УДК 656.7.071: 656.7.052.002.5 (045)

Кузьмин В.Н., к.т.н.,
Соломенцев А.В., д.т.н.,
Залисский М.Ю., к.т.н.

ИСПОЛЬЗОВАНИЕ МНОГОСЕГМЕНТНОЙ РЕГРЕССИИ ДЛЯ ОЦЕНКИ ДОЛГОВЕЧНОСТИ КОНСТУКТИВНЫХ ЭЛЕМЕНТОВ СИСТЕМ

Национальный авиационный университет

arec@nau.edu.ua
avsolomentsev@ukr.net
maximus2812@ukr.net

Рассматривается подход, основанный на применении многосегментной регрессии для возможности оценки результатов испытаний долговечности конструктивных элементов систем на примере образцов из алюминиевого сплава. Приведены аналитические соотношения для построения многосегментной регрессии, которые позволяют решать задачи прогнозирования долговечности образцов

Ключевые слова: многосегментная регрессия, аппроксимация, метод наименьших квадратов, точка переключения, прогнозирование

Введение

Многочисленные исследования показывают, что рассеяние характеристик механических свойств материалов должно учитываться при создании современных машин и сооружений. Для построения моделей этих характеристик применяют различные статистические методы [1].

Среди этих статистических методов широкое распространение получили: методы проверки статистических гипотез, многофакторный дисперсионный анализ, регрессионный и корреляционный анализ, графо-аналитические метод с использованием вероятностных сеток для различных законов распрелеления и т. д.

Полученные модели рассеяние характеристик механических свойств материалов могут быть использованы для анализа их долговечности.

Постановка задачи

Анализ литературы в сфере статистического анализа долговечности материалов [1; 2] показал, что вопросам оценки предельных свойств различных материалов и систем уделяется широкое внимание. Поэтому в теории надежности в настоящее время интенсивно развиваются новые методы статистической обработки данных, которые направлены на повышение достоверности расчетов текущих и прогнозируемых значений характеристик долговечности материалов.

Зачастую статистический анализ предполагает, что случайные величины обычно описываются одним законом распределения (нормальным, логнормальным, Вейбулла и другими). В случае графо-аналитического статистического анализа данных на соответствующих сетчатках вероятностей эти данные чаще всего аппроксимируют единой прямой линией. Однако на практике такая аппроксимация не позволяет решать задачи прогнозирования с высокой точностью.

Возможны варианты обработки эмпирических данных в рамках решения задачи сглаживания с использованием, например, аппарата сплайн-аппроксимации, многосегментной регрессии и т.д. При использовании сплайн-аппроксимации экспериментальных данных, как правило, рассматривают одинаковые модели аппроксимирующей функции на отдельных участках, а также равные по длительности эти участки. При этом минимум среднеквадратической ошибки аппроксимации находят, рассматривая весь интервал наблюдения за экспериментальными данными.

В этой статье будет рассматриваться пример решения актуальной научно-технической задачи использования многосегментной регрессии на сетчатке логнормального распределения для реального примера, представленного статистическими данными об испытании образцов из алюминиевого сплава на консольный изгиб с вращением, взятом из [1]. При этом нахождение точек переключениями между отдельными сегментами регрессии будет осуществляться путем определения и минимизации среднеквадратических ошибок аппроксимации не на всём интервале наблюдения за данными.

Основная часть

Для анализа статистических свойств экспериментальных данных, которые обладают свойством нелинейности, предложено использовать многосегментную регрессии [2].

Во время использования многосегментных регрессий возникает задача определения оптимальных абсцисс точек переключения [3]. Методику построения многосегментной регресссии и ее оптимизацию рассмотрим на реальном примере статистических данных относителшьно испытаний образцов из алюминиевого сплава на консольный изгиб с вращением. Статистические данные приведены в таблице 1.

Таблица 1. Вариационный ряд значений логарифмов долговечности образцов из алюминиевого сплава

Номер образца	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Логарифм долговечности	5,8669	5,9164	6,0216	6,0386	6,0426	6,0445	6,0645	6,0799	6,0821	6,1062
Номер образца	11	12	13	14	15	16	17	18	19	20
Логарифм долговечности	6,1082	6,1183	6,1186	6,1238	6,1605	6,1685	6,1746	6,1801	6,1892	6,1951
Номер образца	21	22	23	24	25	26	27	28	29	30
Логарифм долговечности	6,2071	6,2100	6,2297	6,2608	6,3115	6,3310	6,3688	6,3746	6,3811	6,3829
Номер образца	31	32	33	34	35	36	37	38	39	40
Логарифм долговечности	6,3829	6,3918	6,3967	6,4076	6,4089	6,4094	6,4216	6,4328	6,4342	6,4620
Номер образца	41	42	43	44	45	46	47	48	49	50
Логарифм долговечности	6,4630	6,4646	6,4704	6,4713	6,4842	6,4975	6,4984	6,5179	6,5200	6,5214
Номер образца	51	52	53	54	55	56	57	58	59	60
Логарифм долговечности	6,5224	6,5431	6,5464	6,5578	6,5603	6,5607	6,5654	6,5655	6,5793	6,5823
Номер образца	61	62	63	64	65	66	67	68	69	70
Логарифм долговечности	6,5916	6,5957	6,6096	6,6471	6,6474	6,6739	6,6739	6,6780	6,6896	6,6916
Номер образца	71	72	73	74	75	76	77	78	79	80
Логарифм долговечности	6,7000	6,7086	6,7132	6,7178	6,7197	6,7275	6,7275	6,7392	6,7432	6,7538
Номер образца	81	82	83	84	85	86	87	88	89	90
Логарифм долговечности	6,7809	6,7975	6,7975	6,7988	6,8038	6,8142	6,8169	6,8649	6,8717	6,8977
Номер образца	91	92	93	94	95	96	97	98	99	100
Логарифм долговечности	6,9051	6,9110	6,9189	6,9299	6,9545	7,0824	7,1682	7,2603	7,2775	7,4586

Графический вид исходных данных, нанесённых на логнормальную сетчатку вероятностей, приведен на рис. 1 (по оси абсцисс отложены логарифмы долговечности, а по оси ординат – квантили нормального распределения). Значение квантилей изменяется в пределах [–3,09; 3,09]. Следует также отметить, что такое представление эмпирических данных широко используется в теории надёжности [1].

Для принятия решения о возможности аппроксимации эмпирических данных единой прямой линией выполним тестирование этих данных на линейность. Для этого аппроксимируем данные с помощью линейной регрессии на всем интервале наблюдения.

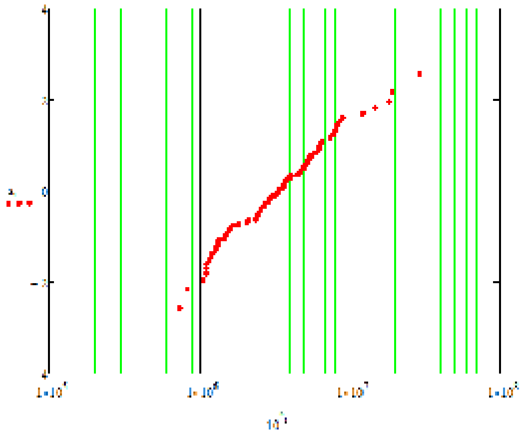

Рис. 1. Графический вид исходных данных, нанесённых на логнормальную сетчатку вероятностей

Полученное уравнение имеет вид:

$$x = 0{,}30907y + 6{,}52357\,.$$

При этом представление зависимости x от y является более удобным для решения задач прогнозирования (x имеет смысл долговечности, а y – вероятности).

Графическое изображение исходных данных и их аппроксимация с помощью линейной регрессии изображены на рис. 2.

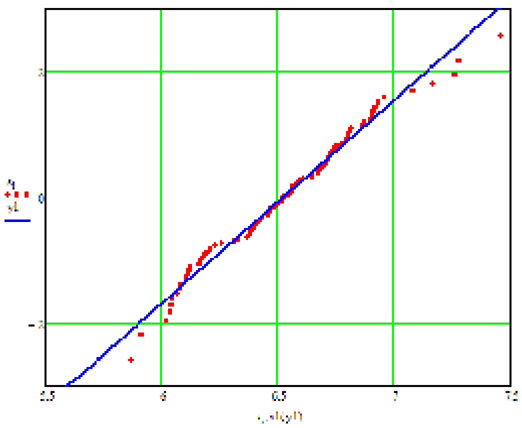

Рис. 2. Аппроксимация статистических данных с помощью линейной регрессии

Визуальный анализ рис. 2 показывает, что эмпирические данные неудовлетворительно согласуются с линейной регрессией, поскольку шесть первых точек и пять последних точек значительно отклоняются от прямой. Для принятия решения о возможности линейной аппроксимации можно провести более детальный анализ на нелинейность по методу, предложенному в [4].

Для этого рассчитывают куммулятивную кривую остатков относительно прямой линии и определяют ее размах, который в данном случае составляет 0,922. Стандартное отклонение составляет 0,041. Отношение размаха к стандартному отклонению – 22,618. Согласно таблице критических значений для данного критерия [5] можно утверждать, что с вероятностью в пределах (0,99; 0,999) эмпирические данные не могут быть аппроксимированы единой прямой линией.

Визуальный анализ эмпирических данных показывает, что можно выделить три прямолинейных сегмента. Поэтому выполним аппроксимацию с помощью трёхсегментной линейной регрессии.

Аппроксимацию будем осуществлять согласно следующей процедуре.

1. Из эмпирических данных отбрасываются пять последних точек и для оставшейся совокупности строят пять вариантов двухсегментной полигональной регрессии. Находим суммы квадратов отклонений для всех вариантов аппроксимации.

2. Определяется оптимальная ордината первой точки переключения. Для этого по пяти значениям суммы квадратов отклонений и соответсвующим им точкам переключения проводят параболу по методу наименьших квадратов. После чего определяется минимумальное значение параболы, абсцисса которой сооответствует оптимальной точке переключения.

3. Отбрасываются все начальные точки, находящиеся до первой оптимальной точки переключения. Через оставшиеся значения совокупности строят пять вариантов двухсегментной регрессии. Находят суммы квадратов отклонений для всех пяти вариантов аппроксимации.

4. Находят оптимальную ординату второй точки переключения.

5. На основе полученных данных можна построить трехсегментную регрес-

сию для всего интервала наблюдения с использованием функций Хевисайда.

В результе применения данной процедуры получены следующие значения ординат оптимальных точек переключения: –1,254 и 1,372.

Полученное уравнение трёхсегментной регрессии имеет вид

$$x = 6{,}3666 + 0{,}1937y + 0{,}1167((y + \\ + 1{,}2540)h(y + 1{,}2540)) + \\ + 0{,}1292((y - 1{,}3720)h(y - 1{,}3720)),$$

где $h(y)$ – функция Хевисайда.

Графическое изображение исходных статистических данных и их аппроксимация с помощью трехсегментной регрессии изображены на рис. 3.

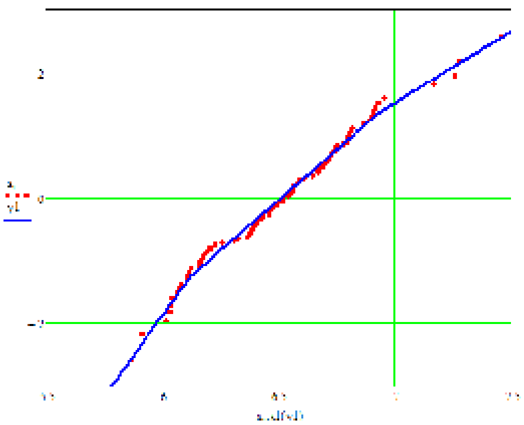

Рис. 3. Аппроксимация статистических данных с помощью трехсегментной регрессии

Проведем сравнительные расчеты прогнозных величин для вероятностей достижения предельного состояния 0,001 и 0,999 для двух альтернативных методов аппроксимации (единой прямой линией и трехсегментной регрессией). Для случая аппроксимации единой прямой линией прогнозные значения составляют $3{,}703 \cdot 10^5$ и $3{,}010 \cdot 10^7$. Для трехсегментной аппроксимации эти значения составляют $5{,}863 \cdot 10^5$ и $4{,}943 \cdot 10^7$ соотвественно.

Как видно из полученных значений трехсегментная аппроксимация позволяет существенно уточнить прогнозные значения (в среднем на 50 %).

Выводы

Для аппроксимации исследуемых эмпирических данных относительно испытаний образцов из алюминиевого сплава на консольный изгиб с вращением была использована многосегментная регрессия, в процессе построения которой использовалась итерационная процедура. Отличительная особенность этой процедуры состоит в том, что задача оптимизации точек переключения решалась не для всего интервала наблюдения, а для соседних пар его отдельных сегментов.

Исследуемые статистические данные носили нелинейный характер, поэтому проведенная оптимизация ординат точек переключения для случая трехсегментной аппроксимации эмпирических данных позволила получить более точные прогнозные значения предельных свойств оценок долговечности.

Полученные результаты могут быть использованы в подсистемах обработки данных о надёжностных свойствах материалов, конструкций и систем во время их эксплуатации.

Список литературы

1. Степанов М.Н. Статистические методы обработки результатов механических испытаний. Справочник. – М.: Машиностроение, 1985. – 232 с.

2. Байхельт Ф., Франкен П. Надёжность и техническое обслуживание. Математический подход: Пер. с нем. – М.: Радио и связь, 1988. – 392 с.

3. Wagner A.K., Soumerai S.B., Zhang F., Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther. 2002, 27 (4): PP. 299-309.

4. Kuzmin V.N. New Statistical Method for Identification of Nonlinearity of Empirical Data // Computer data analysis and modeling. Proceedings of the Fifth International Conference. June, 8 – 12, 1998, Minsk, Volume 1: A-M, PP. 159 – 164.

5. Кузьмин В.М., Соломенцев О.В., Заліський М.Ю. Методичний підхід для тестування статистичних даних на лінійність // Проблеми інформатизації та управління. – 2015. – № 4 (52). – С. 63 – 67.

Статью представлено в редакцию 01.04.2016