Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторной работе

Двумерная случайная величина: выборочные коэффициенты корреляции и эллипсы рассеивания

По дисциплине «Теория вероятностей и Математическая статистика»

Выполнил
Студент гр. 3630201/80101 _______ М. Д. Маляренко

Руководитель
к.ф.-м.н., доцент _______ А. Н. Баженов

«___»__________ 2020г.

Санкт-Петербург
2020

Содержание

Список иллюстраций

Список таблиц

1 Постановка задачи

Сгенерировать двумерные выборки размерами 20, 60, 100 для нормального двумерного распределения $N(x, y, 0, 0, 1, 1, \rho)$ с коэффициентами корреляции $\rho = 0,\ 0.5,\ 0.9$. Изобразить сгенерированные точки на плоскости и нарисовать эллипс рассеивания. Каждую выборку сгенерировать 1000 раз и вычислить среднее значение, среднее значение квадрата и дисперсию

1. Выборочных коэффициентов корреляции Пирсона
2. Выборочных коэффициентов корреляции Спирмена
3. Выборочных квадратных коэффициентов

Повторить все вычисления для смеси нормальных распределений:

$$f(x, y) = 0.9N(x, y, 0, 0, 1, 1, 0.9) + 0.1N(x, y, 0, 0, 10, 10, -0.9) \tag{1}$$

2 Теория

2.1 Двумерное нормальное распределение

Двумерная случайная величина (X, Y) называется распределённой нормально (или просто нормальной), если её плотность вероятности определена формулой

$$N(x, y, \overline{x}, \overline{y}, \sigma_x, \sigma_y, \rho) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \\ \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\overline{x})^2}{\sigma_x^2} - 2\rho\frac{(x-\overline{x})(y-\overline{y})}{\sigma_x\sigma_y} + \frac{(y-\overline{y})^2}{\sigma_y^2}\right]\right\} \quad (2)$$

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями $\overline{x}, \overline{y}$ и средними квадратическими отклонениями σ_x, σ_y соответственно [1]. Параметр ρ называется коэффициентом корреляции.

2.2 Корреляционный момент (ковариация) и коэффициент корреляции

Корреляционный момент, иначе ковариация, двух случайных величин X и Y:

$$K = cov(X, Y) = M\left[(X-\overline{x})(Y-\overline{y})\right] \quad (3)$$

Коэффициент корреляции ρ двух случайных величин X и Y:

$$\rho = \frac{K}{\sigma_x\sigma_y} \quad (4)$$

2.3 Выборочные коэффициенты корреляции

2.3.1 Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n}\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\frac{1}{n}\sum(x_i-\overline{x})^2\frac{1}{n}\sum(y_i-\overline{y})^2}} = \frac{K}{s_X s_Y}, \quad (5)$$

где K, s_X^2, s_Y^2 – выборочные ковариация и дисперсии с.в. X и Y.

2.3.2 Выборочный коэффициент ранговой корреляции Спирмена

Обозначим ранги, соответствующие значениям переменной X, через u, а ранги, соответствующие значениям переменной Y, — через v.

Выборочный коэффициент ранговой корреляции Спирмена :

$$r_S = \frac{\frac{1}{n}\sum(u_i-\overline{u})(v_i-\overline{v})}{\sqrt{\frac{1}{n}\sum(u_i-\overline{u})^2\frac{1}{n}\sum(v_i-\overline{v})^2}}, \quad (6)$$

где $\overline{u} = \overline{v} = \frac{1+2+\ldots+n}{n} = \frac{n+1}{2}$ – среднее значение рангов.

2.3.3 Выборочный квадрантный коэффициент корреляции

Выборочный квадрантный коэффициент корреляции

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n}, \tag{7}$$

где n_1, n_2, n_3, n_4 – количества точек с координатами (x_i, y_i), попавшими соответственно в I, II, III и IV квадранты декартовой системы с осями $x' = x - med\, x, y' = y - med\, y$ и с центром в точке с координатами $(med\, x, med\, y)$.

2.4 Эллипсы рассеивания

Уравнение проекции эллипса рассеивания на плоскость xOy:

$$\frac{(x-\overline{x})^2}{\sigma_x^2} - 2\rho\frac{(x-\overline{x})(y-\overline{y})}{\sigma_x\sigma_y} + \frac{(y-\overline{y})^2}{\sigma_y^2} = const. \tag{8}$$

Центр эллипса (8) находится в точке с координатами $(\overline{x}, \overline{y})$; оси симметрии эллипса составляют с осью Ox углы, определяемые уравнением

$$tg 2\alpha = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}. \tag{9}$$

3 Реализация

Расчёты и построение графиков производились в среде аналитических вычислений Maxima. Для построения выборок двумерного нормального распределения, вычисления выборочных коэффициентов корреляции Пирса, Спирмена и квадратного выборочного коэффициента корреляции были написаны функции соответственно `random2_normal`, `r`, `r_S` и `r_Q`. Для построения графиков использовалась интегрированная в среду утилита gnuplot. Полный текст скрипта для среды Maxima представлен в репозитории GitHub.

4 Результаты

4.1 Выборочные коэффициенты корреляции

В Таблице 1 представлены выборочные коэффициенты корреляции Прирса (5), Спирмена (6) и квадратный коэффициент корреляции (7) для выборок размера 20, 60 и 100 элементов двумерного нормального распределения $N(x, y, 0, 0, 1, 1, \rho)$ с коэффициентами корреляции $\rho = 0,\ 0.5,\ 0.9$.

		r			r_Q			r_S		
		$E(z)$	$E(z^2)$	$D(z)$	$E(z)$	$E(z^2)$	$D(z)$	$E(z)$	$E(z^2)$	$D(z)$
$\rho = 0$	$N = 20$	0.0	0.05	0.05	0.0	0.05	0.05	0.0	0.05	0.05
	$N = 60$	0.0	0.02	0.02	0.0	0.02	0.02	0.0	0.02	0.02
	$N = 100$	0.0	0.01	0.01	0.0	0.01	0.01	0.0	0.01	0.01
$\rho = 0.5$	$N = 20$	0.5	0.27	0.03	0.3	0.16	0.05	0.5	0.25	0.03
	$N = 60$	0.5	0.25	0.01	0.33	0.12	0.01	0.5	0.24	0.01
	$N = 100$	0.49	0.253	0.006	0.33	0.120	0.009	0.48	0.234	0.006
$\rho = 0.9$	$N = 20$	0.89	0.800	0.003	0.68	0.50	0.03	0.87	0.749	0.005
	$N = 60$	0.89	0.809	0.007	0.70	0.505	0.009	0.88	0.782	0.001
	$N = 100$	0.90	0.806	0.004	0.71	0.506	0.005	0.88	0.7837	0.0007

Таблица 1: Выборочные коэффициенты корреляции для двумерного нормального распределения

В Таблице 2 представлены выборочные коэффициенты корреляции Пирса, Спирмена и квадратный коэффициент корреляции для выборок смеси двумерных нормальных распределений (1) размера 20, 60 и 100 элементов.

	r			r_Q			r_S		
	$E(z)$	$E(z^2)$	$D(z)$	$E(z)$	$E(z^2)$	$D(z)$	$E(z)$	$E(z^2)$	$D(z)$
$N = 20$	-0.3	0.5	0.4	0.5	0.32	0.04	0.5	0.29	0.08
$N = 60$	-0.6	0.49	0.09	0.6	0.34	0.01	0.5	0.26	0.03
$N = 100$	-0.7	0.52	0.03	0.56	0.318	0.007	0.5	0.24	0.02

Таблица 2: Выборочные коэффициенты корреляции для смеси двумерных нормальных распределений

4.2 Эллипсы рассеивания

На Рис. 1 - 3 представлены графики выборок размера 20, 60 и 100 двумерного нормального распределения $N(x, y, 0, 0, 1, 1, \rho)$ с коэффициентом корреляции $\rho = 0,\ 0.5,\ 0.9$, а также теоретические эллипсы рассеивания, рассчитанными по формуле (8).

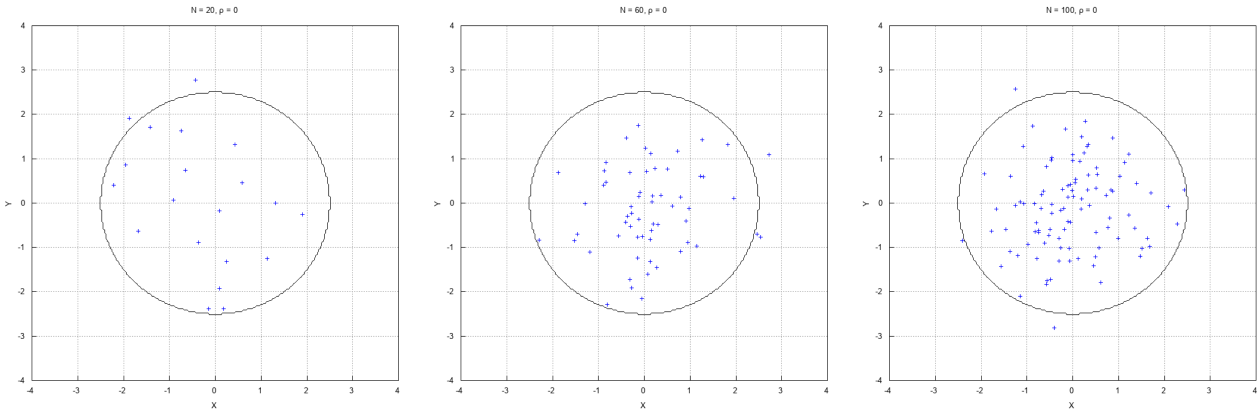

Рис. 1: Эллипсы рассеивания для выборок нормального распределения с $\rho = 0$

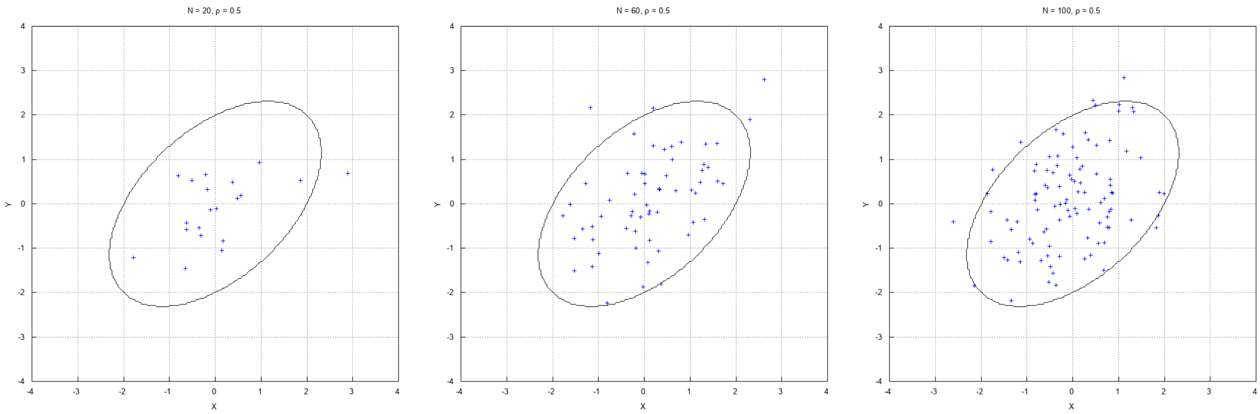

Рис. 2: Эллипсы рассеивания для выборок нормального распределения с $\rho = 0.5$

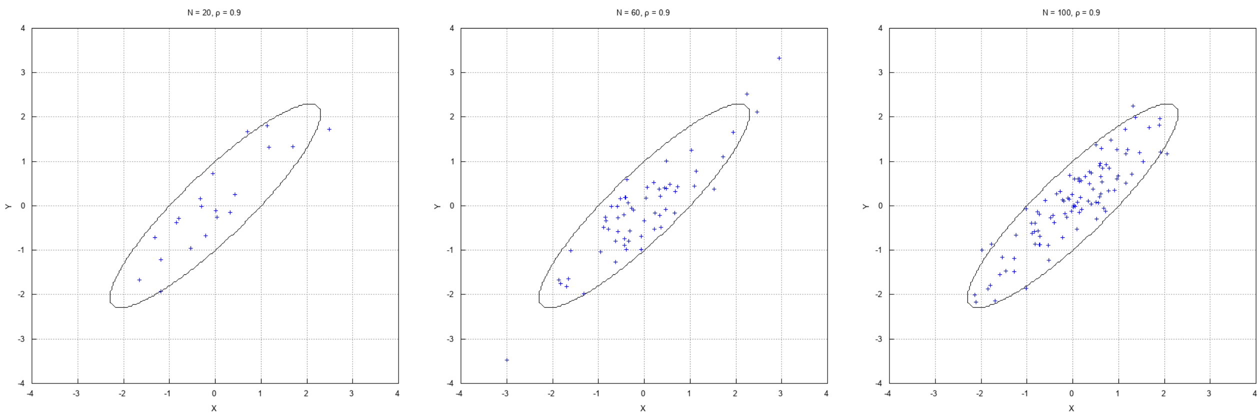

Рис. 3: Эллипсы рассеивания для выборок нормального распределения с $\rho = 0.9$

Заключение

В результате выполнения лабораторной работы были построены выборки двумерного нормального распределения с коэффициентами корреляции $\rho = 0,\ 0.5,\ 0.9$ и для смеси нормальных распределений. По оценкам выборочных коэффициентов корреляции можно сказать, что квадратный выборочный коэффициент имеет наибольшее отклонение от теоретического коэффициента корреляции, с увеличением мощности выборки все выборочные коэффициенты корреляции стремятся к своему теоретическому значению.

По графикам двумерного нормального распределения видно, что чем больше коэффициент корреляции, тем более узкий эллипс рассеивания, в пределе при модуле коэффициента корреляции равном единице эллипс вырождается в прямую.

Список литературы

[1] Теоретическое приложение к лабораторным работам №5-8 по дисциплине «Математическая статистика». – СПб.: СПбПУ, 2020. – 19 с

Приложение А. Репозиторий с исходным кодом

Исходный код скрипта для среды аналитических вычислений Maxima находится в репозитории GitHub – URL `https://github.com/malyarenko-md/TeorVer`