СБОРНИК НАУЧНЫХ ТРУДОВ НГТУ. – 2002. – № 3(29). – 27-32
УДК 59.2

МОДЕЛИРОВАНИЕ ПРОЦЕССОВ И УСТРОЙТСВ

ПРОВЕРКА ГИПОТЕЗ О ДИСПЕРСИИ ПРИ НАРУШЕНИИ ПРЕДПОЛОЖЕНИЙ О НОРМАЛЬНОСТИ

О.Н. ВАНЮКЕВИЧ♥, Б.Ю. ЛЕМЕШКО♣

Методами статистического моделирования исследуются распределения классических статистик, используемых при проверке гипотез о дисперсии, для случаев когда наблюдаемый закон отличается от нормального. Показано, что распределения статистик зависят от наблюдаемого закона. Строятся модели распределений статистик для некоторых наблюдаемых законов.

1. ВВЕДЕНИЕ

В прикладной статистике, особенно при статистическом контроле качества, очень часто сталкиваются с задачами проверки гипотез о дисперсии наблюдаемой случайной величины. Проверяемая гипотеза имеет вид $H_0: \ \sigma = \sigma_0$, где σ_0 – номинальное значение среднеквадратичного отклонения. При этом математическое ожидание случайной величины μ может быть известным или оцениваться по этой же выборке. В классической математической статистике предполагается, что наблюдаемая выборка принадлежит нормальному закону распределения. В этом случае при известном μ статистика

$$S = \left(\sum_{i=1}^{n} (x_i - \mu)^2 \right) \Big/ \sigma_0^2 , \qquad (1)$$

где x_i, $i = \overline{1,n}$, выборка объема n, подчиняется в пределе χ_n^2-распределению с числом степеней свободы n [1]. При неизвестном μ аналогичная статистика

♥ Магистрант кафедры прикладной математики

♣ Профессор кафедры прикладной математики, д-р техн. наук

$$S = \left(\sum_{i=1}^{n} (x_i - \hat{\mu})^2 \right) \Big/ \sigma_0^2 \ , \qquad (2)$$

где $\hat{\mu}$ – оценка максимального правдоподобия математического ожидания вычисляемая по данной выборке, в пределе подчиняется χ^2_{n-1} –распределению

Естественно, что на практике далеко не всегда выполняется предположение о нормальности наблюдаемого закона. Относительно того, чтó произойдет с распределениями статистик (1)–(2), если наблюдаемый закон будет другим, в литературных источниках можно найти лишь одно утверждение: распределения статистик будут иными, нежели в случае нормального закона. Цель данной работы заключалась в исследовании методами статистического моделирования распределений статистик (1)–(2) в ситуациях, когда наблюдаемый закон существенно отличается от нормального.

Исследование распределений статистик. Распределения статистик (1)–(2) исследовались при принадлежности наблюдаемых выборок нормальному и четырем другим законам распределения:

логистическому с плотностью

$$f(x) = \frac{\pi}{\theta_1\sqrt{3}} \exp\left(-\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}} \right) \Big/ \left[1 + \exp\left(-\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}} \right) \right]^2 ,$$

Лапласа с плотностью

$$f(x) = \frac{\theta_0}{2} \exp\left(-\theta_0 \left| x - \theta_1 \right| \right),$$

минимального значения с плотностью

$$f(x) = \frac{1}{\theta_1} \exp\left(\frac{x-\theta_0}{\theta_1} - \exp\left(\frac{x-\theta_0}{\theta_1} \right) \right),$$

максимального значения с плотностью

$$f(x) = \frac{1}{\theta_1} \exp\left(-\frac{x-\theta_0}{\theta_1} - \exp\left(-\frac{x-\theta_0}{\theta_1} \right) \right).$$

При статистическом моделировании была использована методика, хорошо зарекомендовавшая себя при исследовании распределений статистик критериев согласия при проверке сложных гипотез [2, 3]. В случае нормального наблюдаемого закона получаемые в результате моделирования эмпирические распределения статистик (1)–(2) при различных объемах выборок n с высоким достигаемым уровнем значимости [4, 5] согласовывались с χ^2_n- и χ^2_{n-1}-распределениями соответственно.

При отличии наблюдаемого закона от нормального, распределения статистик (1)–(2) не являются χ^2-распределениями. При этом, чем больше наблюдаемый закон отличается от нормального, тем в большей степени распределения статистик (1)–(2) отличаются от соответствующих χ^2-распределений. По рис. 1, на котором представлены построенные функции распределения $G(S|H_0)$ статистики (1) при объеме выборок $n = 30$, можно проследить, как меняется распределение статистики в зависимости от вида наблюдаемого закона. На рисунке цифрой *1* отмечено распределение статистики в случае нормального закона, цифрой *2* – в случае логистического, цифрой *3* – в случае распределения Лапласа, цифрой *4* – в случае распределений минимального и максимального значения. На рисунке приводится случай, когда μ известно. Подобная же картина наблюдается для распределений статистики (2).

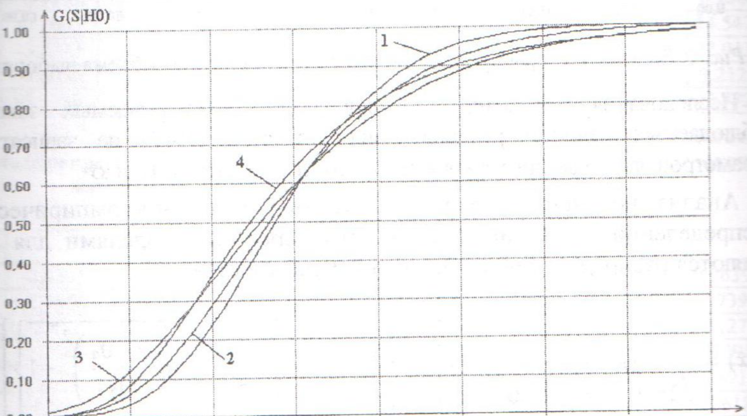

Рис. 1. Распределения статистики (1) в зависимости от наблюдаемого закона

С увеличением объемов выборок n характер различий между функциями распределениями статистик (1)-(2) в зависимости от наблюдаемого закона сохраняется. На рис. 2 показана зависимость функций распределения статистики (1) и (2) от n при наблюдении логистического закона. Для сравнения с классическим случаем на рисунке цифрой *1* отмечено χ^2_{n-1}-распределение, *2* – χ^2_n-распределение, *3* – распределение статистики (2), *4* – распределение статистики (1). Аналогичные тенденции прослеживаются для всех четырех законов, относительно которых проводились исследования.

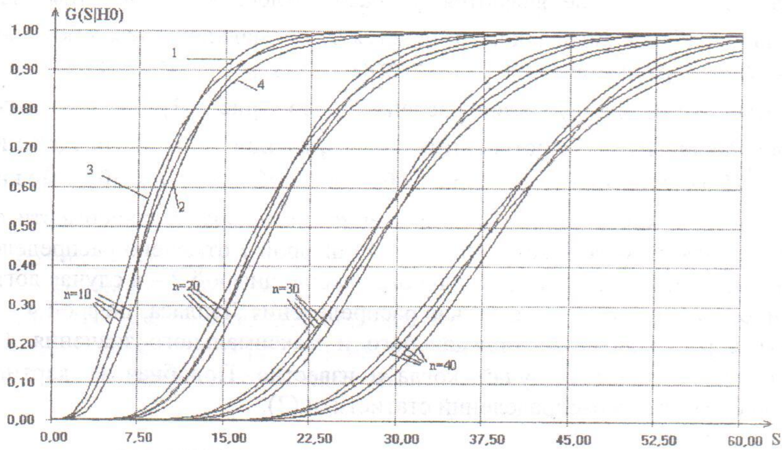

Рис. 2. Зависимость распределений статистик (1) и (2) от объема выборки

Исследования показали, что для всех рассматриваемых законов наблюдаемых величин распределения статистик (1)-(2) не зависят от параметров наблюдаемого закона, а значит, и от значений μ и σ.

Анализ полученных в результате моделирования эмпирических распределений статистик показал, что хорошими моделями для них являются распределения *Su*-Джонсона с плотностью

$$f(x)=\frac{\theta_1}{\sqrt{2\pi}\sqrt{(x-\theta_3)^2+\theta_2^2}}\exp\left\{-\frac{1}{2}\left[\theta_0+\theta_1\ln\left\{\frac{x-\theta_3}{\theta_2}+\sqrt{\left(\frac{x-\theta_3}{\theta_2}\right)^2+1}\right\}\right]^2\right\}.$$

В табл. 1…3 представлены значения параметров распределений *Su*-Джонсона, при которых данное распределение может использоваться в качестве модели предельного закона для статистики (1) или (2) при заданном объеме выборки n.

Таблица 1

Модели предельных распределений статистик (1)-(2) при наблюдаемом логистическом законе (значения параметров распределения Su-Джонсона)

n	Для статистики (1)				Для статистики (2)			
10	-7,1472	1,9618	0,4823	-0,6204	-7,2884	1,9485	0,4016	-0,8096
20	-8,101	2,7949	2,2005	-1,0484	-8,8944	2,776	1,5582	-1,271
30	-9,5814	3,7514	5,4148	-5,4719	-9,8203	3,7269	4,8312	-5,3169
40	-8,7706	3,6	6,7153	0,4674	-8,7877	3,6187	6,6299	0,2476
50	-10,042	4,054	8,0454	0,7117	-9,7136	3,8627	7,2455	2,648
60	-3,6164	3,2174	24,6455	24,5859	-2,7265	2,9803	26,1846	30,1689
70	-8,4695	4,7777	23,2483	2,3225	2,4675	1,5205	20,907	32,8737
80	-9,9165	5,5815	29,0151	-4,3129	-10,539	5,6833	27,2255	-6,8532
90	-11,391	5,0739	17,7035	5,7275	-11,538	4,9793	16,0102	6,9557
100	-5,1505	4,458	43,7646	35,9405	-5,6013	4,5217	41,1594	32,4735

Таблица 2

Модели предельных распределений статистик (1)–(2) при наблюдаемом законе распределения Лапласа (значения параметров распределения Su-Джонсона)

n	Для статистики (1)				Для статистики (2)			
10	-6.8936	1.8532	0.4728	-1.3951	-6.7663	1.8499	0.4613	-1.4621
20	-5.4134	2.1506	2.977	-0.3785	-5.4342	2.1184	2.6548	-0.1253
30	-9.4325	3.0198	2.8922	-4.6528	-9.0781	2.9358	2.8370	-4.0213
40	-8.0906	2.8029	4.0183	0.9219	-8.8015	2.8295	3.1718	0.4223
50	-10.264	3.4782	5.5200	-5.0129	-9.7243	3.3919	5.7348	-3.7298
60	-6.9173	3.3826	14.1346	3.3197	-5.7370	3.2873	17.9457	6.2863
70	-4.2003	3.6301	35.9671	16.0584	-4.3210	3.7128	36.0763	14.5474
80	-3.8772	3.5097	38.2614	25.7941	-4.2663	3.6307	37.0208	22.0438
90	-9.0050	22.284	22.2835	-0.2800	-8.2615	4.1942	23.5068	3.2125
100	-3.5280	3.5239	46.6587	42.3844	-3.1144	3.3483	46.2507	46.7727

Таблица 3

Модели предельных распределений статистик (1)–(2) при наблюдаемых законах максимального и минимального экстремального значения (значения параметров распределения Su-Джонсона)

n	Для статистики (1)				Для статистики (2)			
10	-2,4291	1,4207	2,3229	2,1608	-5,0805	1,6547	0,7263	-0,2639
20	-2,9656	1,753	4,7068	5,3886	-3,3569	1,9806	5,2998	3,0739
30	-2,812	2,0014	8,7513	10,8722	-3,3598	2,276	9,4395	7,3127
40	-3,2368	2,1093	10,1068	15,1711	-3,8429	2,4051	11,0083	10,7094
50	-6,1326	3,147	11,9104	7,3078	-6,4644	3,5455	15,1726	1,7891
60	-4,549	2,7494	14,5863	20,5565	-3,5343	2,819	21,3096	22,3618
70	-8,3168	3,1461	7,2875	15,9986	-8,6221	3,4305	9,0478	10,7072
80	-2,9202	2,7003	28,4078	39,937	-2,4734	2,7557	33,3375	42,1869
90	-5,0311	3,1145	22,4844	33,1897	-3,9492	3,1666	31,558	36,46
100	-2,6492	2,8188	36,42	58,0124	-2,19	2,7958	40,1695	62,008

ЗАКЛЮЧЕНИЕ

Результаты исследований показали, что распределения статистик (1)-(2), используемых при проверке гипотез о дисперсии, существенно зависят от наблюдаемого закона распределения случайных величин. Построенные модели распределений статистик могут использоваться при проверке соответствующих гипотез при объемах выборок от 10 до 100 в ситуациях, когда наблюдаемые случайные величины принадлежат законам: логистическому, минимального и максимального значений.

1. **Боровков А.А**. Математическая статистика. Дополнительные главы: Учебное пособие для математ. и физ. спец. вузов. – М.: Наука, 1984.

2. **Лемешко Б.Ю., Постовалов С.Н**. О распределениях статистик непараметрических критериев согласия при оценивании по выборкам параметров наблюдаемых законов // Заводская лаборатория. – 1998. – Т. 64. – № 3. – С. 61-72.

3. **Лемешко Б.Ю., Постовалов С.Н**. Применение непараметрических критериев согласия при проверке сложных гипотез // Автометрия. – 2001. – № 2. – С. 88-102.

4. **Денисов В.И., Лемешко Б.Ю., Постовалов С.Н**. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. – Новосибирск: Изд-во НГТУ, 1998. – Ч. I: Критерии типа χ^2.

5. **Лемешко Б.Ю., Постовалов С.Н**. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. – Новосибирск: Изд-во НГТУ, 1999. – Ч. II: Непараметрические критерии.