: https://doi.org/10.25728/ubs.2019.78.10/
6. *МЭК 60605-4:2001 Испытания аппаратуры на надежность. Часть 4. Статистические методы для экспоненциального распределения. Точечные оценки, доверительные, предикционные и толерантные интервалы (IEC 60605-4:2001 Equipment reliability testing – Part 4: Statistical procedures for exponential distribution – Point estimates, confidence intervals, prediction intervals and tolerance intervals).*
7. РУСИН А.Ю., АБДУЛХАМЕД М. *Обработка информации в системе испытаний промышленного оборудования на надежность* // Интернет-журнал «Технологии техносферной безопасности». – 2014. – №4(56). – С. 4–7. – URL: http://ipb.mos.ru/ttb.
8. ТИХОВ М.С., АГЕЕВ В.В., БОРОДИНА Т.С. *Оценивание параметров распределения Вейбулла по случайно цензурированным выборкам* // Математическое моделирование. Оптимальное управление. Вестник Нижегородского университета им. Н.И. Лобачевского. – 2010. – №4(1). – С. 141–145.
9. ALIZADEH M., REZAEI S., BAGHERI S.F. *On the Estimation for the Weibull* // Distribution Annals of Data Science. – 2015. – Vol. 2. – P. 373–390.
10. MOEINIA A., JENABB K., MOHAMMADIC M., FOUMANID M. *Fitting the three-parameter Weibull distribution with Cross Entropy* // Applied Mathematical Modelling. – May, 2013. – Vol. 37, Iss. 9.1. – P. 6354–6363.
11. ASMUSSEN S., KORTSCHAK D. *Error Rates and Improved Algorithms for Rare Event Simulation with Heavy Weibull Tails* // Methodology and Computing in Applied Probability. – 2015. – Vol. 17. – P. 441–461.

12. COUSINEAU D. *Fitting the three-parameter weibull distribution: review and evaluation of existing and new methods* // IEEE Trans. Diel. Elect. Insul. – 2009. – Vol. 16, No. 1. – P. 281–288.
13. MARKOVIĆ D., JUKIĆ D. *On nonlinear weighted total least squares parameter estimation problem for the three-parameter Weibull density* // Applied Mathematical Modelling. – July, 2010. – Vol. 34, Iss. 7. – P. 1839–1848.
14. HAGEY T.J., PUTHOFF J.B., CRANDELL K.E., AUTUMN K., HARMON L.J. *Modeling observed animal performance using the Weibull distribution* // J. of Experimental Biology. – 2016. – Vol. 219. – P. 1603–1607.
15. NAGATSUKA H., BALAKRISHNAN N. *An Efficient Method of Parameter and Quantile Estimation for the Three-Parameter Weibull Distribution Based on Statistics Invariant to Unknown Location Parameter* // Communications in Statistics – Simulation and Computation. – 2015. – Vol. 44, Iss. 2. – P. 295–318.
16. ÖRKCÜ H.H., ÖZSOY V.S., AKSOY E. *Estimating the parameters of 3-p Weibull distribution using particle swarm optimization: A comprehensive experimental comparison* // Appl. Math. Computation. – 2015. – Vol. 268. – P. 201–226.
17. SALAHADDIN A.A. *Comparative study of four methods for estimating Weibull parameters for Halabja, Iraq* // Int. J. of Physical Sciences. – February, 2013. – Vol. 8(5). – P. 186–192.
18. SHAHANI A.R., BABAEI M. *Helicopter blade reliability: Statistical data analysis and modeling* // Aerospace Sci. Technol. – 2016. – Vol. 55, No. 1. – P. 43–48.
19. ZHANG XIANGPO *Estimation of Three-Parameter Weibull Distribution Based on Artificial Fish-Swarm Algorithm* // ICoMS. – 2018. – P. 34–40
20. ZHAO J., PENG G., ZHANG H. *Schedule and cost integrated estimation for complex product modeling based on Weibull distribution* // Proc. IEEE 19th Int. Conf. Computer Supported Cooperative Work in Design. – 2015. – Article number 7230971. – P. 276–280.

EVALUATION OF THE WEIBULL DISTRIBUTION PARAMETERS FOR SMALL VOLUME OF THE SAMPLE IN THE PROBLEMS OF OPTIMIZATION OF PLANNING REPAIR WORKS ON ENGINEERING NETWORKS

Andrey Krygin, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Cand.Sc. (Moscow, Profsoyuznaya st., 65, andreyakr@yandex.ru).

Abstract: In the previously developed economic and statistical methodology for assessing the state of the engineering network section and information support when deciding to extend its service life, there are serious requirements for the initial data. One of the conditions for the correct operation of the methodology is the presence of complete statistics of damage and repairs carried out, starting from the moment the site is commissioned. However, for many sites such statistics are available only for the last few years. In order to expand the scope for cases of incomplete damage statistics, a separate study was carried out to estimate the Weibull distribution parameters with a small sample size using the example of heat supply network sections. This work is devoted to the results of the study. Three classes of sites were identified: with fairly complete damage statistics, with a small amount of statistics, and without statistics. For each class, an algorithm was proposed for estimating the necessary parameters for the Weibull distribution, which were further used in the methodology for estimating the state of a site, and a criterion was determined by which a site could be assigned to a particular class.

Keywords: estimation of Weibull distribution parameters with a small sample size, estimation of averaged values of parameters, optimization of repairs of heating network sections.

УДК 338.49
ББК 31.38
DOI: 10.25728/ubs.2020.84.9

Статья представлена к публикации членом редакционной коллегии В.Г. Лебедевым.

Поступила в редакцию 17.04.2019.
Опубликована 31.03.2020.