Непрерывная двумерная случайная величина

Пример решения задачи

Задание. Дана плотность распределения $f(x,y)$ системы (X,Y) двух непрерывных случайных величин в треугольнике ABC.

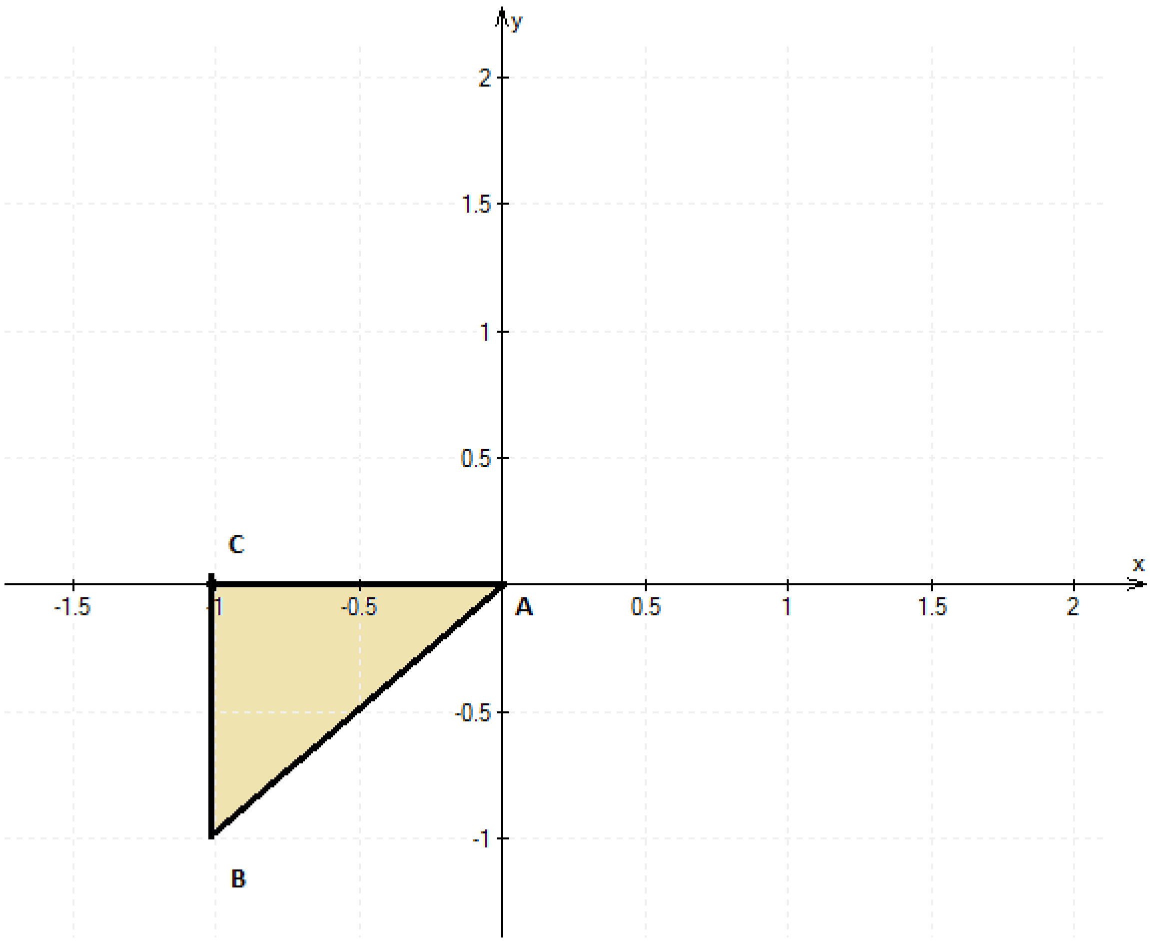
1.1. Найдите константу c.
1.2. Найдите $f_X(x)$, $f_Y(y)$ - плотности распределения с.в. X и с.в. Y. Выясните, зависимы или нет с.в. X и Y. Сформулируйте критерий независимости системы непрерывных случайных величин.
1.3. Найдите математическое ожидание и дисперсию с.в. X и с.в. Y. Поясните смысл найденных характеристик.
1.4. Найдите коэффициент корреляции с.в. X и Y. Являются ли случайные величины коррелированными? Сформулируйте свойства коэффициента корреляции.
1.5. Запишите уравнение регрессии с.в. Y на X и постройте линию регрессии в треугольнике ABC.
1.6. Запишите уравнение линейной среднеквадратичной регрессии с.в. Y на X и постройте эту прямую в треугольнике ABC.

$f(x,y)=c\sqrt{xy}$	A(0;0), B(-1;-1), C(-1;0)

Решение.

Найдем c из условия нормировки: $\iint\limits_D f(x,y)dxdy=1$, здесь за D мы для краткости обозначили весь треугольник $\triangle ABC$, на котором функция плотности распределения принимает ненулевые значения.

Сделаем чертеж для удобства расчетов.



Нижняя прямая, ограничивающая область, имеет уравнение $y = x$.

Получаем:

$$\iint_D f(x,y)dxdy = c\iint_D \sqrt{xy}dxdy = c\int_{-1}^{0}\sqrt{x}dx\int_{x}^{0}\sqrt{y}dy = c\int_{-1}^{0}\sqrt{x}dx\left(\frac{2}{3}\sqrt{y^3}\right)\Big|_x^0$$

$$= c\int_{-1}^{0}\sqrt{x}\left(-\frac{2}{3}\sqrt{x^3}\right)dx = -\frac{2}{3}c\int_{-1}^{0}x^2dx = -\frac{2}{9}cx^3\Big|_{-1}^{0} = -\frac{2}{9}c = 1.$$

Отсюда получаем $c = -\frac{9}{2}$, $f(x,y) = -\frac{9}{2}\sqrt{xy}$, $(x,y)\in D$.

Найдем $f_X(x)$, $f_Y(y)$ - плотности распределения с.в. X и с.в. Y.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = -\frac{9}{2}\int_{x}^{0}\sqrt{xy}dy = -\frac{9}{2}\frac{2}{3}\left(\sqrt{xy^3}\right)\Big|_x^0 = 3x^2,\ x\in[-1;0].$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = -\frac{9}{2}\int_{-1}^{y}\sqrt{xy}dy = -\frac{9}{2}\frac{2}{3}\left(\sqrt{yx^3}\right)\Big|_{-1}^{y} = 3\left(\sqrt{-y} - y^2\right),\ y\in[-1;0].$$

Выясним, зависимы или нет с.в. X и Y. Рассмотрим произведение:

$$f_X(x)f_Y(y)=3x^2 3\left(\sqrt{-y}-y^2\right)=9x^2\left(\sqrt{-y}-y^2\right)\neq -\frac{9}{2}\sqrt{xy}=f(x,y),$$

поэтому X и Y зависимы.

Сформулируем критерий независимости системы непрерывных случайных величин.
Теорема. *Для того, чтобы случайные величины X и Y были независимы, необходимо и достаточно, чтобы плотность совместного распределения системы (X, Y) была равна произведению плотностей распределения составляющих.*

$$f(x,y)=f_1(x)f_2(y)$$

Найдем математическое ожидание и дисперсию с.в. X и с.в. Y.

Найдем математические ожидания:

$$MX=\int_{-\infty}^{\infty} f_X(x)x\,dx=3\int_{-1}^{0} x^3dx=\frac{3}{4}x^4\Big|_{-1}^{0}=-\frac{3}{4}.$$

$$MY=\int_{-\infty}^{\infty} f_Y(y)y\,dx=3\int_{-1}^{0}\left(\sqrt{-y}-y^2\right)y\,dx=3\int_{-1}^{0}\left(\sqrt{(-y)^3}-y^3\right)dx=3\left(\frac{2}{5}(-y)^{5/2}-\frac{1}{4}y^4\right)\Big|_{-1}^{0}=$$

$$=-3\left(\frac{2}{5}-\frac{1}{4}\right)=-\frac{9}{20}.$$

Найдем дисперсии:

$$DX=\int_{-\infty}^{\infty} f_x(x)x^2dx-(MX)^2=3\int_{-1}^{0}x^4dx-\left(-\frac{3}{4}\right)^2=\frac{3}{5}x^5\Big|_{-1}^{0}-\frac{9}{16}=\frac{3}{5}-\frac{9}{16}=\frac{3}{80}.$$

$$DY=\int_{-\infty}^{\infty} f_y(y)y^2dx-(MY)^2=3\int_{-1}^{0}\left(\sqrt{-y}-y^2\right)y^2dx-\left(-\frac{9}{20}\right)^2=3\int_{-1}^{0}\left(\sqrt{(-y)^5}-y^4\right)dx-\frac{81}{400}=$$

$$=3\left(-\frac{2}{7}(-y)^{7/2}-\frac{1}{5}y^5\right)\Big|_{-1}^{0}-\frac{81}{400}=-3\left(-\frac{2}{7}+\frac{1}{5}\right)-\frac{81}{400}=\frac{9}{35}-\frac{81}{400}=\frac{153}{2800}.$$

Средние квадратические отклонения: $\sigma_X=\sqrt{\frac{3}{80}}$, $\sigma_Y=\sqrt{\frac{153}{2800}}$.

Математическое ожидание характеризует среднее ожидаемое значение случайной величины, дисперсия – квадрат величины рассеивания относительно математического ожидания. Точка $(MX;MY)=\left(-\frac{3}{4};-\frac{9}{20}\right)$ называется центром распределения величины (X,Y).

Найдем коэффициент корреляции с.в. X и Y. Вычислим ковариацию:

$$K_{XY}=\iint_D (x-MX)(y-MY)f(x,y)dxdy=-\frac{9}{2}\iint_D \sqrt{xy}\,(x+3/4)(y+9/20)\,dxdy=\frac{9}{400}.$$

Значит, коэффициент корреляции равен

$$r_{XY}=\frac{K_{XY}}{\sigma_X\sigma_Y}=\frac{9/400}{\sqrt{3/80}\cdot\sqrt{153/2800}}=\frac{\sqrt{1785}}{85}\approx 0,497.$$

Случайные величины являются коррелированными, так как коэффициент отличен от нуля. Сформулируем свойства коэффициента корреляции: он находится в пределах от -1 до 1.

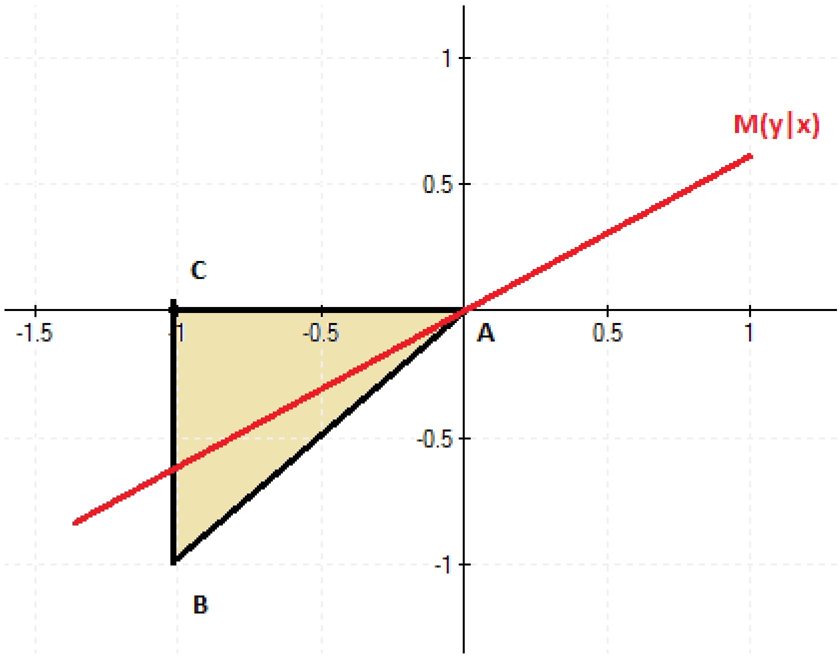
Запишем уравнение регрессии с.в. Y на X. Найдем условную плотность распределения:

$$f(y\,|\,x)=\frac{f(x,y)}{f_X(x)}=\frac{-\frac{9}{2}\sqrt{xy}}{3x^2}=-\frac{3}{2}\sqrt{\frac{y}{x^3}}.$$

Тогда уравнение регрессии найдем по формуле:

$$M\left(y\,|\,x\right)=\int_{-\infty}^{\infty} yf(y\,|\,x)dy=\int_x^0 -\frac{3}{2}y\sqrt{\frac{y}{x^3}}dy=-\frac{3}{2}\frac{2}{5}\sqrt{\frac{y^5}{x^3}}\Bigg|_x^0=\frac{3}{5}x=0,6x.$$

Построим линию регрессии в треугольнике ABC.



Запишем уравнение линейной среднеквадратичной регрессии с.в. Y на X по формуле:

$$g(X)=MY+r_{XY}\frac{\sigma_Y}{\sigma_X}(X-MX).$$

Получаем:

$$g(X) = -\frac{9}{20} + 0,497\frac{\sqrt{153/2800}}{\sqrt{3/80}}\left(X + \frac{3}{4}\right),$$
$$g(X) = -0,00005 + 0,6x \approx 0,6x.$$

Построим эту прямую в треугольнике ABC.