Санкт-Петербургский политехнический университет Петра Великого Физико-механический институт Высшая школа прикладной математики и вычислительной физики

Отчёт
по лабораторным работам №5
по дисциплине
«Математическая статистика»

Выполнил студент:
Басалаев Даниил Александрович
группа:
5030102/10201
Проверил:
доцент Баженов Александр Николаевич

Санкт-Петербург

1 Постановка задачи

Сгенерировать двумерные выборки размерами 20, 60, 100 для нормального двумерного распределения $N(x, y, 0, 0, 1, 1, \rho)$. Коэффициент корреляции ρ взять равным 0, 0.5, 0.9. Каждая выборка генерируется 1000 раз и для неё вычисляются: среднее значение, среднее значение квадрата и дисперсия коэффициентов корреля- ции Пирсона, Спирмена и квадрантного коэффициента корреляции. Повторить все вычисления для смеси нормальных распределений:

$$f(x, y) = 0.9 * N(x, y, 0, 0, 1, 1, 0.9) + 0.1 * N(x, y, 0, 0, 10, 10, -0.9).$$

Изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.

2 Теоретическое обоснование

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

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями x, y и средними квадратическими отклонениями σ_x, σ_y соответственно.

Параметр ρ называется коэффициентом корреляции.

2.2 Корреляционный момент (ковариация) и коэффициент корреляции

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

$$K = \mathbf{cov}(X, Y) = \mathbf{M}[(X - \bar{x})(Y - \bar{y})] \tag{1}$$

Коэффициент корреляции ρ двух случайных величин X и Y:

$$\rho = \frac{K}{\sigma_x\sigma_y} \tag{2}$$

2.3 Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2\frac{1}{n}\sum(y_i - \bar{y})^2}} = \frac{K}{s_X s_Y}, \tag{3}$$

где K, s_X^2, x_Y^2 — выборочные ковариации и дисперсии случайных величин X и Y.

2.4 Выборочный квадрантный коэффициент корреляции

Выборочный квадрантный коэффициент корреляции

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n}, \tag{4}$$

где n_1, n_2, n_3, n_4 — количество точек с координатами (x_i, y_i), попавшими, соответственно, в I, II, III, IV квадранты декартовой системы с осями $x' = x - \mathbf{med}x$, $y' = y - \mathbf{med}y$.

2.5 Выборочный коэффициент ранговой корреляции Спирмена

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

Выборочный коэффициент ранговой корреляции Спирмена:

$$r_S = \frac{\frac{1}{n}\sum(u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\frac{1}{n}\sum(u_i - \bar{u})^2 \frac{1}{n}\sum(v_i - \bar{v})^2}}, \tag{5}$$

где $\bar{u} = \bar{v} = \frac{1+2+...+n}{n} = \frac{n+1}{2}$ — среднее значение рангов.

2.6 Эллипсы рассеивания

Уравнение проекции такого эллипса на плоскость xOy:

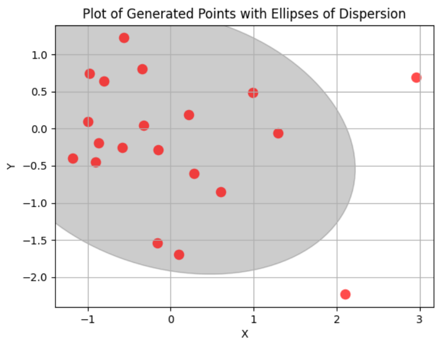
$$\frac{(x-\bar{x})^2}{\sigma_x^2} - 2\rho\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2} = const \tag{6}$$

Уравнение эллипса 6 можно проанализировать обычными методами аналитической геометрии. Применяя их, убеждаемся, что центр эллипса 6 находится в точке с координатами $(\bar{x}, \bar{y})$; что касается направления осей симметрии эллипса, то они составляют с осью Ox углы, определяемые уравнением

$$tg(2\alpha) = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \tag{7}$$

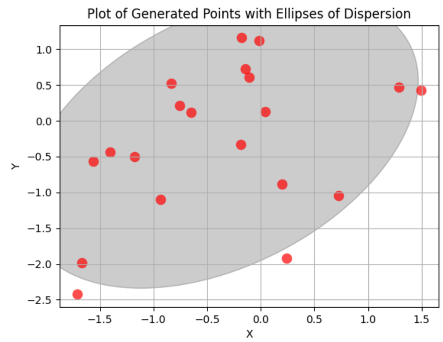
3 Графики

3.1 Нормальное распределение выборка 20



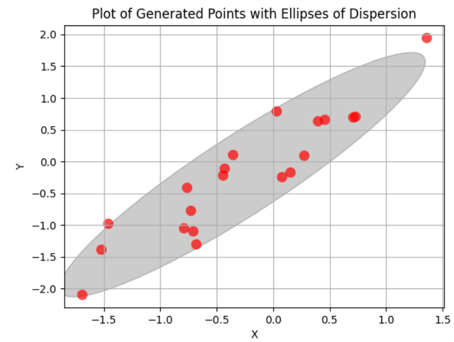
ρ = 0

```
mean:
    pearson: -0.003632
    spearman: -0.001427
    quadrant: 0.24965
mean square:
    pearson: 0.051583
    spearman: 0.051291
    quadrant: 0.067693
variance:
    pearson: 0.05157
    spearman: 0.051289
    quadrant: 0.005367
```



ρ = 0.5

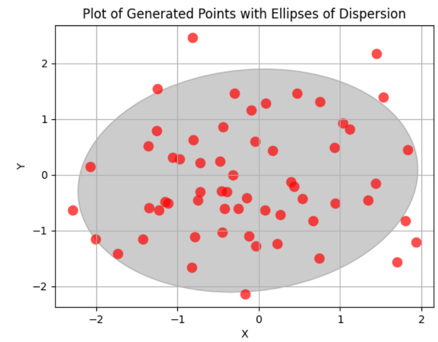
```
mean:
    pearson: 0.484857
    spearman: 0.449146
    quadrant: 0.33195
mean square:
    pearson: 0.26472
    spearman: 0.235238
    quadrant: 0.115098
variance:
    pearson: 0.029634
    spearman: 0.033506
    quadrant: 0.004907
```



ρ = 0.9

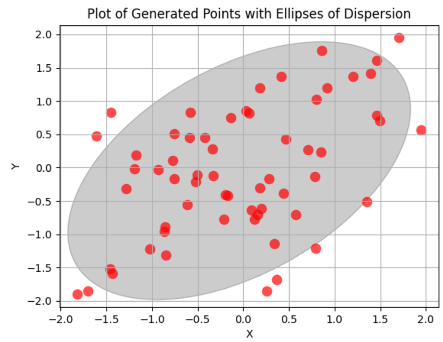
```
mean:
    pearson: 0.895924
    spearman: 0.865934
    quadrant: 0.43085
mean square:
    pearson: 0.805089
    spearman: 0.754686
    quadrant: 0.190603
variance:
    pearson: 0.00241
    spearman: 0.004845
    quadrant: 0.004971
```

3.2 Нормальное распределение выборка 60



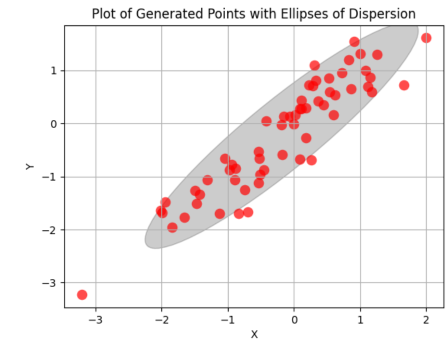
ρ = 0

```
mean:
    pearson: 0.007866
    spearman: 0.007409
    quadrant: 0.251617
mean square:
    pearson: 0.016391
    spearman: 0.016891
    quadrant: 0.065106
variance:
    pearson: 0.016329
    spearman: 0.016836
    quadrant: 0.001795
```



ρ = 0.5

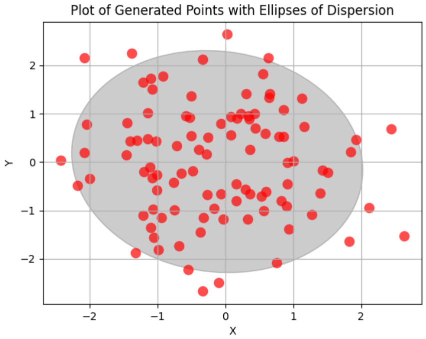
```
mean:
    pearson: 0.498591
    spearman: 0.479156
    quadrant: 0.3338
mean square:
    pearson: 0.258045
    spearman: 0.240227
    quadrant: 0.113071
variance:
    pearson: 0.009451
    spearman: 0.010636
    quadrant: 0.001649
```



ρ = 0.9

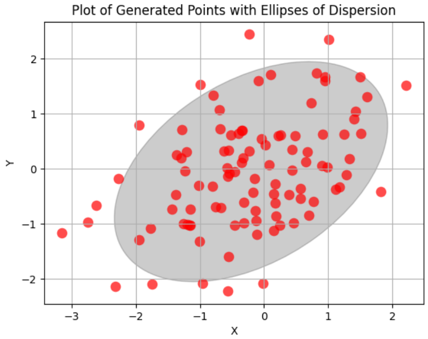
```
mean:
    pearson: 0.897508
    spearman: 0.881108
    quadrant: 0.426767
mean square:
    pearson: 0.806169
    spearman: 0.777515
    quadrant: 0.183676
variance:
    pearson: 0.000648
    spearman: 0.001164
    quadrant: 0.001546
```

3.3 Нормальное распределение выборка 100



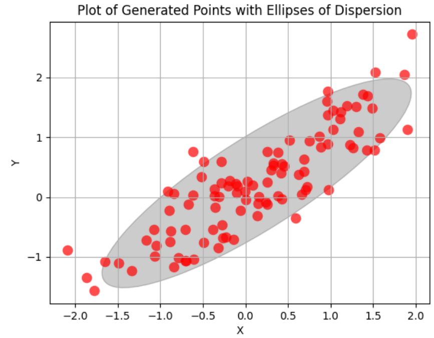
ρ = 0

```
mean:
    pearson: 0.003951
    spearman: 0.003704
    quadrant: 0.25088
mean square:
    pearson: 0.010635
    spearman: 0.010941
    quadrant: 0.064047
variance:
    pearson: 0.010619
    spearman: 0.010928
    quadrant: 0.001106
```



ρ = 0.5

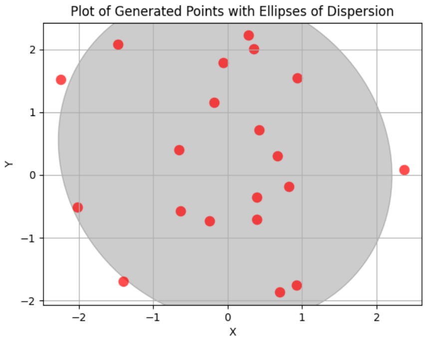
```
mean:
    pearson: 0.498584
    spearman: 0.479436
    quadrant: 0.33436
mean square:
    pearson: 0.236173
    spearman: 0.236173
    quadrant: 0.112819
variance:
    pearson: 0.005635
    spearman: 0.006314
    quadrant: 0.001022
```



ρ = 0.9

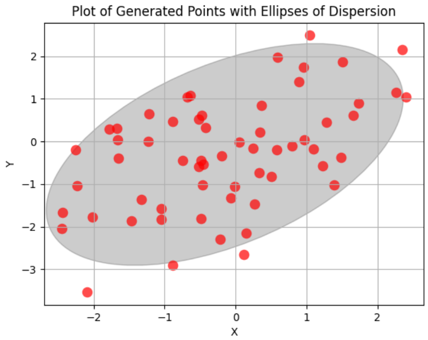
```
mean:
    pearson: 0.899356
    spearman: 0.886931
    quadrant: 0.42872
mean square:
    pearson: 0.809238
    spearman: 0.184687
    quadrant: 0.184687
variance:
    pearson: 0.000398
    spearman: 0.000625
    quadrant: 0.000886
```

3.4 Смешанное распределение f(x, y) = 0.9N(x,y,0,0,1,10.9) + 0.1N(x,y,0,0,10,10,-0.9)



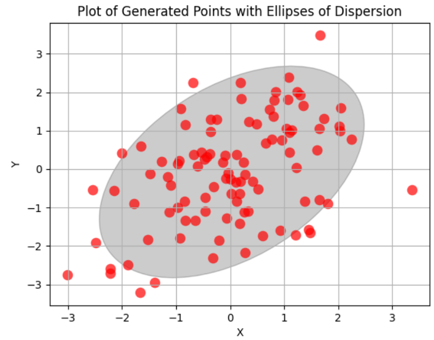
size = 20

```
mean:
    pearson: 0.343782
    spearman: 0.325947
    quadrant: 0.309800
mean square:
    pearson: 0.159329
    spearman: 0.148703
    quadrant: 0.101345
variance:
    pearson: 0.041143
    spearman: 0.042462
    quadrant: 0.005369
```



size = 60

```
mean:
    pearson: 0.358987
    spearman: 0.339783
    quadrant: 0.307767
mean square:
    pearson: 0.142658
    spearman: 0.129829
    quadrant: 0.096448
variance:
    pearson: 0.013786
    spearman: 0.014376
    quadrant: 0.001728
```



size = 100

```
mean:
    pearson: 0.367462
    spearman: 0.352392
    quadrant: 0.31016
mean square:
    pearson: 0.142802
    spearman: 0.132416
    quadrant: 0.09729
variance:
    pearson: 0.007774
    spearman: 0.008237
    quadrant: 0.001091
```

4 Выводы

При увеличении размера выборки наблюдается повышение точности оценок, что проявляется в уменьшении дисперсий коэффициентов корреляции. Этот эффект соответствует основным принципам центральной предельной теоремы и закона больших чисел. Увеличение коэффициента корреляции ρ сопровождается ростом средних значений коэффициентов Пирсона, Спирмена и квадратичного коэффициента корреляции вследствие прямой зависимости между ρ и другими коэффициентами корреляции.

Метод наименьших квадратов проявляет эффективность в условиях, когда данные не содержат значительных выбросов, в то время как метод наименьших модулей демонстрирует превосходство в случае наличия значительных возмущений. При выборе метода следует учитывать особенности данных: при наличии выбросов предпочтительнее использовать метод наименьших модулей из-за его устойчивости к выбросам.

5 GitHub

`https://github.com/11AgReS1SoR11/MatStat/tree/main/Laba5`