Week 8: Joint, Marginal, Conditional Continuous Distributions, Expectations, Covariance and Correlations

MATH697

Sahir Bhatnagar

October 24, 2017

McGill University

Joint Continuous Distributions

Introduction

- Once we have a handle on **discrete** joint distributions, it is straightforward to consider **continuous** joint distributions

- Once we have a handle on **discrete** joint distributions, it is straightforward to consider **continuous** joint distributions
- We simply make the now-familiar substitutions of integrals for sums and PDFs for PMFs

Introduction

- Once we have a handle on **discrete** joint distributions, it is straightforward to consider **continuous** joint distributions
- We simply make the now-familiar substitutions of integrals for sums and PDFs for PMFs
- Remembering that the probability of any individual point is now 0

Joint Continuous CDF

- In order for X and Y to have a continuous joint distribution, we require that the joint CDF

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

be differentiable with respect to x and y.

Joint Continuous CDF

- In order for X and Y to have a continuous joint distribution, we require that the joint CDF

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

  be differentiable with respect to x and y.
- The partial derivative with respect to x and y is called the joint PDF.

Joint Continuous CDF

- In order for X and Y to have a continuous joint distribution, we require that the joint CDF

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

  be differentiable with respect to x and y.
- The partial derivative with respect to x and y is called the joint PDF.
- The joint PDF determines the joint distribution, as does the joint CDF.

Joint Continuous PDF

Definition 1 (Joint Continuous PDF)

If X and Y are continuous with joint CDF $F_{X,Y}$, their **joint PDF** is the derivative of the joint CDF with respect to x and y:

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) \tag{1}$$

We require $f_{X,Y}(x, y) \geq 0$ for all x and y and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$$

Joint Continuous PDF

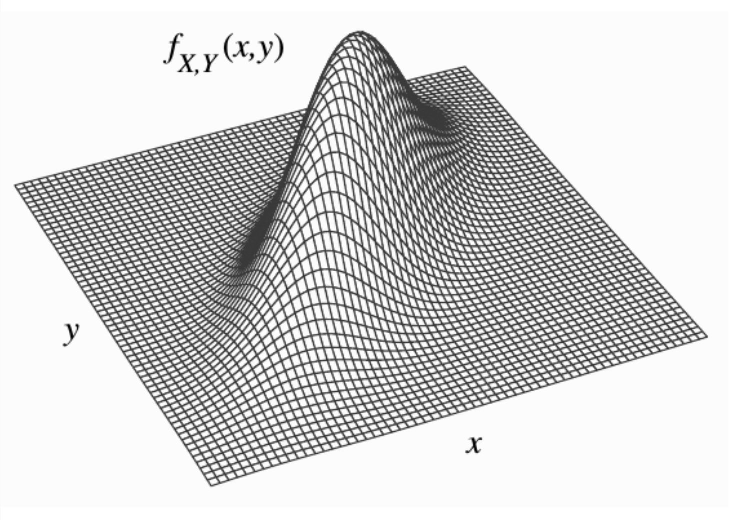


Figure 1

Joint Continuous PDF Example

Example 2 (Joint Continuous PDF)

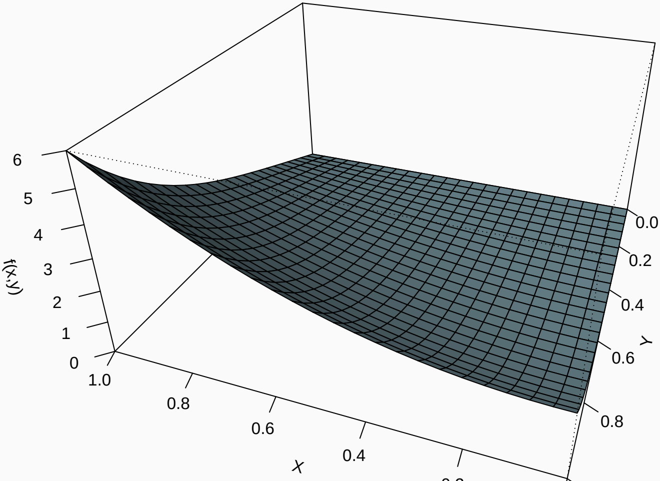
Let X and Y be continuous with joint PDF f given by

$$f_{X,Y}(x,y) = \begin{cases} 4x^2y + 2y^5 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases}$$

1. Verify f is indeed a density function
2. Compute $P\big(0.5 \leq X \leq 0.7, 0.2 \leq Y \leq 0.9\big)$

Joint Continuous PDF Example

```
x <- seq(0, 1,len = 25) ; y <- x ; f <- function(x, y) 4*x^2*y + 2*y^5
z <- outer(x, y, f)
persp(x, y, z, theta = 200, phi = 30, expand = 0.5, col = ”lightblue”, ltheta = 120,
      shade = 0.75, ticktype = ”detailed”, xlab = ”X”, ylab = ”Y”, zlab = ”f(x,y)”)
```



Definition 3 (Marginal Continuous PDF)

If X and Y are continuous RVs with joint PDF $f_{X,Y}(x,y)$, the *marginal* PDF of X is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy, \quad x \in \mathbb{R} \tag{2}$$

Similarly the *marginal* PDF of Y is:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx, \quad y \in \mathbb{R} \tag{3}$$

Marginal Continous PDF (General Case)

- We have mainly been looking at the joint distribution of two RVs

Marginal Continous PDF (General Case)

- We have mainly been looking at the joint distribution of two RVs
- However marginalization works analogously with any number of variables

- We have mainly been looking at the joint distribution of two RVs
- However marginalization works analogously with any number of variables
- For example, if we have the joint PDF of X, Y, Z, W but want the joint PDF of X, W, we just have to integrate over all possible values of Y and Z:

$$f_{X,W}(x, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z,W}(x, y, z, w) dydz$$

Marginal Continous PDF (Rule of Thumb)

Integrate over the unwanted variables to get the joint PDF of the wanted variables

Example 4 (Joint Continuous PDF)

Let X and Y be continuous with joint PDF f given by

$$f_{X,Y}(x,y) = \begin{cases} 4x^2y + 2y^5 & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

1. Find the marginals of X and Y
2. Are X and Y independent?

Continuous PDF Example

Example 5 (Joint Continuous PDF)

Let X and Y be continuous with joint PDF f given by

$$f_{X,Y}(x,y) = \begin{cases} 120x^3y & x \geq 0, y \geq 0, x + y \leq 1 \\ 0 & \text{else} \end{cases}$$

1. Verify that f is indeed a valid joint PDF
2. Find the marginals of X and Y

Definition 6 (Conditional Continuous PDF)

If X and Y are continuous with joint PDF $f_{X,Y}$, the **conditional PDF** of Y given $X = x$ is:

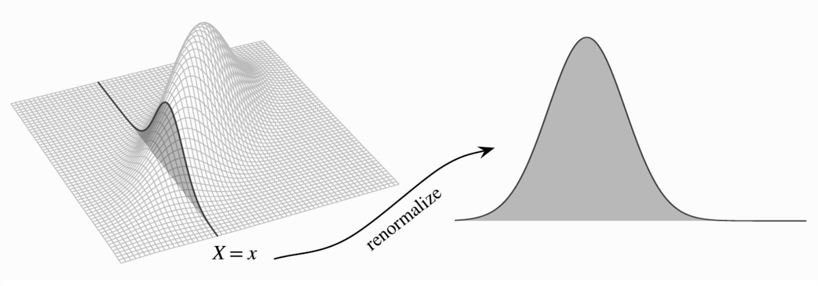
$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \tag{4}$$

The **conditional PDF** of X given $Y = y$ is:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{5}$$

Conditional Continuous PDF

Conditional PDF of Y given $X = x$. The conditional PDF $f_{Y|X}(y \mid x)$ is obtained by renormalizing the slice of the joint PDF at the fixed value x



Conditional Continuous PDF - A word of caution

- If X is continuous then we will have $P(X = x) = 0$

Conditional Continuous PDF - A word of caution

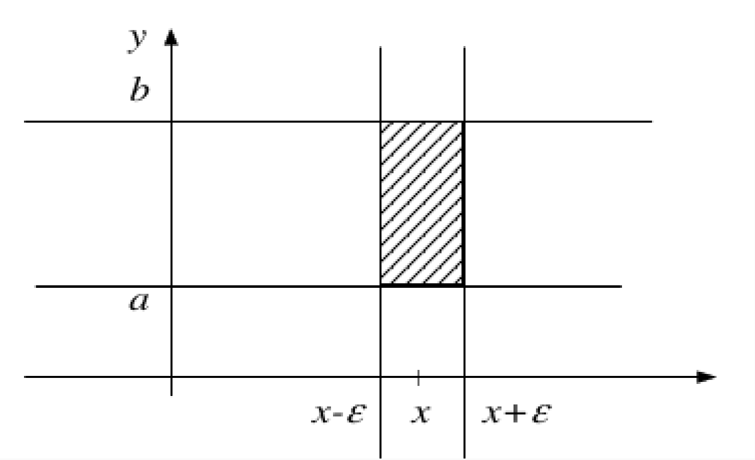
- If X is continuous then we will have $P(X = x) = 0$
- We instead condition on a small interval, $x - \epsilon < X \leq x + \epsilon$, where $\epsilon > 0$, and take the limit as $\epsilon \to 0$

Conditional Continuous PDF - A word of caution

- If X is continuous then we will have $P(X = x) = 0$
- We instead condition on a small interval, $x - \epsilon < X \leq x + \epsilon$, where $\epsilon > 0$, and take the limit as $\epsilon \to 0$

Conditional Continuous PDF - A word of caution

- If X is continuous then we will have $P(X = x) = 0$
- We instead condition on a small interval, $x - \epsilon < X \leq x + \epsilon$, where $\epsilon > 0$, and take the limit as $\epsilon \to 0$



Conditional Continuous PDF - A word of caution

- We will not fuss over this technicality

Conditional Continuous PDF - A word of caution

- We will not fuss over this technicality
- Fortunately, many important results such as Bayes' rule work in the continuous case exactly as one would hope

Joint PDF from Conditional and Marginal

- Recall multiplicative rule from probability

$$P(X \cap Y) = P(X \mid Y)P(Y)$$

Joint PDF from Conditional and Marginal

- Recall multiplicative rule from probability

$$P(X \cap Y) = P(X \mid Y)P(Y)$$

- Similarly we can recover the joint PDF $f_{X,Y}$ from the conditional PDF $f_{X|Y}$ and the marginal f_Y:

$$f_{X,Y}(x, y) = f_{X|Y}(x \mid y)f_Y(y)$$

Joint PDF from Conditional and Marginal

- Recall multiplicative rule from probability

$$P(X \cap Y) = P(X \mid Y)P(Y)$$

- Similarly we can recover the joint PDF $f_{X,Y}$ from the conditional PDF $f_{X|Y}$ and the marginal f_Y:

$$f_{X,Y}(x, y) = f_{X|Y}(x \mid y)f_Y(y)$$

- Alternatively we have:

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

Definition 7 (Continuous form of Bayes' rule and Law of Total Probability)

If X and Y are continuous then:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X|Y}(x \mid y) f_Y(y)}{f_X(x)} \tag{6}$$

and

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) dy \tag{7}$$

- In Equation (7), if we plugged in the other expression for $f_{X,Y}(x, y) = f_{Y|X}(x \mid y) f_X(x)$ we get

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \\ &= \int_{-\infty}^{\infty} f_{Y|X}(x \mid y) f_X(x) dy \\ &= f_X(x) \int_{-\infty}^{\infty} f_{Y|X}(x \mid y) dy \end{aligned} \tag{8}$$

Equation (8) implies that $\int_{-\infty}^{\infty} f_{Y|X}(x \mid y) dy = 1$, confirming the fact that conditional PDFs must integrate to 1.

Example 8 (Joint Continuous PDF)

Let X and Y be continuous with joint PDF f given by

$$f_{X,Y}(x,y) = \begin{cases} 4x^2y + 2y^5 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases}$$

1. Compute $P\big(0.2 \leq Y \leq 0.3 \big| X = 0.8\big)$
2. Compute $P\big(0.2 \leq Y \leq 0.3\big)$

Summary of Bayes' Rule

	Y discrete	Y continuous
X discrete	$P(Y = y \mid X = x) = \frac{P(X=x \mid Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y \mid X = x) = \frac{P(X=x \mid Y=y)f_Y(y)}{P(X=x)}$
X continuous	$P(Y = y \mid X = x) = \frac{f_X(x \mid Y=y)P(Y=y)}{f_X(x)}$	$f_{Y \mid X}(y \mid x) = \frac{f_{X \mid Y}(x \mid y)f_Y(y)}{f_X(x)}$

Summary of Law of Total Probability

	Y discrete	Y continuous
X discrete	$P(X = x) =$ $\sum_y P(X = x \mid Y = y)P(Y = y)$	$P(X = x) =$ $\int_{-\infty}^{\infty} P(X = x \mid Y = y)f_Y(y)dy$
X continuous	$f_X(x) =$ $\sum_y f_X(x \mid Y = y)P(Y = y)$	$f_X(x) =$ $\int_{-\infty}^{\infty} f_{X\|Y}(x \mid y)f_Y(y)dy$

Definition 9 (Independence of Continuous RVs)

If X and Y are jointly continuous, then X and Y are independent if and only if their joint density function $f_{X,Y}$ can be chosen to satisfy

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), \qquad x, y \in \mathbb{R} \tag{9}$$

Independence of Continuous RVs - Example

Example 10 (Independence of Continuous RVs)

Let X and Y have joint density

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{8080}\left(12xy^2 + 6x + 4y^2 + 2\right) & 0 \le x \le 6, 3 \le y \le 5 \\ 0 & else \end{cases}$$

1. Compute the marginal densities for X and Y
2. Show that X and Y are independent

Example 11 (Unit Disk)

Let X and Y be a completely random point in the unit disk $\{(x, y) : x^2 + y^2 \leq 1\}$ with joint PDF

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \leq 1 \\ 0 & else \end{cases}$$

1. Can $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ where $f_X(x) = 1$ and $f_Y(y) = 1/\pi$ for $0 \leq x \leq 1, 0 \leq y \leq 1$, demonstrating that X and Y are independent?
2. Try out the point (0.9,0.9) to support your answer above

Theorem 12 (MGFs of Independent Continuous RVs)

Let X and Y be independent random variables with MGF $M_X(t)$ and $M_Y(t)$, respectively. Then the MGF of the random variable $Z = X + Y$ is given by

$$M_Z(t) = M_X(x)M_Y(y) \tag{10}$$

Proof: on board

MGFs For Deriving Distributions

- Let $Z = X + Y$

MGFs For Deriving Distributions

- Let $Z = X + Y$
- Sometimes Theorem (12) can be used to easily derive the distribution of Z if we know the distributions of X and Y

MGFs For Deriving Distributions

- Let $Z = X + Y$
- Sometimes Theorem (12) can be used to easily derive the distribution of Z if we know the distributions of X and Y

- Let $Z = X + Y$
- Sometimes Theorem (12) can be used to easily derive the distribution of Z if we know the distributions of X and Y

Example 13 (Distribution of the sum of Normal RVs)

Let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\gamma, \tau^2)$ be independent normal random variables. From Assignment 3, we know that the MGFs of X and Y are

$$M_X(t) = \exp\{\mu t + \sigma^2 t^2/2\} \qquad M_Y(t) = \exp\{\gamma t + \tau^2 t^2/2\}$$

Thus from Theorem (12), the MGF of $Z = X + Y$ is

$$M_Z(t) = M_X(t) M_Y(t) = \exp\{(\mu + \gamma)t + (\sigma^2 + \tau^2)t^2/2\}$$

MGF of a Normal RV with mean $\mu + \gamma$ and variance $\sigma^2 + \tau^2$

Expectations

Theorem 14 (Expectation of Continuous RVs)

Let g be a function from $\mathbb{R}^2 \to \mathbb{R}$. If X and Y are jointly continuous, with joint PDF $f_{X,Y}$, then

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy \tag{11}$$

Example 15 (Expected distance between two Uniforms)

For independent $X \sim Unif(0,1)$ and $Y \sim Unif(0,1)$, find $E(|X - Y|)$

Covariance and Correlation

Covariance and Correlation

- The mean and variance provided single-number summaries of the distribution of a **single RV**

- The mean and variance provided single-number summaries of the distribution of a **single RV**
- Covariance is a single-number summary of the **joint distribution of two RVs**

Covariance and Correlation

- The mean and variance provided single-number summaries of the distribution of a **single RV**
- Covariance is a single-number summary of the **joint distribution of two RVs**
- Roughly speaking, covariance measures a tendency of two RVs to go up or down together, relative to their expected values

Covariance and Correlation

- The mean and variance provided single-number summaries of the distribution of a **single RV**
- Covariance is a single-number summary of the **joint distribution of two RVs**
- Roughly speaking, covariance measures a tendency of two RVs to go up or down together, relative to their expected values
- Positive covariance between X and Y indicates that when X goes up, Y also tends to go up

Covariance and Correlation

- The mean and variance provided single-number summaries of the distribution of a **single RV**
- Covariance is a single-number summary of the **joint distribution of two RVs**
- Roughly speaking, covariance measures a tendency of two RVs to go up or down together, relative to their expected values
- Positive covariance between X and Y indicates that when X goes up, Y also tends to go up
- Negative covariance indicates that when X goes up, Y tends to go down.

Definition 16 (Covariance)

The Covariance between RVs X and Y is

$$
\begin{aligned}
Cov(X, Y) &= E\big((X - EX)(Y - EY)\big) && (12)\\
&= E\big((X - \mu_X)(Y - \mu_Y)\big) && (13)
\end{aligned}
$$

Covariance and Correlation

Definition 16 (Covariance)

The Covariance between RVs X and Y is

$$\begin{aligned} Cov(X, Y) &= E\big((X - EX)(Y - EY)\big) && (12) \\ &= E\big((X - \mu_X)(Y - \mu_Y)\big) && (13) \end{aligned}$$

Proposition 17 (Alternative definition of Covariance)

$$Cov(X, Y) = E(XY) - E(X)E(Y) \tag{14}$$

Proof: Assignment 4

Definition 18 (Correlation)

The Correlation between RVs X and Y is

$$\rho_{XY} = R^2 = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \tag{15}$$

$$= \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{16}$$

Definition 18 (Correlation)

The Correlation between RVs X and Y is

$$\rho_{XY} = R^2 = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \tag{15}$$

$$= \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{16}$$

- Covariance depends on the units in which X and Y are measured

Definition 18 (Correlation)

The Correlation between RVs X and Y is

$$\rho_{XY} = R^2 = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \tag{15}$$

$$= \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{16}$$

- Covariance depends on the units in which X and Y are measured
- if X is in centimeters rather than meters, the covariance is multiplied a hundredfold

Covariance and Correlation

Definition 18 (Correlation)

The Correlation between RVs X and Y is

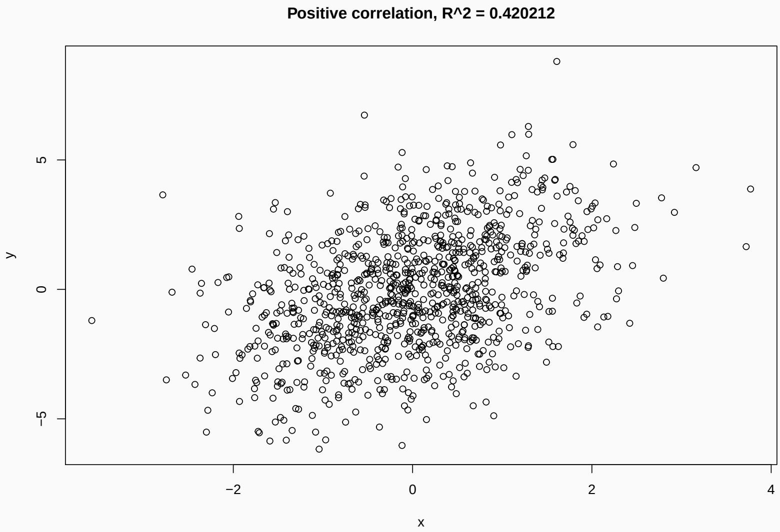
$$\rho_{XY} = R^2 = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \tag{15}$$

$$= \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{16}$$

- Covariance depends on the units in which X and Y are measured
- if X is in centimeters rather than meters, the covariance is multiplied a hundredfold
- Correlation is easier to interpret since it is a **unitless version of covariance**

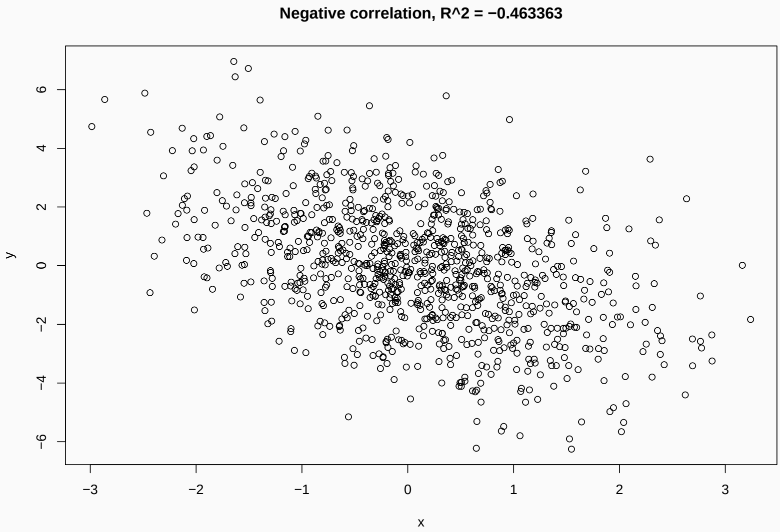
Visualization of Positive Correlation

```
x <- rnorm(1e3); y <- x + 2 * rnorm(1e3)
plot(x ,y, main = sprintf("Positive correlation, R^2 = %g",cor(x,y)))
```



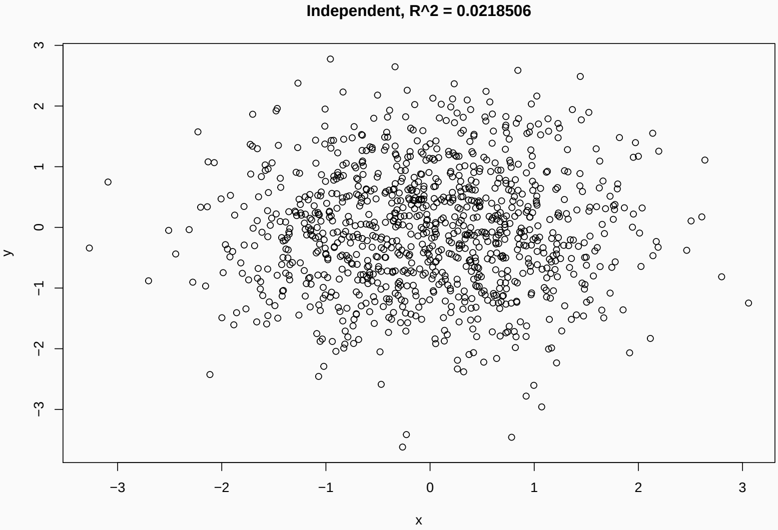
Visualization of Negative Correlation

```
x <- rnorm(1e3); y <- -x + 2 * rnorm(1e3)
plot(x ,y, main = sprintf("Negative correlation, R^2 = %g",cor(x,y)))
```



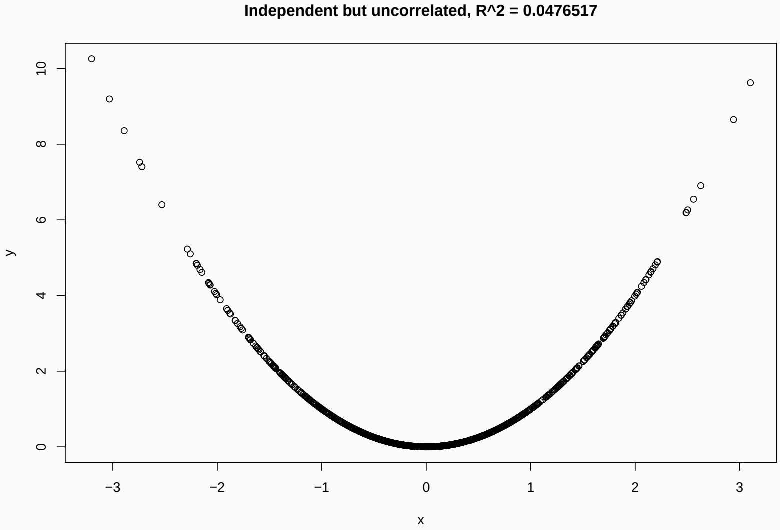
Visualization of Independence

```r
x <- rnorm(1e3); y <- rnorm(1e3)
plot(x ,y, main = sprintf("Independent, R^2 = %g",cor(x,y)))
```



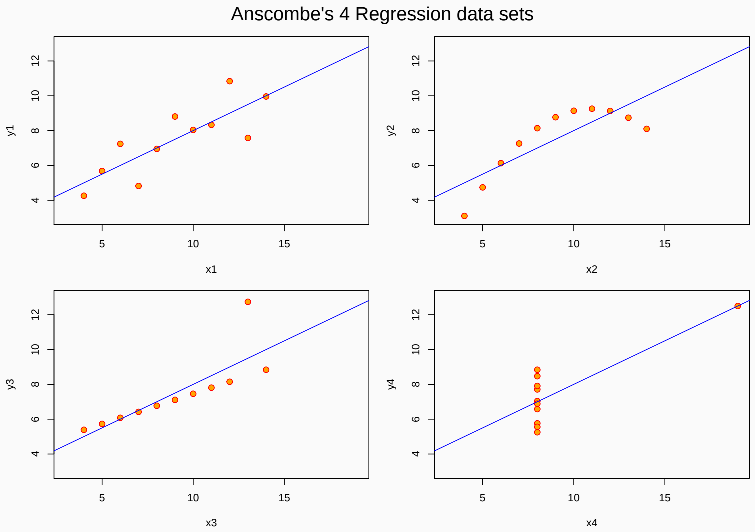
Visualization of Dependent but Uncorrelated

```
x <- rnorm(1e3); y <- x^2
plot(x ,y, main = sprintf("Independent but uncorrelated, R^2 = %g",cor(x,y,method = "s
```

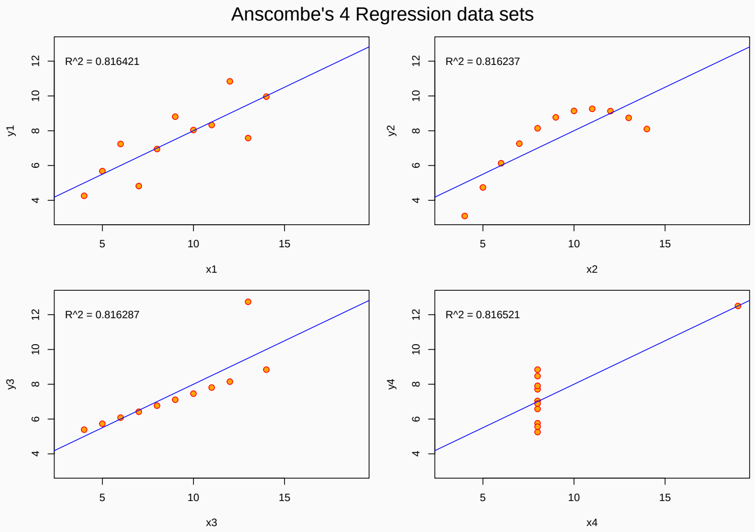


Anscombes Quartet

```r
data("anscombe")
```



```r
data("anscombe")
```



- If X and Y tend to move in the same direction, then $X - EX$ and $Y - EY$ will tend to be either both positive or both negative, so $(X - EX)(Y - EY)$ will be positive on average, giving a positive covariance

- If X and Y tend to move in the same direction, then $X - EX$ and $Y - EY$ will tend to be either both positive or both negative, so $(X - EX)(Y - EY)$ will be positive on average, giving a positive covariance
- If X and Y tend to move in opposite directions, then $X - EX$ and $Y - EY$ will tend to have opposite signs, giving a negative covariance

- If X and Y tend to move in the same direction, then $X - EX$ and $Y - EY$ will tend to be either both positive or both negative, so $(X - EX)(Y - EY)$ will be positive on average, giving a positive covariance
- If X and Y tend to move in opposite directions, then $X - EX$ and $Y - EY$ will tend to have opposite signs, giving a negative covariance
- If X and Y are independent, then their covariance is zero. We say that RVs with **zero covariance are uncorrelated**.

Theorem 19 (Independence and Not Correlated)

If X and Y are independent, then they are uncorrelated.

proof: on board

Theorem 19 (Independence and Not Correlated)

If X and Y are independent, then they are uncorrelated.

proof: on board

> *The converse of this theorem is false: just because X and Y are uncorrelated does not mean they are independent*

Properties of Covariance

1. $Cov(X, X) = Var(X)$
2. $Cov(X, Y) = Cov(Y, X)$
3. $Cov(X, c)$ = 0 for any constant c
4. $Cov(aX, Y) = aCov(X, Y)$ for any constant a
5. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
6. $Cov(X+Y, Z+W) = Cov(X, Z)+Cov(X, W)+Cov(Y, Z)+Cov(Y, W)$
7. $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$

Proof: Assignment 4

Properties of Correlation

Theorem 20 (Correlation Bounds and linear relationships)

For any random variables X and Y

1. $-1 \leq \rho_{XY} \leq 1$
2. $\rho_{XY} = \pm 1$ *if and only if there exist numbers $a \neq 0$ and b such that $P\big(Y = aX + b\big) = 1$ (i.e. $\rho_{XY} = \pm 1$ implies X and Y are linearly related, and if X and Y are linearly related, then $\rho_{XY} = \pm 1$)*
3. *If $\rho_{XY} = 1$, then $a > 0$, and if $\rho_{XY} = -1$ then $a < 0$ (i.e. a and ρ_{XY} have the same sign)*

proof: on board

Example 21 (Correlation)

Let the joint PDF of (X, Y) be

$$f(x,y) = \begin{cases} 1 & 0 < x < 1, x < y < x+1 \\ 0 & else \end{cases}$$

Find the correlation of X and Y