

А. Ю. СИДОРЕНКО, канд. техн. наук, доц. НТУ «ХПИ»;
А. С. МАЗМАНИШВИЛИ, д-р физ.-мат. наук, проф. СГУ, Сумы

НАХОЖДЕНИЕ КВАНТИЛЕЙ ХИ-КВАДРАТ РАСПРЕДЕЛЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПРИ НАЛИЧИИ КОРРЕЛЯЦИИ ОПЫТНЫХ ДАННЫХ

Розглянута задача математичної статистики, що відноситься до корельованих послідовностям (часовим відлікам) з заданим фіксованим рівнем статистичного зв'язку між ними. Представлені таблиці ліво- та правосторонніх квантилів розподілу хі-квадрат при наявності кореляції, а також графіки залежності квантилів від параметрів розподілу. Показано, що отримані результати можуть бути застосовані в задачах прийняття статистичних рішень.

Ключові слова: корельовані відліки, критичне значення, ліво- та правосторонні квантилі, статистичні дані, розподіл хі-квадрат.

Рассмотрена задача математической статистики, относящаяся к коррелированным последовательностям (временным отсчетам) с заданным фиксированным уровнем статистической связи между ними. Представлены таблицы лево- и правосторонних квантилей распределения хи-квадрат при наличии корреляции, а также графики зависимости квантилей от параметров распределения. Показано, что полученные результаты могут быть применены в задачах принятия статистических решений.

Ключевые слова: коррелированные отсчеты, критическое значение, лево- и правосторонние квантили, статистические данные, хи-квадрат распределение.

The problems of mathematical statistics, related to correlated sequences (to time counts) with a given fixed level of statistical correlation between them, are considered. Tables are presented left- and right-side quantiles of distribution of the chi-squared at presence of correlation, and also graphic arts of dependence of quantiles from the parameters of distribution. It is shown that the results obtained can be applied in the tasks of making statistical decisions.

Keywords: correlated sequences, critical value, left- and right-side quantiles, statistical data, chi-squared's distribution.

Введение. В математической статистике принято считать задачу решенной, если предъявлена плотность распределения вероятностей или производящая функция случайной величины. Алгоритмы решения задач принятия статистических решений содержат также заключительный этап сравнения двух величин – наблюдаемого критерия и порогового значения (квантиля). Если критерий (1) определяется данными наблюдений, то квантиль χ_{α}^2 связан с принятой гипотетической моделью. В известных законах распределения [1–3] исходно допускается предположение о некоррелированности наблюдаемых отсчетов. Но при рассмотрении некоторых распространенных практических задач часто сталкиваются с тем, что экспериментальные данные коррелируют между собой. В связи с этим

возникает задача о степени влияния корреляции данных на χ^2_α распределения в задачах математической статистики рассматриваемых величин.

Целью данной статьи является построение и анализ лево- и правосторонних критических значений распределения χ^2 в зависимости от изменения параметров, при этом в качестве отсчетов будет использована коррелированная последовательность.

Алгоритм расчета. Рассмотрение будет проведено при следующем предположении: используемые ниже наборы отсчетов $\{x_0, x_1, \dots, x_N\}$ – совокупность сечений объемом $N+1$, взятых с постоянным временным шагом τ из реализации нормального марковского процесса Орнштейна-Уленбека [4, 5] $x(t)$ с интенсивностью σ и коэффициентом корреляции q .

Для интервала длительностью T рассмотрим плотность распределения вероятностей $f_J(\eta)$ случайных значений функционала

$$J = \frac{1}{N+1} \sum_{n=0}^N x_n^2, \quad (1)$$

которая выражается с помощью обратного преобразования Лапласа

$$f_J(\eta) = \frac{1}{2\pi i} \int \exp(\lambda \eta) Q_{N+1}(\lambda) d\lambda \quad (2)$$

через производящую функцию отсчетов

$$Q_{N+1}(\lambda) = \left\langle \exp\left(-\frac{\lambda}{N+1} \sum_{n=0}^N x_n^2\right) \right\rangle, \quad (3)$$

взятых с равномерным шагом $\tau = T/N$. Здесь угловые скобки $\langle . \rangle$ обозначают операцию нахождения математического ожидания относительно случайной величины (1). В работе [6] было доказано, что для нахождения плотности распределения вероятностей $f_J(\eta)$ аддитивного функционала J (1) можно использовать алгоритм N -кратной свертки

$$f_J(\eta) = (\varphi_0 * \varphi_1 * \dots * \varphi_N)(\eta), \quad (4a)$$

$$\varphi_n(\eta) = \sqrt{\frac{\lambda_n(N+1)}{2\pi\sigma^2\eta}} \exp\left(-\frac{N+1}{2\sigma^2} \lambda_n \eta\right), \quad n = 0, 1, \dots, N. \quad (4b)$$

Алгоритмы решения задач принятия статистических решений содержат заключительный этап сравнения двух величин – наблюдаемого критерия и

порогового значения (квантиля). Если критерий (1) определяется данными наблюдений, то квантиль η_α связан с принятой гипотетической моделью.

Критическая точка η_α распределения случайных значений η функционала качества J отвечает равенству

$$\Pr(J > \eta_\alpha) = \int_{\eta_\alpha}^{\infty} f_J(\eta) d\eta = \alpha, \quad (5)$$

где α – уровень значимости.

Основные результаты. На основе описанного метода было создано программное средство, позволяющее построить таблицы квантилей функционала при изменении различных параметров распределения.

С целью верификации результатов рассмотрено известное в математической статистике распределение χ^2 (табл. 1) с числом степеней свободы $N+1$ (так как индекс числа отсчетов функционала (1) начинается с 0) в качестве частного случая распределения квадратичного функционала J (табл. 2). Сравнение распределения χ^2 и полученного распределения квадратичного функционала (1) возможно только в том случае, когда коэффициент корреляции q будет равен 0.

Таблица 1 – Двусторонние критические точки распределения χ^2

N	α							
	0.01	0.025	0.05	0.1	0.90	0.95	0.975	0.99
4	0.30	0.48	0.71	1.06	7.78	9.49	11.1	13.3
5	0.55	0.83	1.15	1.61	9.24	11.1	12.8	15.1
6	0.87	1.24	1.64	2.20	10.6	12.6	14.4	16.8
10	2.56	3.25	3.94	4.87	16.0	18.3	20.5	23.2
11	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7
15	5.23	6.26	7.26	8.55	22.3	25.0	27.5	30.6
16	5.81	6.91	7.96	9.31	23.5	26.3	28.8	32.0

Таблица 2 – Двусторонние критические точки функционала (1), где $q=0$, $\sigma=1$

N	α							
	0.01	0.025	0.05	0.1	0.90	0.95	0.975	0.99
3	0.30	0.48	0.71	1.06	7.78	9.49	11.1	13.3
5	0.87	1.24	1.64	2.20	10.6	12.6	14.4	16.8
7	1.65	2.18	2.73	3.49	13.4	15.5	17.5	20.1
10	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7
15	5.81	6.91	7.96	9.31	23.5	26.3	28.8	32.0
20	8.90	10.3	11.6	13.2	29.6	28.9	35.5	38.9

Сравнение указанных значений из табл. 1 и табл. 2 позволяет убедиться в правильности построения таблиц лево- и правосторонних квантилей.

В табл. 3 представлены пороговые значения, в которых изменяющимся параметром является коэффициент корреляции q .

Таблица 3 – Двусторонние критические точки функционала (1), где $N = 3$, $\sigma = 1$

q	α							
	0.01	0.025	0.05	0.1	0.90	0.95	0.975	0.99
0.1	0.11	0.25	0.31	0.36	2.15	2.59	3.01	3.55
0.3	0.10	0.24	0.29	0.34	2.19	2.68	3.17	3.82
0.5	0.09	0.21	0.26	0.30	2.27	2.87	3.48	4.32
0.7	0.06	0.16	0.20	0.24	2.41	3.19	3.99	5.09
0.9	0.03	0.08	0.10	0.13	2.68	3.69	4.74	6.16
0.95	0.02	0.05	0.07	0.09	2.77	3.84	4.96	6.48
0.99	0.01	0.02	0.03	0.05	2.85	3.98	5.16	6.75

С уменьшением коэффициента корреляции график плотности распределения имеет тенденцию локализоваться вокруг среднего значения функционала J (1), а значения правосторонних квантилей должны увеличиваться, а левосторонних – уменьшаться. В приведенной табл. 3 можно наблюдать эту тенденцию.

В табл. 4 представлены лево- и правосторонние квантили с изменяющимся параметром $N + 1$ – числом отсчетов. Анализ данной таблицы на правильность построения показывает, что, во-первых, с увеличением уровня значимости α пороговые значения увеличиваются, во-вторых, с увеличением объема выборки N квантили также возрастают.

Таблица 4 – Двусторонние критические точки функционала (1), где $q = 0.9$, $\sigma = 1$

N	α							
	0.01	0.025	0.05	0.1	0.90	0.95	0.975	0.99
3	0.03	0.08	0.10	0.13	2.68	3.69	4.74	6.16
5	0.04	0.10	0.13	0.16	2.58	3.52	4.50	5.84
7	0.05	0.12	0.15	0.18	2.49	3.38	4.30	5.56
10	0.06	0.14	0.18	0.21	2.39	3.20	4.04	5.19
15	0.08	0.18	0.21	0.25	2.25	2.95	3.69	4.69
20	0.10	0.20	0.24	0.28	2.15	2.77	3.42	4.30
25	0.11	0.22	0.26	0.30	2.06	2.63	3.20	3.99

На рис. 1, a – b представлены зависимости $a)$ – левосторонних и $b)$ – правосторонних квантилей функционала J (1) от коэффициента корреляции q . Анализ зависимостей, построенных на рис. 1, показывает влияние коэффициента корреляции q на правосторонние пороговые значения η_α .

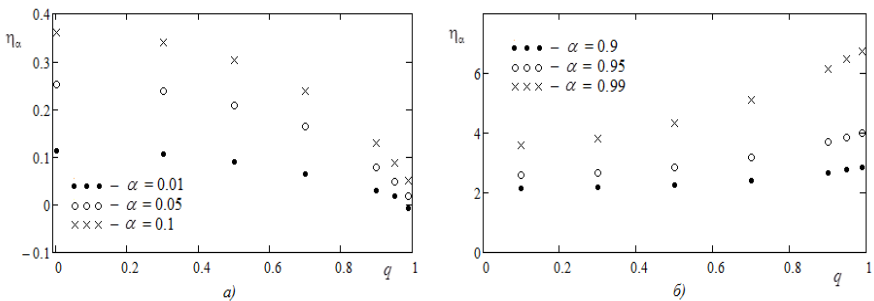


Рис. 1 – Зависимость a – левостороннего и b – правостороннего квантилей η_α от коэффициента корреляции q для $N=3$, $\sigma=1$

Видно, что с увеличением коэффициента корреляции q возрастают правосторонние пороговые значения (квантили) η_α , а левосторонние – уменьшаются. Далее, анализируя рис. 1, можно убедиться в важности выбора уровня значимости α . Чем больше уровень значимости, тем больше лево- и правосторонние пороговые значения η_α . Следовательно, можно сделать вывод, что при $q > 0$ коэффициент корреляции значительно влияет на поведение критических значений функционала (1).

На рис. 2, a – b представлены зависимости $a)$ – левосторонних и $b)$ – правосторонних квантилей функционала J (1) от параметра N .

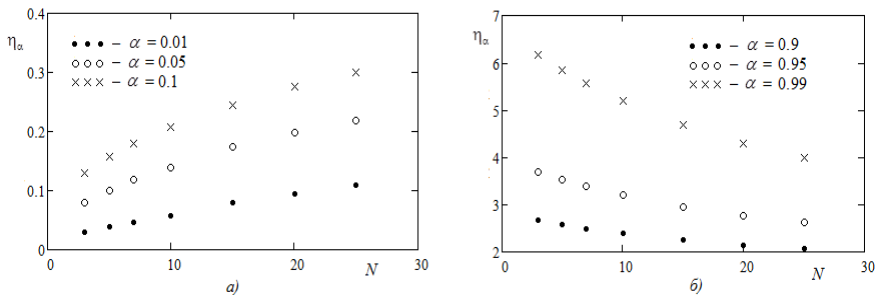


Рис. 2 – Зависимость a – левостороннего и b – правостороннего квантилей η_α от количества числа отсчетов N для $q=0.9$, $\sigma=1$

Из рис. 2, a – b можно сделать вывод о том, что приведенные зависимости отвечают общим закономерностям, характерным для критических значений функционала (1). А именно, с увеличением числа отсчетов N возрастают левосторонние пороговые значения (квантили) η_α , а правосторонние –

уменьшаются. Также, при увеличении уровня значимости лево- и правосторонние пороговые значения η_α функционала (1) возрастают, о чем свидетельствует рис. 3.

На рис. 3 приведены зависимости *а)* – левосторонних и *б)* – правосторонних квантилей функционала J (1) от уровня значимости α .

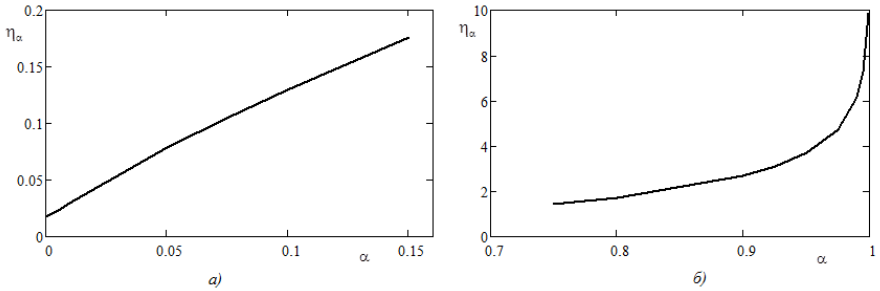


Рис. 3 – Зависимость *а* – левостороннего и *б* – правостороннего квантилей η_α от уровня значимости α для $q = 0.9$, $N = 3$, $\sigma = 1$

В производственных процессах, в которых изначально задаются уровень значимости и коэффициент корреляции результатов (например, при шлифовании), необходимо задавать технологический квантиль, определяющий завершение процесса. На основании полученных результатов можно сделать вывод о том, что учет корреляции представляет не только теоретический, но и практический интерес.

Выводы. Полученные таблицы и зависимости справедливы для любых допустимых значений коррелятора q , в том числе и нулевых. Решение задач принятия статистических решений содержат заключительный этап сравнения двух величин – наблюдаемого критерия и порогового значения (квантиля). Учет коррелированности опытных данных может существенно повлиять на вид принимаемого решения.

Список литературы: 1. Гнеденко Б. В. Курс теории вероятностей / Б. В. Гнеденко. – М. : Наука, 1961. – 406 с. 2. Королюк В. С. Справочник по теории вероятностей и математической статистике / В. С. Королюк – К. : Наукова думка. – 1978. – 584 с. 3. Мазманишвили А. С. Континуальное интегрирование как метод решения физических задач / А. С. Мазманишвили. – К. : Наукова думка, 1987. – 224 с. 4. Корн Г. Справочник по математике для научных работников и инженеров / Г. Корн, Т. Корн. – М. : Наука. – 1987. – 360 с. 5. Венцель Е. С. Теория случайных процессов и ее инженерные приложения / Е. С. Венцель, Л. А. Овчаров. – М. : Высшая школа. – 2000. – 383 с. 6. Сидоренко А. Ю. Применение методов статистического анализа при обработке опытных данных / А. Ю. Сидоренко, А. С. Мазманишвили // Вісник НТУ «ХП». – 2012. – № 2(976). – С. 114–121. 7. Тихонов В. И. Марковские процессы / В. И. Тихонов, М. А. Миронов. – М. : Сов. радио, 1977. – 488 с.