ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 23 МАТЕМАТИКИ Выпуск 4
2016

Д. П. К о ж а н (Москва, Московский Банк ПАО Сбербанк). **Преобразование зависимости. Проверка значимости коэффициента корреляции для любых видов распределений.**

При построении многомерных контрольных карт, а также в банковской деятельности при прогнозировании, анализе связности компаний возникает задача исследования парной коррелируемости компонент двумерной выборки. Задача проверки значимости коэффициента корреляции Пирсона r решена (Р. Фишером) аналитически *только в случае гауссовости компонент* [1, с. 391, 2, с. 66].

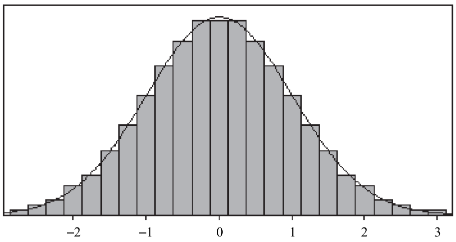

Рис. 1. Гауссовский случай

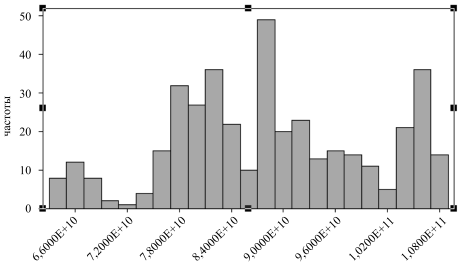

Рис. 2. Распределение реального финансового показателя

На практике большинство распределений имеет вид графика на рис. 2. Отличаясь от гауссовых, они практически никогда не входят в конкретное семейство [3, гл. 4 и 15]. Нахождение вида распределения во втором случае проблематично: можно попросту увязнуть в поиске расщепления на смеси, а, в случае выборок малых объемов (сотен), добраться до результата и вовсе нереально. Да и тратить на это время в практических задачах нерационально.

На основе обобщенного М. Розенблатом [4] фундаментального результата Н. В. Смирнова о равномерности на [0,1] отображения случайной величины своей функцией распределения, С. Я. Шатских ввел *преобразование независимости* случайных векторов [5]. Сформулируем по аналогии *преобразование зависимости*, которое позволит проверять значимость *коэффициента корреляции непараметрически*, т. е. для любых видов распределений.

Если $X \sim F_X(x)$, $Y \sim G_Y(y)$, то $\xi = \Phi^{-1}(F(X)) \sim \mathbf{N}(0,1)$, $\eta = \Phi^{-1}(G(X)) \sim \mathbf{N}(0,1)$, где $\Phi^{-1}(t)$ — обратная к стандартной нормальной. При таком отображении коррелируемость сохранится (так как она «убирается» только преобразованием Розенблата [4]), и значимость проверяется уже для (ξ,η) с использованием распределения Стьюдента [1, с. 391].

Парадоксально, что этот важнейший результат покрывает все классы прикладных распределений при помощи так редко [3, гл. 4 и 15] встречающегося в практике гауссовского.

В связи с вышеизложенным предлагаются два практических способа решения задачи.

Первый способ. По двум исходным выборкам из произвольных распределений строим эмпирические функции. По теореме Гливенко-Кантелли при $n \to \infty$ ЭФР сходятся почти наверное к своим теоретическим аналогам (равномерная сходимость на всей оси с вероятностью 1). Для непрерывных функций скорость сходимости ЭФР к своему теоретическому аналогу имеет порядок $1/\sqrt{n}$, поэтому, начиная с объема в 400, будет достигнута приемлемая для прикладных задач точность (более 95%). Полученные образы выборок являются выборками из равномерного распределения. Возвращаем по ним уже как по значениям функции распределения квантили двух стандартных нормальных распределений. А затем по ним вычисляем r и проверяем его значимость. Минус этого способа — лишний шаг, которого лишен

Второй способ. Сначала также, как и в первом способе, отображаем исходную двумерную выборку в двумерную равномерную, затем вычисляем по равномерным r и проверяем его значимость, используя

Алгоритм проверки значимости для равномерных распределений. Решение сводится к случаю $\rho = 0$, так как на практике мы сравниваем выборочный коэффициент корреляции r с нулем. Генерируем выборку объема 10 000 из двумерного равномерного распределения на [0,1] из независимых компонент. Строим кросс-корреляционную функцию, далее по ее значениям — гистограмму и эмпирическую функцию распределения, табулируем их. На уровне значимости 5% оно гауссово с MX=0, DX =0,0087. Эмпирический критерий проверки значимости построен.

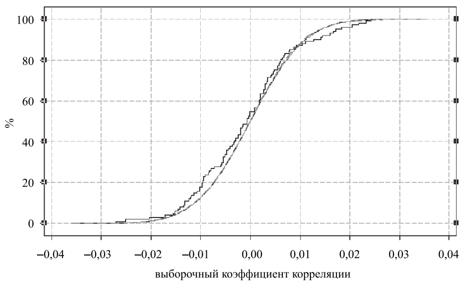

Рис. 3. ЭФР для выборок объемов 100 и 10000, N(0,0087)

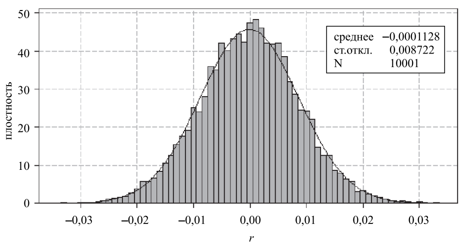

Рис. 4. Гистограмма выборочного коэффициента корреляции и гауссова плотность

СПИСОК ЛИТЕРАТУРЫ

1. *Кендалл М.*, *Стюарт А.* Статистические выводы и связи. М.: Наука, 1973.
2. *Айвазян С. А.*, *Енюков И. С.*, *Мешалкин Л. Д.* Прикладная статистика: Исследование зависимостей. М.: Финансы и статистика, 1985, 487 с.
3. *Орлов А. И.* Эконометрика. Учебник. М.: изд-во «Экзамен», 2002.
4. *Rosenblatt M.* Remarks on a multivariate transformation. — Ann. Math. Statist., 1952, v. 23, №3, p. 470–472.
5. *Шатских С. Я.* Об одном варианте преобразования независимости. — Теория вероятн. и ее примен., 1992, т. 37, в. 4, с. 815–816.