Мат. VI международной конференции "Актуальные проблемы электронного приборостроения" АПЭП-2002. Новосибирск, 2002. - Т.6. - С.39-42.

О РАСПРЕДЕЛЕНИЯХ СТАТИСТИК КРИТЕРИЯ РЕНЬИ[1]

Б.Ю. Лемешко, Е.В. Чимитова

Новосибирский государственный технический университет
Новосибирск, Россия. E-mail: headrd@ fpm.ami.nstu.ru

Критерии Реньи [1] используются для проверки согласия эмпирического распределения $F_n(x)$ с теоретическим $F(x)$ по цензурированным данным. Выборка называется цензурированной (слева или справа), если область определения случайной величины разбита на два интервала, в одном из которых известны индивидуальные наблюдения, а во втором – известно лишь число наблюдений, попавших в этот интервал. Особенно часто с цензурированными выборками сталкиваются при статистическом анализе данных типа времени жизни, когда эксперимент проходит в условиях ограниченности по времени (цензурирование I типа) или по количеству наблюдений (цензурирование II типа).

Статистики критерия Реньи в случае цензурирования слева ($a \le F(x) \le 1$, где a - степень цензурирования, $0 \le a < 1$) задаются следующими выражениями [1]:

$$R_n^+(a,1) = \sup_{F(x)\ge a} \frac{F_n(x) - F(x)}{F(x)},\ R_n^-(a,1) = \sup_{F(x)\ge a} \frac{F(x) - F_n(x)}{F(x)},\ R_n(a,1) = \max\left\{R_n^+(a,1), R_n^-(a,1)\right\},$$

и в случае цензурирования справа ($0 \le F(x) \le 1-a$): $R_n^+(0,1-a) = \sup\limits_{F(x)\le 1-a} \dfrac{F_n(x) - F(x)}{1 - F(x)}$,

$$R_n^-(0,1-a) = \sup_{F(x)\le 1-a} \frac{F(x) - F_n(x)}{1 - F(x)},\ R_n(0,1-a) = \max\left\{R_n^+(0,1-a), R_n^-(0,1-a)\right\}.$$

Случайные величины $R_n^+(a,1)$, $R_n^-(a,1)$, $R_n^+(0,1-a)$ и $R_n^-(0,1-a)$ распределены одинаково (аналогичное утверждение справедливо и для статистик $R_n(0,1-a)$ и $R_n(a,1)$) и, как показал Реньи, имеют место предельные соотношения:

$$\lim_{n\to\infty} P\left\{\sqrt{\frac{na}{1-a}} R_n^+(a,1) < x\right\} = 2\Phi(x) - 1, \quad \lim_{n\to\infty} P\left\{\sqrt{\frac{na}{1-a}} R_n(a,1) < x\right\} = L(x),\ x > 0,$$

где $\Phi(x)$ – функция стандартного нормального распределения,

$$L(x) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left\{-\frac{(2k+1)^2 \pi^2}{8x^2}\right\}.$$

При малых объемах выборок n можно воспользоваться точной формулой [2]:

$$P\left\{R_n^+(a,1) < x\right\} = L_1(x,n,a) = 1 - \frac{x}{1+x} \sum_{k=0}^{S} C_n^k \left(\frac{x + k/n}{1+x}\right)^{k-1} \left(\frac{1 - k/n}{1+x}\right)^{n-k},$$

где $\dfrac{x}{1+x} \le 1-a$, $S = n - [n(1+x)a] - 1$, если $n(1+x)a$ – нецелое число и $S = n - n(1+x)a$, если $n(1+x)a$ – целое (цензурирование слева). Статистика $R_n^-(0,1-a)$ (при цензурировании справа) имеет то же распределение с заменой a на $1-a$ в выражении для S.

Целью данной работы явилось, во-первых, исследование скорости сходимости распределений статистик Реньи к соответствующим предельным законам, во-вторых, исследование распределений статистик $R_n^-(0,1-a)$ и $R_n^+(a,1)$ при малых объемах выборок в зависимости от степени цензурирования.

[1] Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 00-01-00913)

Для этого эмпирические распределения статистик Реньи моделировались при различных проверяемых гипотезах. На рис. 1 приведены эмпирические распределения статистики $R_n(a,1)$ для случая проверки простой гипотезы о согласии с экспоненциальным законом с параметром масштаба $\sigma = 1$ по цензурированным слева выборкам, с вероятностью попадания в наблюдаемую область равной 10% (I тип цензурирования). Из рисунка видно, что при такой степени цензурирования близкими к предельному оказываются распределения статистики при полных значениях объема выборки $n > 1000$. При меньших n распределения статистики существенно отличаются от предельного. Кроме того, исследования показали, что при цензурировании II типа распределения статистики $R_n(a,1)$ еще хуже согласуются с предельным распределением $L(x)$.

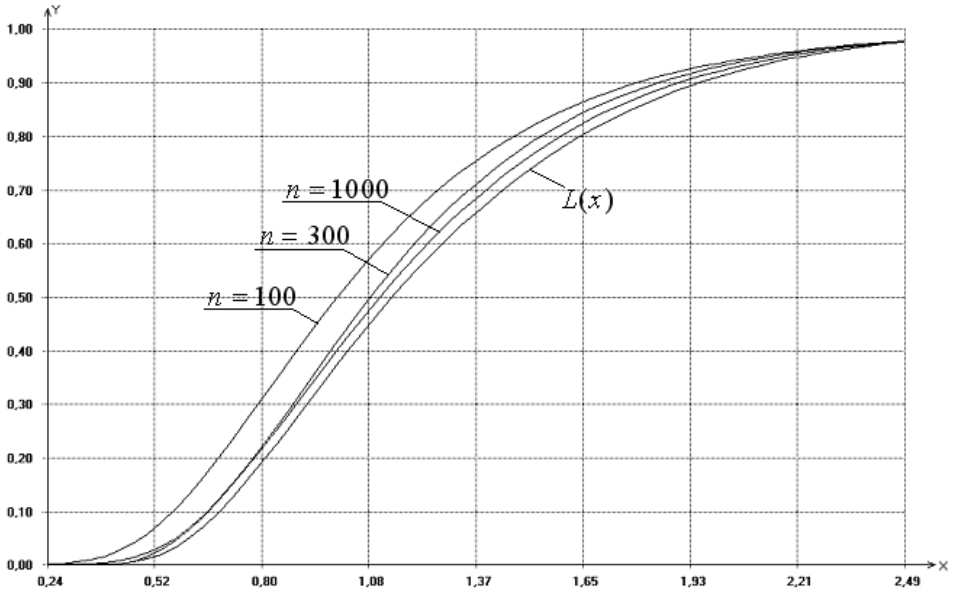

Рис. 1. Распределения статистики $R_n(a,1)$ при 10% наблюдаемой области, цензурирование слева I типа

Рис. 2 иллюстрирует зависимость распределений статистики $R_n(a,1)$ от степени цензурирования. Из рисунка видно, что наилучшее согласие с предельным распределением $L(x)$ достигается при 50% наблюдаемой области определения случайной величины, а при малой или, наоборот, высокой степени цензурирования распределения статистики существенно отличаются от предельного. Необходимо отметить, что в случаях проверки гипотез о согласии с другими законами распределениями, прослеживаются те же закономерности для распределений статистики.

Результаты исследования распределений статистики $\sqrt{\frac{na}{1-a}}R_n^+(a,1)$ в сравнении с соответствующим предельным распределением $2\Phi(x)-1$ выявили аналогичную картину.

Исследования показали, что распределения рассмотренных классических статистик Реньи существенно зависят от полного объема выборки и медленно сходятся к предельному с ростом n. Для практического же использования предпочтительней те статистики, которые менее зависимы от n. Тогда предельными законами статистики можно пользоваться и при малых объемах выборок. Анализ результатов исследования статистик Реньи наталкивает на мысль о необходимости введения некоторой поправки в выражения для статистик, которая сделала бы модифицированные статистики менее зависимыми от объема выборки. Например, предложенная простая модификация статистики $R_n^{new}(a,1) = \sqrt{\frac{n(a+\sqrt[10]{n}/n)}{1-(a+\sqrt[10]{n}/n)}}R_n(a,1)$ при $n \to \infty$ также имеет в каче-

стве предельного функцию распределения $L(x)$. Но распределения этой статистики с ростом n гораздо быстрее сходятся к $L(x)$. На рис. 3 представлены эмпирические функции распределения модифицированной статистики $R_n^{new}(a,1)$ в случае проверки простой гипотезы о согласии с логарифмически нормальным законом с параметром сдвига $\mu = 0$ и параметром масштаба $\sigma = 1$ по цензурированным слева выборкам при $n = 250$ и различной величине доступной наблюдению области. На этом рисунке распределения статистики для наглядности приведены не сглаженными.

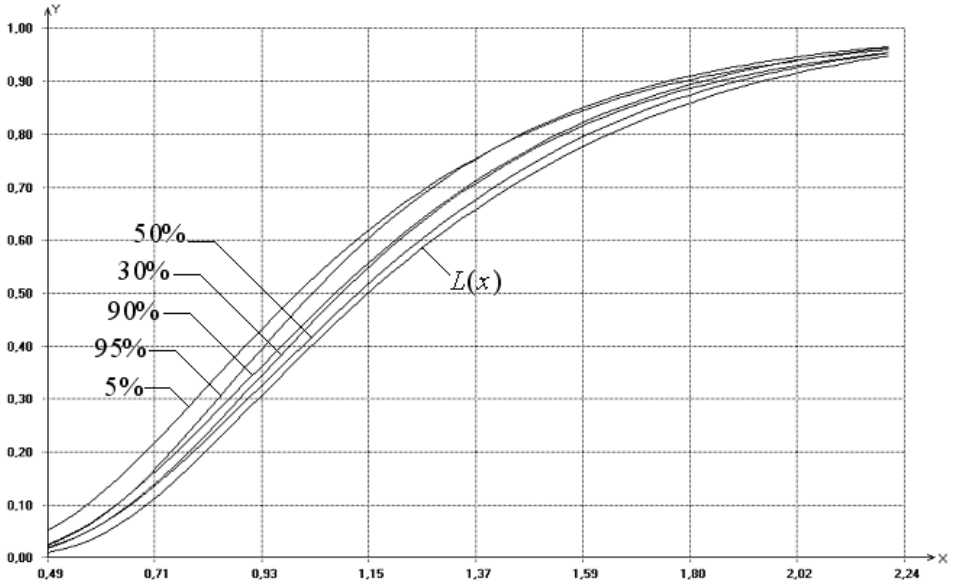

Рис.2. Распределения статистики $R_n(a,1)$, $n = 100$, цензурирование слева, I тип

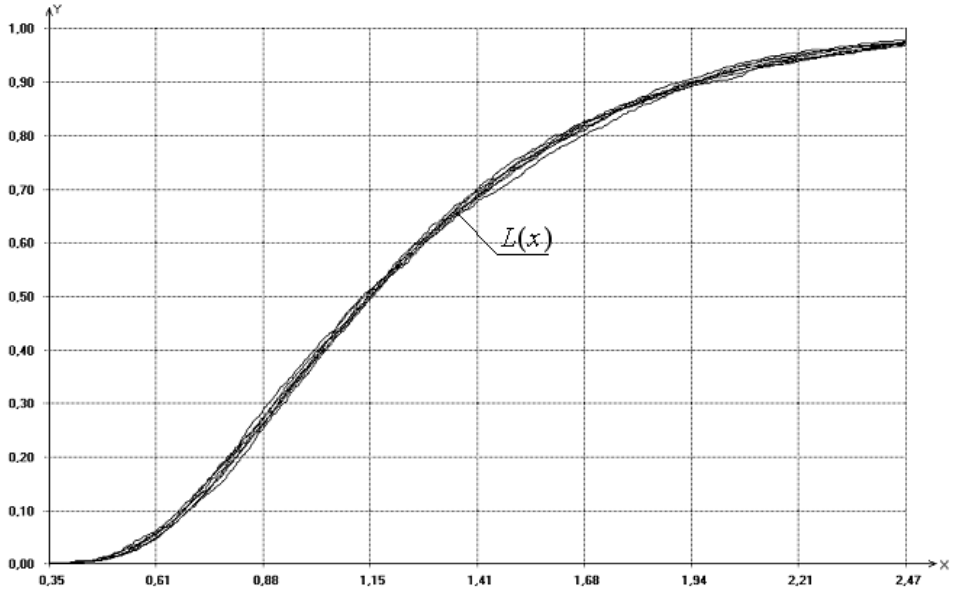

Рис. 3. Распределения статистики $R_n^{new}(a,1)$ при величине наблюдаемой области 5, 10, 30, 50, 70, 80, 95%, $n = 250$, I тип цензурирования

В данной работе при малых объемах выборок нами исследовались также распределения статистик $R_n^-(0,1-a)$ и $R_n^+(a,1)$. Функция $L_1(x,n,a)$ при $n = 10 \div 100$ представляет собой кусочно-гладкую функцию со скачками. При увеличении объема выборки и величины наблюдаемой области величины скачков уменьшаются, и график функции становится все более гладким.

На рис. 4 представлены распределения статистики $R_n^+(a,1)$ для случая проверки простой гипотезы о согласии с логарифмически нормальным законом ($\mu = 0$, $\sigma = 1$) по цензурированным слева выборкам (цензурирование I типа). Из рисунка видно, что при $a = 0.5$ функция распределения $L_1(x,n,a)$ при объеме выборки $n = 50$ становится достаточно гладкой, и эмпирическое распределение статистики хорошо согласуется с $L_1(x,n,a)$. Однако, при дальнейшем увеличении объема выборки распределения $G(R_n^+(a,1)\,|\,H_0)$ становятся все дальше от $L_1(x,n,a)$. Так, при $n = 70$ разница между эмпирическим и теоретическим распределениями статистики $R_n^+(a,1)$ становится уже существенной.

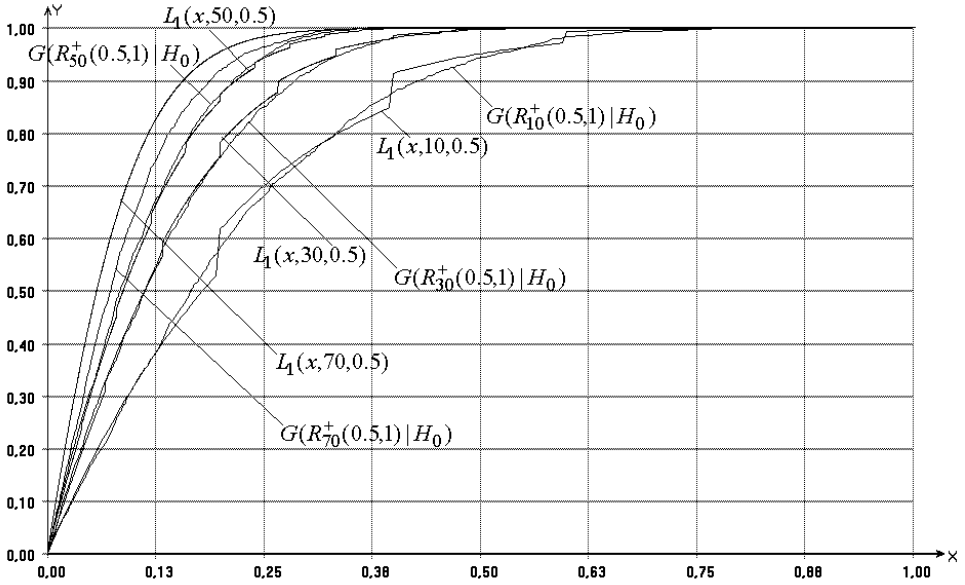

Рис. 4. Распределения статистики $R_n^+(a,1)$ и соответствующие функции $L_1(x,n,a)$, при *n*=10, 30, 50, 70 и степени цензурирования 50%

Таким образом, проведенные методами статистического моделирования исследования распределений статистик типа Реньи при проверке простых гипотез показали:

- близость распределений статистик к предельным сильно зависит от объема выборки n;
- при одном и том же объеме выборки распределение соответствующей статистики тем ближе к предельному, чем ближе степень цензурирования a к величине 0,5;
- при малых n "точные" распределения $L_1(x,n,a)$ плохо согласуются с действительными распределениями статистик вида $R_n^+(a,1)$.

Все это говорит о том, что проблема адекватности моделей по цензурированным выборкам ограниченного объема требует дополнительных исследований даже при проверке простых гипотез. Возможно построение модифицированных статистик типа Реньи, предельными распределениями которых можно пользоваться при меньших объемах выборок.

ЛИТЕРАТУРА

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
2. Смирнов Н.В. Вероятности больших значений непараметрических односторонних критериев согласия // Тр. матем. ин-та им. Стеклова В.А. АН СССР, 64, М., 1961. – С. 185-210.