

МОДЕЛЬ СЛУЧАЙНЫХ ОШИБОК ИЗМЕРЕНИЙ ОСТАТОЧНЫХ НАПРЯЖЕНИЙ ЖЕЛЕЗНОДОРОЖНЫХ КОЛЕС КОНТАКТНЫМ УЛЬТРАЗВУКОВЫМ МЕТОДОМ

Аннотация. На основании более десяти тысяч измерений окружных остаточных напряжений железнодорожных колес прибором *Debbie* рассмотрены различные параметрические модели распределения случайной ошибки измерений. Показано, что распределение Лапласа аппроксимирует неизвестный закон распределения вероятностей случайных ошибок лучше, чем нормальное распределение Гаусса. Применение модели рассмотрено в задаче обнаружения выбросов.

Ключевые слова: модель, случайная ошибка, остаточные напряжения, железнодорожное колесо, ультразвуковой метод, выброс.

Постановка задачи. Для использования на практике вероятностного подхода к оценке погрешностей, как средств, так и результатов измерений, прежде всего, необходимо установить для данной конкретной погрешности вид аналитической модели закона распределения [1]. Необходимость в этом возникает, в частности, при разработке методик контроля качества изделий, которые измеряются со случайными ошибками. В [1] указана ошибочность предположения о том, что случайные погрешности всегда должны иметь нормальное распределение. В связи с этим возникает задача идентификации формы их распределения по экспериментальным данным.

Возможность идентификации формы распределения по экспериментальным данным ограничена, прежде всего, малостью объема выборки, особенно когда идентификация должна осуществляться по измерениям, которые выполняются не на образце или мере с известным номинальным значением измеряемого параметра, а на однотипном множестве объектов, причём в каждой точке объекта производится небольшое количество повторных измерений. Обоснованный выбор модели в таком случае должен осуществляться по множеству выборок малого объема (5-10 каждая). Такая задача возникла при об-

работке измерений остаточных напряжений железнодорожных колёс, выполненных ультразвуковым прибором Debbie, и рассматривается далее.

Экспериментальные измерения. Одним из основных факторов надежности колеса является величина остаточных напряжений. Как отмечается в европейском стандарте EN 13262 [2] и директиве немецкой железной дороги RIL 907.0801 [3], окружные остаточные напряжения могут измеряться прибором Debbie. Их величина рассчитывается прибором по значению разности времени прохождения двух поперечных ультразвуковых волн, одна из которых поляризована в окружном, а другая в радиальном направлении. Измерения осуществляются контактным методом ввода ультразвуковых колебаний с помощью пьезопреобразователя (ПЭП), как показано на рис.1а. Для создания акустического контакта используется мед либо эпоксидная смола. Качество колеса оценивают по измерениям в нескольких точках, равномерно расположенных по окружности обода (рис.1б).

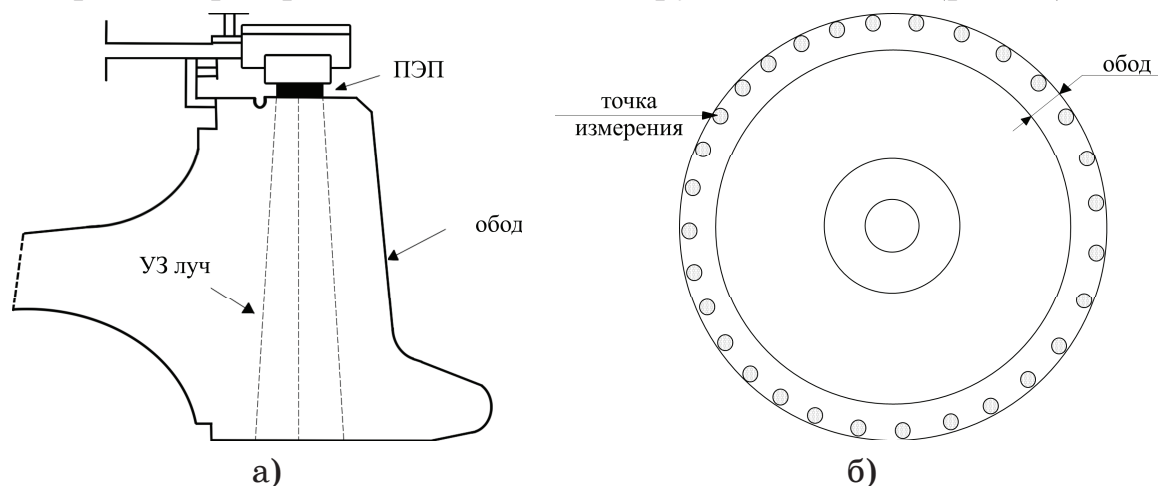


Рисунок 1 – Схема измерений остаточных напряжений:

- а) положение датчика на позиции измерений;
- б) расположение точек измерений

Как показали исследования, если после измерения датчик повторно переустановить на исходную позицию и заново повторить измерение, то полученное значение будет несколько отличаться от предыдущего. Таким образом, при повторных измерениях в одной и той же точке обода колеса, наблюдается разброс, вызванный случайной ошибкой измерений. Эта ошибка обусловлена неконтролируемыми случайными условиями измерительного процесса: нестабильность

акустического контакта, смещение положения датчика, опыт оператора и т.д.

Исследование случайных ошибок было выполнено на множестве из 89 колес K LW[®]. Данную группу образуют чистовые колеса из стали R7T по UIC 812-3, выбранные из различных плавок, произведенных в течение одного года. Номинальное значение шероховатости поверхности обода колес, что важно для контактного ввода ультразвуковых колебаний, составляло 6.3 Ra.

Каждое колесо измерялось в m точках по окружности обода. В каждой точке выполнялась серия из n повторных измерений с переустановкой датчика на исходную позицию после каждого измерения. Для группы, состоящей из N колес, измерения остаточных напряжений представляют собой трехмерный массив, обозначенный как $\{x_{j,i,k}, \forall x \in \mathbb{Z}\}$, $j = [1, N]$, $i \in [1, m]$, $k \in [1, n]$, $i, j, k \in \mathbb{N}$. Для рассматриваемых 89 колес количество точек по окружности было $m=25$, а количество повторных измерений было либо $n=5$ либо $n=10$.

Параметрическая модель случайной ошибки. Рассматривается аддитивная модель измерений со случайной ошибкой:

$$x_{j,i,k} = H_{j,i} + \varepsilon_{j,i,k}, \quad (1)$$

где $H_{j,i} \in \mathbb{R}$ - неизвестное значение остаточного напряжения в i -й точке обода j -го колеса; $\varepsilon_{j,i,k}$ - значение случайной величины ошибки ε при k -м повторном измерении в i -й точке обода j -го колеса из группы.

Если $f(\varepsilon)$ - неизвестная плотность распределения вероятностей ошибки, то из (1) следует, что $f(\varepsilon) = f(\varepsilon_{j,i,k}) = f(x_{j,i,k} - H_{j,i})$. Таким образом неизвестное значение $H_{j,i}$ можно рассматривать как сдвиг, который получают случайные ошибки при измерениях в i -й точке обода j -го колеса. Тогда оценкой этого неизвестного значения остаточного напряжения $H_{j,i}$ будет оценка сдвига $H_{j,i}^*$.

Рассмотрим задачу выбора параметрической модели случайной ошибки ε , т.е. плотности распределения вероятностей $g(\varepsilon)$, которая в некотором смысле хорошо аппроксимирует неизвестную плотность ошибки $f(\varepsilon)$:

$$f(\varepsilon) \approx g(\varepsilon) = g\left(\varepsilon, \left\{0, \hat{\theta}\right\}\right), \quad (2)$$

где $g(\varepsilon, \{\mu, \theta\})$ - параметрическая модель плотности распределения вероятностей ошибки, у которой число параметров равно p ; μ - параметр сдвига, $\theta \in \mathbb{R}^{p-1}$ - вектор параметров распределения без учёта параметра сдвига μ ; $\hat{\theta}$ - оценка вектора параметров θ , вычисленная по измерениям $\{x_{j,i,k}\}$. В формуле (2) принято, что ошибки сами имеют нулевой сдвиг $\mu = 0$. Если это условие не выполняется, то всегда может быть введена соответствующая поправка.

Если для ошибок выбран конкретный вид параметрической модели плотности g , то из (1) следует, что модель условной плотности вероятности распределения измерений в точке обода будет иметь вид:

$$w(x / j, i) = g\left(x - H_{j,i}, \left\{0, \hat{\theta}\right\}\right) = g\left(x, \left\{H_{j,i}, \hat{\theta}\right\}\right), \quad (3)$$

где $w(x / j, i) = g\left(x, \left\{H_{j,i}, \hat{\theta}\right\}\right)$ - модель условной плотности вероятности распределения измерений в i -й точке обода j -го колеса.

Свойство (3) позволяет определить условную логарифмическую функцию правдоподобия для вычисления оценок по измерениям группы из N колёс:

$$\ell(H, \theta / N, m, n) = \sum_{j=1}^N \sum_{i=1}^m \sum_{k=1}^n \ln g\left(x_{j,i,k}, \left\{H_{j,i}, \theta\right\}\right), \quad (4)$$

где $H = \{H_{j,i}\}$, $H \in \mathbb{R}^{N \times m}$ - массив переменных (неизвестных значений напряжений).

Оценки максимального условного правдоподобия соответствуют условию

$$\left\{\hat{H}, \hat{\theta}\right\} := \arg \max \left\{ \ell(H, \theta / N, m, n); H \in \mathbb{R}^{N \times m}, \theta \in \mathbb{R}^{p-1} \right\},$$

где $\hat{H} = \{H_{j,i}^*\}$, $\hat{H} \in \mathbb{R}^{N \times m}$ - массив оценок остаточных напряжений. Таким образом, путём максимизации функции (4), кроме оценок параметров плотности ошибки $\hat{\theta}$ можно определить ещё и оценки остаточных напряжений, т.е. матрицу \hat{H} .

Из множества моделей некоторого параметрического семейства $\{\mu, \theta\} \in \Theta$ наилучшую модель в смысле минимума расстояния Кульбака-Лейблера можно выбрать по информационным критериям [4]. Для этой цели можно использовать информационный критерий Акаике (Akaike), скорректированный на конечную длину выборок измерений. Для измерений N колес он запишется в виде:

$$AIC_c = AIC + \frac{2(Nm + p - 1)(Nm + p)}{Nm(n - 1) - p}, \quad (5)$$

где $AIC = -2\ell(\hat{N}, \hat{\theta} / N, m, n) + 2(Nm + p - 1)$. Среди моделей параметрического семейства Θ следует выбрать ту модель, для которой значение AIC_c наименьшее.

Рассмотрено два класса обобщающих распределений. Это класс распределений Пирсона тип-IV и обобщённое нормальное распределение.

Распределение Пирсона тип-IV задаётся плотностью

$$g(\varepsilon, \mu, \theta) = \frac{\Gamma(b) s^{2b-1}}{\sqrt{\pi} \Gamma\left(b - \frac{1}{2}\right)} \left| \frac{\Gamma\left(b + \frac{h}{2} \sqrt{-1}\right)}{\Gamma(b)} \right|^2 \left[s^2 + (\varepsilon - \mu)^2 \right]^{-b} \exp \left[-h \arctg \left(\frac{\varepsilon - \mu}{s} \right) \right],$$

где $\theta = \{b, h, s\}$ - вектор параметров за исключением параметра сдвига μ ; При значении $h = 0$ и замене $b = (v + 1) / 2$, $s = \sqrt{v\sigma^2}$ из плотности распределения Пирсона тип-IV получают симметричное нестандартизированное распределение Стьюдента с параметрами μ и $\theta = \{s, v\}$. Предельным случаем нестандартизированного распределения Стьюдента, когда $v \rightarrow \infty$, является нормальный закон распределения Гаусса с параметрами μ и $\theta = \{\sigma\}$.

Обобщённое нормальное распределение задаётся плотностью

$$g(\varepsilon, \mu, \theta) = \frac{k}{2\alpha \Gamma(1/k)} \exp \left(- \left| \frac{\varepsilon - \mu}{\alpha} \right|^k \right),$$

где $\theta = \{k, \alpha\}$. Это распределение симметричное и включает в себя как распределение Лапласа (при $k = 1$), так и нормальное распределение (при $k = 2$) с математическим ожиданием μ и дисперсией рав-

ной $\alpha^2 / 2$. Так как производная плотности данного распределения может иметь разрыв, то для вычисления оценок максимального правдоподобия использовался алгоритм, приведенный в [5].

Из представленных обобщающих распределений можно вывести и другие, более частные случаи. Для вычисления оценок необходимо, в общем случае, использовать численные методы, однако для распределения Лапласа и нормального закона могут быть получены аналитические формулы. Можно показать, что для нормального закона распределения формулы для вычисления оценок максимального правдоподобия имеют вид:

$$H_{j,i}^* = \bar{x}_{j,i}; \quad \sigma^{*2} = \frac{1}{Nmn} \sum_{j=1}^N \sum_{i=1}^m \sum_{k=1}^n (x_{j,i,k} - \bar{x}_{j,i})^2, \quad (6)$$

где $\bar{x}_{j,i}$ - среднее значение измерений в i -й точке обода j -го колеса.

Для распределения Лапласа формулы оценок максимального правдоподобия, основываясь на [6], можно записать следующим образом:

$$H_{j,i}^* = \text{med}_{j,i}^*; \quad \alpha^* = \frac{1}{Nmn} \sum_{j=1}^N \sum_{i=1}^m \sum_{k=1}^n |x_{j,i,k} - \text{med}_{j,i}^*|, \quad (7)$$

где $\text{med}_{j,i}^*$ - оценка медианы измерений в i -й точке обода j -го колеса;

α^* - оценка параметра масштаба α распределения Лапласа.

Из формул (6) и (7) можно получить основные представления о свойствах оценок, определяемых по множеству малых выборок. С одной стороны, несмотря на то, что функция правдоподобия выражается через все измерения группы из N колёс, оценка параметра сдвига $H_{j,i}^*$ в i -й точке обода j -го колеса зависит только от сдвига в этой конкретной точке. Это связано с тем, что смешанные производные функции правдоподобия по двум параметрам сдвига из разных точек равны нулю. С другой стороны, оценка $\hat{\theta}$, параметров плотности ошибки, зависит от всех измерений N колёс. Для оценки σ^{*2} , в формуле (6), можно показать, что её математическое ожидание и дисперсия определяются выражениями:

$$E[\sigma^{*2}] = \frac{n-1}{n} \sigma^2, \quad \text{Var}[\sigma^{*2}] = \frac{2(n-1)\sigma^4}{Nmn^2}. \quad (8)$$

Как следует из (8), смещение оценок параметров распределения случайной ошибки можно уменьшить лишь путём увеличения количества измерений n в каждой точке, в то время как дисперсия оценок уменьшается как при увеличении n , так m и N .

В таб.1 приведены рассчитанные значения оценок $\hat{\theta}$ для колёс четырёх плавков с наибольшим количеством измерений: 21039 ($N=7$, $m=25$, $n=10$), 22108 ($N=15$, $m=25$, $n=5$), 31174 ($N=10$, $m=25$, $n=5$), 32174 ($N=9$, $m=25$, $n=5$), а в таб.2 представлены значения нормированного показателя $AICc / Nm$.

Из анализа результатов таб.2 следует, что у нормального распределения значение показателя $AICc$ самые максимальные, а из класса обобщённого нормального распределения ошибок, наилучшей моделью является распределение Лапласа. Из семейства распределения Пирсона тип-IV для плавков 21039 и 31174 наилучшей моделью оказалось само это распределение, что, видимо, объясняется статистически значимой асимметрией распределения ошибок измерений. В свою очередь для плавков 22108 и 32174 наилучшей моделью оказалось нестандартизированное распределение Стьюдента.

Таблица 1

Оценки максимального правдоподобия для различных плавков

| Модель | Плавка | | | |
|---|--|--|--|--|
| | 21039 | 22108 | 31174 | 32174 |
| Распределение Пирсона тип-IV | $s^* = 3.39;$ $b^* = 1.85;$ $h^* = 0.56$ | $s^* = 4.74;$ $b^* = 2.69;$ $h^* = 0.14$ | $s^* = 4.45;$ $b^* = 2.73;$ $h^* = 1.76$ | $s^* = 2.72;$ $b^* = 1.64;$ $h^* = 0.09$ |
| Нестандартизированное распределение Стьюдента | $\sigma^* = 2.06;$ $v^* = 2.54$ | $\sigma^* = 2.27;$ $v^* = 4.36$ | $\sigma^* = 2.23;$ $v^* = 4.24$ | $\sigma^* = 1.81;$ $v^* = 2.28$ |
| Обобщённое нормальное распределение | $\alpha^* = 2.55;$ $k^* = 1.02$ | $\alpha^* = 3.37;$ $k^* = 1.45$ | $\alpha^* = 3.56;$ $k^* = 1.59$ | $\alpha^* = 2.72;$ $k^* = 1.12$ |
| Нормальное распределение | $\sigma^* = 3.64$ | $\sigma^* = 3.00$ | $\sigma^* = 2.92$ | $\sigma^* = 3.33$ |
| Распределение Лапласа | $\alpha^* = 2.39$ | $\alpha^* = 2.12$ | $\alpha^* = 2.08$ | $\alpha^* = 2.17$ |

Нормированные значения показателя АІСс*

| Модель | Плавка | | | |
|---|---------|---------|---------|---------|
| | 21039 | 22108 | 31174 | 32174 |
| Распределение Пирсона тип-IV | 53.4314 | 28.9217 | 27.1290 | 27.3922 |
| Нестандартизированное распределение Стьюдента | 53.5310 | 28.9146 | 27.2244 | 27.3808 |
| Обобщённое нормальное распределение | 53.5936 | 28.9666 | 27.2916 | 27.4951 |
| Нормальное распределение | 56.4267 | 29.2356 | 27.4047 | 28.7431 |
| Распределение Лапласа | 53.5194 | 28.4986 | 26.7471 | 27.1780 |

* в таблице подчёркнуты минимальные значения в рамках семейства распределений.

На рис.2 для этих четырёх плавков представлены графики плотностей распределения ошибок, которые были признаны наиболее удачными моделями из двух рассмотренных параметрических семейств. Штрихованной линией показаны плотности распределения Лапласа, а сплошной, в зависимости от плавки, либо плотность распределения Пирсона тип-IV, либо плотность нестандартизированного распределения Стьюдента.

Анализ моделей на рис.2, показывает, что распределения ошибок похожие для рассмотренных плавков. Значительная асимметрия распределения наблюдается только для плавки 31174, что возможно вызвано некоторым отличием условий выполнения измерений в сравнении с остальными плавками.

Если предположить, что распределение случайной ошибки не зависит от плавки, то оценки параметров распределения могут быть подсчитаны по всему массиву измерений 89 колёс. Такие оценки приведены в таб.3 наряду со значением показателя АІСс / Nm .

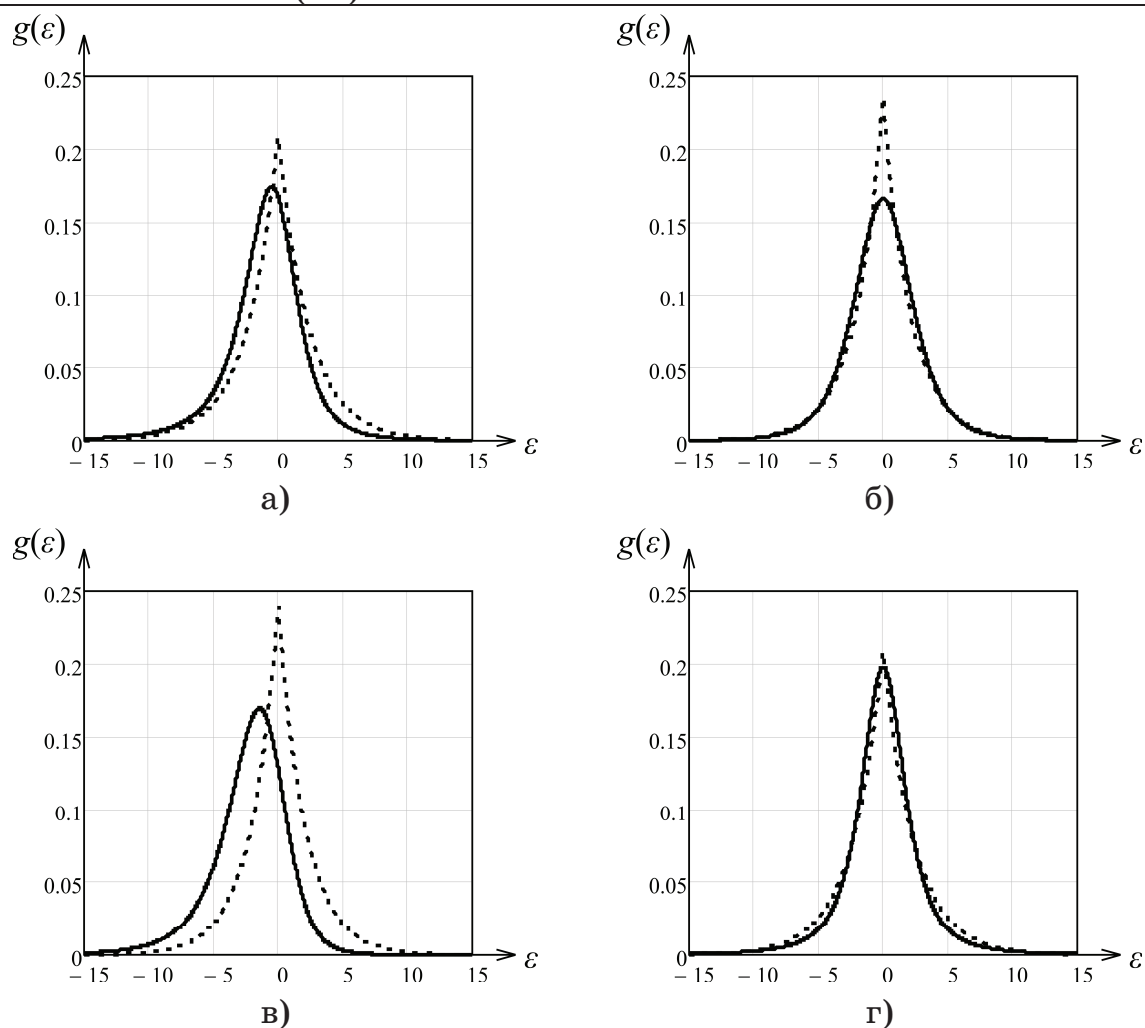


Рисунок 2 – Модели плотностей вероятности распределения ошибок измерений для плавков: а) 21039, б) 22108; в) 31174; г) 32174

Таблица 3

Значения оценок максимального правдоподобия и нормированного показателя AICc для группы из 89 колёс

| Модель | Оценки параметров | AICc / Nm |
|---|---|-----------|
| Распределение Пирсона тип-IV | $s^* = 3.61;$ $b^* = 2.01;$ $h^* = 0.314$ | 33.6495 |
| Нестандартизированное распределение Стьюдента | $\sigma^* = 2.08;$ $\nu^* = 2.97$ | 33.6653 |
| Обобщённое нормальное распределение | $\alpha^* = 3.07;$ $k^* = 1.26$ | 33.7127 |
| Нормальное распределение | $\sigma^* = 3.16$ | 34.4542 |
| Распределение Лапласа | $\alpha^* = 2.18$ | 33.2527 |

Как следует из данных таб.3, наилучшие модели случайной ошибки по критерию Акаике – это распределение Пирсона тип-IV и распределение Лапласа. На рис.3 показаны графики плотностей данных распределений в одном масштабе с гистограммами оценок значений ошибок (остатков), которые рассчитывались по формуле:

$$\varepsilon_{j,i,k}^* = x_{j,i,k} - H_{j,i}^*$$

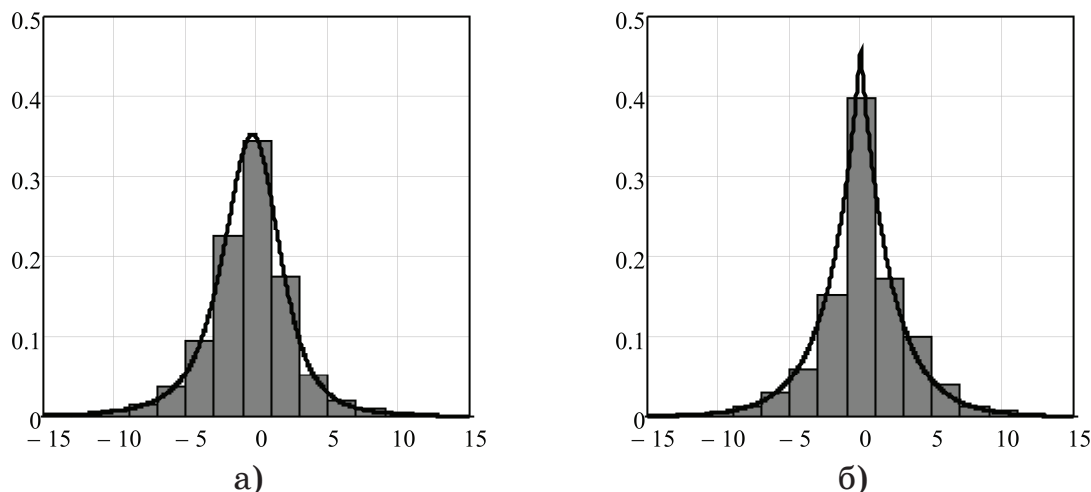


Рисунок 3 – Модели плотностей вероятности распределения ошибок и гистограммы остатков для измерений 89 колёс:

а) распределение Пирсона тип-IV; б) распределение Лапласа

Из определения модели (1) следует, что разность

$$\Delta\varepsilon_{j,i,k} = x_{j,i,2k} - x_{j,i,2k-1} = \varepsilon_{j,i,2k} - \varepsilon_{j,i,2k-1} \quad (9)$$

представляет собой последовательность некоррелированных случайных величин, плотность распределения которых полностью определяется плотностью распределения ошибки ε и может быть представлена в виде интеграла свёртки:

$$f(\Delta\varepsilon) = \int_{-\infty}^{\infty} f(\varepsilon)f(\varepsilon - \Delta\varepsilon) d\varepsilon. \quad (10)$$

Путём замены в (10) неизвестной плотности $f(\varepsilon)$ на модель $g(\varepsilon)$ получим модель $g(\Delta\varepsilon)$ плотности распределения разности (9). Таким образом, модели плотностей разности на основе распределений Лапласа и распределения Пирсона тип-IV можно сравнить с гистограммой разностей, представленной на рис.4.

Как следует из рис.4, модели плотности (10), выраженные через распределения Лапласа и Пирсона тип-IV, практически совпадают

и аппроксимируют гистограмму разностей примерно с одинаковой точностью.

От выбранной модели закона распределения случайной ошибки зависит тип оценок остаточных напряжений $\{H_{j,i}^*\}$. Например, для нормального закона – это средние значения измерений в точках, для распределения Лапласа – медианы, а для распределения Пирсона тип-IV, они не выражаются аналитически, а определяются численно из условия максимума функции правдоподобия.

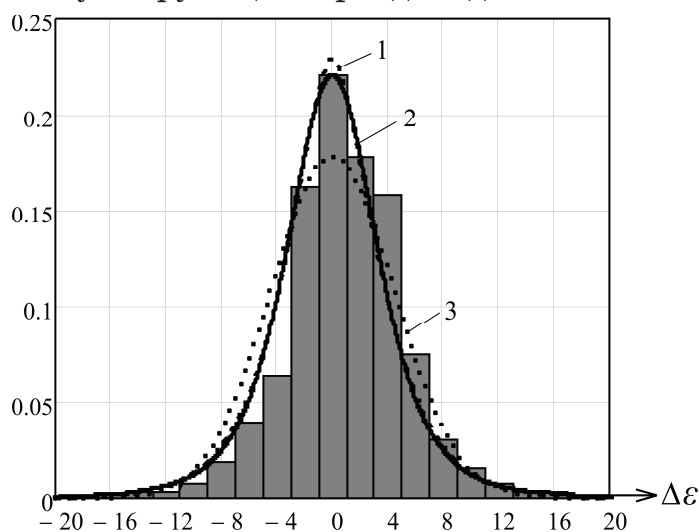


Рисунок 4 – Гистограмма разности $\Delta\varepsilon$ и её модели выраженные через распределения: 1 –Лапласа; 2 –Пирсона тип-IV; 3 –Гаусса

Сравнить два типа оценок $H_{(I)}^*$ и $H_{(II)}^*$ можно по среднеквадратическому значению их разностей

$$\sigma_{(I)-(II)}^{*2} = \frac{1}{Nm} \sum_{j=1}^N \sum_{i=1}^m (H_{(I),j,i}^* - H_{(II),j,i}^*)^2.$$

Величина $\sigma_{(I)-(II)}^*$ при сравнении медиан и средних значений равна $\sigma_{(I)-(II)}^* = 1.0$; для оценок, полученных на основе распределения Пирсона тип-IV, и средних значений $\sigma_{(I)-(II)}^* = 1.1$, а для медиан и оценок на основе распределения Пирсона тип-IV $\sigma_{(I)-(II)}^* = 1.0$. Таким образом, в среднем, оценки отличаются на достаточно малую величину около ± 2 МПа. Поэтому среднее значение измерений в точке можно использовать в качестве оценки остаточного напряжения в ней.

Критерий обнаружения выбросов. Правило обнаружение выбросов при многократных измерениях может быть получено из Лапласовой модели случайной ошибки. Если x - случайная величина с распределением Лапласа, параметр сдвига которой μ и масштаб α , то случайная величина

$$\zeta = \frac{2}{\alpha} \sum_{k=1}^n |x_k - \mu|$$

будет иметь распределение хи-квадрат с $2n$ степенями свободы независимо от значений параметров μ и α [7]. Как показывает компьютерное моделирование, если неизвестных сдвиг μ , обусловленный остаточными напряжениями, заменить средним значением \bar{x} , то случайная величина

$$\zeta^* = \frac{2}{\alpha} \sum_{k=1}^n |x_k - \bar{x}| = \frac{2}{\alpha} \sum_{k=1}^n |\Delta \varepsilon_k^*|,$$

приблизительно будет иметь распределение хи-квадрат с $(2n-1)$ степенями свободы. В этом случае правило для обнаружения выбросов в i -й точке обода можно записать в виде:

$$z_i < z_0; \quad z_i = \frac{1}{n} \sum_{k=1}^n |\Delta \varepsilon_{i,k}^*|; \quad z_0 = \frac{\alpha^*}{2n} \chi_{P,2n-1}^2,$$

где $\chi_{P,2n-1}^2$ - квантиль уровня $(1-P)$ распределения хи-квадрат с $(2n-1)$ степенями свободы, P - доверительная вероятность критерия, $\alpha^* = 2.18$ - оценка параметра α . Если выполняется условие $(z_i < z_0)$, то принимается, что выборка измерений остаточных напряжений в i -й точке обода колеса не содержит выбросов.

Выводы. По измерениям остаточных напряжений группы железнодорожных колес проведено обоснование выбора модели плотности вероятности распределения случайных ошибок из двух параметрических семейств: обобщенного нормального распределения (распределения ошибок) и распределения Пирсона тип-IV. Установлено, что если пренебречь асимметрией ошибок, то в качестве модели можно использовать распределение Лапласа, а оценками остаточных напряжений в точках могут быть средние значения измерений в этих точках.

Рассмотренный подход может быть применен для выбора модели распределения случайных ошибок измерений однотипных объектов контроля.

ЛИТЕРАТУРА

1. Новицкий П.В. Оценка погрешностей результатов измерений. / П.В. Новицкий, И.А. Зограф. — [2-е изд.].— Л: Энергоатомиздат, 1991— 304 с.
2. Railroad applications. Wheelsets and bogies. Wheels. Product requirements: DIN EN 13262 [European standard] — DIN Deutsches Institut für Normung e. V., Berlin. Beuth Verlag GmbH, 10772 Berlin, Germany, 2006 — 48p.
3. Richtlinie der Deutschen Bahn AG Ermittlung der Eigenspannungen im Radkranz von Vollrädern mit Ultraschall: RIL 907.0801.
4. Konishi S. Information Criteria and Statistical Modeling / S.Konishi, G.Kitagawa. — New York: Springer, 2008. — 274 p.
5. Shaoquan YU. A Review of Estimating the Shape Parameter of Generalized Gaussian Distribution / YU Shaoquan, Anyi Zhang, Hongwei LI // Journal of Computational Information Systems. — 2012. — №8 (21). — P. 9055–9064.
6. R.M. Norton. The Double Exponential Distribution: Using Calculus to Find a Maximum Likelihood Estimator / Robert M. Norton // The American Statistician (American Statistical Association). — 1984. — № 38 (2). —P. 135–136.
7. Johnson N.L. Continuous Univariate Distributions / N.L. Johnson, S. Kotz, N. Balakrishnan. — [2-nd edition]. — New York: Wiley, 1994. — 756 p.