

Абрамович М. С., Мицкевич М.Н., Пыжик Н.Н.

**АЛГОРИТМИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ СТАТИСТИЧЕСКОГО
ОЦЕНИВАНИЯ ПОКАЗАТЕЛЕЙ НАДЕЖНОСТИ АВТОТРАНСПОРТНЫХ СРЕДСТВ ПО
ЦЕНЗУРИРОВАННЫМ ВЫБОРКАМ**

Рассмотрен подход к статистическому оцениванию показателей надежности автотранспортных средств по выборкам ограниченного объема. Приводятся описание программного комплекса оценивания показателей надежности.

Введение

Важное место в комплексе мероприятий по обеспечению надежности автомобильных транспортных средств (АТС) – автомобилей, прицепного состава, их сборочных единиц и деталей, далее – объектов, занимают статистические методы оценки показателей надежности [1].

По используемым для вычисления статистическим методам показатели надежности и эффективности использования АТС можно классифицировать на две группы: первая – показатели безотказности, долговечности и вторая – комплексные показатели надежности, технико-экономические и технико-эксплуатационные показатели использования АТС.

К первой группе относятся характеристики ресурса до и между отказами объектов. Показатели этой группы определяются на основе аппроксимации эмпирических функций распределения случайной величины (ресурса, наработки объекта) теоретическими законами распределения. При оценке параметров законов распределения по результатам эксплуатации (и испытаний) наиболее общим случаем является такой, когда часть объектов за период испытаний имеет отказы (наработки на момент отказа известны), а часть – не имеет (известны наработки работоспособных объектов – наработки до цензурирования). В качестве теоретических законов распределения, как правило, используются следующие: нормальный, экспоненциальный, трехпараметрические логнормальный и Вейбулла [1,2]. Оценив по выборке параметры закона распределения, можно на основании их построить оценки показателей надежности.

Изменение показателей второй группы в зависимости от срока эксплуатации АТС (пробег с начала эксплуатации) определяется на основе аппроксимации эмпирических значений показателей уравнениями регрессии заданного вида.

На практике не всегда представляется возможность сформировать выборки объектов достаточно большого объема, но тем не менее, уже по выборкам ограниченного объема требуется оценить надежность объектов до начала серийного выпуска, чтобы по возможности предотвратить тиражирование отказов.

В настоящей работе рассмотрен подход построения оценок параметров законов распределения по многократно цензурированным выборкам объектов с использованием метода максимального правдоподобия. Для построения точечных и интервальных оценок показателей надежности в случае выборок ограниченного объема предложен алгоритм, основанный на методе бутстрепа [5].

Оценка показателей надежности для конкретных моделей АТС требует больших вычислительных и информационных ресурсов, связанных с созданием и поддержкой баз данных с первичной информацией о надежности, выбором и обработкой информации из баз данных, процедурами вычисления показателей надежности, формированием выходных форм отчетов. Все перечисленные функции реализованы в программном комплексе, описанном в работе.

1. Оценивание показателей ресурса

Характеристики ресурса АТС включают средний и гамма-процентный ресурс, среднее квадратическое отклонение ресурса, вероятность безотказной работы, интенсивность отказов, которые вычисляются на основании оценок параметров законов распределения.

1.1. Оценивание параметров распределений по цензурированным выборкам

Предположим, что сведения о надежности изделий представлены наработками работоспособных и отказавших объектов:

$$t_1, t_2, \dots, t_r; \tau_1, \tau_2, \dots, \tau_n, \quad (1)$$

где $t_i (i=1, \dots, r)$ – наработки изделий до отказа, $\tau_j (j=1, \dots, n)$ – наработки изделий до цензурирования, объем рассматриваемой выборки $N = r + n$.

По результатам эксплуатационных испытаний вида (1) требуется оценить показатели надежности объектов. В выборке (1) в общем случае значения наработок до отказа и наработок до цензурирования не равны между собой и могут чередоваться случайным образом. Такая модель выборки называется многократно цензурированной случайной выборкой [1].

В случае применения параметрических методов оценки обычно предполагается известным закон распределения многократно цензурированной случайной выборки $F(t, \Theta_1, \Theta_2, \dots, \Theta_l)$ с точностью до вектора неизвестных параметров $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_l)$. Оценки показателей надежности вычисляются в два этапа: на первом – по имеющейся выборке (1) оценивают параметры распределения $F(t, \Theta_1, \Theta_2, \dots, \Theta_l)$, на втором – по найденным оценкам параметров вычисляют оценки показателей надежности как функции оценок параметров распределения.

Рассмотрим подход для вычисления оценок параметров распределения по цензурированной выборке с использованием метода максимального правдоподобия. Из механизма формирования выборки (1) следует, что отказы и цензурирования независимы. Тогда функция правдоподобия для этой выборки может быть записана в виде:

$$L = \prod_{i=1}^r f(t_i) \times \prod_{j=1}^n [1 - F(\tau_j)].$$

Оценки максимального правдоподобия параметров распределения (МП-оценки) получаются решением уравнений правдоподобия:

$$\frac{\partial L}{\partial \Theta_i} = 0, i = 1, \dots, l.$$

Для нормального и экспоненциального законов распределения МП-оценки параметров распределений по цензурированным выборкам подробно рассмотрены в [1]. Рассмотрим алгоритм построения МП-оценок для трехпараметрического распределения Вейбулла с плотностью

$$f(t, a, b, c) = \frac{b}{a} \left(\frac{t-c}{a} \right)^{b-1} e^{-((t-c)/a)^b}, t > c.$$

Функция правдоподобия для нахождения оценок параметров распределения Вейбулла запишется в виде:

$$L = \prod_{i=1}^r \frac{b}{a^b} (t_i - c)^{b-1} e^{-\left(\frac{t_i-c}{a}\right)^b} \times \prod_{j=1}^n \left[e^{-\left(\frac{\tau_j-c}{a}\right)^b} \right],$$

а ее логарифм:

$$\ln L = r \ln b + (b-1) \sum_{i=1}^r \ln(t_i - c) - br \ln a - \sum_{i=1}^r \left(\frac{t_i - c}{a} \right)^b - \sum_{j=1}^n \left(\frac{\tau_j - c}{a} \right)^b. \quad (2)$$

Зафиксировав параметр сдвига c , используем необходимое условие экстремума и на основании (2) получим следующую систему уравнений для нахождения оценок параметров a и b :

$$\frac{1}{\hat{b}} = \frac{\sum_{i=1}^r (t_i - c)^{\hat{b}} \ln(t_i - c) + \sum_{j=1}^n (\tau_j - c)^{\hat{b}} \ln(\tau_j - c)}{\sum_{i=1}^r (t_i - c)^{\hat{b}} + \sum_{j=1}^n (\tau_j - c)^{\hat{b}}} - \frac{\sum_{i=1}^r \ln(t_i - c)}{r}, \quad (3)$$

$$\hat{a} = \left(\frac{\sum_{i=1}^r (t_i - c)^{\hat{b}} + \sum_{j=1}^n (\tau_j - c)^{\hat{b}}}{r} \right)^{1/\hat{b}}. \quad (4)$$

Уравнение (3) решается методом последовательных приближений. Начальное приближения \hat{b}_0 выбирается следующим образом [1]:

$$\hat{b}_0 = \frac{r+1}{\prod_{i=1}^r t_i \ln \frac{i-1}{t_{(1)}} (0.23r + 3.75)}, \quad (5)$$

где $t_{(1)}$ - минимальное значение наработки до отказа.

K -ое приближение определяется по формуле:

$$\hat{b}_k = \left(\frac{\sum_{i=1}^r (t_i - c)^{\hat{b}_{k-1}} \ln(t_i - c) + \sum_{j=1}^n (\tau_j - c)^{\hat{b}_{k-1}} \ln(\tau_j - c) - \sum_{i=1}^r \ln(t_i - c)}{\sum_{i=1}^r (t_i - c)^{\hat{b}_{k-1}} + \sum_{j=1}^n (\tau_j - c)^{\hat{b}_{k-1}} - \frac{\sum_{i=1}^r \ln(t_i - c)}{r}} \right)^{-1}. \quad (6)$$

Итеративный процесс продолжается до тех пор, пока не будет достигнута требуемая точность ε :

$$\left| \frac{\hat{b}_k - \hat{b}_{k-1}}{\hat{b}_k} \right| \leq \varepsilon. \quad (7)$$

В качестве оценки параметра b принимается $\hat{b} = \hat{b}_k$. Оценка параметра a определяется в соответствии с (4). Для этих значений проверяется достаточное условие максимума функции правдоподобия:

$$r \left(\sum_{i=1}^r \left(\frac{t_i - c}{a} \right)^b \ln^2 \left(\frac{t_i - c}{a} \right) + \sum_{j=1}^n \left(\frac{\tau_j - c}{a} \right)^b \ln^2 \left(\frac{\tau_j - c}{a} \right) \right) - 2r \sum_{i=1}^r \ln \left(\frac{t_i - c}{a} \right) - \left(\sum_{i=1}^r \ln \left(\frac{t_i - c}{a} \right) \right)^2 > 0.$$

Для оценки параметра сдвига c трехпараметрического распределения Вейбулла построим следующую процедуру.

1) Выбирается некоторое натуральное число K . Пусть $\Delta = x_{(2)} - x_{(1)}$, где $x_{(i)}$ - i -я порядковая статистика выборки. Отметим, что в качестве правой границы изменения параметра c можно взять минимальный элемент выборки $x_{(1)}$, а в качестве левой - 0. Тогда $K = \left\lceil \frac{x_{(1)}}{\Delta} \right\rceil$.

2) В цикле меняется значение параметра c : $c_m = x_{\min} - m \cdot \Delta, m = 1, \dots, K$. В качестве текущей оценки для параметра c принимается значение \hat{c}_m и на основании (6), (7) определяются оценки \hat{a}_m, \hat{b}_m и $\ln L_m$.

3) Определяется номер шага $m^* = \arg \max_m (\ln L_m)$. В качестве оценок параметров распределения Вейбулла принимаются значения $\hat{a} = \hat{a}_{m^*}, \hat{b} = \hat{b}_{m^*}, \hat{c} = \hat{c}_{m^*}$.

1.2. Построение точечных и интервальных оценок показателей надежности

Как отмечалось выше, оценки показателей надежности вычисляются на основании оценок параметров распределения. Так, например, если выборка (1) подчиняется закону распределения Вейбулла, то оценка ресурса \hat{L} вычисляется следующим образом [1]:

$$\hat{L} = \hat{a} \Gamma \left(1 + \frac{1}{\hat{b}} \right) + \hat{c}.$$

Известно [3], что из-за потерь информации при цензурировании выборки уменьшается точность оценивания и имеет место смещение традиционных оценок параметров. При этом величина смещения зависит от степени цензурирования и от объема выборки.

Количество информации Фишера по цензурированной выборке определяется следующим соотношением [4]:

$$J_c(\theta) = \frac{1}{P_1(\theta)} \left[\frac{\partial P_1(\theta)}{\partial \theta} \right] \left[\frac{\partial P_1(\theta)}{\partial \theta} \right]^T + \int_{x_1}^{x_2} \left[\frac{\partial \ln f(x, \theta)}{\partial \theta} \right] \left[\frac{\partial \ln f(x, \theta)}{\partial \theta} \right]^T f(x, \theta) dx + \frac{1}{P_2(\theta)} \left[\frac{\partial P_2(\theta)}{\partial \theta} \right] \left[\frac{\partial P_2(\theta)}{\partial \theta} \right]^T,$$

где $P_1(\theta)$ и $P_2(\theta)$ - вероятности попадания в область цензурирования слева и справа, а наблюдаемая область лежит в пределах от x_1 до x_2 .

В [4] методом статистического моделирования исследована эффективность оценивания параметров законов распределения по цензурированной выборке по отношению к оцениванию по полной выборке. Соотношение (5) позволяет судить о потерях информации при оценивании параметров распределения в зависимости от степени цензурирования.

Эффективность оценивания параметра определяется соотношением:

$$S = \frac{\det J_c(\theta)}{\det J(\theta)},$$

где $J(\theta)$ - информационная матрица Фишера, построенная по полной выборке, $J_c(\theta)$ - по цензурированной, \det - определитель матрицы.

Результаты, проведенного в [4] исследования показали, что при умеренной доле цензурирования (до 40% от объема полной выборки) имеет место высокая эффективность оценивания параметров законов распределения, т.е. сохраняющаяся в цензурированной выборке информация является достаточной для оценивания параметров распределения с требуемой точностью. Таким образом, основной проблемой при оценке параметров законов распределения и на основании их показателей надежности при умеренном цензурировании является ограниченность объема анализируемой выборки.

Как отмечается в [5], оценки показателей надежности, построенные по выборкам ограниченного объема, являются смещенными.

Для построения более точных оценок рассмотрим процедуру, основанную на методе бутстрепа [5] и состоящую из следующих шагов.

1. По цензурированной выборке объема N вычисляются МП-оценки $\hat{\Theta}_1, \hat{\Theta}_2, \dots, \hat{\Theta}_l$ параметров закона распределения $F(t, \Theta_1, \Theta_2, \dots, \Theta_l)$.

2. Генерируются M независимых выборок объема N распределения $F(t, \hat{\Theta}_1, \hat{\Theta}_2, \dots, \hat{\Theta}_l)$. В качестве параметров для моделирования используются оценки параметров, полученные на шаге 1.

3. Для каждой из M выборок вычисляются МП-оценки параметров закона распределения и на основании их оценка показателя надежности. Формируется выборка оценок показателей надежности $R = \{\hat{R}_i, i = 1, \dots, M\}$.

4. В качестве искомого значения оценки показателя надежности принимается среднее значение выборки R .

Для нахождения непараметрических доверительных интервалов показателей надежности по полученной выборке оценок показателей надежности $R = \{\hat{R}_i, i = 1, \dots, M\}$ формируется вариационный ряд этих оценок $V = \{\hat{R}_{(i)}, i = 1, \dots, M\}$. Доверительные границы для показателя надежности R определяются следующими значениями вариационного ряда [5]: $[\hat{R}_{(l)}, \hat{R}_{(u)}]$, где

$$l = \left\lfloor \frac{1-\nu}{2} M \right\rfloor, u = \left\lceil \frac{1+\nu}{2} M \right\rceil, [\cdot] - \text{целая часть числа, } \nu - \text{доверительная вероятность.}$$

2. Оценивание показателей надежности и эффективности

Показатели надежности и эффективности АТС включают:

комплексные показатели надежности – коэффициенты готовности, оперативной готовности, технического использования [6];

техничко – экономические показатели использования АТС – затраты на запасные части, оплату труда ремонтных рабочих, компенсацию простоя [6];

техничко – эксплуатационные показатели использования АТС – средняя эксплуатационная и техническая скорость, коэффициенты использования пробега и грузоподъемности (статический и динамический), коэффициент выпуска, транспортная работа, объем перевозок, производительность и др. [6].

Оценивание этих показателей осуществляется за отдельные отчетные периоды или с накоплением от начала эксплуатации (в зависимости от физического смысла и цели

исследований). Аппроксимация эмпирических значений показателей проводится по следующим типам кривых: $kx+d$, dx^k , de^{kx} , $d+\frac{k}{x}$, $d+k\ln x$, $\frac{1}{d+kx}$, $de^{k/x}$, $\frac{1}{d+ke^{-x}}$, $(d+k\ln x)x$,

где x – значение пробега или продолжительности эксплуатации АТС.

Для вычисления МНК-оценок параметров проводится линеаризация кривых. Для проверки значимости уравнения регрессии применяется F – критерий.

Выбор кривой, наиболее точно аппроксимирующей эмпирические данные, состоит из следующих шагов.

1. Из представленного набора аппроксимирующих кривых исключаются те, различие которых с эмпирическими данными являлось значимым по F – критерию;

2. Из оставшихся кривых исключаются такие, экстраполяция которых некорректна (значения в отрицательной области, появление точек разрыва), и характер изменения которых противоречит физическому смыслу величины (например, коэффициент готовности должен монотонно убывать с ростом значения пробега, а затраты на запасные части – возрастать);

3. В качестве кривой, наиболее адекватно описывающей эмпирические данные выбирается та, для которой среднеквадратическая ошибка прогноза минимальна.

Построенные уравнения зависимости показателей надежности и эффективности от пробега используются для прогнозирования этих показателей в краткосрочном периоде. На рис. 1 представлены эмпирические и теоретические значения удельной стоимости запасных частей, использованных для ремонта АТС, в зависимости от пробега. Для вычисления теоретических значений в приведенном примере была использована линейная зависимость $y=kx+b$.

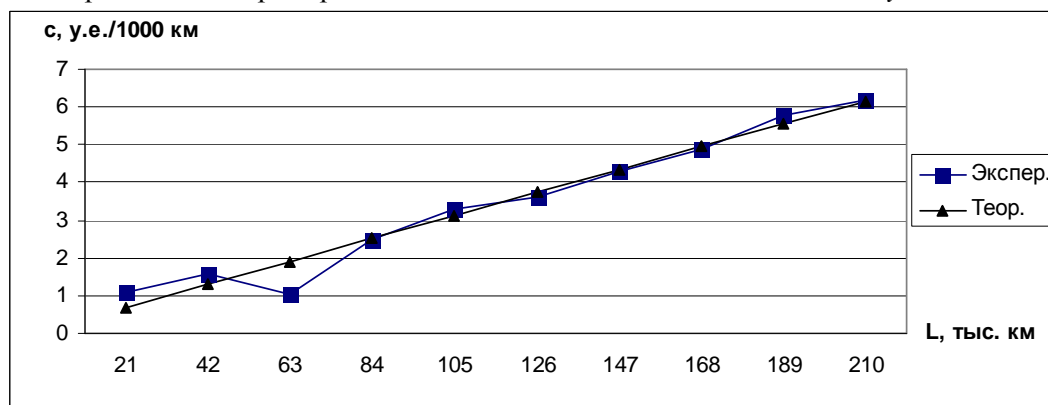


Рис. 1. Зависимость удельной стоимости запасных частей от пробега АТС

3. Структура программного комплекса оценки показателей надежности

Программный комплекс оценки показателей надежности, структура которого представлена на рис. 2, предназначен для сбора и обработки информации о надежности, вычисления показателей надежности и эффективности АТС и формирования соответствующих отчетов.

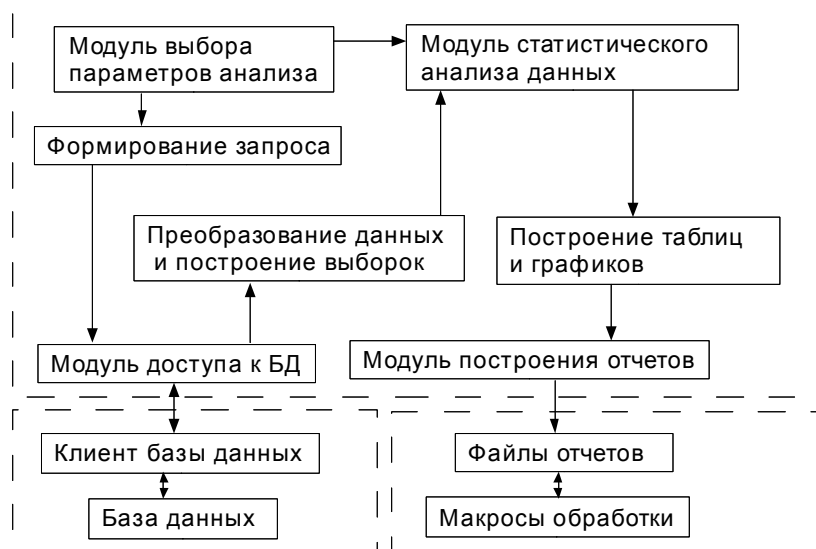


Рис. 2. Структура программного комплекса

Модуль выбора параметров анализа включает в себя выбор объекта статистического анализа (деталь, сборочная единица, АТС в целом), выбор модели, контрольной группы АТС, расчетного периода и вычисляемого показателя надежности.

База данных об отказах автомобилей МАЗ, разработанная в среде Oracle, содержит справочную информацию, а также отчетные данные по отказам деталей за определенный период – с начала эксплуатации на дату наличия данных (информационная карта о надежности). Для работы с информационными картами разработан модуль автоматизированного рабочего места сбора и обработки информации о надежности.

Модуль доступа к базе данных осуществляет:

- редактирование информации базы данных (добавление, исправление, удаление) – по всей исходной накопленной информации,

- редактирование справочников и классификаторов объектов (АТС, деталей, сборочных единиц), неисправностей (по характеру, причинам, влиянию на безопасность использования, сложности и способам устранения);

- получение по запросу исходных данных или результатов расчета.

Создание и корректировка базы данных, классификаторов и справочников объектов и неисправностей проводится по мере возникновения неисправностей объектов.

В модуле статистического анализа данных реализованы алгоритмы вычисления показателей надежности и эффективности. В этом модуле предусмотрен расчет и вывод промежуточных результатов для обеспечения возможности оперативного анализа.

Модуль построения отчетов позволяет выполнять:

- подготовку таблиц по заданным показателям, объектам, классификационным группам объектов и признаков классификации показателей;

- построение графиков изменения показателей надежности;

- передачу данных в файл Excel и запуск макросов обработки отчетов;

- формирование отчетов.

Разработанный программный комплекс используется в управлении главного конструктора ОАО «МАЗ» для оценки показателей надежности серийных и прогнозирования показателей надежности проектируемых моделей автомобилей семейства МАЗ.

Работа частично поддержана ГКПНИ «Инфотех» (задание Инфотех 51).

Литература

1. Аронов И.З., Бурдасов Е.И. Оценка надежности по результатам сокращенных испытаний. – М.: Издательство стандартов. – 1987. – 184 с.
2. Скрипник В.М., Назин А.Е., Приходько Ю.Г., Благовещенский Ю.Н. Анализ надежности технических систем по цензурированным выборкам. – М. Радио и связь. – 1988. – 184 с.
3. Лемешко Б.Ю. Об оценивании параметров распределений и проверке гипотез по цензурированным выборкам // Методы менеджмента качества. – 2001. – № 4. – С. 32-38.
4. Лемешко Б.Ю., Гильдебрант С.Я., Постовалов С.Н. К оцениванию параметров надежности по цензурированным выборкам // Заводская лаборатория. Диагностика материалов. – 2001. – Т. 67. – № 1. С. 52-64.
5. Diccio T.J., Efron B. Bootstrap Confidence Intervals // Statistical Science. – 1996. – Vol. 11. – № 3. P. 189-228.
6. Алексеева И.М., Ганченко О.И., Петрова Е.В. Статистика автомобильного транспорта. – М.: Экзамен. – 2005. – 352 с.

Абрамович М. С., Мицкевич М.Н., Пыжик Н.Н. Алгоритмическое и программное обеспечение статистического оценивания показателей надежности автотранспортных средств

Для временных рядов, наблюдаемых с шумом, разработаны критерии обнаружения “скачкообразных” изменений, основанные на вейвлет-преобразовании. Относительно шума предполагается, что он имеет более тяжелые “хвосты”, чем у нормального распределения. Методом статистического моделирования исследована эффективность критериев.

Рис. 2. Библиогр. – 6 назв.

M. S. ABRAMOVICH, M. M. MITSKEVICH, N. N. PYGYK

JUMP DETECTION IN TIME SERIES MEAN USING HAAR WAVELET TRANSFORMATION

Summary

Wavelet-based tests for time series with noise were developed. Noise is assumed to have more heavy tails than a Gaussian. The efficiency of these tests is analyzed by statistical modeling.

Абрамович Михаил Семенович – заведующий лабораторией статистического анализа и моделирования Научно-исследовательского института прикладных проблем математики и информатики, кандидат физ.-мат. наук, доцент. Раб. тел.: 2095481. Дом. тел.: 2505235. e-mail: abramovichms@bsu.by.

Мицкевич Михаил Николаевич – младший научный сотрудник Научно-исследовательского института прикладных проблем математики и информатики. Раб. тел.: 2095481. Дом. тел.: 2121638. e-mail: mitskevich_m@mail.ru.

Пыжик Николай Николаевич – начальник лаборатории надежности и техобслуживания ОАО «МАЗ». Раб. тел. 2172261. Дом. тел. 3450500.