

6. Multivariate Distributions

Spring 2021

Matthew Blackwell

Gov 2002 (Harvard)

Where are we? Where are we going?

- Distributions of one variable: how to describe and summarize uncertainty about one variable.

Where are we? Where are we going?

- Distributions of one variable: how to describe and summarize uncertainty about one variable.
- Today: **distributions of multiple variables** to describe relationships between variables.

Where are we? Where are we going?

- Distributions of one variable: how to describe and summarize uncertainty about one variable.
- Today: **distributions of multiple variables** to describe relationships between variables.
- Later: use data to **learn** about probability distributions.

Why multiple random variables?

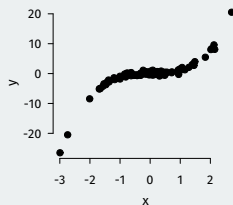
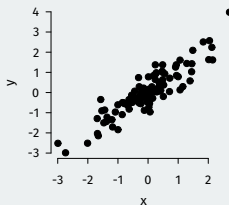
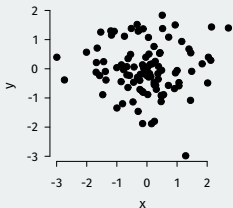
1. How to measure the relationship between two variables X and Y ?

Why multiple random variables?

1. How to measure the relationship between two variables X and Y ?
2. What if we have many observations of the same variable, X_1, X_2, \dots, X_n ?

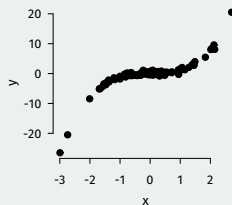
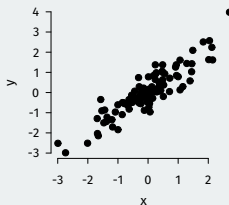
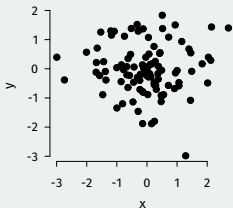
1/ Distributions of Multiple Random Variables

Joint distributions



- The **joint distribution** of two r.v.s, X and Y , describes what pairs of observations, (x, y) are more likely than others.

Joint distributions



- The **joint distribution** of two r.v.s, X and Y , describes what pairs of observations, (x, y) are more likely than others.
- Shape of the joint distribution \rightsquigarrow the relationship between X and Y

Definition

The **joint probability mass function (p.m.f.)** of a pair of discrete r.v.s, (X, Y) describes the probability of any pair of values:

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

Definition

The **joint probability mass function (p.m.f.)** of a pair of discrete r.v.s, (X, Y) describes the probability of any pair of values:

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

- Properties of a joint p.m.f.:

Definition

The **joint probability mass function (p.m.f.)** of a pair of discrete r.v.s, (X, Y) describes the probability of any pair of values:

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

- Properties of a joint p.m.f.:
 - $f_{X,Y}(x, y) \geq 0$

Definition

The **joint probability mass function (p.m.f.)** of a pair of discrete r.v.s, (X, Y) describes the probability of any pair of values:

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

- Properties of a joint p.m.f.:
 - $f_{X,Y}(x, y) \geq 0$ (probabilities can't be negative)

Definition

The **joint probability mass function (p.m.f.)** of a pair of discrete r.v.s, (X, Y) describes the probability of any pair of values:

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

- Properties of a joint p.m.f.:
 - $f_{X,Y}(x, y) \geq 0$ (probabilities can't be negative)
 - $\sum_x \sum_y f_{X,Y}(x, y) = 1$

Definition

The **joint probability mass function (p.m.f.)** of a pair of discrete r.v.s, (X, Y) describes the probability of any pair of values:

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

- Properties of a joint p.m.f.:
 - $f_{X,Y}(x, y) \geq 0$ (probabilities can't be negative)
 - $\sum_x \sum_y f_{X,Y}(x, y) = 1$ (something must happen)

Definition

The **joint probability mass function (p.m.f.)** of a pair of discrete r.v.s, (X, Y) describes the probability of any pair of values:

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

- Properties of a joint p.m.f.:
 - $f_{X,Y}(x, y) \geq 0$ (probabilities can't be negative)
 - $\sum_x \sum_y f_{X,Y}(x, y) = 1$ (something must happen)
 - \sum_x is shorthand for sum over all possible values of X

Example: Gay marriage and gender

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$
Female $X = 1$	0.30	0.21
Male $X = 0$	0.22	0.27

- Joint p.m.f. can be summarized in a cross-tab:

Example: Gay marriage and gender

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$
Female $X = 1$	0.30	0.21
Male $X = 0$	0.22	0.27

- Joint p.m.f. can be summarized in a cross-tab:
 - Each is the probability of that combination, $p_{X,Y}(x,y)$

Example: Gay marriage and gender

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$
Female $X = 1$	0.30	0.21
Male $X = 0$	0.22	0.27

- Joint p.m.f. can be summarized in a cross-tab:
 - Each is the probability of that combination, $p_{X,Y}(x,y)$
- Probability that we randomly select a woman who supports gay marriage?

Example: Gay marriage and gender

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$
Female $X = 1$	0.30	0.21
Male $X = 0$	0.22	0.27

- Joint p.m.f. can be summarized in a cross-tab:
 - Each is the probability of that combination, $p_{X,Y}(x,y)$
- Probability that we randomly select a woman who supports gay marriage?

$$p_{X,Y}(1, 1) = \mathbb{P}(X = 1, Y = 1) = 0.3$$

Marginal distributions

- Can we get the distribution of just one of the r.v.s alone?

Marginal distributions

- Can we get the distribution of just one of the r.v.s alone?
 - Called the **marginal distribution** in this context.

Marginal distributions

- Can we get the distribution of just one of the r.v.s alone?
 - Called the **marginal distribution** in this context.
- Computing **marginal p.m.f.** from the joint p.m.f.:

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

Marginal distributions

- Can we get the distribution of just one of the r.v.s alone?
 - Called the **marginal distribution** in this context.
- Computing **marginal p.m.f.** from the joint p.m.f.:

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

- Intuition: sum over the probability that $Y = y$ and $X = x$ for all possible values of x

Marginal distributions

- Can we get the distribution of just one of the r.v.s alone?
 - Called the **marginal distribution** in this context.
- Computing **marginal p.m.f.** from the joint p.m.f.:

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

- Intuition: sum over the probability that $Y = y$ and $X = x$ for all possible values of x
 - Called **marginalizing out** X .

Marginal distributions

- Can we get the distribution of just one of the r.v.s alone?
 - Called the **marginal distribution** in this context.
- Computing **marginal p.m.f.** from the joint p.m.f.:

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

- Intuition: sum over the probability that $Y = y$ and $X = x$ for all possible values of x
 - Called **marginalizing out** X .
 - Works because values of X are disjoint.

Example: marginals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	
Male $X = 0$	0.22	0.27	
Marginal			

- What's $\mathbb{P}(Y = 1)$?

Example: marginals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	
Male $X = 0$	0.22	0.27	
Marginal			

- What's $\mathbb{P}(Y = 1)$?
 - Probability that a man supports gay marriage plus the probability that a woman supports gay marriage.

$$\mathbb{P}(Y = 1)$$

Example: marginals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	
Male $X = 0$	0.22	0.27	
Marginal			

- What's $\mathbb{P}(Y = 1)$?
 - Probability that a man supports gay marriage plus the probability that a woman supports gay marriage.

$$\mathbb{P}(Y = 1) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 0, Y = 1)$$

Example: marginals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	
Male $X = 0$	0.22	0.27	
Marginal	0.30 + 0.22		

- What's $\mathbb{P}(Y = 1)$?
 - Probability that a man supports gay marriage plus the probability that a woman supports gay marriage.

$$\mathbb{P}(Y = 1) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 0, Y = 1) = 0.3 + 0.22$$

Example: marginals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	
Male $X = 0$	0.22	0.27	
Marginal	0.52		

- What's $\mathbb{P}(Y = 1)$?
 - Probability that a man supports gay marriage plus the probability that a woman supports gay marriage.

$$\mathbb{P}(Y = 1) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 0, Y = 1) = 0.3 + 0.22 = 0.52$$

Example: marginals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	$0.30 + 0.21$
Male $X = 0$	0.22	0.27	$0.22 + 0.27$
Marginal	0.52	$0.21 + 0.27$	

- What's $\mathbb{P}(Y = 1)$?
 - Probability that a man supports gay marriage plus the probability that a woman supports gay marriage.

$$\mathbb{P}(Y = 1) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 0, Y = 1) = 0.3 + 0.22 = 0.52$$

- Works for all marginals.

Example: marginals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- What's $\mathbb{P}(Y = 1)$?
 - Probability that a man supports gay marriage plus the probability that a woman supports gay marriage.

$$\mathbb{P}(Y = 1) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 0, Y = 1) = 0.3 + 0.22 = 0.52$$

- Works for all marginals.

Conditional p.m.f.

Definition

The **conditional probability mass function** or conditional p.m.f. of Y conditional on X is

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

for all values x s.t. $\mathbb{P}(X = x) > 0$.

Conditional p.m.f.

Definition

The **conditional probability mass function** or conditional p.m.f. of Y conditional on X is

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

for all values x s.t. $\mathbb{P}(X = x) > 0$.

- This is a valid univariate probability distribution!

Conditional p.m.f.

Definition

The **conditional probability mass function** or conditional p.m.f. of Y conditional on X is

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

for all values x s.t. $\mathbb{P}(X = x) > 0$.

- This is a valid univariate probability distribution!
 - $P(Y = y | X = x) \geq 0$ and $\sum_y \mathbb{P}(Y = y | X = x) = 1$

Conditional p.m.f.

Definition

The **conditional probability mass function** or conditional p.m.f. of Y conditional on X is

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

for all values x s.t. $\mathbb{P}(X = x) > 0$.

- This is a valid univariate probability distribution!
 - $P(Y = y | X = x) \geq 0$ and $\sum_y \mathbb{P}(Y = y | X = x) = 1$
- Can define the **conditional expectation** of this p.m.f.:

$$E[Y | X = x] = \sum_y y \mathbb{P}(Y = y | X = x)$$

Example: conditionals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- Probability of favoring gay marriage conditional on **male**?

Example: conditionals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- Probability of favoring gay marriage conditional on **male**?

$$\mathbb{P}(Y = 1 \mid X = 0)$$

Example: conditionals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- Probability of favoring gay marriage conditional on **male**?

$$\mathbb{P}(Y = 1 \mid X = 0) = \frac{\mathbb{P}(X = 0, Y = 1)}{\mathbb{P}(X = 0)}$$

Example: conditionals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- Probability of favoring gay marriage conditional on **male**?

$$\mathbb{P}(Y = 1 \mid X = 0) = \frac{\mathbb{P}(X = 0, Y = 1)}{\mathbb{P}(X = 0)} = \frac{0.22}{0.22 + 0.27}$$

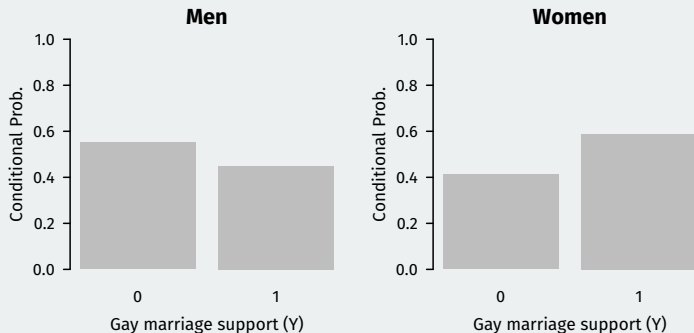
Example: conditionals for gay marriage

	Support Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.30	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- Probability of favoring gay marriage conditional on **male**?

$$\mathbb{P}(Y = 1 \mid X = 0) = \frac{\mathbb{P}(X = 0, Y = 1)}{\mathbb{P}(X = 0)} = \frac{0.22}{0.22 + 0.27} = 0.449$$

Example: conditionals for gay marriage



- Two values of $X \rightsquigarrow$ two **univariate** conditional distributions of Y

- Bayes' rule for r.v.s:

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y)}{\mathbb{P}(X = x)}$$

- Bayes' rule for r.v.s:

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y)}{\mathbb{P}(X = x)}$$

- Law of total probability for r.v.s:

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y)$$

Definition

For two r.v.s X and Y , the **joint cumulative distribution function** or joint c.d.f. $F_{X,Y}(x, y)$ is a function such that for finite values x and y ,

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

Definition

For two r.v.s X and Y , the **joint cumulative distribution function** or joint c.d.f. $F_{X,Y}(x,y)$ is a function such that for finite values x and y ,

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

- Well-defined for discrete and continuous X and Y .

Definition

For two r.v.s X and Y , the **joint cumulative distribution function** or joint c.d.f. $F_{X,Y}(x,y)$ is a function such that for finite values x and y ,

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

- Well-defined for discrete and continuous X and Y .
- For discrete we simply have:

$$F_{X,Y}(x,y) = \sum_{i \leq x} \sum_{j \leq y} \mathbb{P}(X = i, Y = j)$$

Continuous r.v.s

- One continuous r.v.: prob. of being in a subset of the real line.

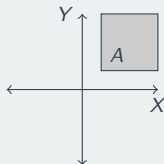


Continuous r.v.s

- One continuous r.v.: prob. of being in a subset of the real line.



- Two continuous r.v.s: probability of being in some subset of the 2-dimensional plane.



Continuous joint p.d.f.

Definition

If two continuous r.v.s X and Y with joint c.d.f. $F_{X,Y}$, their **joint p.d.f.** $f_{X,Y}(x,y)$ is the derivative of $F_{X,Y}$ with respect to x and y ,

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

- Integrate over both dimensions to get the probability of a region:

$$\mathbb{P}((X, Y) \in A) = \iint_{(x,y) \in A} f_{X,Y}(x,y) dx dy.$$

Continuous joint p.d.f.

Definition

If two continuous r.v.s X and Y with joint c.d.f. $F_{X,Y}$, their **joint p.d.f.** $f_{X,Y}(x,y)$ is the derivative of $F_{X,Y}$ with respect to x and y ,

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

- Integrate over both dimensions to get the probability of a region:

$$\mathbb{P}((X, Y) \in A) = \iint_{(x,y) \in A} f_{X,Y}(x,y) dx dy.$$

- $\{(x,y) : f_{X,Y}(x,y) > 0\}$ is called the **support** of the distribution.

Properties of the joint p.d.f.

- Joint p.d.f. must meet the following conditions:

Properties of the joint p.d.f.

- Joint p.d.f. must meet the following conditions:
 1. $f_{X,Y}(x,y) \geq 0$ for all values of (x,y) ,

Properties of the joint p.d.f.

- Joint p.d.f. must meet the following conditions:
 1. $f_{X,Y}(x,y) \geq 0$ for all values of (x,y) , (nonnegative)

Properties of the joint p.d.f.

- Joint p.d.f. must meet the following conditions:
 1. $f_{X,Y}(x,y) \geq 0$ for all values of (x,y) , (nonnegative)
 2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$,

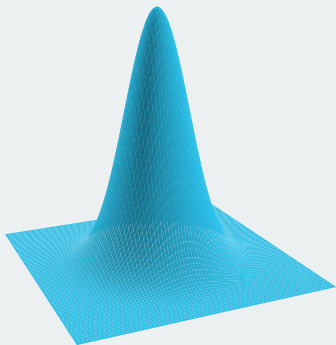
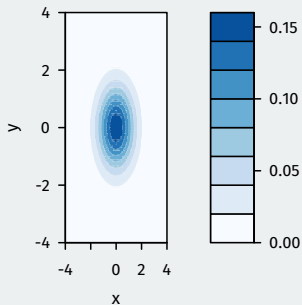
Properties of the joint p.d.f.

- Joint p.d.f. must meet the following conditions:
 1. $f_{X,Y}(x,y) \geq 0$ for all values of (x,y) , (nonnegative)
 2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$, (probabilities “sum” to 1)

Properties of the joint p.d.f.

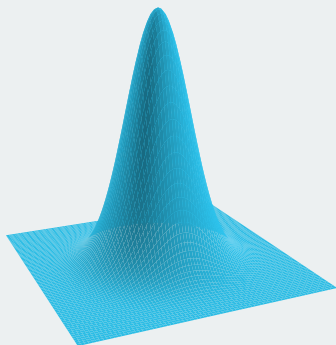
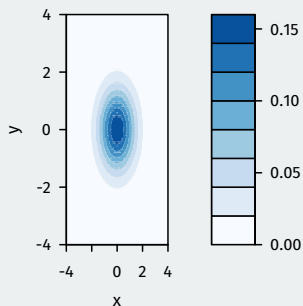
- Joint p.d.f. must meet the following conditions:
 1. $f_{X,Y}(x,y) \geq 0$ for all values of (x,y) , (nonnegative)
 2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$, (probabilities “sum” to 1)
- $\mathbb{P}(X = x, Y = y) = 0$ for similar reasons as with single r.v.s.

Joint densities are 3D



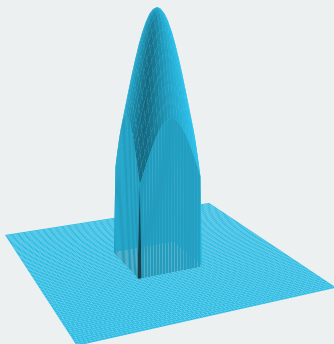
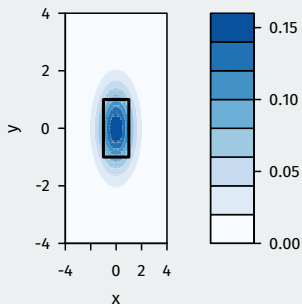
- X and Y axes are on the “floor,” height is the value of $f_{X,Y}(x,y)$.

Joint densities are 3D



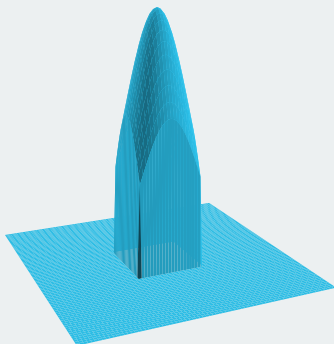
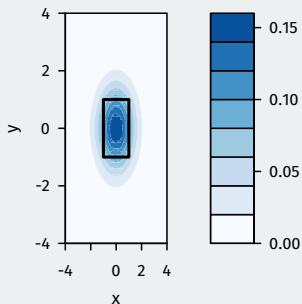
- X and Y axes are on the “floor,” height is the value of $f_{X,Y}(x,y)$.
- Remember $f_{X,Y}(x,y) \neq \mathbb{P}(X = x, Y = y)$.

Probability = volume



- $\mathbb{P}((X, Y) \in A) = \iint_{(x,y) \in A} f_{X,Y}(x,y) dx dy$

Probability = volume



- $\mathbb{P}((X, Y) \in A) = \iint_{(x,y) \in A} f_{X,Y}(x,y) dx dy$
- Probability = volume above a specific region.

Continuous marginal distributions

- We can recover the marginal PDF of one of the variables by integrating over the distribution of the other variable:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Continuous marginal distributions

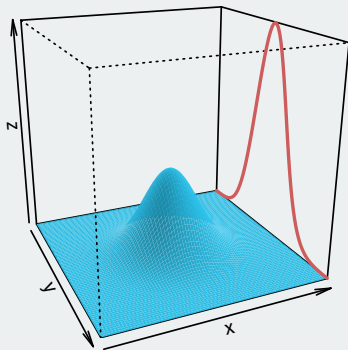
- We can recover the marginal PDF of one of the variables by integrating over the distribution of the other variable:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

- Works for either variable:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

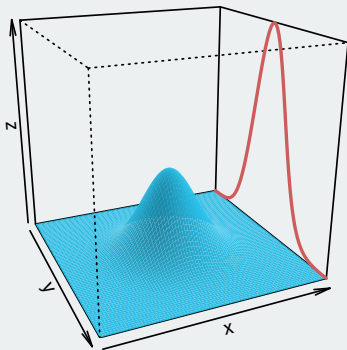
Visualizing continuous marginals



- Marginal integrates (sums, basically) over other r.v.:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Visualizing continuous marginals



- Marginal integrates (sums, basically) over other r.v.:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx$$

- Pile up/flatten all of the joint density onto a single dimension.

Continuous conditional distributions

Definition

The **conditional p.d.f.** of a continuous random variable is

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

for all values x s.t. $f_X(x) > 0$.

- Implies

$$\mathbb{P}(a < Y < b | X = x) = \int_a^b f_{Y|X}(y|x) dy$$

Continuous conditional distributions

Definition

The **conditional p.d.f.** of a continuous random variable is

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

for all values x s.t. $f_X(x) > 0$.

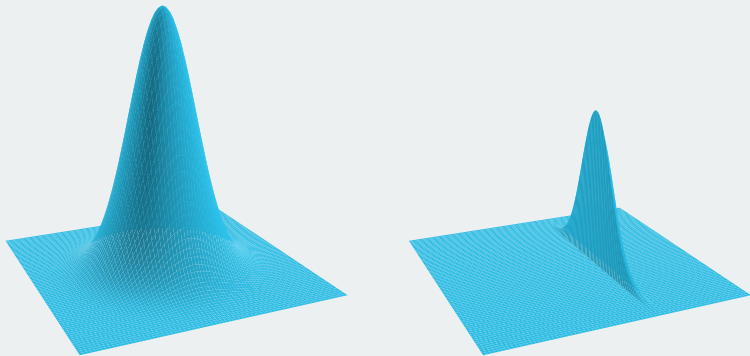
- Implies

$$\mathbb{P}(a < Y < b | X = x) = \int_a^b f_{Y|X}(y|x) dy$$

- Based on the definition of the conditional p.m.f./p.d.f., we have the following factorization:

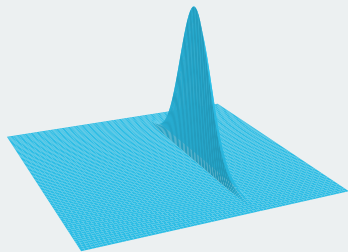
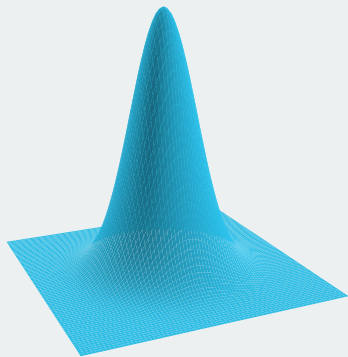
$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x)$$

Conditional distributions as slices



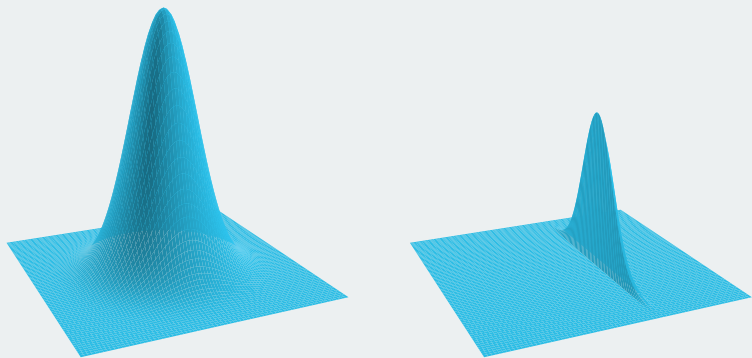
- $f_{Y|X}(y|x_0)$ is the conditional p.d.f. of Y when $X = x_0$

Conditional distributions as slices



- $f_{Y|X}(y|x_0)$ is the conditional p.d.f. of Y when $X = x_0$
- $f_{Y|X}(y|x_0)$ is proportional to joint p.d.f. along x_0 : $f_{X,Y}(y, x_0)$

Conditional distributions as slices



- $f_{Y|X}(y|x_0)$ is the conditional p.d.f. of Y when $X = x_0$
- $f_{Y|X}(y|x_0)$ is proportional to joint p.d.f. along x_0 : $f_{X,Y}(y, x_0)$
- Normalize by dividing by $f_X(x_0)$ to ensure proper p.d.f.

Independence

Independence

Two r.v.s Y and X are **independent** (which we write $X \perp\!\!\!\perp Y$) if for all sets A and B :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

- Knowing the value of X gives us no information about the value of Y .

Independence

Independence

Two r.v.s Y and X are **independent** (which we write $X \perp\!\!\!\perp Y$) if for all sets A and B :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

- Knowing the value of X gives us no information about the value of Y .
- If X and Y are independent, then:

Independence

Independence

Two r.v.s Y and X are **independent** (which we write $X \perp\!\!\!\perp Y$) if for all sets A and B :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

- Knowing the value of X gives us no information about the value of Y .
- If X and Y are independent, then:
 - $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ (joint is the product of marginals)

Independence

Independence

Two r.v.s Y and X are **independent** (which we write $X \perp\!\!\!\perp Y$) if for all sets A and B :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

- Knowing the value of X gives us no information about the value of Y .
- If X and Y are independent, then:
 - $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ (joint is the product of marginals)
 - $F_{X,Y}(x,y) = F_X(x)F_Y(y)$

Independence

Independence

Two r.v.s Y and X are **independent** (which we write $X \perp\!\!\!\perp Y$) if for all sets A and B :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

- Knowing the value of X gives us no information about the value of Y .
- If X and Y are independent, then:
 - $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ (joint is the product of marginals)
 - $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
 - $f_{Y|X}(y|x) = f_Y(y)$ (conditional is the marginal)

Independence

Independence

Two r.v.s Y and X are **independent** (which we write $X \perp\!\!\!\perp Y$) if for all sets A and B :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

- Knowing the value of X gives us no information about the value of Y .
- If X and Y are independent, then:
 - $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ (joint is the product of marginals)
 - $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
 - $f_{Y|X}(y|x) = f_Y(y)$ (conditional is the marginal)
- **Conditional independence** implies similar to conditional distributions:

$$\mathbb{P}(X \in A, Y \in B | Z) = \mathbb{P}(X \in A | Z)\mathbb{P}(Y \in B | Z)$$

2/ Expectations of Joint Distributions

Properties of joint distributions

- Single r.v.: summarized $f_X(x)$ with $\mathbb{E}[X]$ and $\mathbb{V}[X]$

Properties of joint distributions

- Single r.v.: summarized $f_X(x)$ with $\mathbb{E}[X]$ and $\mathbb{V}[X]$
- With 2 r.v.s: how strong is the dependence is between X and Y ?

Properties of joint distributions

- Single r.v.: summarized $f_X(x)$ with $\mathbb{E}[X]$ and $\mathbb{V}[X]$
- With 2 r.v.s: how strong is the dependence is between X and Y ?
- First: **expectations** over joint distributions.

Expectations over multiple r.v.s

- **2-d LOTUS:** take expectations over the joint distribution.

Expectations over multiple r.v.s

- **2-d LOTUS:** take expectations over the joint distribution.
- With discrete X and Y :

$$\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) f_{X, Y}(x, y)$$

Expectations over multiple r.v.s

- **2-d LOTUS:** take expectations over the joint distribution.
- With discrete X and Y :

$$\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) f_{X, Y}(x, y)$$

- With continuous X and Y :

$$\mathbb{E}[g(X, Y)] = \int_x \int_y g(x, y) f_{X, Y}(x, y) dx dy$$

Expectations over multiple r.v.s

- **2-d LOTUS:** take expectations over the joint distribution.
- With discrete X and Y :

$$\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) f_{X, Y}(x, y)$$

- With continuous X and Y :

$$\mathbb{E}[g(X, Y)] = \int_x \int_y g(x, y) f_{X, Y}(x, y) dx dy$$

- Marginal expectations:

$$\mathbb{E}[Y] = \sum_x \sum_y y f_{X, Y}(x, y)$$

Applying 2D LOTUS

Theorem

If X and Y are independent r.v.s, then

$$E[XY] = E[X]E[Y].$$

Applying 2D LOTUS

Theorem

If X and Y are independent r.v.s, then

$$E[XY] = E[X]E[Y].$$

- Proof for discrete X and Y :

Theorem

If X and Y are independent r.v.s, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

- Proof for discrete X and Y :

$$\mathbb{E}[XY] = \sum_x \sum_y xy f_{X,Y}(x, y)$$

Theorem

If X and Y are independent r.v.s, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

- Proof for discrete X and Y :

$$\begin{aligned}\mathbb{E}[XY] &= \sum_x \sum_y xy f_{X,Y}(x,y) \\ &= \sum_x \sum_y xy f_X(x)f_Y(y)\end{aligned}$$

Applying 2D LOTUS

Theorem

If X and Y are independent r.v.s, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

- Proof for discrete X and Y :

$$\begin{aligned}\mathbb{E}[XY] &= \sum_x \sum_y xy f_{X,Y}(x,y) \\ &= \sum_x \sum_y xy f_X(x) f_Y(y) \\ &= \left(\sum_x x f_X(x) \right) \left(\sum_y y f_Y(y) \right)\end{aligned}$$

Theorem

If X and Y are independent r.v.s, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

- Proof for discrete X and Y :

$$\begin{aligned}\mathbb{E}[XY] &= \sum_x \sum_y xy f_{X,Y}(x,y) \\ &= \sum_x \sum_y xy f_X(x) f_Y(y) \\ &= \left(\sum_x x f_X(x) \right) \left(\sum_y y f_Y(y) \right) \\ &= \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

3/ Covariance and Correlation

Why (in)dependence?

- Independence assumptions are **everywhere** in statistics.

Why (in)dependence?

- Independence assumptions are **everywhere** in statistics.
 - Each response in a poll is considered **independent** of all other responses.

Why (in)dependence?

- Independence assumptions are **everywhere** in statistics.
 - Each response in a poll is considered **independent** of all other responses.
 - In a randomized control trial, treatment assignment is **independent** of background characteristics.

Why (in)dependence?

- Independence assumptions are **everywhere** in statistics.
 - Each response in a poll is considered **independent** of all other responses.
 - In a randomized control trial, treatment assignment is **independent** of background characteristics.
- Lack of independence is a blessing or a curse:

Why (in)dependence?

- Independence assumptions are **everywhere** in statistics.
 - Each response in a poll is considered **independent** of all other responses.
 - In a randomized control trial, treatment assignment is **independent** of background characteristics.
- Lack of independence is a blessing or a curse:
 - Two variables not independent \rightsquigarrow potentially interesting relationship.

Why (in)dependence?

- Independence assumptions are **everywhere** in statistics.
 - Each response in a poll is considered **independent** of all other responses.
 - In a randomized control trial, treatment assignment is **independent** of background characteristics.
- Lack of independence is a blessing or a curse:
 - Two variables not independent \rightsquigarrow potentially interesting relationship.
 - In observational studies, treatment assignment is usually **not independent** of background characteristics.

Defining covariance

- How do we measure the strength of the dependence between two r.v.?

Defining covariance

- How do we measure the strength of the dependence between two r.v.?

Covariance

The **covariance** between two r.v.s, X and Y is defined as:

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Defining covariance

- How do we measure the strength of the dependence between two r.v.?

Covariance

The **covariance** between two r.v.s, X and Y is defined as:

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

- How often do high values of X occur with high values of Y ?

Defining covariance

- How do we measure the strength of the dependence between two r.v.?

Covariance

The **covariance** between two r.v.s, X and Y is defined as:

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

- How often do high values of X occur with high values of Y ?
- Properties of covariances:

Defining covariance

- How do we measure the strength of the dependence between two r.v.?

Covariance

The **covariance** between two r.v.s, X and Y is defined as:

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

- How often do high values of X occur with high values of Y ?
- Properties of covariances:
 - $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Defining covariance

- How do we measure the strength of the dependence between two r.v.?

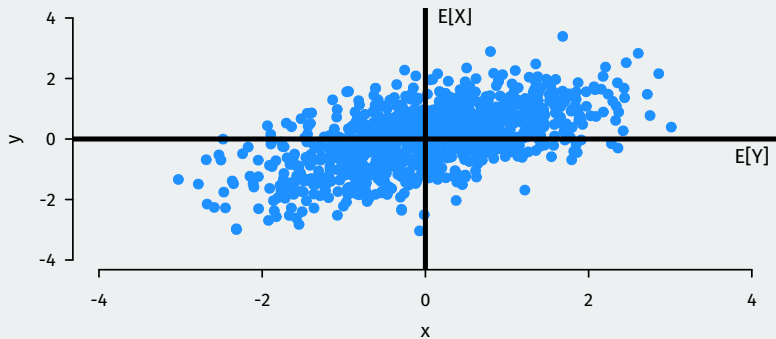
Covariance

The **covariance** between two r.v.s, X and Y is defined as:

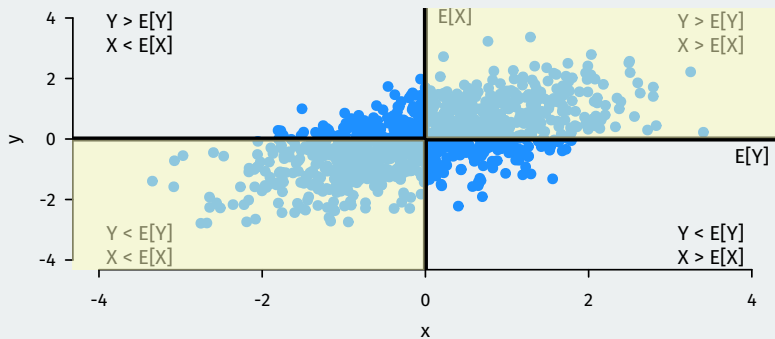
$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

- How often do high values of X occur with high values of Y ?
- Properties of covariances:
 - $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
 - If $X \perp\!\!\!\perp Y$, then $\text{Cov}[X, Y] = 0$

Covariance intuition

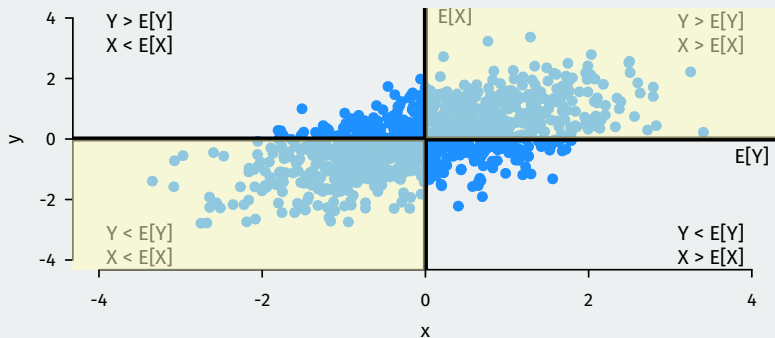


Covariance intuition



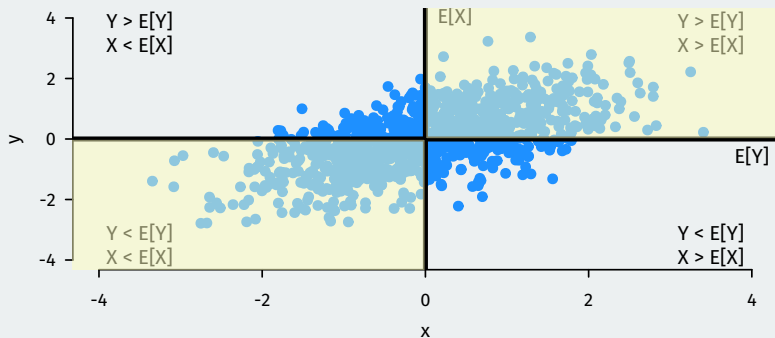
- Large values of X tend to occur with large values of Y :

Covariance intuition



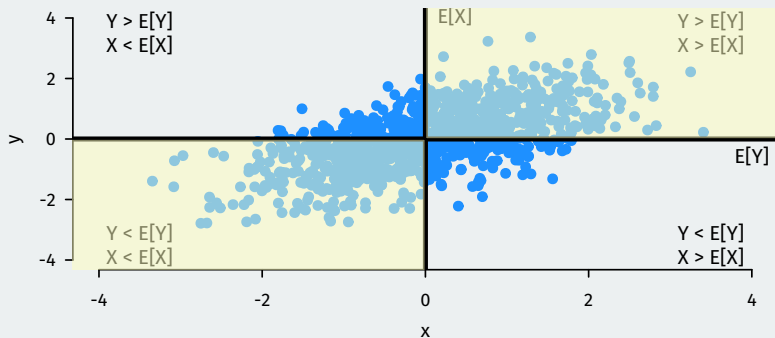
- Large values of X tend to occur with large values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{pos. num.}) \times (\text{pos. num.}) = +$

Covariance intuition



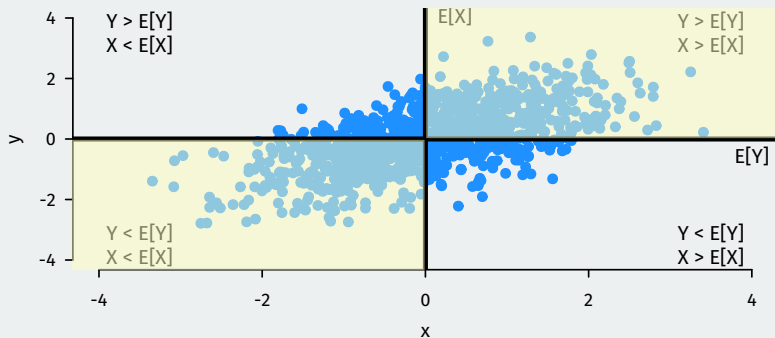
- Large values of X tend to occur with large values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{pos. num.}) \times (\text{pos. num.}) = +$
- Small values of X tend to occur with small values of Y :

Covariance intuition



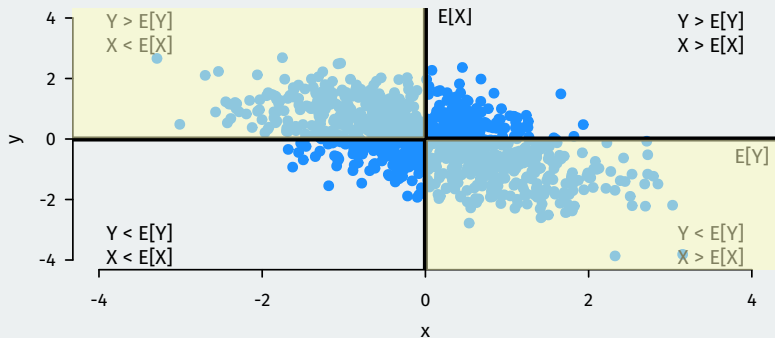
- Large values of X tend to occur with large values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{pos. num.}) \times (\text{pos. num.}) = +$
- Small values of X tend to occur with small values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{neg. num.}) \times (\text{neg. num.}) = +$

Covariance intuition



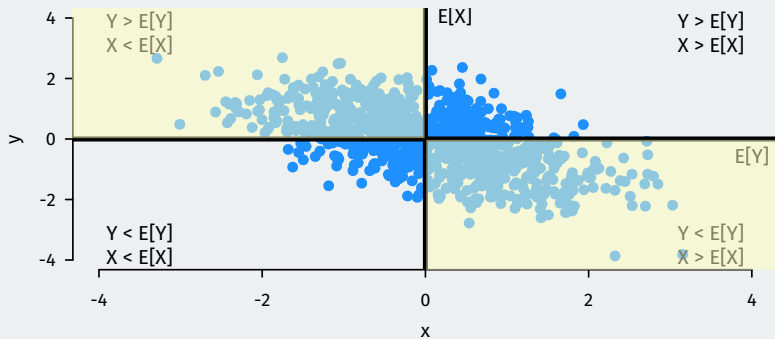
- Large values of X tend to occur with large values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{pos. num.}) \times (\text{pos. num.}) = +$
- Small values of X tend to occur with small values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{neg. num.}) \times (\text{neg. num.}) = +$
- If these dominate \rightsquigarrow positive covariance.

Covariance intuition



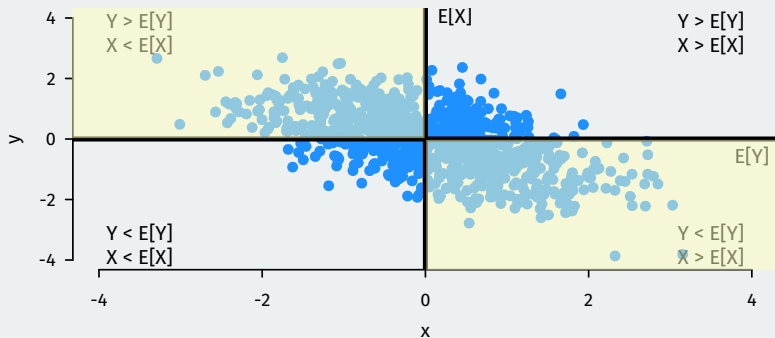
- Large values of X tend to occur with small values of Y :

Covariance intuition



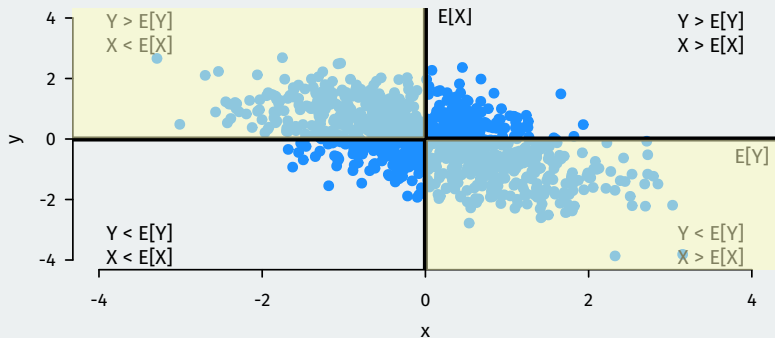
- Large values of X tend to occur with small values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{pos. num.}) \times (\text{neg. num.}) = -$

Covariance intuition



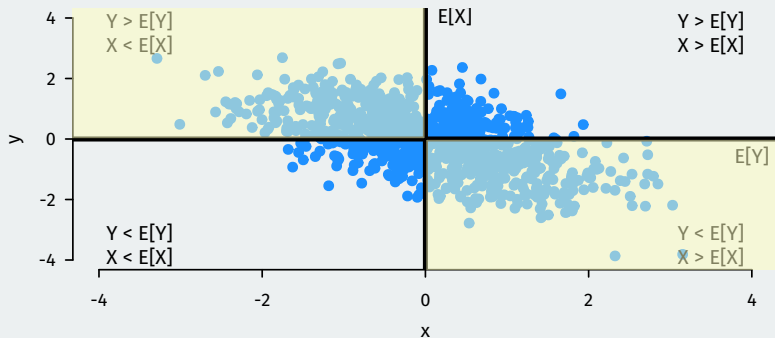
- Large values of X tend to occur with small values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{pos. num.}) \times (\text{neg. num.}) = -$
- Small values of X tend to occur with large values of Y :

Covariance intuition



- Large values of X tend to occur with small values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{pos. num.}) \times (\text{neg. num.}) = -$
- Small values of X tend to occur with large values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{neg. num.}) \times (\text{pos. num.}) = -$

Covariance intuition



- Large values of X tend to occur with small values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{pos. num.}) \times (\text{neg. num.}) = -$
- Small values of X tend to occur with large values of Y :
 - $(X - E[X])(Y - E[Y]) = (\text{neg. num.}) \times (\text{pos. num.}) = -$
- If these dominate \rightsquigarrow negative covariance.

Properties of variances and covariances

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Properties of covariances:

Properties of variances and covariances

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Properties of covariances:

1. $\text{Cov}[X, X] = \mathbb{V}[X]$

Properties of variances and covariances

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Properties of covariances:

1. $\text{Cov}[X, X] = \mathbb{V}[X]$

2. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$

Properties of variances and covariances

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Properties of covariances:
 1. $\text{Cov}[X, X] = \mathbb{V}[X]$
 2. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
 3. $\text{Cov}[X, c] = 0$ for any constant c

Properties of variances and covariances

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Properties of covariances:
 1. $\text{Cov}[X, X] = \mathbb{V}[X]$
 2. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
 3. $\text{Cov}[X, c] = 0$ for any constant c
 4. $\text{Cov}[aX, Y] = a\text{Cov}[X, Y]$.

Properties of variances and covariances

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Properties of covariances:
 1. $\text{Cov}[X, X] = \mathbb{V}[X]$
 2. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
 3. $\text{Cov}[X, c] = 0$ for any constant c
 4. $\text{Cov}[aX, Y] = a\text{Cov}[X, Y]$.
 5. $\text{Cov}[X + Y, Z] = \text{Cov}[X, Z] + \text{Cov}[Y, Z]$

Properties of variances and covariances

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Properties of covariances:

1. $\text{Cov}[X, X] = \mathbb{V}[X]$

2. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$

3. $\text{Cov}[X, c] = 0$ for any constant c

4. $\text{Cov}[aX, Y] = a\text{Cov}[X, Y]$.

5. $\text{Cov}[X + Y, Z] = \text{Cov}[X, Z] + \text{Cov}[Y, Z]$

6. $\text{Cov}[X + Y, Z + W] = \text{Cov}[X, Z] + \text{Cov}[Y, Z] + \text{Cov}[X, W] + \text{Cov}[Y, W]$

Covariances and variances

- Can now state a few more properties of variances.

Covariances and variances

- Can now state a few more properties of variances.
- Variance of a sum:

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}[X, Y]$$

Covariances and variances

- Can now state a few more properties of variances.
- Variance of a sum:

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}[X, Y]$$

- More generally for n r.v.s X_1, \dots, X_n :

$$\mathbb{V}[X_1 + \dots + X_n] = \mathbb{V}[X_1] + \dots + \mathbb{V}[X_n] + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$$

Covariances and variances

- Can now state a few more properties of variances.
- Variance of a sum:

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}[X, Y]$$

- More generally for n r.v.s X_1, \dots, X_n :

$$\mathbb{V}[X_1 + \dots + X_n] = \mathbb{V}[X_1] + \dots + \mathbb{V}[X_n] + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$$

- If X and Y independent, $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.

Covariances and variances

- Can now state a few more properties of variances.
- Variance of a sum:

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}[X, Y]$$

- More generally for n r.v.s X_1, \dots, X_n :

$$\mathbb{V}[X_1 + \dots + X_n] = \mathbb{V}[X_1] + \dots + \mathbb{V}[X_n] + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$$

- If X and Y independent, $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.
 - Beware: $\mathbb{V}[X - Y] = \mathbb{V}[X] + \mathbb{V}[Y]$ as well.

Zero covariance doesn't imply independence

- We saw that $X \perp\!\!\!\perp Y \rightsquigarrow \text{Cov}[X, Y] = 0$.

Zero covariance doesn't imply independence

- We saw that $X \perp\!\!\!\perp Y \rightsquigarrow \text{Cov}[X, Y] = 0$.
- Does $\text{Cov}[X, Y] = 0$ imply that $X \perp\!\!\!\perp Y$?

Zero covariance doesn't imply independence

- We saw that $X \perp\!\!\!\perp Y \rightsquigarrow \text{Cov}[X, Y] = 0$.
- Does $\text{Cov}[X, Y] = 0$ imply that $X \perp\!\!\!\perp Y$? **No!**

Zero covariance doesn't imply independence

- We saw that $X \perp\!\!\!\perp Y \rightsquigarrow \text{Cov}[X, Y] = 0$.
- Does $\text{Cov}[X, Y] = 0$ imply that $X \perp\!\!\!\perp Y$? **No!**
- **Counterexample:** $X \in \{-1, 0, 1\}$ with equal probability and $Y = X^2$.

Zero covariance doesn't imply independence

- We saw that $X \perp\!\!\!\perp Y \rightsquigarrow \text{Cov}[X, Y] = 0$.
- Does $\text{Cov}[X, Y] = 0$ imply that $X \perp\!\!\!\perp Y$? **No!**
- **Counterexample:** $X \in \{-1, 0, 1\}$ with equal probability and $Y = X^2$.
- Covariance is a measure of **linear dependence**, so it can miss non-linear dependence.

Correlation

- Correlation is a scale-free measure of linear dependence.

Correlation

- Correlation is a scale-free measure of linear dependence.

Definition

The **correlation** between two r.v.s X and Y is defined as:

$$\rho = \rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\text{V}[X]\text{V}[Y]}} = \text{Cov}\left(\frac{X - \mathbb{E}[X]}{SD[X]}, \frac{Y - \mathbb{E}[Y]}{SD[Y]}\right)$$

Correlation

- Correlation is a scale-free measure of linear dependence.

Definition

The **correlation** between two r.v.s X and Y is defined as:

$$\rho = \rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\text{V}[X]\text{V}[Y]}} = \text{Cov} \left(\frac{X - \mathbb{E}[X]}{SD[X]}, \frac{Y - \mathbb{E}[Y]}{SD[Y]} \right)$$

- Covariance after dividing out the scales of the respective variables.

Correlation

- Correlation is a scale-free measure of linear dependence.

Definition

The **correlation** between two r.v.s X and Y is defined as:

$$\rho = \rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\text{V}[X]\text{V}[Y]}} = \text{Cov}\left(\frac{X - \mathbb{E}[X]}{SD[X]}, \frac{Y - \mathbb{E}[Y]}{SD[Y]}\right)$$

- Covariance after dividing out the scales of the respective variables.
- Correlation properties:

Correlation

- Correlation is a scale-free measure of linear dependence.

Definition

The **correlation** between two r.v.s X and Y is defined as:

$$\rho = \rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\text{V}[X]\text{V}[Y]}} = \text{Cov}\left(\frac{X - \mathbb{E}[X]}{SD[X]}, \frac{Y - \mathbb{E}[Y]}{SD[Y]}\right)$$

- Covariance after dividing out the scales of the respective variables.
- Correlation properties:
 - $-1 \leq \rho \leq 1$

Correlation

- Correlation is a scale-free measure of linear dependence.

Definition

The **correlation** between two r.v.s X and Y is defined as:

$$\rho = \rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\text{V}[X]}\sqrt{\text{V}[Y]}} = \text{Cov}\left(\frac{X - \mathbb{E}[X]}{SD[X]}, \frac{Y - \mathbb{E}[Y]}{SD[Y]}\right)$$

- Covariance after dividing out the scales of the respective variables.
- Correlation properties:
 - $-1 \leq \rho \leq 1$
 - $|\rho(X, Y)| = 1$ if and only if X and Y are perfectly correlated with a deterministic linear relationship: $Y = a + bX$.

4/ Random vectors

Multivariate random vectors

- When we have many r.v.s, we sometimes group them into **random vectors** $X = (X_1, \dots, X_m)^T$

Multivariate random vectors

- When we have many r.v.s, we sometimes group them into **random vectors** $X = (X_1, \dots, X_m)^T$
 - X is a function from the sample space to \mathbb{R}^m

Multivariate random vectors

- When we have many r.v.s, we sometimes group them into **random vectors** $X = (X_1, \dots, X_m)^T$
 - X is a function from the sample space to \mathbb{R}^m
 - x is now a length- m vector and potential value of X

Multivariate random vectors

- When we have many r.v.s, we sometimes group them into **random vectors** $X = (X_1, \dots, X_m)^T$
 - X is a function from the sample space to \mathbb{R}^m
 - x is now a length- m vector and potential value of X
 - Generalizes all ideas from 2 variables to m

Multivariate random vectors

- When we have many r.v.s, we sometimes group them into **random vectors** $X = (X_1, \dots, X_m)^T$
 - X is a function from the sample space to \mathbb{R}^m
 - x is now a length- m vector and potential value of X
 - Generalizes all ideas from 2 variables to m
- Joint distribution function: $F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X_1 \leq x_1, \dots, X_m \leq x_m)$.

Multivariate random vectors

- When we have many r.v.s, we sometimes group them into **random vectors** $X = (X_1, \dots, X_m)^T$
 - X is a function from the sample space to \mathbb{R}^m
 - x is now a length- m vector and potential value of X
 - Generalizes all ideas from 2 variables to m
- Joint distribution function: $F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X_1 \leq x_1, \dots, X_m \leq x_m)$.
 - Discrete: joint p.m.f. $\mathbb{P}(X = x)$.

Multivariate random vectors

- When we have many r.v.s, we sometimes group them into **random vectors** $X = (X_1, \dots, X_m)^T$
 - X is a function from the sample space to \mathbb{R}^m
 - x is now a length- m vector and potential value of X
 - Generalizes all ideas from 2 variables to m
- Joint distribution function: $F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X_1 \leq x_1, \dots, X_m \leq x_m)$.
 - Discrete: joint p.m.f. $\mathbb{P}(X = x)$.
 - Continuous: joint p.d.f.

$$f(x) = \frac{\partial^m}{\partial x_1 \dots \partial x_m} F(x)$$

Multivariate random vectors

- When we have many r.v.s, we sometimes group them into **random vectors** $X = (X_1, \dots, X_m)^T$
 - X is a function from the sample space to \mathbb{R}^m
 - x is now a length- m vector and potential value of X
 - Generalizes all ideas from 2 variables to m
- Joint distribution function: $F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X_1 \leq x_1, \dots, X_m \leq x_m)$.
 - Discrete: joint p.m.f. $\mathbb{P}(X = x)$.
 - Continuous: joint p.d.f.

$$f(x) = \frac{\partial^m}{\partial x_1 \dots \partial x_m} F(x)$$

- Expectation of a random vector is just the vector of expectations:

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \dots, \mathbb{E}[X_m])^T$$

Covariance matrices

- Covariance matrix generalizes (co)variance to this setting:

$$\mathbb{V}[\mathbf{X}] = \mathbb{E} [(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]$$

Covariance matrices

- Covariance matrix generalizes (co)variance to this setting:

$$\mathbb{V}[\mathbf{X}] = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]$$

- We usually write $\mathbb{V}[\mathbf{X}] = \Sigma$ and it is a $m \times m$ **symmetric** matrix:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_m^2 \end{pmatrix}$$

where, $\sigma_j^2 = \mathbb{V}[X_j]$ and $\sigma_{ij} = \text{Cov}(X_i, X_j)$.

Covariance matrices

- Covariance matrix generalizes (co)variance to this setting:

$$\mathbb{V}[\mathbf{X}] = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]$$

- We usually write $\mathbb{V}[\mathbf{X}] = \Sigma$ and it is a $m \times m$ **symmetric** matrix:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_m^2 \end{pmatrix}$$

where, $\sigma_j^2 = \mathbb{V}[X_j]$ and $\sigma_{ij} = \text{Cov}(X_i, X_j)$.

- Symmetric ($\Sigma = \Sigma^T$) because $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$.

Linear transformations of random vectors

Theorem

If $X \in \mathbb{R}^m$ with $m \times 1$ expectation μ and $m \times m$ covariance matrix Σ , and \mathbf{A} is a $q \times m$ matrix, then $\mathbf{A}X$ is a random vector with mean $\mathbf{A}\mu$ and covariance matrix $\mathbf{A}\Sigma\mathbf{A}^T$.