СИСТЕМЫ И ПРОЦЕССЫ УПРАВЛЕНИЯ

УДК006.91

РАСЧЕТ СТАТИСТИК КРИТЕРИЯ ГРАББСА ДЛЯ АРКСИНУСОИДАЛЬНЫХ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ

САФАРЯН Г.Г., СЕРГИЕНКО М.П.

Исследуется возможность применения критериев типа Граббса для выявления аномальных наблюдений, принадлежащих выборкам из генеральной совокупности, распределенной по закону арксинуса. Получены распределения статистик Граббса для случаев существенного отклонения наблюдений в сторону как больших, так и меньших значений, а также процентные точки для различных объемов выборки при уровне значимости 0,95.

Актуальность исследования

В метрологической практике при осуществлении этапа предварительной обработки данных необходимым является использование статистических критериев выделения аномальных наблюдений (выбросов). Игнорирование этой процедуры приводит, как правило, к некорректным результатам, поскольку в большинстве случаев используемые классические статистические методы чувствительны к имеющимся аномальным наблюдениям.

Наиболее часто для проверки наблюдений на аномальность используют простые критерии Граббса [1-4]. Эти критерии предусматривают возможность проверки наличия в выборке одного (наибольшего или наименьшего) либо двух (двух наибольших или двух наименьших) аномальных наблюдений и применяются для проверки на аномальность наблюдений, распределенных по нормальному закону. Однако многие физические величины не подчиняются нормальному закону распределения, и использование [2] в этих случаях не является корректным. Особый интерес в связи с этим вызывают величины, имеющие арксинусоидальное распределение (погрешности при измерениях параметров круговых величин, погрешности от наводок и помех на выходе средств измерительной техники, от силовых цепей промышленной частоты и т.п.). Поэтому необходимо исследовать распределение статистик критерия Граббса при арксинусоидальных законах распределения наблюдений.

Целью данной работы является нахождение и исследование распределений статистик типа Граббса для выборок, принадлежащих арксинусоидальной генеральной совокупности. В связи с поставленной целью выделены основные *задачи*: расчет процентных точек при уровне значимости $\alpha = 0{,}05$ для выявления одного, двух и трех аномальных наблюдений (выбросов).

Проверка на один выброс

При проверке на выброс наибольшего выборочного значения проверяется гипотеза, заключающаяся в том, что наблюдение $X_1, X_2, ..., X_{n-1}$ из построенного по выборке вариационного ряда $X_1, X_2, ..., X_n$ принадлежит арксинусоидальному закону распределения, а наибольшее наблюдение X_n принадлежит другому закону распределения, существенно сдвинутому вправо. В этом случае статистика критерия Граббса имеет вид

$$G = \frac{(X_n - \overline{X})}{S}, \qquad (1)$$

где

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, \qquad (2)$$

$$S = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2}. \qquad (3)$$

При проверке наименьшего выборочного значения X_1 на выброс проверяется гипотеза, предполагающая, что X_1 принадлежит другому закону распределения, существенно сдвинутому влево. В этом случае статистика Граббса имеет вид

$$G = \frac{(\overline{X} - X_1)}{S}. \qquad (4)$$

Максимальное или минимальное наблюдение считается выбросом, если значение соответствующей статистики превысит критическое: $G \geq G_{1-\alpha}$, где α – заданный уровень значимости.

Распределение статистик (1) и (4) в [1-3] не приводится. Вид условных распределений $F(G)$ этих статистик в зависимости от объема выборки при арксинусоидальном законе распределения наблюдаемых величин показан на рис. 1. Статистики (1) и (4) распределены одинаково.

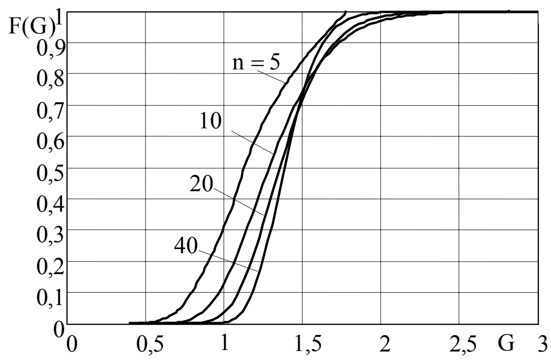

Рис.1. Зависимость распределения статистик (1) и (4) критерия Граббса от объема выборки

Поскольку решение об аномальности проверяемого максимального или минимального выборочного значения принимается по правой части распределения статистики, были рассчитаны верхние процентные точки для различных объемов выборки.

Процентные точки для всех рассматриваемых случаев были построены по моделируемым выборкам статистик. Объем каждой выборки составлял 10^5 смоделированных значений статистики с усреднением 50 раз. СКО полученных процентных точек не превысило 10^{-4}.

Для уровня значимости $\alpha = 0{,}05$ (доверительная вероятность $P = 0{,}95$) верхние точки $G_{0,95}$ в зависимости от объема выборки n показаны на рис. 2.

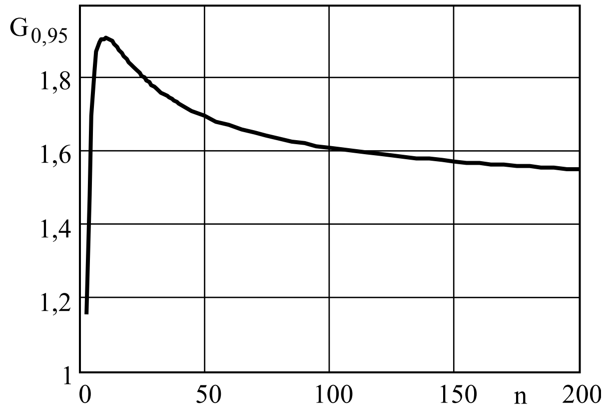

Рис.2. Зависимость статистики $G_{0,95}$ от объема выборки n

Зависимость статистики $G_{0,95}$, показанная на рис.2, существенно отличается от аналогичной зависимости, построенной для выборок, принадлежащих нормальной генеральной совокупности [2], которая монотонно возрастает с увеличением числа наблюдений n.

Проверка на два выброса

В этом случае проверяется гипотеза о том, что некоторому другому закону принадлежат наблюдения: 1) X_n и X_{n-1}; 2) X_1 и X_2; 3) X_n и X_1.

При проверке на выброс одновременно двух наибольших наблюдений статистика критерия Граббса имеет вид

$$G = \frac{S^2_{n-1,n}}{S^2_0}, \qquad (5)$$

где

$$S^2_0 = \sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad (6)$$

$$S^2_{n-1,n} = \sum_{j=1}^{n-2}(X_j - \overline{X}_{n-1,n})^2, \qquad (7)$$

в которых

$$\overline{X}_{n-1,n} = \frac{1}{n-2}\sum_{j=1}^{n-2} X_j, \qquad (8)$$

а $\overline{X}$ определяется из выражения (2).

При проверке на выброс одновременно двух наименьших наблюдений X_1 и X_2 статистика критерия Граббса принимает вид

$$G = \frac{S^2_{1,2}}{S^2_0}, \qquad (9)$$

где S^2_0 определяется из выражения (6):

$$S^2_{1,2} = \sum_{j=3}^{n}(X_j - \overline{X}_{1,2})^2, \qquad (10)$$

$$\overline{X}_{1,2} = \frac{1}{n-2}\sum_{j=3}^{n} X_j. \qquad (11)$$

Оба значения X_n, X_{n-1} или X_1, X_2 считаются выбросами, если значение соответствующей статистики окажется ниже критического $G < G_\alpha$.

Вид условных распределений $F(G)$ статистик G (5) и (9) в зависимости от объема выборки показан на рис. 3.

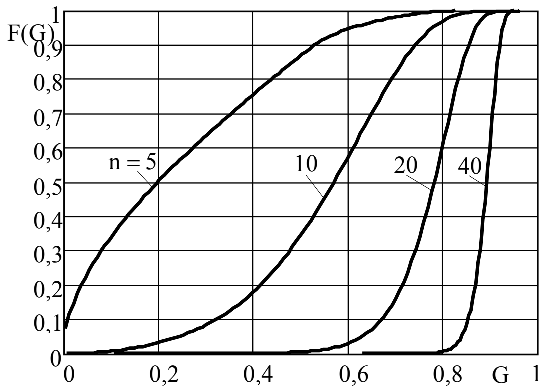

Рис. 3. Зависимость распределения статистик (5) и (9) критерия Граббса от объема выборки n

Решение об аномальности одновременно двух наибольших или двух наименьших наблюдений принимается по левой части распределения статистики. Для уровня значимости $\alpha = 0{,}05$ (доверительная вероятность $P = 0{,}95$) нижние точки $G_{0,05}$ в зависимости от объема выборки n показаны на рис.5 линией 1.

При проверке на выброс одновременно наибольшего и наименьшего наблюдений статистика Граббса имеет вид

$$G = \frac{S^2_{1,n}}{S^2_0}, \qquad (12)$$

где S^2_0 определяется из выражения (6):

$$S^2_{1,n} = \sum_{j=2}^{n-1}(X_j - \overline{X}_{1,n})^2, \qquad (13)$$

$$\overline{X}_{1,n} = \frac{1}{n-2}\sum_{j=2}^{n-1} X_j. \qquad (14)$$

Оба значения X_n и X_1 считаются выбросами при заданном уровне доверия α, если значение соответствующей статистики, вычисленное по выборке, окажется ниже критического $G < G_\alpha$.

Вид условных распределений $F(G)$ статистики G (12) в зависимости от объема выборки показан на рис. 4.

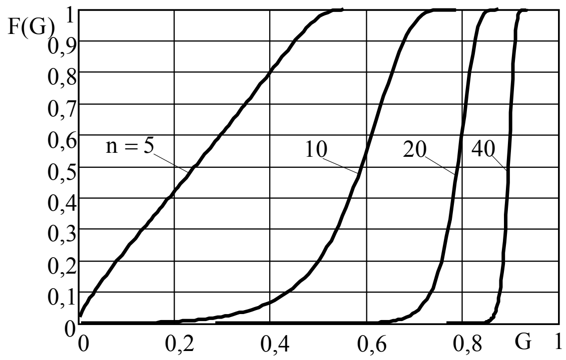

Рис. 4. Зависимость распределения статистик (12) критерия Граббса от объема выборки n

Решение об аномальности одновременно наибольшего и наименьшего наблюдений принимается по левой части распределения статистики. Для уровня значимости $\alpha = 0{,}05$ нижние точки $G_{0,05}$ в зависимости от объема выборки n показаны на рис. 5 линией 2.

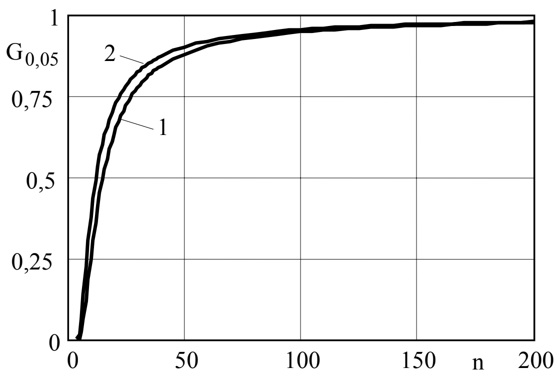

Рис. 5. Зависимость статистики $G_{0,95}$ от объема выборки n

Проверка на три выброса

В случае проверки наблюдаемой выборки на три выброса проверяется гипотеза о том, что некоторому другому закону, отличному от закона арксинуса, принадлежат наблюдения: 1) X_n, X_{n-1} и X_{n-2}; 2) X_1, X_2 и X_3; 3) X_n, X_{n-1} и X_1; 4) X_n, X_1 и X_2.

При проверке на выброс одновременно трех наибольших наблюдений статистика критерия Граббса имеет вид

$$G = \frac{S^2_{n-2,n-1,n}}{S^2_0}, \qquad (15)$$

где S^2_0 определяется в соответствии с выражением (6):

$$S^2_{n-2,n-1,n} = \sum_{j=1}^{n-3} (X_j - \overline{X}_{n-2,n-1,n})^2, \qquad (16)$$

$$\overline{X}_{n-2,n-1,n} = \frac{1}{n-3} \sum_{j=1}^{n-3} X_j. \qquad (17)$$

При проверке на выброс одновременно трех наименьших наблюдений X_1, X_2 и X_3 статистика критерия Граббса принимает вид

$$G = \frac{S^2_{1,2,3}}{S^2_0}, \qquad (18)$$

где S^2_0 определяется из выражения (6):

$$S^2_{1,2,3} = \sum_{j=4}^{n} (X_j - \overline{X}_{1,2,3})^2, \qquad (19)$$

$$\overline{X}_{1,2,3} = \frac{1}{n-3} \sum_{j=4}^{n} X_j. \qquad (20)$$

Три значения X_n, X_{n-1} и X_{n-2} или X_1, X_2 и X_3 считаются выбросами, если значение соответствующей статистики окажется ниже критического $G < G_\alpha$.

Вид условных распределений $F(G)$ статистик G (15) и (18) одинаков и показан на рис.6 в зависимости от объема выборки.

Решение об аномальности одновременно трех наибольших или трех наименьших наблюдений принимается по левой части распределения статистики. Для уровня значимости $\alpha = 0{,}05$ нижние точки $G_{0,05}$ в зависимости от объема выборки n показаны на рис.6.

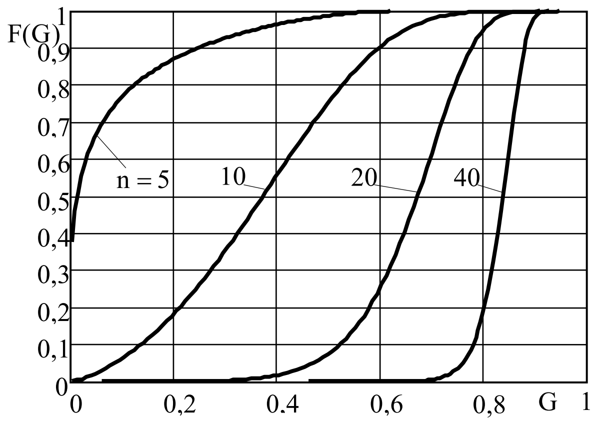

Рис. 6. Зависимость распределения статистик (15) и (19) критерия Граббса от объема выборки n

При проверке на выброс одновременно двух наибольших и одного наименьшего наблюдений статистика критерия Граббса имеет вид

$$G = \frac{S^2_{n-1,n,1}}{S^2_0}, \qquad (21)$$

где S^2_0 определяется в соответствии с выражением (6):

$$S^2_{n-1,n,1} = \sum_{j=2}^{n-2} (X_j - \overline{X}_{n-1,n,1})^2, \qquad (22)$$

$$\overline{X}_{n-1,n,1} = \frac{1}{n-3} \sum_{j=2}^{n-2} X_j. \qquad (23)$$

При проверке на выброс одновременно одного наибольшего и двух наименьших наблюдений статистика критерия Граббса

$$G = \frac{S^2_{1,2,n}}{S^2_0}, \qquad (24)$$

где S_0^2 определяется из выражения (6):

$$S_{1,2,n}^2 = \sum_{j=3}^{n-1} (X_j - \overline{X}_{1,2,n})^2 , \qquad (25)$$

$$\overline{X}_{1,2,n} = \frac{1}{n-3} \sum_{j=3}^{n-1} X_j . \qquad (26)$$

Три значения X_n, X_{n-1} и X_1 или X_1, X_2 и X_n считаются выбросами, если значение соответствующей статистики окажется ниже критического для заданного уровня α: $G < G_\alpha$.

Вид условных распределений F(G) статистик G (21) и (24) одинаков и показан на рис.7 в зависимости от объема выборки.

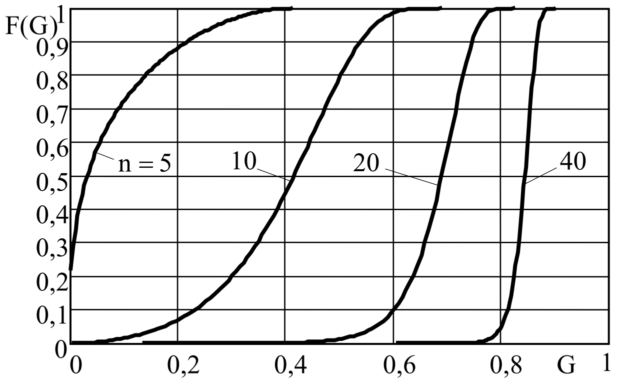

Рис. 7. Зависимость распределения статистик (21) и (24) критерия Граббса от объема выборки n

Решение об аномальности одновременно двух наибольших и одного наименьшего или одного наибольшего и двух наименьших наблюдений принимается по левой части распределения статистики. Для уровня значимости $\alpha = 0{,}05$ нижние точки $G_{0,05}$ в зависимости от объема выборки n показаны на рис. 8 линией 2.

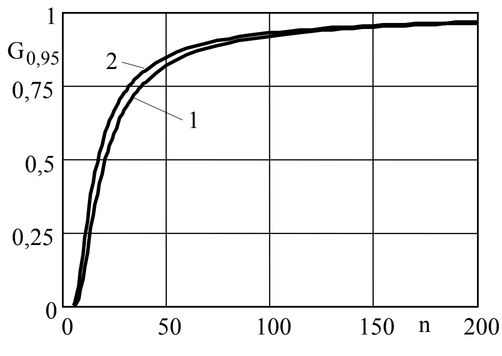

Рис. 8. Зависимость статистики $G_{0,95}$ от объема выборки n

Выводы

Полученные результаты позволяют обнаруживать до трех наблюдений в исследуемой выборке, не принадлежащих арксинусоидальному закону распределения и существенно сдвинутых в сторону наибольших или наименьших значений.

Научная новизна исследований заключается в получении распределения статистик типа Граббса для выявления аномальных наблюдений при арксинусоидальном распределении генеральной совокупности.

Практическая значимость результатов заключается в том, что проведенные исследования позволяют выявлять до пяти аномальных наблюдений при различных комбинациях их отклонений в сторону больших и малых значений. Получены процентные точки статистик критерия Граббса для уровней значимости $\alpha = 0{,}05$.

Литература: 1. *Frank E. Grubbs, Glenn Beck*. Extension of sample sizes and percentage points for significance tests of outlying observations// Technometrics, 1972. Vol. 14, No.4. P. 847–854. **2.** *ГОСТ Р ИСО 5725-2-2002*. Точность (правильность и прецезионность) методов и результатов измерений. Часть 2. М.: Изд-во стандартов. 51 с. **3.** *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. М.: Наука, 1983. 416 с. **4.** *Frank E. Grubbs*. Procedures for Detecting Outlying Observations in Samples// Technometrics, 1969. Vol. 11. No.1. P. 1 – 21.

Поступила в редколлегию 13.07.2006

Рецензент: д-р техн. наук, проф. Крюков А.М.

Сафарян Григорий Гагикович, инженер кафедры МИТ ХНУРЭ. Научные интересы: исследование погрешностей вычислительных операций при цифровой обработке сигналов, статистическая обработка результатов измерений. Адрес: Украина, 61166, Харьков, пр. Ленина,14, тел. 702-1331.

Сергиенко Марина Петровна, канд. техн. наук, с.н.с. кафедры МИТ ХНУРЭ. Научные интересы: динамические измерения. Адрес: Украина, 61166, Харьков, пр. Ленина,14, тел. 702-1331.