

ИССЛЕДОВАНИЕ РАСПРЕДЕЛЕНИЙ СТАТИСТИК И МОЩНОСТИ КРИТЕРИЕВ СОГЛАСИЯ ПО ЦЕНЗУРИРОВАННЫМ ВЫБОРКАМ

Чимитова Е.В., Лемешко Б.Ю.

НГТУ, г. Новосибирск

E-mail: chim@mail.ru

При исследовании случайных величин типа “времени жизни”, например в задачах теории надежности, в медицинских или биологических исследованиях, нередко возникает задача обработки цензурированных выборок. Выборка называется односторонне цензурированной, если область определения случайной величины разбита на два интервала, в одном из которых известны индивидуальные наблюдения (обозначим данный интервал через M), а во втором – известно лишь число наблюдений, попавших в этот интервал. Для проверки согласия эмпирического распределения $F_n(x)$ с теоретическим $F(x)$ по цензурированным данным можно использовать критерии типа Реньи, Колмогорова, критерий Мозеса, а также критерии типа ω^2 Мизеса и Ω^2 Андерсона-Дарлинга.

Целью данной работы является исследование методами компьютерного моделирования распределений статистик рассматриваемых критериев в зависимости от объемов выборок и степени цензурирования, а также сравнение данных критериев по мощности.

Статистика двустороннего критерия **Реньи** $S_R = \sup_{x \in M} \frac{|F_n(x) - F(x)|}{F(x)}$ в случае

проверки простых гипотез о согласии в пределе при $n \rightarrow \infty$ подчиняется закону

распределения
$$L(S) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp \left\{ -\frac{(2k+1)^2 \pi^2}{8S^2} \right\} \quad [1].$$
 Результаты

статистического моделирования показали, что распределения статистики типа Реньи существенно зависят от степени цензурирования. Достаточно хорошее согласие эмпирических распределений статистики Реньи с предельным законом $L(S)$ достигается только при $n > 1000$ и степени цензурирования, близкой к 0.5. Данный результат говорит о том, что при значительной степени цензурирования и объеме выборки меньше 1000 (именно такая ситуация чаще всего имеет место на практике) применение данного критерия может привести к неверным выводам.

В случае проверки простых гипотез предельные распределения статистики типа **Колмогорова** $S_K^c = \sup_{x \in M} |F_n(x, \theta) - F(x, \theta)|$ найдены [2] в зависимости от величины $0 < a < 1$, характеризующей степень цензурирования. Методами компьютерного моделирования показано, что уже при $15 \div 20$ зарегистрированных наблюдениях в выборке достигается достаточно хорошее

согласие распределения статистики Колмогорова с соответствующим предельным законом. Исследованы распределения статистики Колмогорова при проверке сложных гипотез с использованием ОМП по цензурированным выборкам. Для ряда законов распределения, соответствующих проверяемой гипотезе, методами компьютерного моделирования и анализа статистических закономерностей построены аппроксимации предельных распределений статистики Колмогорова.

Статистика **Мозеса** [3] в случае цензурирования справа I типа имеет вид:

$$S_M = \frac{1}{r \cdot F(x_0)} \cdot \sum_{i=1}^{r(x_0)} F(x_i), \quad \text{при цензурировании II типа} -$$

$$S_M = \frac{2 \cdot (n+1)}{[2 \cdot (n+1) \cdot F(x_r) + 1] \cdot r} \cdot \sum_{i=1}^r F(x_i), \quad \text{где } r - \text{число наблюдаемых}$$

(нецензурированных) элементов выборки. Статистика S_M имеет в пределе при $n \rightarrow \infty$ нормальное распределение со средним 0.5 и дисперсией $(12 \cdot r)^{-1}$. Показано, что при I типе цензурирования распределение статистики с ростом n гораздо быстрее сходится к теоретическому закону, чем при II типе цензурирования.

В [4] приведены верхние процентные точки предельных распределений статистик **Крамера-Мизеса-Смирнова** $\omega_n^2 = \int_M (F(x) - F_n(x))^2 dF(x)$ и

$$\text{Андерсона-Дарлинга} \quad \Omega_n^2 = \int_M (F(x) - F_n(x))^2 \frac{1}{F(x)(1-F(x))} dF(x). \quad \text{В данной}$$

работе с использованием методов статистического моделирования исследовались эмпирические распределения данных статистик в зависимости от типа и степени цензурирования. Показано, что характер сходимости распределений статистик типа ω^2 , Ω^2 и статистики типа Колмогорова к соответствующим предельным законам аналогичен.

Среди рассмотренных критериев наиболее предпочтительным по мощности является критерий Мозеса. Но в то же время следует учитывать, что данный вывод не является очень строгим. Иногда критерий имеет преимущества при одних значениях степени цензурирования a и объемах выборок n и уступает при других значениях a и n .

Литература

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1983. - 416 с.
2. Мания Г.М. Статистическое оценивание распределений. - Тбилиси: Изд-во ТГУ, 1974. - 237 с.
3. Приходько Ю.Г. Проверка согласия по цензурированным выборкам // Заводская лаборатория. Диагностика материалов. 1982, № 10. - С. 56-59.

4. Pettitt A.N., Stephens M.A. Modified Cramer von Mises statistics for censored data // Biometrika, 1976. V. 63. N. 2.