Topic 6: Joint Distributions

Rohini Somanathan

Course 003, 2018

Joint distributions

Social scientists are typically interested in the relationship between many random variables. They may be able to change some of these and would like to understand the effects on others.

Examples:

Education and earnings

Height and longevity

Attendance and learning outcomes

Sex-ratios and areas under rice cultivation

Genetic make-up and disease

All these problems use the joint distribution of two or more random variables.

We will define multivariate CDFs, PMFs and PDFs and see how to go from joint to marginal and conditional distributions.

We will also study multivariate extensions of some of special distributions we have considered.

Joint CDFs

Definition (Joint CDF): *The joint CDF of random variables* X *and* Y *is the function* $F_{X,Y}$ *given by:*

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

Like the univariate CDF, this definition of the joint CDF applies both to discrete and continuous random variables.

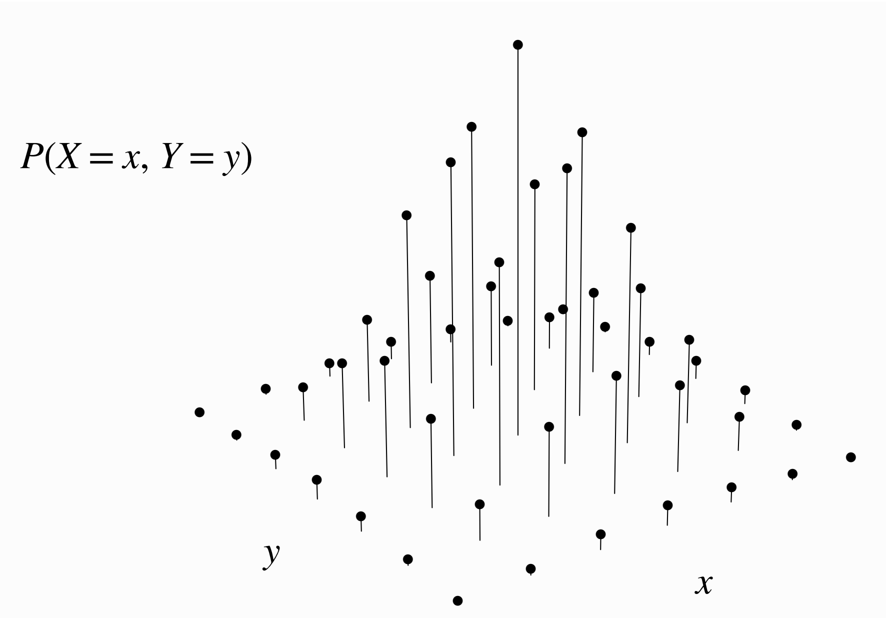
PMFs of discrete random variables

Definition (Joint PMF): *The joint PMF of discrete random variables* X *and* Y *is the function* $p_{X,Y}$ *given by:*

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

We require $p_{X,Y}(x, y)$ *to be non-negative and* $\sum_x \sum_y P(X = x, Y = y) = 1$. *For* n *r.v.s* $X_1, \ldots X_n$*:*

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$



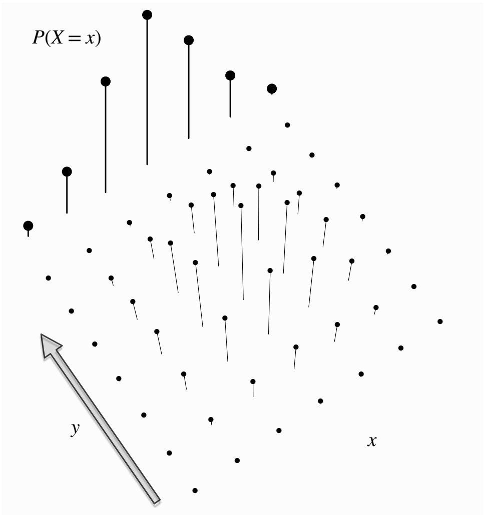
The marginal and conditional PMFs

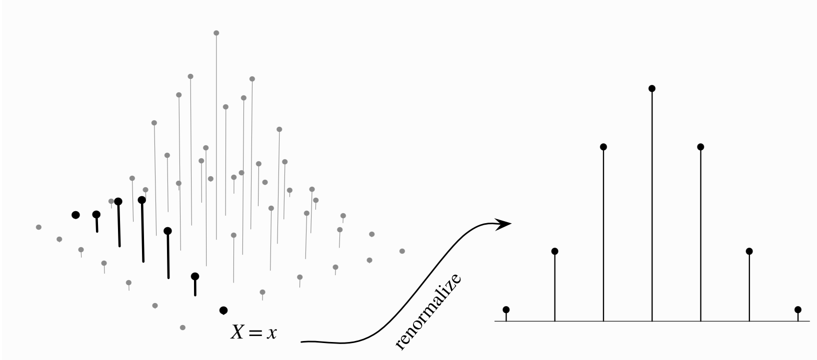
Definition (Marginal PMF): *For discrete r.v.s* $\mathbf{X}$ *and* $\mathbf{Y}$*, the marginal PMF of* $\mathbf{X}$ *is given by:*

$$P(X = x) = \sum_{y} P(X = x, Y = y)$$

Definition (Conditional PMF): *For discrete r.v.s* $\mathbf{X}$ *and* $\mathbf{Y}$*, the conditional PMF of* $\mathbf{Y}$*, given* $\mathbf{X} = \mathbf{x}$ *is given by:*

$$P(Y = y | X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$





 Rohini Somanathan

Example: Gender and education

When X and Y take only a few values, the PMF can be conveniently presented in a table:

education ↓ gender →	male	female
none	.05	.2
primary	.25	.1
middle	.15	.04
high	.1	.03
senior secondary	.03	.02
graduate and above	.02	.01

What are some features of a table like this one? In particular, how do we obtain probabilities of

receiving no education

becoming a female graduate

completing primary school

What else can you learn about this population? What is the marginal distribution of education and gender? Are gender and education are independent?

Can one construct the joint distribution from one of the marginal distributions?

Example: Bernoulli distributions

If X and Y are both Bernoulli, there are only four points in the support of the joint PMF, $p_{X,Y}$ which can be shown in a contingency table like the one below.

	Y = 1	Y = 0	Total
X = 1	.05	.2	.25
X = 0	.03	.72	.75
Total	.08	.92	1

Say X indicates smoking behavior and Y the incidence of lung disease. We see that $X \sim \text{Bern}(0.25)$ and $Y \sim \text{Bern}(0.08)$.

The conditional distribution of Y for smokers is $\text{Bern}(.2)$ and for non-smokers is $\text{Bern}(.04)$

Continuous bivariate distributions

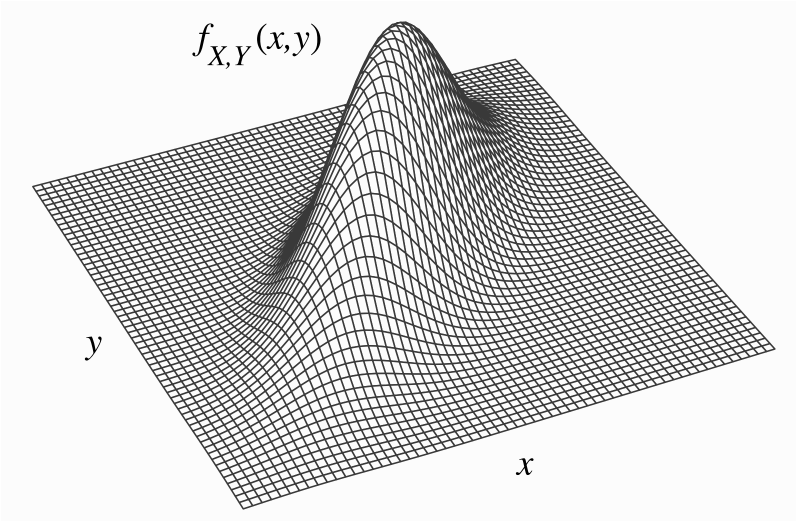
We can extend our definition of a continuous univariate distribution to the bivariate case:

Definition: *Two random variables* X *and* Y *have a continuous joint distribution if there exists a nonnegative function* f *defined over the* xy*-plane such that for any subset* A *of the plane*

$$P[(X,Y) \in A] = \int_A \int f(x,y)\,dx\,dy$$

f *is now called the joint probability density function and must satisfy*

1. $f(x,y) \geq 0$ *for* $-\infty < x < \infty$ *and* $-\infty < y < \infty$
2. $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$



Bivariate densities.. examples

Example 1: Given the following joint density function on X and Y, we'll calculate $P(X \geq Y)$

$$f(x,y) = \begin{cases} cx^2y & \text{for } x^2 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

First find c to make this a valid joint density (notice the limits of integration here)-it will turn out to be 21/4. Then integrate the density over $Y \in (x^2, x)$ and $X \in (-1, 1)$. Now using this density, $P(X \geq Y) = \frac{3}{20}$.

Example 2: A point (X, Y) is selected at random from inside the circle $x^2 + y^2 \leq 9$. To determine $f(x,y)$, we find a constant c such that the volume ($c \times$ area of S) is 1 so $c = \frac{1}{9\pi}$

Using joint CDFs

Given a joint CDF, F(x,y), the

probability that (X, Y) will lie in a specified rectangle in the xy-plane is given by

$$Pr(a < X \leq b \text{ and } c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

Note: The distinction between weak and strict inequalities is important when points on the boundary of the rectangle occur with positive probability.

and distribution functions of X and Y are derived as:

$$Pr(X \leq x) = F_1(x) = \lim_{y\to\infty} F(x, y) \text{ and } Pr(Y \leq y) = F_2(y) = \lim_{x\to\infty} F(x, y)$$

If $F(x, y)$ is differentiable, the joint density is:

$$f(x, y) = \frac{\delta^2 F(x, y)}{\delta x \delta y}$$

Example: Suppose that, for x and $y \in [0, 2]$, we have $F(x, y) = \frac{1}{16}xy(x + y)$, derive the distribution functions of X and Y and their joint density.

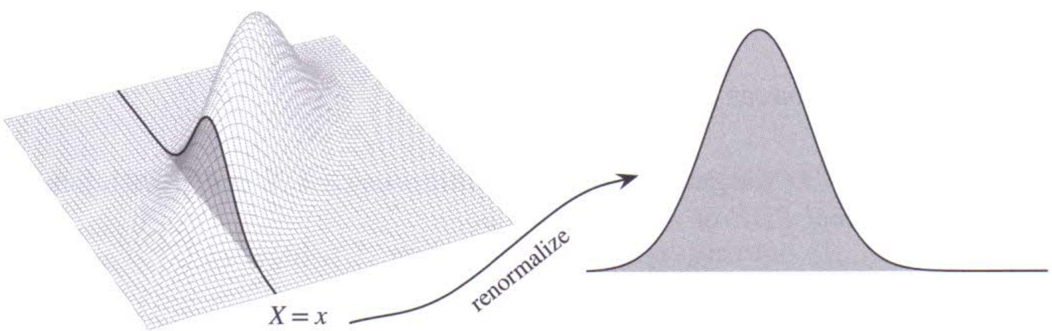
Marginal and conditional densities

For a continuous joint density $f(x, y)$*, the marginal density functions for* X *and* Y *are given by:*

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \quad \textit{and} \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$$

and the conditional probability density function of Y *given* $X = x$ *as*

$$g_2(y|x) = \frac{f(x, y)}{f_1(x)} \quad \textit{for } (-\infty < x < \infty \textit{ and } -\infty < y < \infty)$$



The total area under the function in the cross-section above is $f_1(x)$**, so dividing by this ensures that the conditional pdf integrates to 1.**

Joint densities for independent random variables

Recall that for independent r.v.s, $f(x, y) = f(x)f(y)$.

Example 1: There are two independent measurements X and Y of rainfall at a certain location:

$$g(x) = \begin{cases} 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the probability that $X + Y \leq 1$.

The joint density $4xy$ is got by multiplying the marginal densities because these variables are independent. The required probability of $\frac{1}{6}$ is then obtained by integrating over $y \in (0, 1 - x)$ and $x \in (0, 1)$

Example 2: Given the following density, can we tell whether the variables X and Y are independent?

$$f(x, y) = \begin{cases} ke^{-(x+2y)} & \text{for } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Notice that we can factorize the joint density as the product of $k_1 e^{-x}$ and $k_2 e^{-2y}$ where $k_1 k_2 = k$. To obtain the marginal densities of X and Y, we multiply these functions by appropriate constants which make them integrate to unity. This gives us the two exponential distributions:

$$f_1(x) = e^{-x} \text{ for } x \geq 0 \text{ and } f_2(y) = 2e^{-2y} \text{ for } y \geq 0$$

Dependent random variables..examples

Given the following density densities, let's see why the variables X and Y are dependent:

1. $$f(x, y) = \begin{cases} x + y & \text{for } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

   Notice that we cannot factorize the joint density as the product of a non-negative function of x and another non-negative function of y. Computing the marginals gives us

   $$f_1(x) = x + \frac{1}{2} \text{ for } 0 < x < 1 \text{ and } f_2(y) = y + \frac{1}{2} \text{ for } 0 < y < 1$$

   so the product of the marginals is not equal to the joint density.

2. Suppose we have

   $$f(x, y) = \begin{cases} kx^2y^2 & \text{for } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Independence of continuous r.v.s: a result

Result: *Whenever the space of positive probability density of* $\mathbf{X}$ *and* $\mathbf{Y}$ *is bounded by a curve, rather than a rectangle, the two random variables are dependent. If, on the other hand, the support of* $f(x, y)$ *is a rectangle and the joint density is of the form* $f(x, y) = kg(x)h(y)$*, then* $\mathbf{X}$ *and* $\mathbf{Y}$ *are independent.*

Proof: **For the first part, consider any point (x, y) outside the set where $f(x, y) > 0$. If x and y are independent, we have $f(x, y) = f_1(x)f_2(y)$, so one of these must be zero. Now as we move due south and enter the set where $f(x, y) > 0$, our value of x has not changed, so it could not be that $f_1(x)$ was zero at the original point. Similarly, if we move west, y is unchanged so it could not be that $f_2(y)$ was zero at the original point. So we have a contradiction.**

For the latter part, suppose the support of $f(x, y)$ is given by the rectangle $abcd$ where $-\infty \leq a < b \leq \infty$ and $-\infty \leq c < d \leq \infty$ and $a \leq x \leq b$ and $c \leq y \leq d$. Now the joint density $f(x, y)$ can be written as $k_1 g(x) k_2 h(y)$ where $k_1 = \frac{1}{\int_a^b g(x)dx}$ and $k_2 = \frac{1}{\int_c^d h(y)dy}$.

Conditional and joint densities..an example

Suppose we start with the following density function for a variable X_1:

$$f_1(x) = \begin{cases} e^{-x} \text{ for } x > 0 \\ 0 \quad \text{otherwise} \end{cases}$$

and are told that for any given value of $X_1 = x_1$, two other random variables X_2 and X_3 are independently and identically distributed with the following conditional p.d.f.:

$$g(t|x_1) = \begin{cases} x_1 e^{-x_1 t} \text{ for } t > 0 \\ 0 \quad \text{otherwise} \end{cases}$$

The conditional p.d.f. is now given by $g_{23}(x_2, x_3|x_1) = x_1^2 e^{-x_1(x_2+x_3)}$ for non-negative values of x_2, x_3 (and zero otherwise) and the joint p.d.f of the three random variables is given by:

$$f(x_1, x_2, x_3) = f_1(x_1)g_{23}(x_2, x_3|x_1) = x_1^2 e^{-x_1(1+x_2+x_3)}$$

for non-negative values of each of these variables. We can now obtain the marginal joint p.d.f of X_2 and X_3 by integrating over X_1

Deriving the distributions of functions of an r.v.

We'd like to derive the distribution of X^2, knowing that X has a uniform distribution on $(-1,1)$

The density $f(x)$ of X over this interval is $\frac{1}{2}$ and Y takes values in $[0,1)$.

The distribution function of Y is therefore given by

$$G(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f(x)\,dx = \sqrt{y}$$

We can differentiate this to obtain the density function of Y

$$g(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{for } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Covariance and correlation

Covariance measures the tendency for two r.v.s to go up or down together, relative to their expected values.

Definition (Covariance): *The covariance between r.v.s* X *and* Y *is*

$$Cov(X,Y) = E[(X - EX)(Y - EY)]$$

Definition (Correlation): *The correlation between r.v.s* X *and* Y *is*

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$-1 \leq \rho \leq 1$ **(by the Cauchy Schwarz inequality,** $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$ **written in deviation form.)**

We can expand the above expression for covariance to get

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

If X **and** Y **are independent, then clearly their covariance is zero. The reverse is not true. For example, let** $X \sim N(0,1)$ **and** $Y = X^2$**. Then** $E(XY) = E(X^3) = 0$ **since all odd moments of a Normal distribution are zero, but the variables are dependent.**

Properties of covariance and correlation

The following results can be verified using basic definitions and properties of expectations:

1. $Cov(X, X) = Var(X)$
2. $Cov(X, Y) = Cov(Y, X)$
3. $Cov(aX, Y) = aCov(X, Y)$ for any constant a.
4. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
5. Let $Y = aX + b$ for some constants a and b. If $a > 0$, then $\rho(X, Y) = 1$. If $a < 0$, then $\rho(X, Y) = -1$

   Proof: $Y - \mu_Y = a(X - \mu_x)$, so $Cov(X, Y) = aE[(X - \mu_X)^2] = a\sigma_X^2$ and $\sigma_Y = |a|\sigma_X$, plug these values into the expression for ρ to get the result.

6. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$.

   Proof:

   $Var(X + Y) = E[(X + Y - \mu_X - \mu_Y)^2] = E((X - \mu_x)^2 + (Y - \mu_Y)^2 + 2(X - \mu_x)(Y - \mu_Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

7. If $X_1, \dots, X_n$ are random variables each with finite variance, then $Var(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} Var(X_i) + 2\sum\sum_{i<j} Cov(X_i, X_j)$

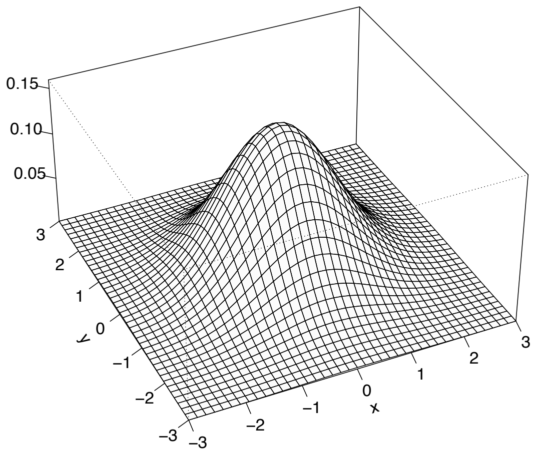
The Multivariate Normal

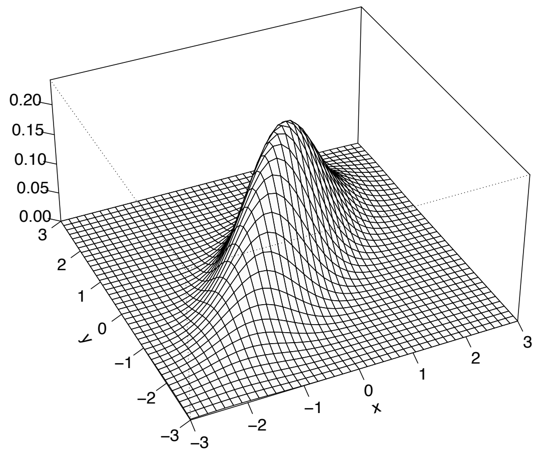
Definition (Multivariate Normal): *A random vector* $\mathbf{X} = (\mathbf{X_1, \dots X_n})$ *is said to have a Multivariate Normal distribution if every linear combination of the* $\mathbf{X_j}$ *has a Normal distribution. That is* $\mathbf{t_1X_1 + \dots t_kX_k}$ *is Normal for all* $\mathbf{t_1, t_2, \dots t_k}$*. An important special case is* $\mathbf{k = 2}$ *and we call this the Bivariate Normal (BVN), whose pdf is given by*

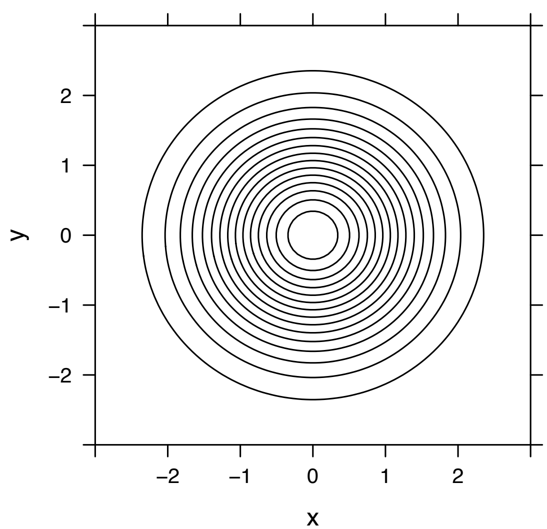
$$f_{X,Y}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\tau} e^{-\frac{1}{2\tau^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}$$

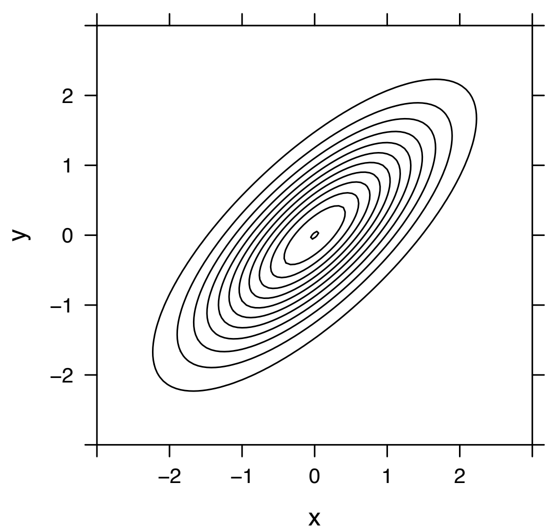
where $\tau = \sqrt{1-\rho^2}$.

With an MVN random vector, uncorrelated implies independent.


0.15
0.10
0.05
y
x
3
2
1
0
−1
−2
−3


0.20
0.15
0.10
0.05
0.00
y
x
3
2
1
0
−1
−2
−3


y
x
2
1
0
−1
−2


y
x
2
1
0
−1
−2