

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

ПРИМЕНЕНИЕ МЕТОДОВ СТАТИСТИЧЕСКОГО АНАЛИЗА ПРИ ОБРАБОТКЕ ОПЫТНЫХ ДАННЫХ

Розглянуто типові задачі математичної статистики, що відносяться до корельованим послідовностям (часовим відлікам) з заданим фіксованим рівнем статистичного зв'язку між ними. Одержані аналітичні вирази, що описують розподіл χ^2 у випадку корельованих відліків. Показано, що отримані результати можуть бути застосовані в задачах прийняття статистичних рішень. Представлені графіки функції та щільності розподілу χ^2 при наявності кореляції.

Ключові слова: статистичні дані, розподіл χ^2 , корельовані відліки, щільність розподілу, функція розподілу.

Рассмотрены типичные задачи математической статистики, отнесенные к коррелированным последовательностям (временным отсчетам) с заданным фиксированным уровнем статистической связи между ними. Получены аналитические выражения, описывающие распределение χ^2 для случая коррелированных отсчетов. Представлены графики функции и распределения χ^2 при наличии корреляции.

Ключевые слова: статистические данные, χ^2 , коррелированные отсчеты, плотность распределения, функция распределения.

The typical problems of mathematical statistics, related to correlated sequences (to time counts) with a given fixed level of statistical correlation between them, are considered. The analytic expressions featuring distributions Chi-Squared's ones in the case of correlated counts are obtained. The charts of functions and distributions Chi-Squared's in the case of correlated counts are introduction.

Keywords: statistical data, Chi-Squared distribution, correlated sequences, the charts of functions and distributions.

Введение. При аналитическом рассмотрении задач теории вероятностей и математической статистики распространено допущение о том, что рассматриваемая задача получила своё разрешение, если построена характеристическая (производящая) функция рассматриваемой случайной величины. Это можно пояснить тем, что при сложении случайных величин искомая плотность распределения композиции есть многократная (по числу слагаемых) свертка парциальных плотностей, в то время как характеристическая функция композиции является произведением парциальных характеристических функций [1, 2], то есть операцией существенно более просто выполнимой, чем многократное интегрирование.

Однако операция обратного преобразования Фурье или обратного преобразования Лапласа вызывает основные трудности в вычислительном отношении. Как числовая процедура преобразование Фурье характеризуется неустойчивостью, степень которой увеличивается с ростом параметра

преобразования. Несмотря на большое число попыток [3, 4], к настоящему времени отсутствуют устойчивые алгоритмы обращения Фурье.

Известные таблицы распределений χ^2 , Стьюдента, Фишера и других [5, 6] относятся по существу к ситуации, когда наблюдаемые данные являются некоррелированными. В этой связи возникает вопрос о степени влияния корреляции данных на распределения рассматриваемых в задачах математической статистики величин (прямые задачи).

Целью данной статьи является построение числовых алгоритмов, обеспечивающих расчет стандартных статистических распределений в законе типа χ^2 , при этом в качестве источника отсчетов будет использована коррелированная последовательность.

Представление с помощью базисных функций и обращение по Лапласу. Рассмотрение будет проведено при следующем предположении: используемые ниже наборы отсчетов $\{x_0, x_1, \dots, x_N\}$ – суть совокупность сечений объемом $N+1$, взятых с постоянным временным шагом τ из реализации нормального марковского процесса Орнштейна-Уленбека (ОУ-процесса) [7, 8] $x(t)$ 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 и далее угловыми скобками $\langle . \rangle$ будем обозначать операцию нахождения математического ожидания относительно указываемых индексом случайных величин. Для функции $Q_{N+1}(\lambda)$ аналитическое представление было получено ранее [8]. Поскольку оно содержит вложенные радикалы, осуществить интегрирование по контуру Бромвича в (2) оказалось малоперспективным. Поэтому запишем операцию нахождения безусловного среднего $Q_{N+1}(\lambda)$ (3) явно

$$Q_{N+1}(\lambda) = \int_{-\infty}^{\infty} dx_0 f(x_0) \int_{-\infty}^{\infty} dx_1 f(x_0; x_1) \dots \int_{-\infty}^{\infty} dx_N f(x_{N-1}; x_N) \exp\left(-\frac{\lambda}{N+1} \sum_{n=0}^N x_n^2\right), \quad (4)$$

где

$$f(x_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x_0^2}{2\sigma^2}\right), \quad f(x_n; x_{n+1}) = \frac{1}{\sqrt{2\pi(1-q^2)}\sigma} \exp\left(-\frac{(x_{n+1} - qx_n)^2}{2\sigma^2(1-q^2)}\right)$$

– равновесная ($n=0$) и переходные ($n=1, 2, \dots, N$) плотности распределения вероятностей для соответствующих сечений используемого ОУ-процесса; q – коэффициент корреляции соседних отсчетов.

После алгебраических преобразований, прямого интегрирования в (4), повторенного $(N+1)$ раз, и учета нормировки получим следующее факторизационное представление для производящей функции

$$Q_{N+1}(\lambda) = \prod_{n=0}^N \sqrt{\frac{\lambda_n(N+1)}{2\lambda\sigma^2 + \lambda_n(N+1)}}. \quad (6)$$

Факторизации производящих функций отвечает свертка совокупности парциальных плотностей вероятностей $\varphi_0(\eta), \varphi_1(\eta), \dots, \varphi_N(\eta)$. Рассмотрим, например, плотность, отвечающую n -ой компоненте

$$\varphi_n(\eta) = \frac{1}{2\pi i} \int \exp(\lambda\eta) \sqrt{\frac{\lambda_n(N+1)}{2\lambda\sigma^2 + \lambda_n(N+1)}} d\lambda. \quad (7)$$

Хотя в каждом из интегралов (7) также присутствует корневая особенность, можно воспользоваться известным результатом ($\eta > 0$)

$$\varphi_n(\eta) = \sqrt{\frac{\lambda_n(N+1)}{2\pi\sigma^2\eta}} \exp\left(-\frac{N+1}{2\sigma^2} \lambda_n \eta\right), \quad n=0, 1, \dots, N \quad (8)$$

и $\varphi_n(\eta) = 0$, если $\eta \leq 0$.

Таким образом, для искомой плотности распределения вероятностей $f_J(\eta)$ случайных значений аддитивного функционала J (1) получаем формулу в виде N -кратной свертки

$$f_J(\eta) = (\varphi_0 * \varphi_1 * \dots * \varphi_N)(\eta). \quad (9)$$

Например, при $N=2$ имеем

$$f_j(\eta) = \int_0^\infty du_0 \int_0^\infty du_1 \varphi_0(\eta - u_0) \varphi_1(u_0 - u_1) \varphi_2(u_1). \quad (10)$$

Проверка алгоритма расчета плотности распределения.

Производящая функция $\mathcal{Q}_{N+1}(\lambda)$ аддитивного квадратичного функционала (1) была получена ранее

Рассмотрим комплекснозначное обобщение формулы (3). Пусть $y(t)$ – еще один независимый процесс Орнштейна-Уленбека, статистически эквивалентный процессу $x(t)$, а процесс

$$z(t) = x(t) + iy(t). \quad (11)$$

есть комплекснозначное их объединение. Тогда удобно вместо функционала (1) рассматривать следующий функционал, основанный на комплекснозначном процессе $z(t)$

$$J_z = \frac{1}{N} \sum_{n=1}^N |z_n|^2, \quad z_n = z(n\tau). \quad (12)$$

В принятых предположениях аддитивный функционал (12) имеет равновесную производящую функцию $\mathcal{Q}_z(\lambda) \equiv \mathcal{Q}_z(\lambda; q, N) = \langle \exp(-\lambda J_z) \rangle$ следующего вида

$$\mathcal{Q}_N(\lambda) = \left(\frac{(1-q^2)R}{(\mu_1 - q^2)^2 a_+^{N-1} - (\mu_2 - q^2)^2 a_-^{N-1}} \right)^{1/2}, \quad (13)$$

где q – коэффициент корреляции соседних отсчетов комплекснозначного ОУ-процесса;

$$a_{\pm} = \frac{1}{2} (1 + q^2 + N^{-1} \lambda \sigma (1 - q^2) \pm R);$$

$$R = \sqrt{(1 + q^2 + N^{-1} \lambda \sigma (1 - q^2))^2 - 4q^2}.$$

Поскольку в силу статистической эквивалентности процессов $x(t)$ и $y(t)$ справедливо

$$\mathcal{Q}_z(\lambda) = \mathcal{Q}_x(\lambda) \mathcal{Q}_y(\lambda) = (\mathcal{Q}_x(\lambda))^2, \quad (14)$$

то приходим к искомой производящей функции для функционала (1)

$$\mathcal{Q}_x(\lambda) = \sqrt{\mathcal{Q}_z(\lambda)}. \quad (15)$$

Так как согласно (15) квадрат производящей функции $Q_x(\lambda) = Q_j(\lambda)$ равен $Q_z(\lambda)$, то автосвертка найденной плотности для функционала (1) должна совпадать (при одинаковых значениях параметров q и N) с плотностью распределения случайных значений функционала (12). Плотность распределения вероятностей $f_z(\eta)$ случайной величины J_z связана с производящей функцией $Q_z(\lambda)$ обратным преобразованием Лапласа.

На рис. 1-2 изображены плотности и функции распределения функционала (12), построенные с помощью автосвертки и производящей функции.

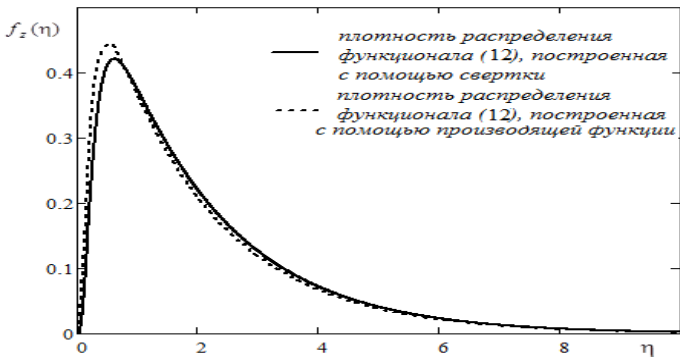


Рис. 1 – Плотности распределения функционала (12) с параметрами: $N = 3$, $\sigma = 1$, $q = 0.9$, $\tau = 0.1$

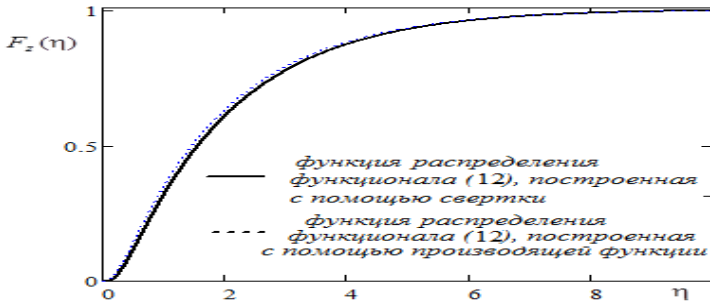


Рис. 2 – Функции распределения функционала (12) с параметрами: $N = 3$, $\sigma = 1$, $q = 0.9$, $\tau = 0.1$

Как видно из рис. 1-2 плотности и функции распределения функционала (12), построенные с помощью автосвертки и производящей функции

практически совпадают. Из этого следует, что вычислительные алгоритмы работают правильно.

Основные результаты. В теории оценивания и принятия решений важной особенностью является учет статистической связи между оценками уже известных распределений. В теории вероятностей и математической статистике известен аналог данному функционалу (1) – это χ^2 -распределение с числом степеней свободы равным $N-1$. Представляет интерес сравнить плотности распределения вероятностей рассматриваемого критерия качества (1) с χ^2 -распределением. На рис. 3 представлены плотности распределения функционала (1) и χ^2 -распределения в случае отсутствия корреляции.

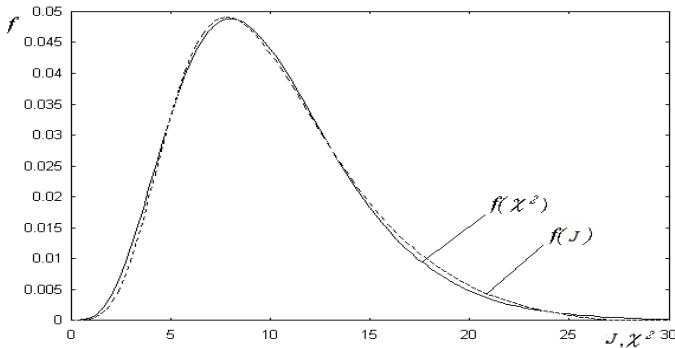


Рис. 3 – Плотность распределения вероятностей $f(J)$ для функционала J и для распределения χ^2 с $N-1$ степенями свободы ($N=10$, $\sigma=1$, $q=0$, $\tau=0.1$)

Как видно из рис. 3 эти плотности практически полностью совпадают, что свидетельствует в пользу правильности полученных формул.

На рис. 4 представлены плотности распределения функционала (1) в случае, когда коэффициент корреляции равен 0.95, и χ^2 -распределения с $N-1$ степенями свободы.

Из рис. 3-4 следует, что плотность распределения функционала (1) совпадает с плотностью χ^2 -распределения только в том случае, когда коэффициент корреляции равен 0, то есть корреляция между соседними сечениями случайного процесса отсутствует. В реальности, однако, значения случайных процессов в точках с соседними отсчетами являются в той или иной степени коррелированными.

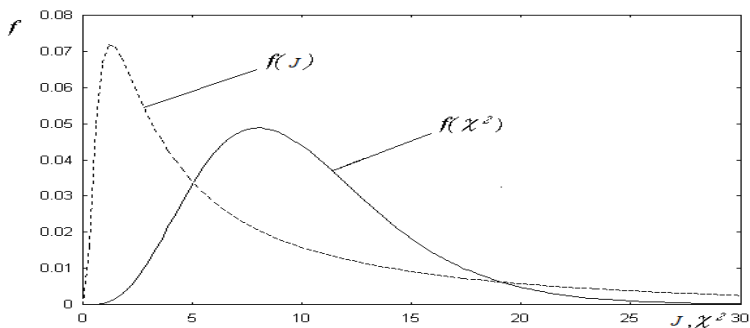


Рис. 4 – Плотность распределения вероятностей $f(J)$ для функционала J и для распределения χ^2 с $N-1$ степенями свободы ($N=10, \sigma=1, q=0.95, \tau=0.1$)

При $q > 0$ коэффициент корреляции влияет на поведение плотности и функции распределения функционала (1). На рис. 5, *a-b* представлены графики плотности и функции распределения функционала (1) при изменении параметра N . Можно сделать вывод о том, что приведенные зависимости отвечают общим закономерностям, характерным для плотности и функции распределения.

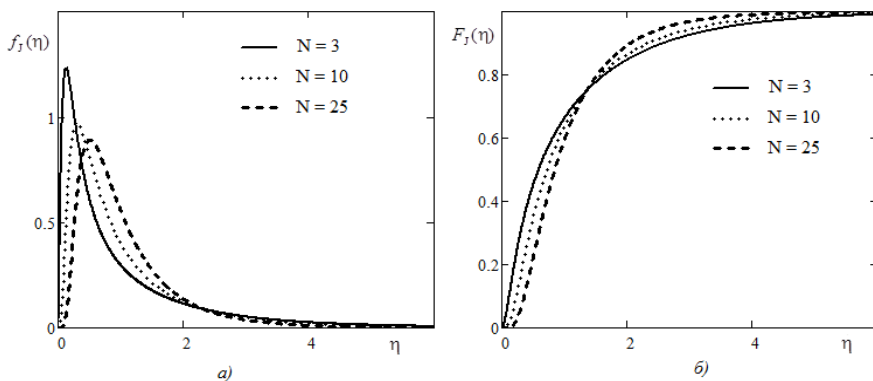


Рис. 5 – Плотности (*a*) и функции (*b*) – распределения вероятностей функционала (1); параметры: $q = 0.9, N = 3, 10, 25, \sigma = 1$

На рис. 6, *a-b* приведены плотности и функции распределения для функционала (1) соответственно. С увеличением уровня корреляции q график функции распределения имеет тенденцию локализоваться вокруг среднего значения функционала (1), поэтому значения правосторонних квантилей должны увеличиваться.

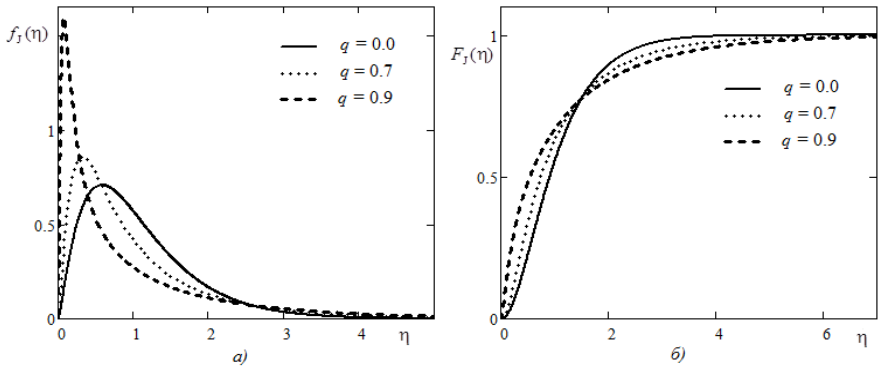


Рис. 6 – Плотности (а) и функции (б) распределения вероятностей функционала (1); параметры: $q = 0.00, 0.7, 0.9$, $N = 3$, $\sigma = 1$

На основании полученных результатов (рис. 5-6) можно сделать вывод о том, что при увеличении параметров: коэффициента корреляции и количества отсчетов плотности распределения χ^2_α становятся более пологими. Таким образом, учет корреляции представляет практический интерес.

Выводы. Полученные выше выражения справедливы для любых допустимых значений коррелятора q , в том числе и нулевых. Построенные числовые алгоритмы, обеспечивают расчет статистических характеристик χ^2 распределения при коррелированной последовательности отсчетов.

В заключение отметим, что алгоритмы решения задач принятия статистических решений содержат заключительный этап сравнения двух величин – наблюдаемого критерия и порогового значения (квантиля). В известных статистических моделях [1-3] исходно допускается предположение о некоррелированности наблюдаемых отсчетов. Учет их коррелированности 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. Derin H. Discrete-Index Markov Of Tipe Random Processes / H. Derin, A. Kelly // Proc. IEEE, vol. 77, № 10, pp.1485-1510, 1989. – P. 63–74. 4. Корн Г. Справочник по математике для научных работников и инженеров / Г. Корн, Т. Корн. – М. : Наука. – 1987. – 360 с. 5. Вентцель Е. С. Теория случайных процессов и ее инженерные приложения / Е. С. Вентцель, Л. А. Овчаров. – М. : Высшая школа. – 2000. – 383 с. 6. Ивченко Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев. – М. : Высшая школа, 1984. – 248 с. 7. Тихонов В. И. Марковские процессы / В. И. Тихонов, М. А. Миронов. – М. : Сов. радио, 1977. – 488 с. 8. Мазманишвили А. С. Континуальное интегрирование как метод решения физических задач / А. С. Мазманишвили. – К. : Наукова думка, 1987. – 224 с.

Надійшла до редколегії 05.12.2012