Lectures

MVE055 / MSG810
Mathematical statistics and discrete mathematics

Moritz Schauer

Last updated September 14, 2022

GU & Chalmers University of Technology

Summary distributions

Discrete distributions		
Bernoulli	Ber(p)	single trial, success with probability p
Binomial	Bin(p)	n of Ber(p) trials
Poisson	Poisson(μ)	number of events (meanμ) in interval/box
Geometric	Geo(p)	Ber(p) trials until first success
Negative binomial	nBin(r, p)	Ber(p) trials until r successes
Hypergeometric	Hyp(N, r, n)	r of N items are marked. How many marked items in a selection of n items?

Summary distributions

Continuous distributions		
Uniform	Uni($[a, b]$)	equally likely random points in $[a, b]$,
Normal	Normal(μ, σ^2)	our "approximately $\mu \pm \sigma$", variation by multiple independent causes
Exponential	Exp(λ)	distance between arrival times at rate λ (mean $1/\lambda$)
Gamma	Gamma(α, β)	"usefull"
χ^2	$\chi^2(n)$	sum of n normal squares

Exponential counts the time between ticks, Poisson the number of ticks.

Bivariate distributions

For two discrete random variables X and Y, that together give a pair of random numbers,

the joint density (probability mass function) is

$$f_{X,Y}(x, y) = \mathsf{P}(X = x, Y = y) = \mathsf{P}(X = x \text{ and } Y = y).$$

Here $f_{X,Y}(x, y) \geq 0$ and $\sum_{\text{all } x,y} f_{X,Y}(x, y) = 1$.

Example

Let X and Y be the number of girls, respectively boys in a randomly chosen Swedish family. The joint density function $f_{X,Y}(x, y)$ is given in the table below.

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$$\sum_{\text{all } x,y} f_{X,Y}(x, y) = 1$$

$\mathrm{P}(X = 0 \text{ and } Y = 1) = f_{X,Y}(0, 1) = 0.16$

$\mathrm{P}(X = 2) = f_{XY}(2, 0) + f_{X,Y}(2, 1) + f_{XY}(2, 2) = 0.08$

Expected values ♡

$$\mathsf{E}[h(X,Y)] = \sum_{\text{all } x,y} h(x,y) f_{X,Y}(x,y).$$

For example:

$$\mathsf{E}[X+Y] = \sum_{\text{all } x,y} (x+y) f_{X,Y}(x,y)$$

with $h(x,y) = x + y$.

Expected number of children

X and Y be the number of girls, respectively boys in a randomly chosen Swedish family.

$\mathsf{E}[X+Y]$ is the expected number of girls + boys = children. So $h(x,y) = x + y$.

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$\mathsf{E}[X+Y] = (0+0)\cdot 0.38 + (1+0)\cdot 0.17 + ... = 1.08$

The distribution of X alone (or Y alone) is called the **marginal distribution** of X (or of Y).

Marginal densities / densities of marginal distributions

Given a pair of discrete random variables (X, Y) with joint density $f_{X,Y}$ density for X and Y are given by

$$f_X(x) = \sum_{\text{all } y} f_{X,Y}(x, y)$$

$$f_Y(y) = \sum_{\text{all } x} f_{X,Y}(x, y).$$

and called **marginal densities** (marginal p.m.f.'s.)

Y / X	0	1	2	3	4	f_X
0	0.38	0.16	0.04	0.01	0.01	0.60
1	0.17	0.08	0.02			0.27
2	0.05	0.02	0.01			0.08
3	0.02	0.01				0.03
4	0.02					0.02
f_Y	0.64	0.27	0.07	0.01	0.01	1

Continuous bivariate distributions

Draw a random point (X, Y) from a square area $[0,1] \times [0,1]$. All points are equally likely. What is the probability that $X > Y$?

Solution idea: Draw the square. The event $\{X > Y\}$ corresponds to all points below the diagonal, with area 1/2.

$$\mathsf{P}(X > Y) = \frac{1}{2}$$

Continuous bivariate random variables

A joint density of a pair of continuous random variables X and Y is a function $f_{X,Y}(x,y)$ such that

$$\mathsf{P}(a \leq X \leq b \text{ and } c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y.$$

Properties of densities:

1. $f_{X,Y}(x,y) \geq 0$,
2. $\int\int f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = 1$

Marginal densities

For a bivariate continuous random variable (X, Y), the probability density functions for X and Y are given by

$$f_X(x) = \int f_{X,Y}(x, y)\mathrm{d}y$$
$$f_Y(y) = \int f_{X,Y}(x, y)\mathrm{d}x$$

Expected value

For a bivariate random variable (X, Y), the expected values of X and Y are given by

$$\mathsf{E}[X] = \begin{cases} \displaystyle\sum_{\text{all } x,\ y} x f_{X,Y}(x,y), & \text{for } X \text{ discrete,} \\ \displaystyle\int\int x f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y, & \text{for } X \text{ continuos,} \end{cases}$$

and

$$\mathsf{E}[Y] = \begin{cases} \displaystyle\sum_{\text{all } x,\ y} y f_{X,Y}(x,y), & \text{for } Y \text{ discrete,} \\ \displaystyle\int\int y f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y, & \text{for } Y \text{ continuous.} \end{cases}$$

Conditional distribution

The *conditional distribution of X given $Y = y$* is defined by its density

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

provided that $f_Y(y) > 0$.

Independent random variables (cont' or discr')

Two random variables X and Y are called independent if their bivariate density can be written as product of the marginal densities:

$$f_{X,Y}(u,v) = f_X(u) f_Y(v).$$

There is no "samvariation", knowing X does not explain Y, etc.

Covariance

Covariance

Covariance between random variables X and Y is defined as $\mathrm{Cov}(X, Y) = \mathsf{E}[(X - \mu_X)(Y - \mu_Y)]$, where $\mu_X = \mathsf{E}[X]$ and $\mu_Y = \mathsf{E}[Y]$.

- According to the definition,

$$\mathrm{Cov}(X, Y) = \begin{cases} \sum\limits_{\text{all } x,\, y} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y), & \text{discrete} \\ \int \int (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) \mathrm{d}x \mathrm{d}y, & \text{cont.} \end{cases}$$

- Note that $\mathrm{Cov}(X, X) = \mathsf{V}(X)$.
- **If** X and Y are independent, **then** $\mathrm{Cov}(X, Y) = 0$ and $\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y]$.
- Unit??

Rules for covariance

$\mathrm{Cov}(X, Y)$ can be calculated as
$\mathrm{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$.

For two random variables X and Y, and two numbers a and b we have

$$\mathsf{V}(aX + bY) = a^2\mathsf{V}(X) + b^2\mathsf{V}(Y) + 2ab \ \mathrm{Cov}(X, Y).$$

Examples:

$$\mathsf{V}(2X) = \mathsf{V}(X + X) = \mathsf{V}(X) + \mathsf{V}(X) + 2 \ \mathrm{Cov}(X, X) = 4\mathsf{V}(X)$$

$$\mathsf{V}(X + Y) = \mathsf{V}(X) + \mathsf{V}(Y) \text{ when } X \text{ and } Y \text{ are independent}$$

(“Fun” thing to do: look up the law of cosines.)

Correlation and independence

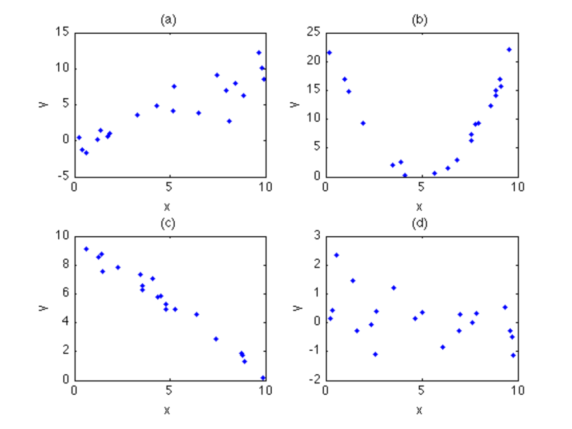
Correlation

The correlation coefficient is defines as

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathsf{V}(X)\mathsf{V}(Y)}}.$$

- A measure of linear relationship (linjär samvariation) of X and Y.
- It holds $-1 \leq \rho \leq 1$.
- X and Y are called uncorrelated if $\rho(X, Y) = 0$ (there is no "linjär samvariation") .
- Unit??

Visualisation



Assume 2d measurements (x_i, y_i). A scatter plot is a two-dimensional plot in which each (x_i, y_i) measurement is represented as a point in the x-y-plane.

Descriptive statistic for bivariate data

The sample covariance is defined as,

$$c_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

and sample correlation coefficient is defined as

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{c_{xy}}{s_x s_y}$$

The sample correlation is a measure of linear dependence.

In the picture $r_{xy} = 0.8067$ i (a), $r_{xy} = 0.2912$ i (b), $r_{xy} = -0.9884$ i (c), och $r_{xy} = 0.3640$ i (d).

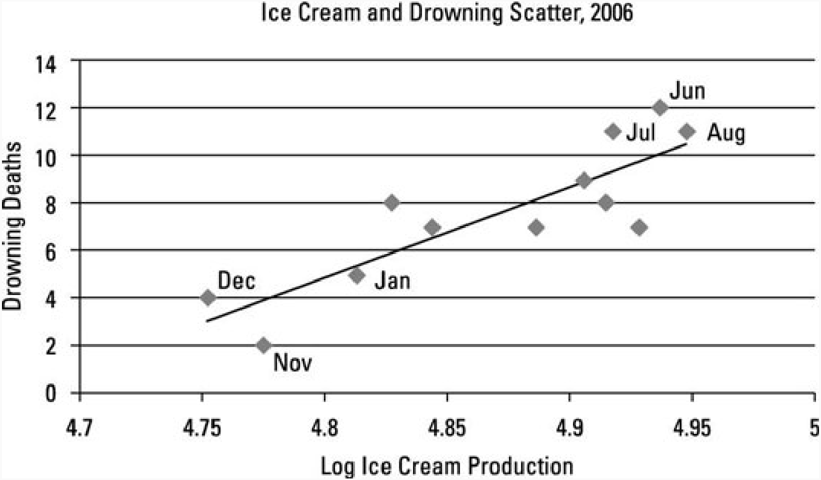
We have the following relationship between dependence and correlation:

- If X and Y are independent, then they are also uncorrelated.
- (Thus if X and Y are uncorrelated, they do not need to be independent.)

This is natural because two random variables are independent if there is no "samvariation" at all, while they are not correlated if there is no "*linjär* samvariation".

Correlation, dependence and causality

- Correlation does not say anything about causality!*
- Sometimes correlation is present but can be explained by a third variable which was not measured.
- Month with high ice cream sales tend to have more drowning accidents. Time to ban ice cream? In this example, an important variable which perhaps was not measured is the sunshine. Such variables are sometimes called confounding variables.

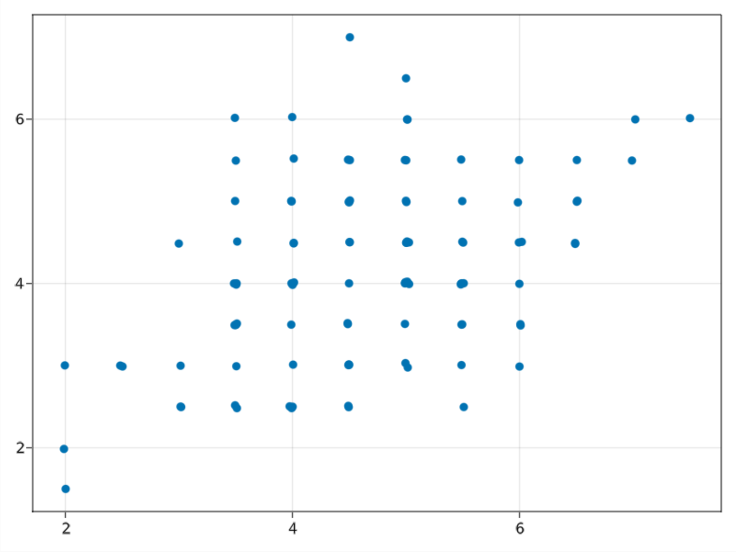


https://twitter.com/dannagal/status/1244082688899919872, September 14, 2022

Causality

- Correlation can also be introduced by selection effects.
- Exam with two questions, one difficult, one easy. A student achieves X out of 10 points on the easy question, Y out of 10 points on the difficult question (random).
- Say X and Y slightly positively correlated. *But* only students with $X + Y \geq 10$ pass. Say I tell you the student has passed.
- Passing students performance on easy questions may now be negatively correlated with performance on the difficult question.

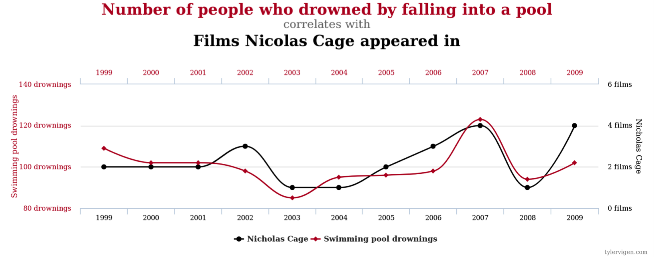
Exam points



Causality

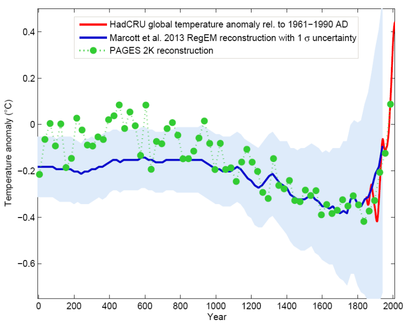
- If we want to know/predict what will change if we perform an action we need insight into causality.
- Will the number of drowning accidents change if we ban ice?
- There are many causal statements in the news!
  - "Do not skip breakfast if you want to reduce the risk of coronary heart disease"
- Be careful...
  - Candidate for a confounding variable: stress.
- We need to understand the science to answer causal questions! We will come back to this later.

Cherry picking



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

Thinking statistics: Global warming



Two millennia of mean surface temperatures according to different reconstructions from climate proxies with the instrumental temperature record overlaid in red.

Stefan Rahmstorf: Paleoclimate: The End of the Holocene.

`http://www.realclimate.org/index.php/archives/2013/09/paleoclimate-the-end-of-the-holocene/`.

Web. 3 Feb. 2019.