

УДК 681.518

В.Ф. МИРГОРОД, Г.С. РАНЧЕНКО

ОАО «Элемент», Одесса, Украина

ВЕРОЯТНОСТНЫЕ ХАРАКТЕРИСТИКИ ИНТЕРВАЛЬНЫХ И ТРЕНДОВЫХ СТАТИСТИК ПРИ НЕГАУССОВЫХ РАСПРЕДЕЛЕНИЯХ ОШИБОК ИЗМЕРЕНИЯ ПАРАМЕТРОВ ГТД

Рассматриваются вероятностные характеристики статистик интервальных оценок и трендовых статистик при анализе параметров регистрации ГТД системах технической диагностики.

техническая диагностика, распределение ошибок, трендовый анализ

Постановка проблемы и цель исследования

Проблемным вопросом повышения эффективности систем технической диагностики (СТД), имеющим в настоящее время высокий уровень технического, алгоритмического и программного совершенства, является соответствие принятых статистических моделей (СМ) формирования данных фактически реализуемым. Актуальной прикладной задачей диагностики полагается установление влияния различий указанных моделей на вероятностные характеристики статистик интервальных оценок и трендовых статистик при анализе реальных временных рядов параметров регистрации ГТД.

Известный теоретически обоснованный и экспериментально подтвержденный факт высокой чувствительности выборочных оценок к наличию в данных существенных отклонений (выбросов) обусловил возникновение в прикладной статистике [1] самостоятельного направления – «робастной» статистики [2, 3], в рамках которого разработаны эффективные алгоритмы цензурирования данных и состоятельных оценок. В то же время в опубликованных работах уделяется недостаточное внимание исследованию влияния отличий принятой СМ (обычно это принадлежность выборки к генеральной совокупности некоррелированных нормально распределенных случайных величин) от фактически реализуемой. Негауссовость распределения погрешностей различных датчиков и измерительных каналов

(ИК) [4], а также результатов обработки косвенных измерений обуславливает необходимость проведения исследований надежности применяемых статистик для выборок указанного характера, особенно в условиях малого объема выборок.

Целью настоящей работы является оценка смещения уровней значимости и значений доверительных вероятностей статистик интервальных оценок и некоторых трендовых статистик в зависимости от эксцесса априорного распределения выборок данных параметров регистрации ГТД.

Основные результаты

Принятая при исследованиях априорная СМ порождения данных выбрана в виде широко применяемой при аппроксимации негауссовских плотностей вероятности (ПВ) модели разложения в ряд Грама-Шарлье [5, 6], первые члены которого позволяют учесть в удобной форме моменты до четвертого порядка включительно, обычно используемые при анализе данных

$$f_{na}(x) = f_n(x) \left[1 + \frac{1}{3!} j_1 (x^3 - 3x) + \frac{1}{4!} j_2 (x^4 - 6x^2 + 3) \right], \quad (1)$$

где $f_{na}(x)$ – ПВ нормального распределения;

$j_1 = \mu_3 / \sigma^3$, $j_2 = \mu_4 / \sigma^4 - 3$ – коэффициенты асимметрии и эксцесса соответственно.

В рамках иллюстрируемых исследований анализировалось влияние эксцесса распределения, поэтому в последующем полагается $j_1 = 0$. Что касается допустимости аппроксимации реальных ПВ в виде (1), то целесообразно принимать $j_2 = \varepsilon \leq 1$, поскольку при больших значениях эксцесса нарушается условие неотрицательности ПВ вида (1).

Диапазон возможных СМ типа (1) иллюстрирует рис. 1, где также изображена для сравнения гауссовская ПВ.

Выбранный путь достижения цели работы заключается в аналитическом определении вида основных статистик интервальных оценок и трендовых статистик для СМ порождения данных (1) в зависимости от эксцесса ε и объема выборки N . Интервальные оценки среднего определяются $t(x, N)$ распределением Стьюдента, дисперсии – $\chi^2(x, N)$ распределением, статистики дисперсионного анализа – $F(x, N_1, N_2)$ распределением Фишера. Центральным вопросом является определение вида статистики χ^2 для распределения (1), поскольку статистики Стьюдента и Фишера основаны на статистике χ^2 . Определение вида искомой статистики $\chi_a^2(x, N, \varepsilon)$ предполагает получение ПВ квадрата СВ, распределенной по закону (1) с последующей N -мерной последовательной сверткой ПВ. Для выпол-

нения указанной в аналитическом виде использован пакет расширения Symbolic Toolbox системы MATLAB, содержащей ядро символьной математики MAPL. Свертка определялась переходом в область изображений ПВ, возведением изображения в N -ю степень и обратным переходом в область оригиналов для получения искомой плотности $\chi_a^2(x, N, \varepsilon)$.

Контроль корректности символьных преобразований осуществлялся одновременным исполнением указанной процедуры для нормального распределения. Зависимость статистики χ^2 от эксцесса распределения в виде (1) иллюстрирует рис. 2.

Как это следует из выполненного анализа и приведенных иллюстраций, даже весьма умеренные отличия эксцесса от его значения для нормального распределения существенно смещают квантили распределения χ^2 , что имеет важное значение для установления интервальных оценок дисперсии на заданном уровне значимости. Если интервальные оценки установлены исходя из СМ о нормальном распределении данных, фактическая принадлежность выборки к СМ вида (1) приводит к смещению уровня значимости, что иллюстрирует рис. 3.

Связанные с χ^2 статистики Стьюдента и Фишера получены соответствующими нелинейными преобразованиями [6].

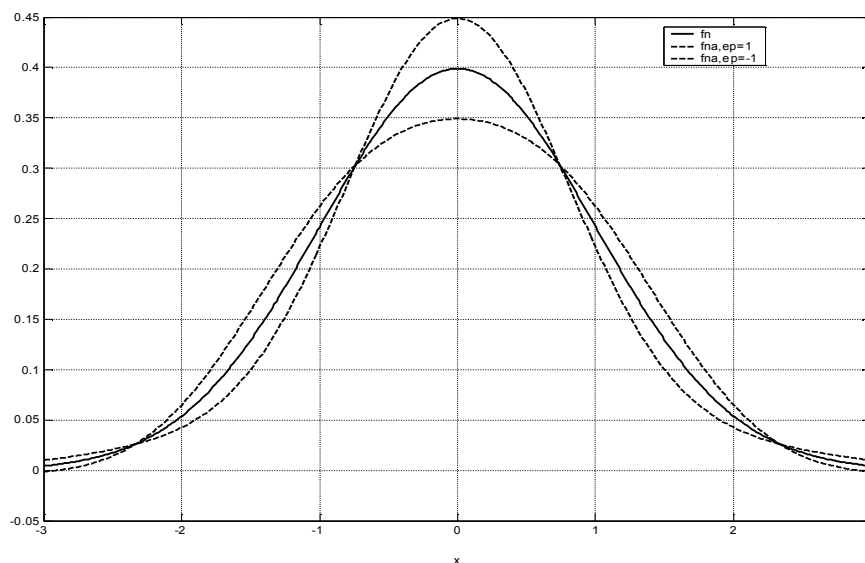


Рис. 1. Априорная плотность вероятности

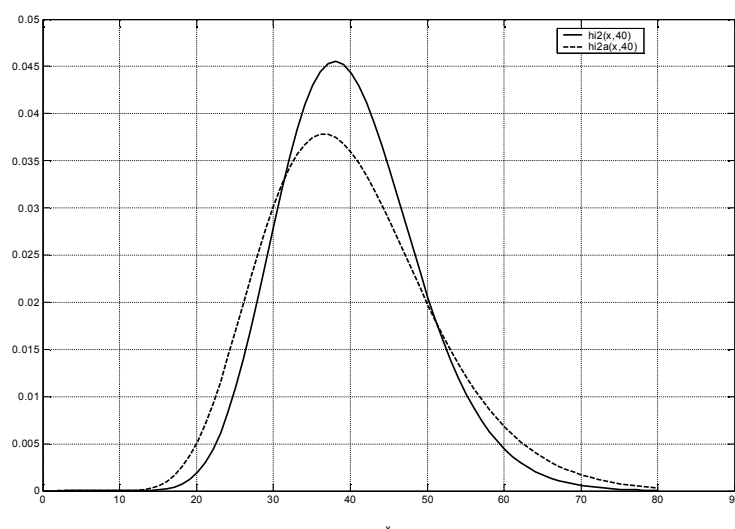
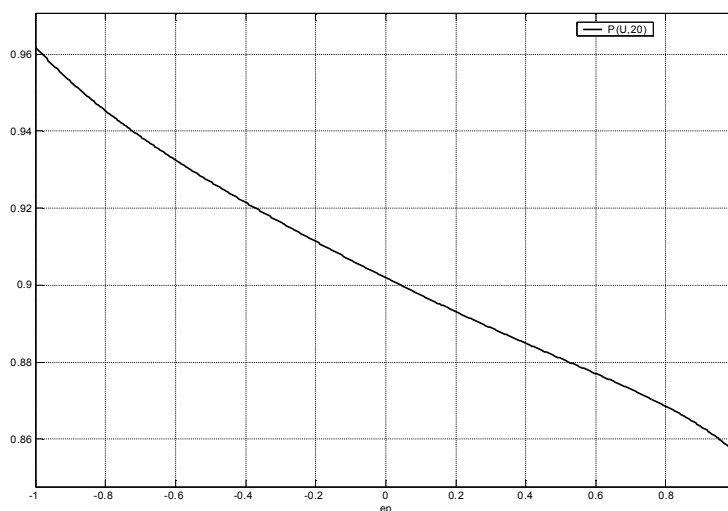
Рис. 2. Зависимость статистики χ^2 от эксцесса

Рис. 3. Зависимость доверительной вероятности от эксцесса

Для статистики Стьюдента соответственно как отношение СВ типа (1) и СВ с распределением $\chi_a^2(x, N, \varepsilon)$, а для статистики Фишера – как отношение двух СВ с распределениями $\chi_a^2(x, N_1, \varepsilon)$ и $\chi_a^2(x, N_2, \varepsilon)$. Детальный анализ полученных результатов показывает, что указанные статистики испытывают существенную деформацию по сравнению с каноническими, если СМ порождения данных имеет вид (1). Тем не менее, статистика Стьюдента $t_a(x, N, \varepsilon)$ остается симметричной, как этого следовало ожидать, а смещение ее квантилей в диапазоне $(0,9 \dots 0,95)$, представляющим интерес для практики, можно считать несущественным, так как уровни значимости изменяются на доли процента.

Следовательно, в рамках ограничений СМ (1), интервальные оценки и тесты, в которых используется статистика Стьюдента (оценка среднего, равенства средних в выборках и другие) не претерпевают существенных изменений. Для статистики Фишера, напротив, смещение квантилей является статистически значимым, особенно при значительном различии степеней свободы N_1 и N_2 . Поэтому тесты, основанные на статистике Фишера и ее модификациях (v -распределение, z -распределение) для выборок типа (1) могут сопровождаться потерей уровней значимости, достигающими 5% и более. В первую очередь это касается оценок равенства дисперсий и равенства средних при заданной дисперсии [5].

Дополнительные исследования проведены для ря-

да трендовых статистик, применяемых в СТД газотурбинных двигателей [7], и других областях. Выполнен анализ чувствительности к негауссовскому распределению временных рядов алгоритмов коммулятивных сумм (АКС), статистики Хальда-Аббе и критерия Фишера. Симметричность исходного распределения обуславливает, как и следовало ожидать, высокую устойчивость АКС к эксцессу ПВ модели порождения данных (1). Для анализа свойств статистики Хальда-Аббе использован установленный [8] факт ее статической эквивалентности r -критерию некоррелированности СВ и известная [5] связь указанного критерия со статистикой Стьюдента в виде

$$t = \sqrt{N-2} \cdot \frac{r}{\sqrt{1-r^2}}.$$

Вычисления пороговых уровней для трендовой статистики Хальда-Аббе по полученным аналитическим выражениям $r(x, N, \varepsilon)$ согласно модели (1) показало высокую степень устойчивости исследуемой статистики к эксцессу распределения в области доверительных вероятностей (0,9... 0,95) и числа степеней свободы (числа отсчетов ряда) $N > 20$. Существенные различия возникают на уровнях (0,6... 0,8), однако эта область не представляет практического интереса.

Трендовая статистика, основанная на критерии Фишера, обычно задается в виде $F(x, N_1, N)$, где N_1 – фиксированное число, например, $N_1=19$ а N возрастает с ростом числа отсчетов ряда. Анализ поведения указанной статистики для СВ вида (1) показывает, что для $N \leq N_1$ наблюдается значительное смещение пороговых уровней и потеря значимости, однако при $N > N_1$ потеря уровней значимости не превышает 1%, что для обычно применяемых доверительных вероятностей не представляется значимым.

Отличительной особенностью проведенного исследования является тот факт, что все интервальные и трендовые статистики для СВ порождения данных (1) получены с помощью упомянутого пакета символьной математики в аналитическом виде, что позволяет выполнить их детальный анализ, необходимые расчеты и табулировать. Громоздкость полученных аналитических выражений не позволяет их привести в рамках данной статьи.

Заключение

Отличие принятой априорной СВ порождения данных от фактически реализуемой приводит, в общем случае, к существенному смещению уровней значимости статистик интервальных оценок и трендовых статистик и, как следствие, снижению достоверности статистических выводов о состоянии объектов в СТД. Поэтому формирование адекватных диагностических моделей поведения контролируемых объектов в частности, ГТД, является необходимым и важным этапом построения современных СТД что и определяет перспективы дальнейших исследований.

Литература

1. Айвазян С.А., Енюков И.С. Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. – М.: Финансы и статистика, 1985. – 488 с.
2. Устойчивые статистические методы оценки данных / Пер. с англ. Ю.И.Малахова; Под ред. Н.Г. Волкова. – М.: Машиностроение, 1984. – 232 с.
3. Питмен Э. Основы теории статистических выводов. – М.: Мир, 1986. – 104 с.
4. Новицкий Л.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат, 1985. – 248 с.
5. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. – М.: Наука, 1973. – 832 с.
6. Тихонов В.И. Статистическая радиотехника. – М.: Радио и связь, 1982. – 624 с.
7. Синтез систем управления и диагностирования газотурбинных двигателей / С.В. Епифанов, Б.И. Кузнецов, И.И. Богаенко и др. – К.: Техніка, 1998. – 312 с.
8. Миргород В.Ф., Ранченко Г.С. Особенности применения трендовых статистик при обработке данных в системах технической диагностики // Технология и конструирование в электронной аппаратуре. – 2005. – № 4. – С. 25-27.

Поступила в редакцию 12.06.2006

Рецензент: д-р техн. наук, проф. С.А. Нестеренко, Институт компьютерных систем ОНПУ, Одесса.

