

Название файла	Длина файла (в байтах)	Предлагаемый метод (%)	Стандартный архиватор ARJ 2.30 (%)
B1.bas	3154	40.4	47.6
B2.bas	1995	38.8	46.7
B3.bas	1503	34.8	48.9
B4.bas	1378	35.9	47.6
B5.bas	1107	38.3	51.8
P1.pas	2604	42.6	45.0
P2.pas	1740	38.3	43.5
P3.pas	1600	38.4	52.5
P4.pas	1560	38.1	52.8
P5.pas	1103	41.4	48.3

Таблица 1. Результаты сжатия текстов программ.

СТАТИСТИЧЕСКИЙ АНАЛИЗ СМЕСЕЙ РАСПРЕДЕЛЕНИЙ ПО ГРУППИРОВАННЫМ ДАННЫМ

Б. Ю. Лемешко, С. Н. Постовалов
НГТУ, Россия.

Рассматриваемая программная система обеспечивает решение таких нетривиальных задач как выбор закона распределения или смесей законов наиболее хорошо описывающих выборочные данные. Работа является дальнейшим развитием программной системы "Оценивание параметров распределений" [1]. Здесь реализованы возможности, которые позволяют находить оценки максимального правдоподобия параметров для смесей из любых 26 наиболее часто используемых в приложениях распределений: экспоненциального, полунормального, Рэлея, Максвелла, модуля многомерного нормального вектора, Парето, Эрланга, Лапласа, нормального, логарифмически нормальных ($\ln$ и $\lg$), Коши, Вейбулла, Накагами, распределения минимального значения, распределения максимального значения, двойного показательного, гамма-распределения, логистического, бета-распределения 1-го рода, стандартного бета-распределения 2-го рода, бета-распределения 2-го рода, распределений

Sb-Джонсона, Sl-Джонсона и Su-Джонсона, семейства экспоненциальных распределений.

Реализация возможностей обработки выборок, состоящих из смесей случайных величин, позволяет существенно расширить область реального использования программной системы.

Функция распределения для смеси имеет вид

$$F(x) = \sum_{i=1}^S w_i F_i(x, \theta_i),$$

где s — число распределений в смеси, w_i — параметры смеси, F_i — 1-я функция распределения, θ_i — вектор ее параметров.

При анализе осуществляется проверка гипотез о согласии по ряду критериев: χ^2 — Пирсона, отношения правдоподобия, Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса. Это дает возможность принимать решения по их совокупности.

По ряду возможностей система не имеет аналогов среди программного обеспечения задач статистического анализа. Во-первых, исходная выборка может быть негруппированной, группированной или частично группированной. Во-вторых, при проверке гипотез по критериям χ^2 — Пирсона или отношения правдоподобия используются полученные таблицы асимптотически оптимального группирования данных, обеспечивающие максимальную мощность критерия при близких альтернативных гипотезах. В-третьих, группирование исходной выборки с последующим оцениванием параметров распределения позволяет получать робастные оценки, менее чувствительные к аномальным результатам наблюдений. В-четвертых, применение системы при обработке результатов дает возможность использовать оптимальное группирование при проведении экспериментов и регистрации их результатов, что позволяет резко сократить объем хранимых и передаваемых по каналам связи данных без существенной потери информации о законе распределения наблюдаемой случайной величины.

В качестве примера на рис. 1 приведены результаты статистического анализа негруппированной выборки (оценки параметров и результаты проверки гипотез о согласии по различным критериям), полученной объединением двух выборок: первая распределена по экспоненциальному закону распределения, вторая — по нормальному закону; каждая объемом по 100 наблюдений (t_0 — параметр сдвига, t_1 — параметр масштаба).

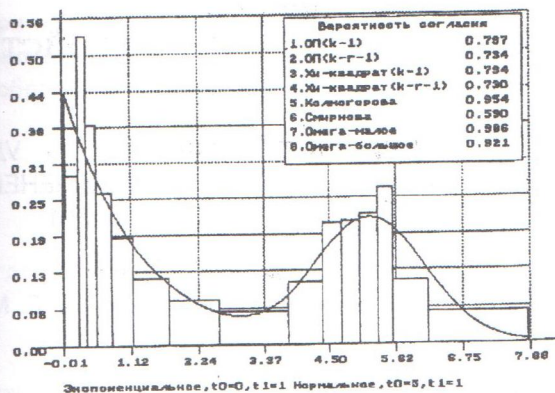


Рис.1 Оценивание параметров смеси распределений

ЛИТЕРАТУРА

1. Денисов В. И., Лемешко Б. Ю., Цой Е. Б. Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов. Новосибирск, 1993, 347 с.