Statistics for Data Science

Lecture 7

Multiple Random Variables

Ken Cameron

Content

- Multiple Random Variables
- Joint and Marginal Distributions
- Conditional Distributions and Independence
- Covariance and Correlation

Multiple Random Variables

- Until now, we've been (mostly) dealing with one random variable at a time.
  - This is known as a 'univariate' model.
- In this lecture we will consider the case where we have more than one random variable.
  - This is known as a 'multivariate' model.

Multiple Random Variables

- When considering a experimental situation it is not unusual to observe more than one random variable simultaneously.
  - For example, if recording heights, we might also be recording weight, temperature, blood pressure.
- Initially we'll deal with the case of just two random variables.
  - This is known as 'bivariate' modelling.

Definition

- Recall:

  - In the univariate case, a random variable was defined to be a function of a sample space S into real numbers.

- Multivariate case:

  - An n-dimensional random vector is a function from a sample space S into R^n, n-dimensional Euclidean space.

2 Dimensional Example

- Consider a point $(x,y) \in R^2$
  - Where R^2 denotes the plane.
- We can defined a two-dimensional random vector:
  - (X, Y)

Joint PMF

- Let (X,Y) be a discrete bivariate random vector.

- The function f(x,y) from R^2 into R is defined by:

  - $f(x,y) = P(X=y, Y=y)$

- Is known as the joint probability mass function.

  - The joint pmf of (X,Y) completely defines the probability distribution of the random vector (X,Y).

Hang on...

- Have n’t we see this notation before?

- Yes, in the very first lecture.

  - I introduced the joint, marginal and conditional probability.

- But without any detailed explanation/proof.

Joint Probability

- The ***joint*** probability is the chance that a collection of events occur together

- We write

$$p(x = X, y = Y, z = Z)$$

- Or simply

$$p(x, y, z)$$

- Exchangeability

$$p(x, y, z) = p(x, z, y) = p(y, z, x) = \cdots$$

Marginal Probability

- The ***marginal*** probability is the chance of observing a random variable in a particular state when two (or more) events are observed simultaneously

- Given variables x and y, the joint is $p(x, y)$
  the marginal is:

$$p(x) = \sum_y p(x, y)$$

Conditional Probability

- The ***conditional*** probability is the chance of observing an event y given that an event x has occurred

- This is written $p(y|x)$, and can be calculated:

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

The generic version of selling newspapers

	Y_1	Y_2	$\dots$	$\dots$	Y_n	*Marginal Probabilities of X*
X_1	p_{11}	p_{12}	$\dots$	$\dots$	p_{1n}	P_1
X_2	p_{21}	p_{22}	$\dots$	$\dots$	p_{2n}	P_2
.	.	.	$\dots$	$\dots$	.	.
.	.	.	$\dots$	$\dots$	.	.
X_m	p_{m1}	p_{m2}	$\dots$	$\dots$	p_{mn}	P_m
Marginal Probabilities of Y	P_1'	P_2'	$\dots$	$\dots$	P_n'	1

Expectation

- We can compute multivariate expectation values in just the same way as we computed them for the univariate case.

- If g(x,y) is a real-valued function defined for all possible values of (x,y) of the discrete random vector (X,Y) then g(X,Y) is itself a random variable and its expected value is...

$$Eg(X,Y) = \sum_{(x,y)\in R^2} g(x,y) f(x,y)$$

Without the sample space.

- Any non-negative function from R^2 into R that:
  - Is non-zero for at most a countable number of (x,y) pairs
  - And sums to 1.

  - Is the joint pmf for some bivariate discrete random vector (X,Y).

- By defining f(x,y) we can define a probability model for (X,Y) without ever working with the fundamental sample space S.

Theorem

- Let (X,Y) be a discrete bivariate random vector with joint pmf $f_{X,Y}(x,y)$

- Then the marginal pmfs of X and Y

  - $f_X(x) = P(X = x)$ and $f_Y(y) = P(Y = y)$

- Are given by:

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

Proof - for $f_X(x)$

- For any $x \in R$, let $Ax = \{(x,y) : -\infty < y < \infty\}$

- That is A_x is the line in the plane with first coordinates equal to x.

- Then, for any $x \in R$

Proof

$$
\begin{aligned}
f_X(x) &= P(X = x) \\
&= P(X = x, -\infty < Y < \infty) \\
&= P((X, Y) \in A_x) \\
&= \sum_{(x,y) \in A_x} f_{X,Y}(x, y) \\
&= \sum_{y \in R} f_{X,Y}(x, y)
\end{aligned}
$$

Conditional Distributions

- Often, when two random variables are observed, their values a related.

- For example,

  - If we were sampling the height of people and their weight,

  - We would expect taller people to weight more.

- i.e. We'd expect, given two random variables X, Y the value of X would tell us something about the value of Y, even if not its exact value.

Conditional Probabilities

- Conditional probabilities regarding Y given knowledge of the value of X can be computed using the joint distribution.

- However, sometimes, knowledge of X will give us no information about Y.

Definition

- Let (X,Y) be a discrete bivariate random vector with
  - Joint pmf $f(x,y)$
  - And marginal $f_X(x)$ and $f_Y(y)$.

- For any x such that
  - $P(X=x) = f_X(x) > 0$
  - The conditional pmf of Y given X=x
  - Is the function of y denoted by $f(y|x)$ and defined by

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

- As we've called f(y|x) a pmf, we should verify that this function of y does indeed define a pmf of a random variable.

  - Firstly, f(y|x) >= 0 for every y since f(y,x) > =0 and $f_X(x) > 0$.

  - Secondly, $\sum_y f(y|x) = \frac{\sum_y f(x,y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$

- Thus, f(y|x) is indeed a pmf and can be used in the usual way to compute probabilities involving Y given the knowledge of X =x occurred.

Independence

- Let X and Y be independent random variables, then:

- For any $A \subset R$ and $B \subset R$, $P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$

  - That is the events $P(X \in A)$ and $P(Y \in B)$ are independent

- Let $g(x)$ be a function only of x and $h(y)$ be a function only of y, then,

  - $E(g(X)h(Y)) = (Eg(X))(Eh(Y))$

Proof

$$
\begin{aligned}
E(g(X)g(Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty} g(x)f_X(x)\int_{-\infty}^{\infty} h(y)f_Y(y)dxdy \\
&= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy\right) \\
&= (Eg(X))(Eh(Y))
\end{aligned}
$$

Independence

- We said mutually exclusive events add
- If they can occur together but are unrelated, we call this *statistically independent*, and say that
$$p(x|y) = p(x)$$
- e.g. probability of getting 6 on two dice: getting a 6 on one die does not affect the probability of getting a 6 on the other
- We can decompose the joint as
$$p(x, y) = p(x|y)p(y)$$
from the definition of the conditional probability
- Therefore statistically independent probabilities multiply:
$$p(x, y) = p(x)p(y)$$

Conditional Independence

- More subtly, it is possible for x and y to be *conditionally independent* if

$$p(x|y,z) = p(x|z)$$

- This means if we know z, knowing y gives no new information about x
- But it is still possible that $p(x|y) \neq p(x)$
- So x and y both depend on z

Example

- Alice (A) and Bob (B) both flip a coin *which may be biased*

- It is *not true* that these are statistically independent

$$p(A = H|B = H) \neq p(A = H)$$

because if Bob flips a head on a biased coin, we would increase the probability of Alice flipping a head

- However if we denote Z as the event “the coin is biased towards heads”, then we can write

$$p(A = H|B = H, Z) = p(A = H|Z)$$

- We can remove Bob from the equation because we know the coin is biased
- An example of “correlation does not imply causation”

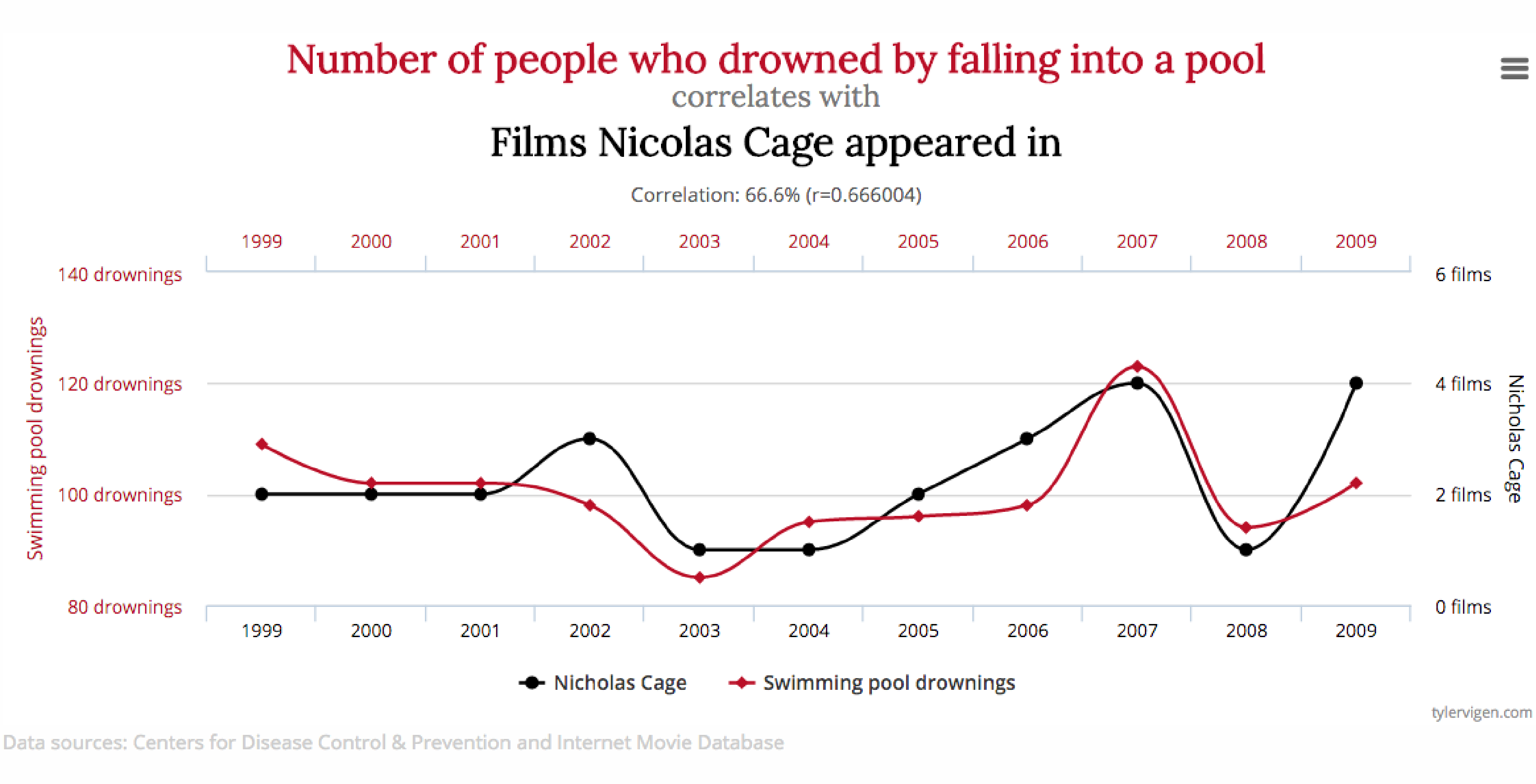
Covariance and Correlation

- We’ve considered the presence or absence of a relationship between two random variables.

- If there is a relationship, it may be strong or weak.

- It would be useful to have a way to quantify the strength of the relationship.

Correlation

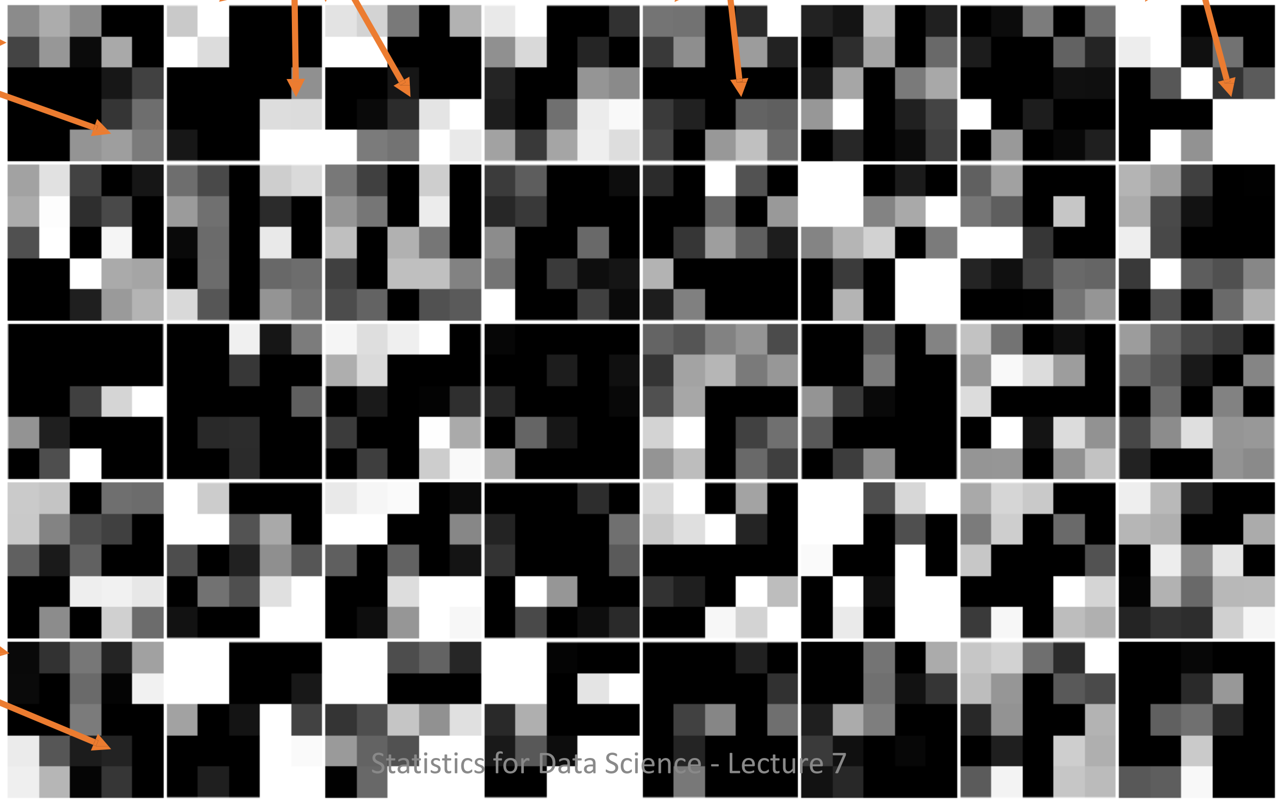


Note: could be a third cause, but more likely to be totally random…!

http://www.tylervigen.com/spurious-correlations

Let's now take a look at non-independent variables

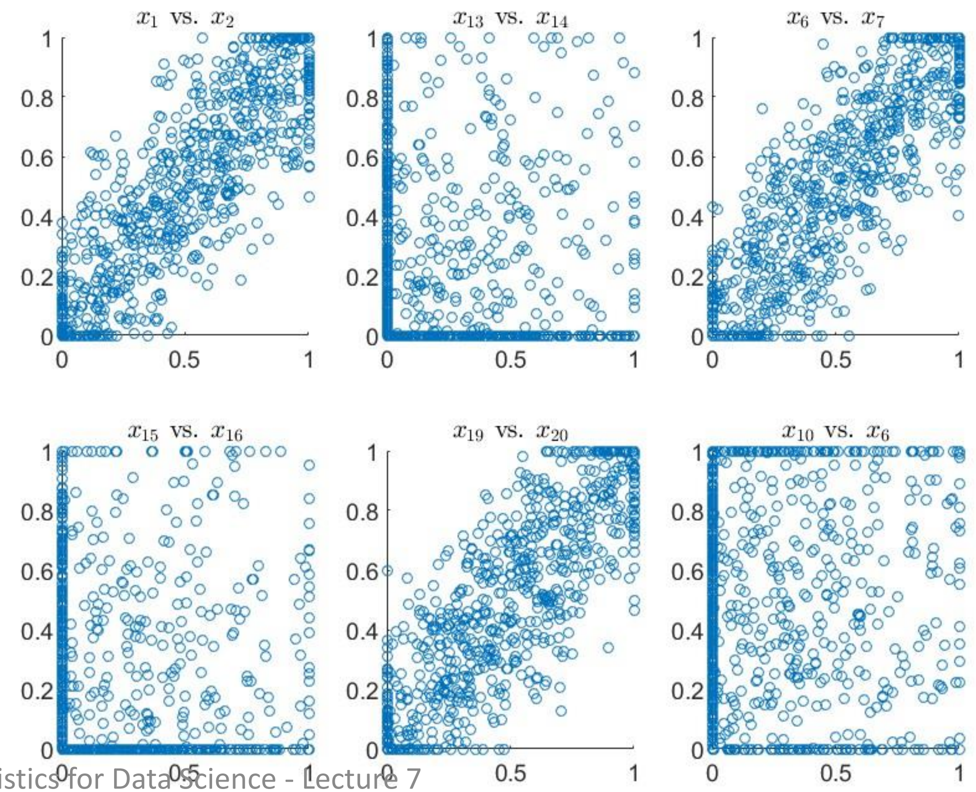
- Example:
  - Let's consider a set of 5x5 gray scale images that look like this:



These regions are either dark together or bright together

Non-independence

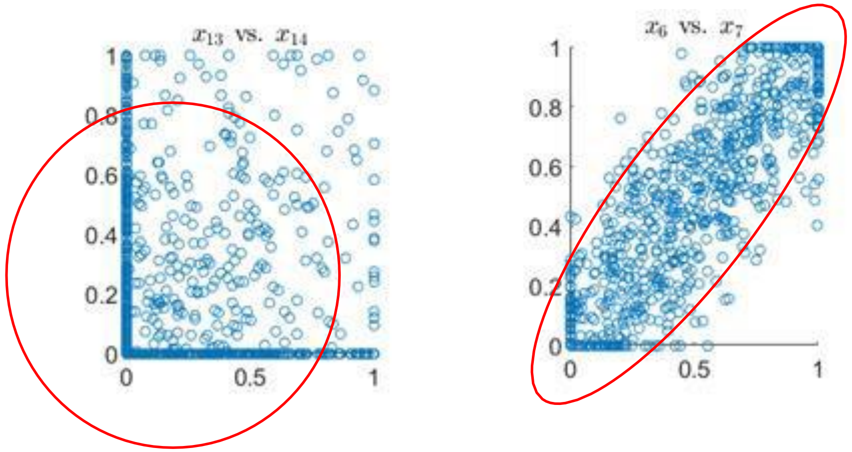
- In this example we have 25 random variables
- We can plot one variable against another to see the relationship



- Which variables are correlated?

Non-independence

- In this example we have 25 random variables
- We can plot one variable against another to see the relationship



- Which variables are non-independent?
  - The non-axis aligned ellipses tell us that there is correlation between variables

Covariance

- The covariance

- $\text{Cov}(X,Y) = E((X - \mu_X)(Y - \mu_Y))$

    - If large values of X tend to be observed with large values of Y

    - And small values of X tend to be observed with small values of Y

    - Then Cov(X,Y) will be positive.

Calculating the covariance of random variables.

For discrete variables with a finite set of equal-probability values, x_k and y_k for k= 1,…,n we can re-write this as:

$$\text{Cov(X,Y)} = \frac{1}{n}\sum_{i=1}^{n}(xi \ - \mu_X)(yi \ - \mu_Y)$$

It can also be formulated without referencing the mean values of each variable:

$$\text{Cov(X,Y)} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j>i}^{n}(xi \ - xj)(yi \ - yj)$$

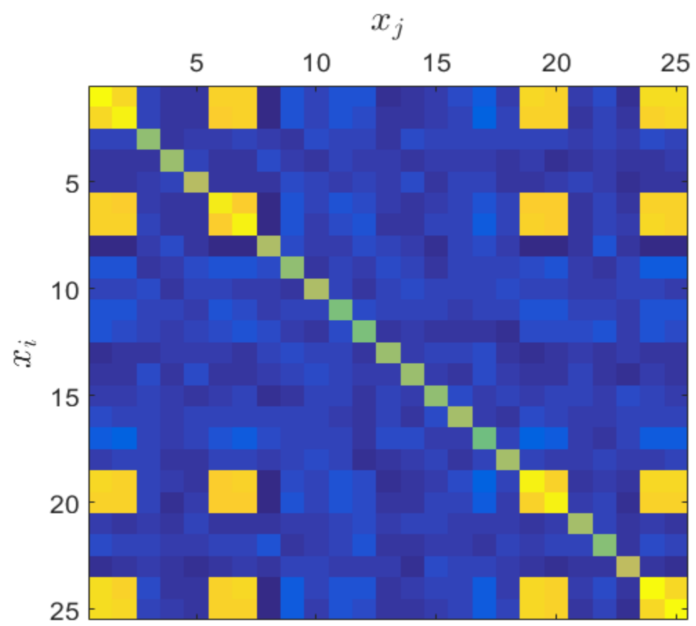
For continuous random variables, the covariance can be expressed as:

$$\text{Cov(X,Y)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - E[X])(y \ - E[Y])f_{XY}(x,y)dxdy$$

Where $f_{XY}(x,y)$ is the joint probability density function of X and Y.

Covariance Matrix (Heatmap)

- Where we have several random variables, a covariance hmatrix gives a visual guide to their relative (in)dependence.



- Note the value of the diagonal.

Correlation

- The correlation coefficient is defined as:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$