УДК 66.0451.091.001.57 : 681.3.06 Х И М И Ч Е С К А Я Т Е Х Н О Л О Г И Я

Академик В.В. КАФАРОВ, В.П. МЕШАЛКИН, Г.И. МАНКО

ИНФОРМАЦИОННЫЙ КРИТЕРИЙ ПРОВЕРКИ ГИПОТЕЗ О ЗАКОНАХ РАСПРЕДЕЛЕНИЯ ХАРАКТЕРИСТИК НАДЕЖНОСТИ ХИМИКО-ТЕХНОЛОГИЧЕСКИХ СИСТЕМ

Одним из этапов системного анализа надежности химико-технологических систем (х.т.с.) является идентификация законов распределения экспериментально полученных значений характеристик надежности ('). При этом производится аппроксимация эмпирических распределений типовыми распределениями (экспоненциальным, нормальным, Релея, Вейбулла и т.п.) и их комбинациями. Проверка гипотез о законах распределения производится обычно известными методами математической статистики по критериям согласия, среди которых наиболее употребительными являются критерий χ^2 и критерий Колмогорова.

Известные критерии согласия основаны на использовании некоторых статистик, представляющих собой функции от выборочных точек (т.е. экспериментально полученных значений характеристик надежности) $x_1, x_2, \ldots, x_n$ и параметров аппроксимирующего закона распределения $\theta_1, \theta_2, \ldots$

$$(1) \qquad y = y(x_1, x_2, \ldots, x_n, \theta_1, \theta_2, \ldots).$$

Статистику y подбирают таким образом, чтобы ее значения измеряли отклонения или отношения некоторых параметров выборки и аппроксимирующего распределения. Вид функциональной зависимости (1) определяет достоинства и недостатки соответствующего критерия.

Так, наличие в статистике критерия χ^2 [2] квадратичной формы определяет существенный недостаток критерия, вызывающий требование осторожности в выводах о верности проверяемой гипотезы. В [3] рассмотрены случаи, когда проверка по критерию χ^2 приводит к отбрасыванию заведомо верной гипотезы. Критерий Колмогорова [2] неприменим, когда параметры гипотетического закона распределения оцениваются по той же выборке, по которой рассчитывается статистика критерия [4]. Кроме того, этот критерий применяется только для непрерывных функций распределения.

В настоящей работе впервые сделана попытка выработать общий формализованный подход к выбору статистики (1), позволяющий разрабатывать новые более совершенные критерии согласия. В основе предлагаемого подхода лежит использование информационных понятий и теорий, в частности понятий неупорядоченности и неорганизованности, введенных в [5] для целей исследования управляющих систем и развитых нами применительно к задачам проверки статистических гипотез.

Определим неупорядоченность характеристики надежности х.т.с. как меру отклонения значений характеристики от тех ее значений, которые обусловливаются гипотетическим законом распределения. Примерами такой меры отклонения является модуль разности реальных и гипотетических вероятностей значений характеристики надежности, а также модуль разности интегральных и дифференциальных функций реального и гипотетического законов распределения. В качестве меры отклонения могут выступать квадрат разности между, реальными и гипотетическими значениями характеристики надежности, отношение реальных вероятностей к гипотетическим и т.д.

Неорганизованностью характеристики надежности х.т.с. будем называть обобщенную за рассматриваемое число возможных ситуаций характеристику неупорядоченности, взвешенную, во-первых, по вероятности p_j j-й ситуации, а во-вторых, по существенности проявления неупорядоченности, которая может определяться выбором некоторой функции ψ от параметра неупорядоченности $\Pi_{\bar{y}j}$. Отсюда неорганизованность определяем выражением неупорядоченности

$$(2) \qquad \bar{O} = \sum_{j=1}^{m} p_j \psi(\Pi_{\bar{y}j}).$$

Как показано в [5], для большинства задач функцию ψ достаточно представить одной из четырех зависимостей: линейной, степенной, логарифмической или экспоненциальной. Соответственно в [5] предлагаются 4 основные выражения для оценки неорганизованности:

$$(3) \qquad \bar{O}_л = k \sum_{j=1}^{m} p_j \Pi_{\bar{y}};$$

(4) $\quad \bar{O}_{c} = k \sum_{j=1}^{m} p_j (\Pi_{\bar{y}})^l;$

(5) $\quad \bar{O}_{\text{лог}} = k \sum_{j=1}^{m} p_j \log \Pi_{\bar{y}j};$

(6) $\quad \bar{O}_{\text{э}} = k \sum_{j=1}^{m} p_j [\exp(\Pi_{\bar{y}j}) - C],$

где $C,\ \kappa,\ l$ - заданные постоянные величины. Очевидно, что $\sum_j p_j = 1.$

Используем неорганизованность в качестве критерия проверки статистических гипотез. Приняв в (3) в качестве параметра неупорядоченности максимум абсолютной величины разности реальной $\Phi(\xi)$ и гипотетической $\widehat{\Phi}(\xi)$ функций распределения характеристики надежности ξ, получим при $\kappa = 1$ выражение (2) критерия Колмогорова. Положив в (4) значение параметра неупорядоченности равным отношению квадрата разности реальных и гипотетических частот попадания значений ξ в j-й интервал диапазона изменения величины ξ к гипотетической частоте и приняв все $p_j = 1/m$, будем иметь при $\kappa = m,\ l = 2$ выражение для критерия χ^2 (1).

Остановимся подробнее на случае логарифмической функции ψ. Примем за параметр неупорядоченности отношение реальных вероятностей распределения (p_j) пк гипотетическим (q_j): $\Pi\bar{y}j = p_j / q_j$. Тогда выражение для оценки неорганизованности примет следующий вид (при $\kappa = 1$) :

(7) $\quad \bar{O}_{\text{лог}} = \sum_{j=1}^{m} p_j \log\left(\frac{p_j}{q_j}\right).$

Целесообразность использования выражения (7) в качестве критерия проверки статистических гипотез становится очевидной, если осмыслить это выражение с информационных позиций. По Бонгарду ([6]) неопределенность задачи с распределением вероятностей ответа $P=\{pj\}$ для наблюдателя, исходящего из гипотезы, что распределение равно $Q = \{qj\}$, оценивается следующим выражением:

(8) $\quad N\left(\frac{p}{q}\right) = -\sum_{j=1}^{m} p_j \log q_j.$

Всякое сообщение, изменяющее значение неопределенности от N_0 до N_1, несет количество полезной информации, равное

(9) $\quad I_{\text{п}} = N_0 - N_1,$

причем величина эта может быть как положительной, так и отрицательной. В последнем случае говорят о дезинформации, вносимой сообщением.

Неопределенность Бонгарда минимальна, когда гипотетическое распределение совпадает с реальным:

(10) $\quad \min_{\{Q\}} N\left(\frac{p}{q}\right) = -\sum_{j=1}^{m} p_j \log p_j.$

Пусть P – распределение вероятностей значений характеристики надежности х.т.с, а Q – аппроксимирующее гипотетическое распределение. В этом случае принятие гипотезы Q создает неопределенность $N(p/q)$, тем большую, чем больше Q отличается от P. Следовательно, принятие гипотезы Q приносит следующее количество полезной информации:

(11) $\quad I_{\text{п}} = N\left(\frac{p}{p}\right) - N\left(\frac{p}{q}\right).$

Поскольку $N\left(\frac{p}{p}\right) \le N\left(\frac{p}{p}\right)$, следует говорить о дезинформации, содержащейся в проверяемой гипотезе:

(12) $\quad D = -I_{\text{п}} = N\left(\frac{p}{q}\right) - N\left(\frac{p}{p}\right) = \sum_{j=1}^{m} p_j \log\left(\frac{p_j}{q_j}\right).$

Таким образом, выражение (7) имеет вполне определенный информационный смысл: оно определяет количество дезинформации, вносимой принятием проверяемой гипотезы.

Для практического применения информационного критерия проверки гипотез о законах распределения характеристик надежности х.т.с. предлагаем использовать следующую статистику:

$$(13)\quad \delta=\frac{1}{n}\sum_{j=1}^{m}\nu_j\ln\frac{\nu_j}{n\hat{p}_j},$$

где n — объем выборки, ν_j – частота результатов наблюдении характеристики надежности ξ в j-м интервале диапазона наблюдения величины ξ, $\hat{p}_j$ — вероятность попадания величины ξ, в j-й интервал, определяемая аппроксимирующим законом распределения.

Информационный критерий (13), обладая всеми достоинствами критерия χ^2 в части универсальности и удобства практического использования, не имеет, однако, указанных выше недостатков традиционно используемых критериев. Кроме того, информационный критерий обладает еще двумя важными преимуществами.

Во-первых, при подсчете информационного критерия согласия значения параметра неупорядоченности взвешиваются по их реальным вероятностям, в то время как в случае критерия χ^2 все эти значения принимаются равновероятными.

Во-вторых, обычно используемые критерии согласия позволяют лишь установить, имеются ли основания отвергнуть проверяемую гипотезу. Информационный же критерий дает количественную оценку объема дезинформации, вносимой принятием гипотезы, и позволяет тем самым сравнивать аппроксимирующие распределения по степени их адекватности экспериментальным данным.

Рассмотрим пример использования информационного критерия δ. В табл. 1 приведены данные о наработке на отказ некоторой х.т.с, полученные при исследовании ее надежности. Экспериментально полученное распределение числа отказов х.т.с. было аппроксимировано гамма-распределением с параметрами $\lambda = 0{,}383$, $r = 2$, распределениеv Релея с $\sigma = 4{,}17$, а также асимметричным треугольным с $t_н = 0$, $t_м = 166$, $t_к = 1400$ (условные обозначения для параметров распределений соответствуют принятым в ([7])).

Таблица 1

Данные о распределениях отказов исследуемой х.т.с.

№№ п.п.	Δt, ч	Эксперим. распредел.	Аппроксимирующее распредел.		
			Г-распр.	Релея	треугольн.
1	**0-200**	**15**	**18,22**	**11,31**	**16,65**
2	**200-400**	**25**	**26,95**	**25,56**	**25,47**
3	**400-600**	**24**	**21,74**	**27,79**	**20,84**
4	**600-800**	**17**	**14,02**	**19,46**	**16,21**
5	**800-1000**	**11**	**8,59**	**10,27**	**11,57**
6	**1000-1200**	**5**	**4,83**	**4,03**	**6,94**
7	**1200-∞**	**3**	**5,65**	**1,58**	**2,32**

*

Проверка гипотез о перечисленных законах распределения методом критерия χ^2 не дала оснований отвергнуть ни одну из проверяемых гипотез. С целью выбора закона, наиболее согласуемого с экспериментальным, для каждого из аппроксимирующих распределений по формуле (13) подсчитано значение критерия δ. Ниже приведены значения информационного критерия ($\delta \cdot 100$) согласия для аппроксимирующих распределений:

Гамма-распределение	**18,62**
Распределение Рэлея	**19,09**
Треугольное распределение	**9,29**

Из этих данных видно, что наименьшее количество дезинформации вносит принятие гипотезы о треугольном распределении.

Московский химико-технологический
институт им. Д.И. Менделеева

Поступило
16 V 1979

ЛИТЕРАТУРА

[1] *В.В. Кафаров, В.Л. Петров и др.,* ДАН, т. 219, № 3, 675 (1974). [2] *Б.Р. Левин,* Теория надежности радиотехнических систем (математические основы). М., "Сов. радио", 1978. [3] *Дж.Э. Юл., М. Кендэл.* Теория статистики, М., Госстатиздат, 1960. [4] *H.W. Lilliefors,* J. Am. Statist. Assoc., v. 62, 399 (1967). [5] *Ю.М. Горский,* Информационные аспекты управления и моделирования, М., "Наука", 1978. * *М.М. Бонгард,* Проблема узнавания, М.,' "Наука", 1967. [7] *Г.В. Дружинин,* Надежность автоматизированных систем, М., 'Энергия", 1977.