УДК 519.2

О. В. Мукан, В. А. Василенко, О. А. Егоров, Т. М. Артеев
Институт компьютерных технологий Национального авиационного университета

Алгоритм проверки гипотезы нормальности случайного вектора

Рассмотрен алгоритм проверки гипотезы о нормальности двухмерного случайного вектора. В качестве меры расхождения выборочного и гипотетического распределения в алгоритме используется кумулянтный критерий. Приведены результаты исследования зависимости мощности кумулянтного критерия от объема выборки. Предложен способ повышения эффективности кумулянтного критерия.

Ключевые слова: *алгоритм, нормальный случайный вектор, гипотеза, критерий, кумулянт.*

Подтверждение гипотезы о принадлежности технического объекта (ТО) к классу изнашиваемых — тесты Вульфа–Холландера, Прошана–Пайка, Холландера–Прошана, Вилкоксона [1] — позволяет случайный процесс (СП) изменения его определяющих парметров (ОП) $\vec{X}(t)=\{X_l(t)\}$, $l=\overline{1,N}$, представить как необратимый нестационарный СП, каждая составляющая которого описывается математической моделью вида:

$$X_l(t)=V_{0,l}+V_l t^{\eta},\ l=\overline{1,N}, \qquad (1)$$

где $V_{0,l}, V_l, l=\overline{1,N}$ — случайные коэффициенты; η — неслучайный показатель степени.

Такая модель процесса достаточно адекватно описывает реальные процессы изменения определяющих параметров объектов, подверженных воздействию старения, коррозии и износа. Для ее математического описания требуется минимальное (по сравнению с известными моделями случайных процессов) количество экспериментальных данных [2–4]. Эта модель дает возможность наиболее просто (аналитически) исследовать надежность технических объектов при изменении их свойств, и, наконец, данная модель учитывает два основных свойства, присущих практически всем процессам изменения определяющих параметров сложных систем — нестационарность и непрерывность [2].

В работе [5] показано, что для определения показателей параметрической надежности изнашиваемых ТО необходимо знать вид плотности распределения вероятности $q_l(v_0, v)$ случайных коэффициентов $V_{0,l}, V_l, l = \overline{1, N}$, модели (1) и ее параметры, а также значение неслучайного показателя степени η.

В нашем случае нет надежных теоретических обоснований для выбора того или иного закона распределения случайных коэффициентов модели. Тем не менее, должны учитываться соображения в пользу нормального распределения. Во-первых, известно, что нормальный закон порождается в результате воздействия большого числа примерно равных по влиянию факторов. Во-вторых, известно также, что нормальный закон распределения содержит минимум информации по сравнению с любым распределением с той же дисперсией. В-третьих, проведенный анализ большого числа параметров изнашиваемых ТО позволил сделать вывод, что, несмотря на некоторые отклонения (особенно на «хвостах» распределения), практически все коэффициенты полиномиальной модели имеют нормальное или близкое к нормальному распределение.

Для проверки гипотезы о нормальности двумерного случайного вектора применимы критерий w^2 и кумулянтные критерии. Требование большого объема N исходных статистических данных ($N > 100$) и плохая формализуемость процедуры разбиения этих данных на разряды делают невозможным использование первого из них. Отсутствие аналитического и табличного описания статистики w^2 для двумерного случая в условиях малой выборки не позволяет здесь применить и критерий w^2. Кумулянтные критерии, использующие в качестве меры расхождения выборочного и гипотетического распределений статистику Q,

$$Q = \gamma^T \times R^{-1} \times \gamma \,,$$

где γ — вектор-столбец кумулянтных коэффициентов и R — выборочная ковариационная матрица, соответствующая вектору γ, в достаточной степени свободны от указанных выше недостатков.

Схема алгоритма *NORM*2 проверки нормальности двумерного случайного вектора с использованием двумерных кумулянтных коэффициентов 4-го порядка, т.е. когда $\gamma^T = (\gamma^*_{40}, \gamma^*_{31}, \gamma^*_{22}, \gamma^*_{13}, \gamma^*_{04})$, приведена на рис. 1.

Работа алгоритма *NORM*2 начинается с ввода массива $\{v_{0,i}, v_{1,i}\}$, $i = \overline{1, I}$, уровня значимости α. В блоке 3 по формуле

$$\alpha^*_{lk} = \frac{1}{I}\sum_{i=1}^{I} v^l_{0,i} \cdot v^k_{1,i}, (l,k) = \overline{0,4}, \;\; l \wedge k \neq 0$$

определяются оценки начальных $\alpha^*_{lk}, (l,k) = \overline{1,4}$ моментов.

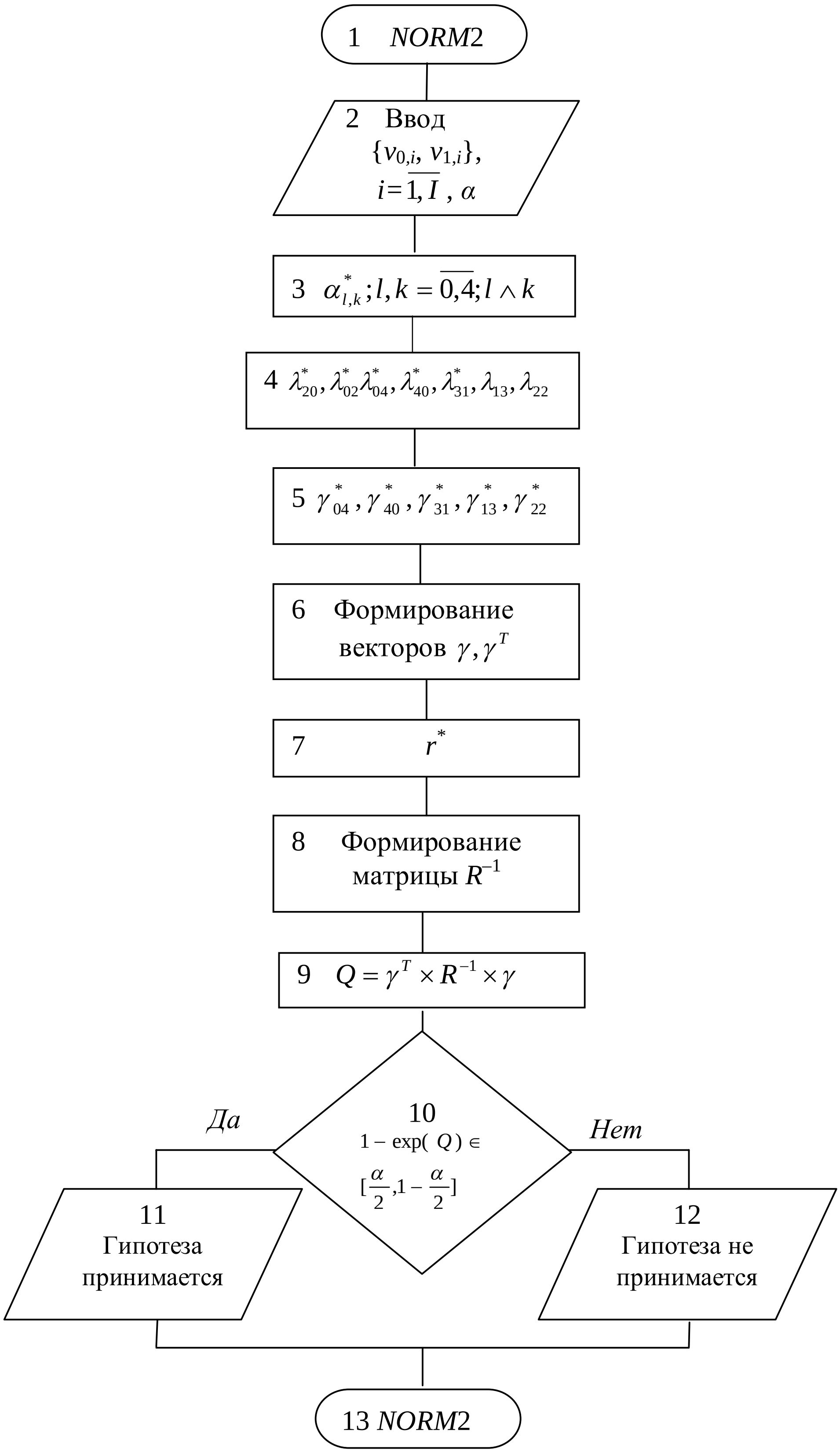

Рис. 1. Схема алгоритма проверки гипотезы о нормальности двумерного вектора

Используя известные

$$\lambda_{02}^* = \alpha_{02}^* - \alpha_{01}^{*2},$$
$$\lambda_{20}^* = \alpha_{20}^* - \alpha_{10}^{*2}$$

и полученные методом кумулянтных скобок соотношения

$$\lambda_{40}^* = \alpha_{40}^* - 3\alpha_{20}^{*2} - 4\alpha_{10}^*\alpha_{30}^* + 12\alpha_{10}^{*2}\alpha_{20}^* - 6\alpha_{10}^{*4},$$

$$\lambda_{31}^* = \alpha_{31}^* - 3\alpha_{10}^*\alpha_{21}^* - \alpha_{01}^*\alpha_{30}^* - 3\alpha_{20}^*\alpha_{11}^* + 6\alpha_{10}^*\alpha_{01}^*\alpha_{20}^* + 6\alpha_{10}^{*2}\alpha_{11}^* - 6\alpha_{10}^{*3}\alpha_{01}^*,$$

$$\lambda_{13}^* = \alpha_{13}^* - 3\alpha_{01}^*\alpha_{12}^* - \alpha_{10}^*\alpha_{03}^* - 3\alpha_{02}^*\alpha_{11}^* + 6\alpha_{01}^*\alpha_{10}^*\alpha_{02}^* + 6\alpha_{01}^{*2}\alpha_{11}^* - 6\alpha_{01}^{*3}\alpha_{10}^*,$$

$$\lambda_{04}^* = \alpha_{04}^* - 3\alpha_{02}^{*2} - 4\alpha_{01}^*\alpha_{03}^* + 12\alpha_{01}^{*2}\alpha_{02}^* - 6\alpha_{01}^*,$$

$$\lambda_{22}^* = \alpha_{22}^* - \alpha_{20}^*\alpha_{02}^* + 2\alpha_{20}^*\alpha_{01}^{*2} + 3\alpha_{10}^{*2}\alpha_{02}^* - 8\alpha_{10}^{*2}\alpha_{01}^{*2} - 3\alpha_{10}^*\alpha_{12}^* + 10\alpha_{20}^*\alpha_{01}^*\alpha_{11}^* - 2\alpha_{01}^*\alpha_{21}^* - 2\alpha_{11}^{*2},$$

в блоке 4 определяются оценки кумулянтов второго $\lambda_{02}^*, \lambda_{20}^*$ и четвертого $\lambda_{04}^*, \lambda_{40}^*, \lambda_{31}^*, \lambda_{13}, \lambda_{22}$ порядков. В блоке 5 по формуле

$$\gamma_{lk}^* = \frac{\lambda_{lk}^*}{\lambda_{02}^{*k/2} \cdot \lambda_{20}^{*l/2}}$$

рассчитываются оценки кумулянтных коэффициентов $\gamma_{04}^*, \gamma_{40}^*, \gamma_{31}^*, \gamma_{13}^*, \gamma_{22}^*$ четвертого порядка.

Дальнейшее функционирование алгоритма *NORM*2 связано с вычислением статистики Q:

$$Q = R_{11}\gamma_{40}^{*2} + 2R_{12}\gamma_{40}^*\gamma_{31}^* + 2R_{13}\gamma_{40}^*\gamma_{31}^* + 2R_{15}\gamma_{40}^*\gamma_{04}^* +$$

$$+ R_{22}\gamma_{31}^{2*} + 2R_{23}\gamma_{31}^*\gamma_{22}^* + 2R_{24}\gamma_{31}^*\gamma_{13}^* + 2R_{25}\gamma_{31}^*\gamma_{04}^* + R_{33}\gamma_{22}^* +$$

$$+ 2R_{34}\gamma_{22}^*\gamma_{13}^* + 2R_{35}\gamma_{22}^*\gamma_{04}^* + R_{44}\gamma_{13}^{*2} + 2R_{45}\gamma_{13}^*\gamma_{24}^* + R_{55}\gamma_{04}^{*2},$$

где $R_{ij}, i = \overline{1, I}, j = \overline{1, J}$ — элементы симметричной матрицы R^{-1}:

$$\frac{I}{24(1-r^{*2})^4} - \frac{Ir^*}{6(1-r^{*2})^4} \cdot \frac{Ir^{*2}}{4(1-r^{*2})^4} - \frac{Ir^{*3}}{6(1-r^{*2})^4} \frac{Ir^{*3}}{24(1-r^{*2})^4}$$

$$\frac{I(1+3r^{*2})}{6(1-r^{*2})^4} \frac{Ir^*(1+r^{*2})}{2(1-r^{*2})4} \frac{Ir^{*2}(r^{*2}+3)}{6(1-r^{*2})^4} - \frac{Ir^{*3}}{6(1-r^{*2})^4}$$

$$\frac{I(1+5r^{*2})}{4(1+r^{*2})(1-r^{*2})^4} - \frac{Ir^*(1+r^{*2})}{2(1-r^{*2})^4} \frac{Ir^{*2}}{4(1-r^{*2})^4}$$

$$\frac{I(1+3r^{*2})}{6(1-r^{*2})^4} \frac{Ir}{6(1-r^{*2})^4}$$

$$\frac{I}{24(1-r^{*2})^4}$$

Здесь

$$r^* = \frac{(\lambda_{02}^* \cdot \lambda_{20}^*)^{-1/2}}{I-1} \sum_{i=1}^{I} v_{0,i} \cdot v_{1,i} - \frac{I}{I-1} \alpha_{01}^* \cdot \alpha_{10}^*$$

оценка коэффициента корреляции.

Так, в блоке 6 осуществляется формирование вектора-столбца γ и вектора-строки $\gamma^T = (\gamma_{40}^*, \gamma_{31}^*, \gamma_{22}^*, \gamma_{13}^*, \gamma_{04}^*)$. В блоке 7 вычисляется оценка r^* коэффициента корреляции. В блоке 8 рассчитываются элементы R_{ij} матрицы R^{-1}, и производится ее формирование. В блоке 9 вычисляется значение статистики Q.

В блоке 10 для заданного уровня значимости α осуществляется принятие решения: если выполняется неравенство

$$\alpha/2 < 1 - \exp(Q) < 1 - \alpha/2,$$

то гипотеза о нормальности вектора $\{v_0, v_1\}$ принимается, в противном случае — отвергается. В блоках 11, 12 печатаются соответствующие сообщения. На этом работа алгоритма *NORM*2 завершается (блок 13).

Для оценки свойств кумулянтного критерия, использующего выборочные кумулянты четвертого порядка, производилось исследование его мощности $M = 1 - \beta$, в зависимости от объема N выборки. В качестве исходной для такого исследования использовалась выборка двумерного вектора $\{X, Y\}$, распределенного по закону:

$$W(x, y) = q(x, y) + af(x, y),$$

где

$$q(x, y) = \frac{1}{2\pi(1 - r^2)^{1/2}} \exp\left[- \frac{-x^2 - 2rxy + y^2}{2(1 - r^2)} \right],$$

$$f(x, y) = \begin{cases} x \cdot y, |x| \le 1, |y| \le 1 \\ 0, |x| < 1, |y| > 1 \end{cases},$$

отличному от нормального, но каждая из составляющих которого имеет нормальное распределение. При проведении исследований объем N выборки, коэффициент корреляции r и уровень значимости α изменялись в пределах: N = 5, 10, 20, 40, 50, 60, 100; r = 0,2, 0,7, 0,9; α = 0,01, 0,05, 0,1. Оценка мощности M^* при фиксированных N, r, α определялись методом статистического моделирования, для чего в каждом варианте производилось 100 опытов по проверке гипотезе о нормальности. Это обеспечило абсолютную ошибку оценки мощности $\Delta = |M - M^*| \le 0{,}1$ с надежностью $\gamma = 0{,}95$.

Полученные в результате эксперимента зависимости приведены на рис. 2–5. Анализ их содержания показывает, что мощность критерия при прочих равных условиях существенно зависит от тесноты стохастической связи между составляющими вектора $\{X, Y\}$. Так, при малой выборке и очень слабой связи ($r \le 0{,}2$) мощность не превышает уровня 0,4 (рис. 2), при слабой связи (0,4 < r < 0,7) — уровня 0,6 (рис. 3). При средней связи (0,7 ≤ r < 0,9) и очень сильной связи (r ≥ 0,5) мощность данного критерия превышает уровень 0,8 практически для всех значений α (рис. 5).

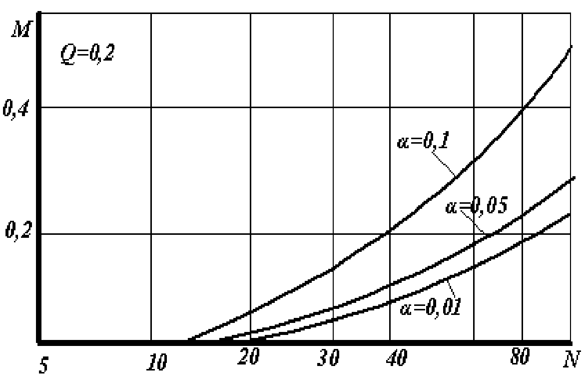

Рис. 2. Зависимости мощности M кумулянтного критерия от объема выборки N при различных уровнях значимости α и коэффициенте корреляции r = 0,2

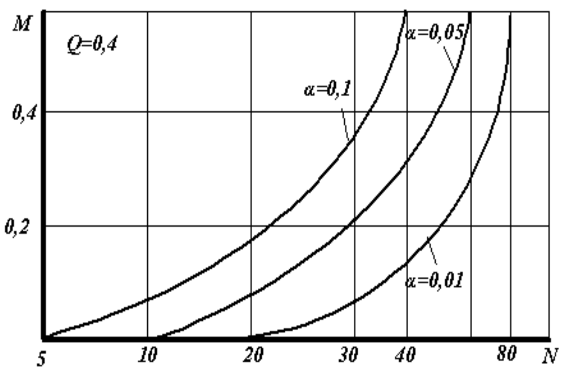

Рис. 3. Зависимости мощности M кумулянтного критерия от объема выборки N при различных уровнях значимости α и коэффициенте корреляции r = 0,4

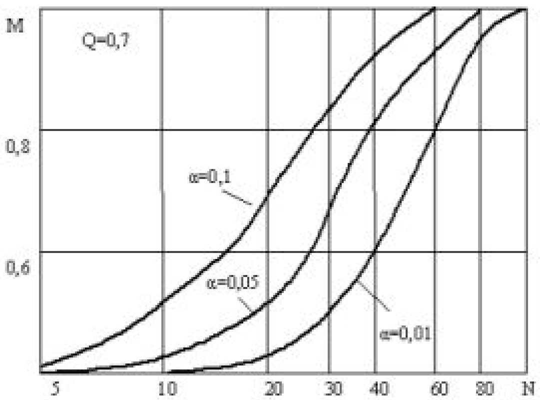

Рис. 4. Зависимости мощности M кумулянтного критерия от объема выборки N при различных уровнях значимости α и коэффициенте корреляции r = 0,7

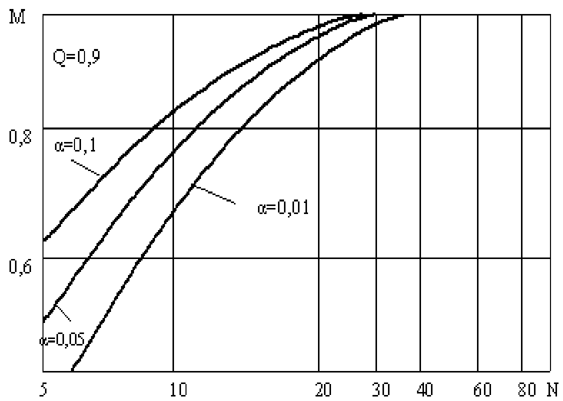

Рис. 5. Зависимости мощности M кумулянтного критерия от объема выборки N при различных уровнях значимости α и коэффициенте корреляции r = 0,9

Низкую эффективность работы кумулянтного критерия при значениях $r < 0{,}4$ нельзя считать недостатком, препятствующим его практическому использованию. Во-первых, потому что ситуации, когда каждая из составляющих случайного вектора распределена по нормальному закону, а распределение вектора отлично от нормального, крайне редки. Во-вторых, негативные последствия этого недостатка можно значительно ослабить, если в случае положительного результата работы алгоритма *NORM*2 осуществлять проверку гипотезы о независимости составляющих вектора $\{v_0, v_1\}$, используя для этих целей высокоэффективный критерий.

1. *Барзилович Е.Ю., Савенков М.В.* Статистические методы оценки состояния авиационной техники. — М.: Транспорт, 1987. — 240 с.

2. *Дружинин Г.В.* Надежность автоматизированных производственных систем. — 4-е изд., пер. и доп. — М.: Энергоатомиздат, 1986. — 480 с.

3. *Абрамов О.В., Бернацкий Ф.И., Здор В.В.* Параметрическая коррекция систем управления. — М.: Энергоатомиздат, 1982. — 176 с.

4. Оптимизация радиоэлектронной аппаратуры / Маслов А.Я., Чернышов А.А., Ведерников В.В. и др. / Под. Ред. А.Я. Маслова, А.А. Чернышева. — М.: Радио и связь, 1982. — 200 с.

5. *Мукан О.В., Бушмелев П.А., Олейник С.А.* Математические модели оценки локальных показателей достоверности многоразового контроля пригодности однопараметрических объектов РЭО // Материалы ВНК училища.: Сб. статей. — К.:КВВАИУ, 1989. — Вып. 1. — С. 23–27.

Поступила в редакцию 05.04.2007