УДК 629.7.01

Критерии нормальности при обработке экспериментальных исследований параметров газотурбинных двигателей на базе методов прикладной математической статистики

Вовк М.Ю.*, Кулалаев В.В.**

Опытно-конструкторское бюро имени А. Люльки,

филиал Уфимского моторостроительного производственного объединения,

ул. Касаткина, 13, стр. 8, Москва, 129301, Россия

e-mail: mvovk@yandex.ru

e-mail: kulalayev.viktor@gmail.com

Аннотация

В статье проведён обобщённый анализ критериев нормальности распределения экспериментальных данных. По результатам анализа разработан новый "векторный" критерий нормальности распределения для расчётов на ЭВМ, который позволяет использовать ограниченный объем экспериментальных данных, что является существенным фактором при проведении испытаний газотурбинных двигателей. Предложенный новый критерий и методика оценок нормальности распределения экспериментальных данных имеет значительно меньший объем вычислений по сравнению с существующими методами. Приведен алгоритм и пример расчётов. Введенный «векторный» критерий нормальности может быть использован при создании прикладных программных глобальных комплексов автоматизированной обработки массивов на ЭВМ экспериментальных данных натурных испытаний любых типов энергетических систем.

Ключевые слова: критерий, модель, регрессивный анализ, закон нормального распределения, статистика, прикладная математика, выборка, алгоритм.

1. Общие положения

Известно, что при испытаниях газотурбинного двигателя имеется значительное количество факторов, влияющих на измеряемые параметры. К таким факторам можно отнести погрешности приборов, изменение геометрии лопаточных машин вследствие раскрутки роторов двигателя, содержание влаги в рабочем теле, непосредственно настройка программ управления двигателем и его элементов (ВНА компрессоров, $F_{кр}$ сопла), изменение теплотворной способности топлива, изменения условий окружающей среды во время испытаний и др. Кроме того, необходимо отметить, что испытания авиационных ГТД как правило ограничено временными рамками, что значительно сокращает количество произведённых замеров. Общий анализ процессов обработки экспериментальных данных стендовых испытаний ГТД статистическими методами можно провести на основании, например, работ [1-20].

При идентификации математической модели ГТД по результатам экспериментов определённой партии двигателей, а также для проведения регрессионного анализа работы двигателей требуется выполнение условия нормального распределения экспериментальных данных. Следовательно, для обеспечения достоверности регрессионного анализа предварительно требуется обязательная проверка закона нормального распределения с помощью критериев нормальности [1-8]. В настоящее время существует 21 критерий согласия, модифицированных для проверки нормальности распределения. Виды критериев согласия с нормальным законом распределения (критериев нормальности) широко представлены в обширной работе [8] и там же приведена таблица мощности различных критериев нормального распределения, [табл. №80, стр. 278]. Следует указать на следующие особенности известных критериев согласия: критерии хорошо работают при величине выборки от 8 до 5000 значений, некоторые критерии требуют применения для расчётов ёмких специальных таблиц, которые предварительно необходимо рассчитывать (например, критерии Шапиро - Уилка, Филлибена, Ла Брека и др.). Практически расчёты всех критериев требуют составления обширных таблиц для субъективного анализа, что приводит к серьезным трудоёмким затратам и требует "не только знаний, практики, опыта, чутья и интуиции" [8] и ограничивает применение ЭВМ.

Необходимость введения критериев нормальности продиктовано тем, что подавляющее большинство используемых в технике экспериментальных исследований опытные данные подчинены закону нормального распределения [3,4,8]. В прикладной математической статистике в практическом отношении характеристики распределения вероятностей рассматривается как центр группирования статистических данных, степень рассеивания и поведение случайных величин в районе центра, а также симметричность относительно центра. Данным требованиям в полной мере соответствует нормальный закон распределения вероятностей. При этом анализ распределения данных статистик, с точки зрения прикладной математики, требует определенных знаний, практики и интуиции при первичном рассмотрении набора опытных данных экспериментов, где субъективный фактор приобретает значительную роль. Методы прикладной математической статистики требуют введения надежных и мощных критериев оценок статистических данных, что сводит субъективные подходы к минимуму.

2. Анализ известных критериев нормальности распределения статистических данных

По данным работы [8] шестая часть публикаций в области математики относиться к проблемам теории вероятностей и математической статистики, и их применениям. Ежегодно появляются более 5000 статей и книг, издаётся около 400 периодических изданий по данной тематике. Однако проблемы обработки данных экспериментов остаются существенными и актуальными. Указанные проблемы связаны, прежде всего, с субъективным подходом к выбору известных, многочисленных критериев оценок набора статистики, что относится и к общим критериям согласия, которые модифицированы для проверки нормальности распределения. Так в работе [9] предложена и исследована форма критерия согласия (критерий хи-квадрат χ^2), когда параметры распределения оцениваются по не группированной выборке. При использовании данного критерия необходимо применять специальную таблицу с коэффициентами c_i модифицированного (χ^2) - критерия, которую необходимо вносить в память ЭВМ при автоматизации расчетов. Такие же проблемы возникают с предложенными критериями Фроцини [10] и критерием типа Колмогорова – Смирнова [11]. В общем случае известен 21 модифицированный критерий нормальности [8]. Введен целый ряд специальных критериев нормальности, направленных на защиту всевозможных альтернатив. К ним отнесены известные основные критерии:

- критерий нормальности *Шапиро – Уилка* (с дополнительными отдельными ёмкими таблицами); энтропийный критерий нормальности *Васичека* весьма чувствительный к выбросам случайных величин;

- *Хегаси – Грина*, который имеет недостатки при равномерной альтернативе данных;

- *Али – Чёрного – Ревеса*, с введением взвешенных квадратов спейсингов – расстояние между соседними порядковыми статистиками, что значительно усложнят алгоритмические расчеты;

- корреляционный критерий *Филлибена* с введением дополнительного коэффициента корреляции r_i между порядковыми статистиками эмпирического ряда наблюдений;

- регрессивный критерий нормальности *Ла Брека*, который не пригоден при ассиметричных альтернативах, к тому же требуется применение дополнительных ёмких таблиц;

- критерий нормальности *Локка – Стурье*, требующий обработки больших объёмов расчетных данных;

- критерий нормальности *Оя*, который требует большого объёмного анализа комбинации порядковых статистик и заполнения сравнительных таблиц;

- критерий среднего абсолютного отклонения – критерий *Гири,* который основан на субъективном отборе α - квантиля стандартного нормального распределения, что практически не поддаётся составлению алгоритмов для ЭВМ;

- критерий *Дэвида – Хартли – Пирсона*, основанном на распределении отношения «размаха» к стандартному распределению с применением построения объёмных дополнительных сравнительных таблиц расчётных критических коэффициентов;

- комбинированный критерий *Шпигельхальтера*, который использует трудоёмкие вычисления сравнительных таблиц трёх известных критериев;

- критерий нормальности *Саркади*, при расчётах которого требуется анализировать $t-$ распределение *Стьюдента* (табулировано) и проводить дополнительный большой объём «ручных» расчётов;

- критерий нормальности *Лина – Мудхолкара*, который предполагает дополнительный расчет коэффициента корреляции и нормальной двумерной совокупности с применением нормализующих преобразований *Вильсона – Хилферти*;

- критерий нормальности *Мартинеса – Иглевича* имеет ограничения и применяется только против симметричных альтернатив, при этом необходимо определять отношения двух оценок дисперсии, что усложняет алгоритм расчётов;

- критерий нормальности *Д`Агостино* требует дополнительно проводить оценки методом правдоподобия с использованием объёмных таблиц и применяется, когда нет сведений об альтернативном распределении, что следует рассматривать как ограничение применимости критерия;

- критерий нормальности *асимметрии и эксцесса* требует применения нескольких таблиц с различных источников и решения логических задач при нормализации преобразований, что затрудняет составления алгоритма расчёта;

- критерий нормальности *Муроты – Такеучи,* в виде нормализующей характеристической функции, предполагает проверки сложных гипотез нормальности с использованием «стьюдентизированную» форму, что ограничивает применение критерия;

- критерий нормальности *Смирнова* связан с большим объёмом вычислений и логических субъективных отборов, что требует высокой математической подготовки и приводит к ограничениям применения в инженерных исследованиях.

Приведенный краткий критический анализ применимости известных критериев нормальности распределения вероятных статистик приводит к **цели** работы, которая заключается в необходимости разработки *нового вида критерия данной группы*, наиболее пригодного для автоматизации обработки объёмных экспериментальных данных при стендовых испытаниях различных типов ГТД с простым алгоритмом реализации на ЭВМ.

3. Построение «векторного» критерия проверки нормальности распределения статистик экспериментов

Теоретической основой введения и построения нового «векторного» критерия нормальности - $K_{\vec{r}}$ является введение вектора псевдо класса, который легко определяется в объёме назначенной выборки экспериментальных данных в зоне исследований по известным координатам точек результатов замеров искомой величины. Для примера, рассмотрим нормальное распределение статистического набора измерений при испытаниях ГТД, представленных на рис. 1.

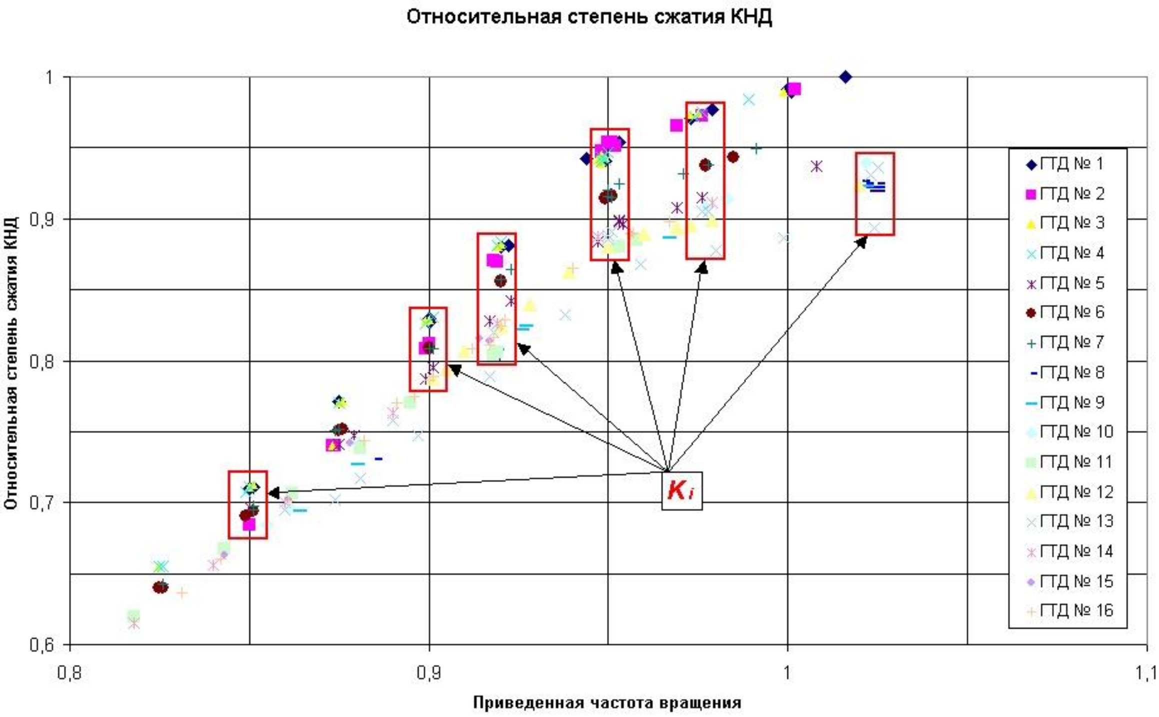

Рис.1. Экспериментальные данные степени повышения давления КНД ГТД.

Назначение зон выборок и интервалов зон.

На рис. 1 представлены экспериментальные значения степени повышения полного давления компрессора низкого давления ГТД. Для проведения анализа на нормальность распределения выделены зоны выборки $[K_i]$. Каждая зона выборок ограничивается по оси абсцисс назначенным интервалом $[N_i]$ для исследования. В данном случае интервал по оси абсцисс выбран как $n_{КНД\ пр}^{\pm 0,5\ \%}$. По оси ординат зона выборки сверху определяется максимальным значением, а снизу минимальным значением в интервале N_i.

В каждой зоне с интервалом $[N_1]$ выбраны точки $A\ (x,\ y_{max})$ и $B\ (x,\ y_{min})$ с координатами относительно условно определенного центра «0» нормального распределения статистик заданной зоны «$(y_{21} - y_{11})/2$», рис. 2. Соответственно вводится «вектор» точки *A (x, y)* псевдо класса (условный вектор, который не отвечает строгому математическому определению) в виде

$$|\square \vec{r}_A| = \sqrt{(x_A - x_{A"0"})^2 + (y_A - y_{A"0"})^2}\,. \tag{1}$$

Такие же вектора определяются для всех точек статистик выбранной для исследования зоны $[K_i]$, данные заносятся в алгоритм расчётов по всем выбранным зонам оценок.

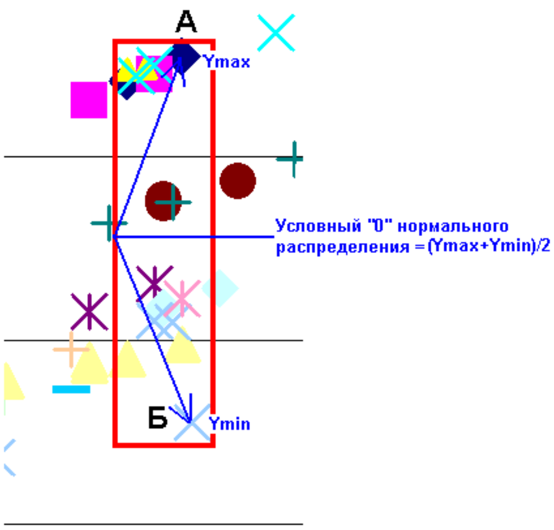

Рис. 2. Определение условного «0» нормального распределения

Алгоритм расчётов оценок нового «векторного» критерия нормальности распределения заключается в следующем:

1) определяются выборки K_i, где $i = 1, 2, 3 \ldots n$ - номер выборки;

2) определяется количество членов объёма выборки m_j, где j=1,2,3,…k количество членов выборки из объёма экспериментальных данных по схеме, представленной на рис. 2;

3) находится среднее арифметическое заданной выборки K_i и определенного объема m_j;

$$\bar{y}_i = \frac{1}{m_i} \cdot \sum_{j=1}^{k} y_i \ ; \tag{2}$$

4) находится среднеквадратичное стандартное отклонение выборки K_i из определенного объема m_j

$$\sigma_i = \sqrt{\frac{1}{m_i} \cdot \sum_{j=1}^{k} (y_i - \bar{y}_i)^2} \ ; \tag{3}$$

5) определяется функциональный вид графика плотности вероятности нормального распределения выборки K_i и определенного объема статистических данных m_j

$$p(y) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \left(\frac{y_i - \bar{y}_i}{\sigma} \right)^2 \right] ; \tag{4}$$

6) наносятся экспериментальные точки на график плотности вероятности нормального распределения исходя из условия, что точка с максимальным значением y_i приравнивается к максимуму значению графика *p(y)*, а минимальное значение y_i приравнивается к нулю. Остальные точки распределяются по графику пропорционально выбранным масштабам по координатам выборки заданного объёма, рис. 3.

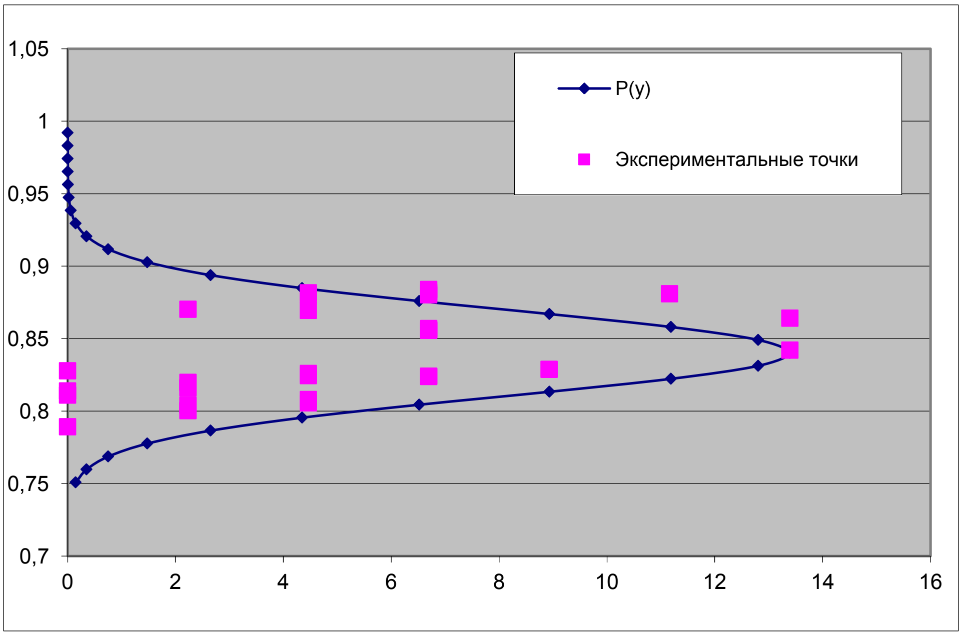

Рис. 3. Графическое представление распределения экспериментальных точек

7) определяем параметры расчётов *критерия* $[K_{\bar{r}}]$ согласно схемы, указанной на рис. 4.

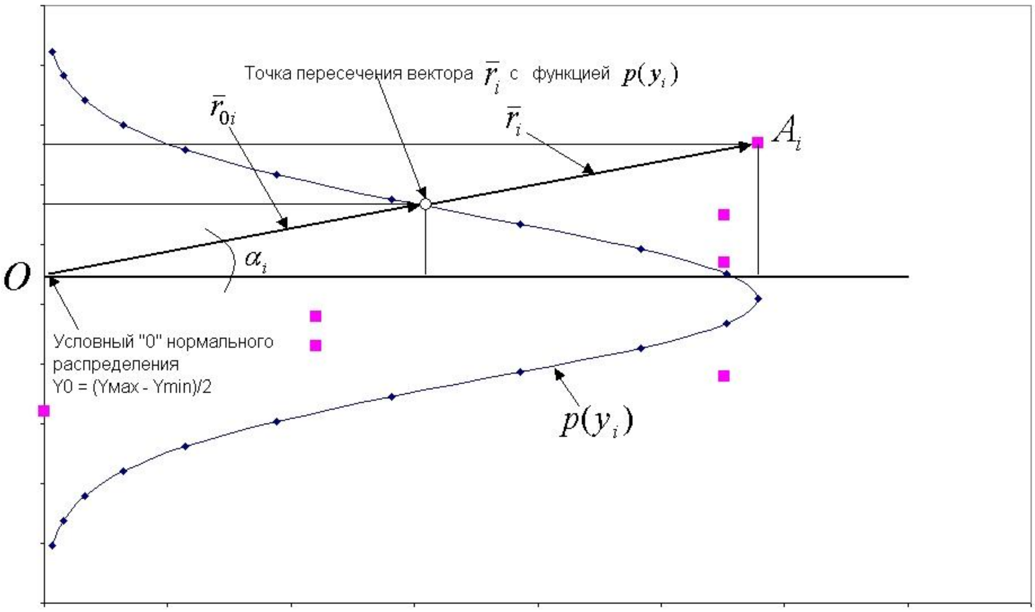

Рис. 4. Схема расчёта параметров критерия для выборки определённого объема.

Здесь α_I - угол наклона вектора *OA* к прямой, проведённой параллельно оси X через условный «0».

На схеме рис.4 видно, что уравнение вектора r_i , направленного на экспериментальную точку A_i приводится к виду

$$\left|\overline{r}_i\right| = \sqrt{y_{Ai}^2 + x_{Ai}^2} = x_{Ai} \cdot \sqrt{1 + \frac{y_{Ai}^2}{x_{Ai}^2}} = x_{Ai} \cdot \sqrt{1 + tg(\alpha)^2}\ . \tag{5}$$

Уравнение вектора r_{0i} , направленного на пересечение вектора r_i с графиком функции *p(y)*, определяется соотношением

$$\left|\overline{r}_{0i}\right| = \sqrt{y_{A0i}^2 + x_{A0i}^2} = x_{A0i} \cdot \sqrt{1 + \frac{y_{A0i}^2}{x_{A0i}^2}} = x_{A0i} \cdot \sqrt{1 + tg(\alpha)^2}\ . \tag{6}$$

Находим модуль «вектора» рассогласования с нормальным распределением по формуле (1), приведенным к виду

$$\left|\square\vec{r}\right| = \left|\vec{r}_i\right| - \left|\vec{r}_{oi}\right|. \tag{7}$$

Далее рассчитываем векторный *критерий* $[K_{\square\vec{r}}]$ для выборки $[N_i]$ и определенного объема $[m_j]$ по алгоритму блок – схемы рис. 5

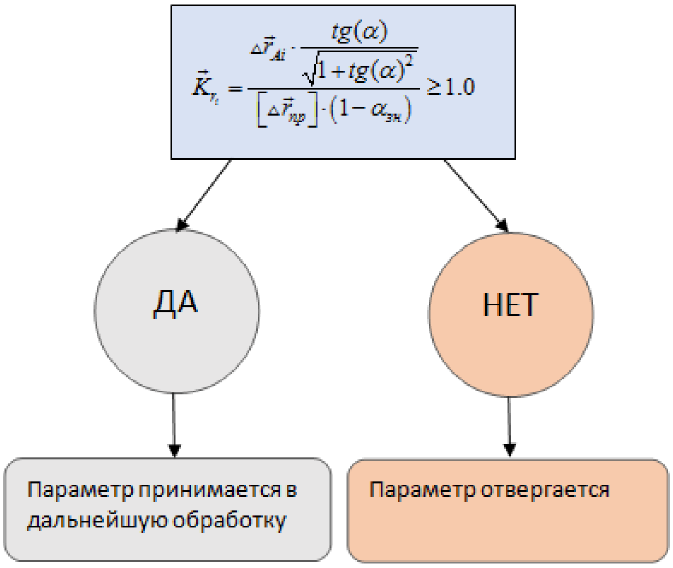

Рис. 5. Расчётная блок - схема ввода *векторного критерия* $[K_{\Delta\vec{r}}]$ для оценки нормальности распределения величин, экспериментальных данных статистики на поле выборки $[N_i]$заданного объема $[m_j]$

На рис.5 обозначены: - $\left[\Delta\vec{r}_{np}\right]$ - предельно допустимая погрешность экспериментальных замеров исследуемой величины в реальных единицах измерений (*задаётся*); $\alpha_{зн} = (0.05; 0.01; 0.01)$ - заданный коэффициент значимости (*назначается*).

Проводятся расчеты по предложенному алгоритму для всех назначенных выборок $[N_i]$ с соответствующими объёмами $[m_j]$.

Алгоритм расчётов координат точек пересечения графика плотности вероятности нормального распределения *p(y)* и вектора $\vec{r}_A$ с углом наклона (α_i) для каждой выборки $[N_i]$ с соответствующими объёмами $[m_j]$ (см. рис. 5) сводится к решению системы уравнений вида:

$$\begin{cases} y_{Aoi} = \dfrac{y_{Ai}}{x_{Ai}} \cdot x_{Aoi}; \\ y_{Aoi} = p(y_{Aoi}) = \dfrac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left\{ -\dfrac{1}{2} \cdot \left(\dfrac{y_{Aoi} - \bar{y}_i}{\sigma} \right)^2 \right\} \end{cases}. \qquad (8)$$

Аналитически приведенная система уравнений не решается, поэтому применяем метод итераций по следующему дополнительному алгоритму:

- назначаем погрешность определения координат точки пресечения графиков (см, рис. 5):

$$y_{Aoi} = y_{Ai} - (\square = 10^{-n}, ед. - задают);$$

$$\tilde{y}_{Aoi} = p(y_{Aoi}) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left\{ -\frac{1}{2} \cdot \left(\frac{y_{Aoi} - \bar{y}_i}{\sigma} \right)^2 \right\}; \qquad (9)$$

$$\tilde{x}_{Aoi} = \frac{\tilde{y}_{Aoi}}{tg\alpha};$$

$$Y'_{Aoi} = \tilde{x}_{Aoi} \cdot tg\alpha,$$

здесь шаг расчётов $\square = 10^{-n}, ед. - задают$, $n = 1,2,3,...$;

- определяем относительную точность расчётов:

$$\frac{\left| p(Y'_{Aoi}) - p(y_{Aoi)} \right|}{p(Y'_{Aoi})} \le (\varepsilon = 10^{-4} - задаётся); \qquad (10)$$

- выполняем *i= 1, 2, 3, …, n* – итераций до получения назначенной точности: если «**Да**», то определяют координату $x_{Aoi} = y_{Aoi} \cdot \frac{x_{Ai}}{y_{Ai}} = \frac{y_{Ai}}{tg\alpha}$. При «**Нет**» расчёты проводят далее по алгоритму, представленному выше (8 - 9).

Таким образом, введенный и предлагаемый авторами *векторный критерий проверки нормальности* [$K_{\vec{r}_i}$] распределения статистики при стендовых испытаниях изделий принимает вид

$$K_{\vec{r}_i} = \frac{\square\vec{r}_{Ai} \cdot \frac{tg(\alpha)}{\sqrt{1+tg(\alpha)^2}}}{\left[\square\vec{r}_{пр}\right] \cdot \left(1-\alpha_{зн}\right)} \geq 1.0 . \qquad (11)$$

Известно, что мощность критерия — это способность не допустить ошибку при выборе гипотезы обработки статистики экспериментальных данных. Мощность критерия определяется эмпирическим путем $(1-\beta)$, где β - вероятность принятия ошибочного решения. Таким образом, ранг предлагаемого «векторного» критерия в настоящее время не определён, однако по мере его применения такое исследование будет проведено в дальнейшем. Сравнительная мощность различных критериев нормальности, применяемых в настоящее время, приведена в работе [8].

Заключение

На основании методов прикладной математической статистики и известных методов регрессивного анализа проведен анализ *критериев нормальности* *распределения* для обработки экспериментальных данных на ЭВМ при выборе оптимального метода экстраполяции данных статистики стендовых испытаний ГТД в программных продуктах:

- предложен *новый «векторный» критерий* нормальности распределения $[K_{\vec{r}_i}]$ (см. зависимость (11), связанный со случайной выборкой с однозначной оценкой адекватности выбранной математической модели при обработке экспериментальных данных стендовых замеров параметров ГТД со значительным сокращением объёма вычислений;
- разработан алгоритм расчетов *нового «векторного» критерия* нормальности распределения $[K_{\vec{r}_i}]$ для реализации в программном продукте ЭВМ при обработке статистики стендовых испытаний ГТД;
- приведен пример расчёта критерия $[K_{\vec{r}_i}]$ на ЭВМ для статистики реальных стендовых испытаний ГТД;
- введенный *новый «векторный» критерий* нормальности распределения $[K_{\vec{r}_i}]$ может быть использован при обработке массивов экспериментальных точек и регрессивного анализа статистик при стендовых испытаниях различных изделий для расчётов на ЭВМ и создания прикладных программных комплексов автоматизированной обработки массивов экспериментальных данных с значительным сокращением времени расчётов, в том числе для обработки экспериментальных данных натурных испытаний ГТД.

Библиографический список

1. Чичков Б.А. Методология оптимального построения и использования диагностических моделей газотурбинных двигателей: Дис. ... доктора техн. наук. – М., 2004. - 147 с.

2. Kvetom K. Formation of empirical regression curves and surfaces using power function // Acta Technica, 1988, no. 2, pp. 141 - 157.

3. Закс Л. Статистическое оценивание. – М.: Статистика, 1976. – 598 с.

4. Химмельблау Д. Анализ процессов статистическими методами. – М.: Мир, 1970. -957 с.

5. Кузьмичева А.О., Мельникова Н.С. Разработка и реализация алгоритма индуктивного порождения регрессионных моделей для системы автоматического управления авиационного газотурбинного двигателя // Труды МАИ. 2010. № 38. URL: http://trudymai.ru/published.php?ID=14155

6. Губин В.И., Осташков В.Н. Статистические методы обработки экспериментальных данных. - Тюмень: Изд-во ТюмГНГУ, 2007. - 202 с.

7. Агапов Е.Г., Битехтин Е.А. Обработка экспериментальных данных в MS Excel. - Хабаровск: Изд-во Тихоокеанского государственного университета, 2012. - 32 с.

8. Кобзарь А.И. Прикладная математическая статистика. - М.: Физматлит, 2006. - 816 с.

9. Dahija R.S., Gurland J. How many classes in the Pearson chi-square test // Journal of the American Statistical Association, 1973, vol. 68, no. 343, pp. 707 - 712.

10. Frozini B.V. A survey of a class of goodness – of – fit statistics // Metron, 1978, 36, no. 1-2, pp. 3 - 49.

11. Мартынов Г.В. Критерий омега - квадрат. - М.: Наука, 1978. – 78 с.

12. Крамер Г. Математические методы статистики. - М.: Мир, 1975. – 648 с.

13. Хостингс Н., Пикок Дж. Справочник по статическим распределениям. – М.: Статистика, 1980. – 95 с.

14. Казакявичюс К.А. Приближенные формулы для статистической обработки результатов механических испытаний // Заводская лаборатория. 1988. Т. 54. № 12. С. 82 - 85.

15. Гумбель Э. Статистика экстремальных значений. - М.: Мир, 1965. – 451 с.

16. Браунли К.А. Статистическая теория и методология в науке и технике. - М.: Наука, 1977. – 408 с.

17. Вульфович Б.А. Оценивание параметров малой выборки // Депонирование во ВНИЭРХ 22.10.91, № 1180 – рх9.

18. Дейвид Г. Порядковые статистики. – М.: Наука, 1979. - 336 с.

19. Кудлаев Э.М. Оценивание параметров распределения Вейбулла - Гнеденко // Техническая кибернетика. 1986. № 6. С. 5 - 18.

20. Shapiro S.S., Wilk M.B., Chen H.J. A comparative study of various testes for Normality // Journal of the Acoustical Society of America, 1968, vol. 63, no. 324, pp. 1343 - 1372.