Lecture 6: Joint distributions

MVE055 / MSG810 Mathematical statistics and discrete mathematics)

Moritz Schauer

Last updated September 14, 2020, 2020

GU & Chalmers University of Technology

What was the mean and the variance of $E \sim \mathrm{Bin}(n, p)$?
$\mathsf{E}[X] = np$. $\quad\quad \mathrm{Var}(X) = np(1-p)$.

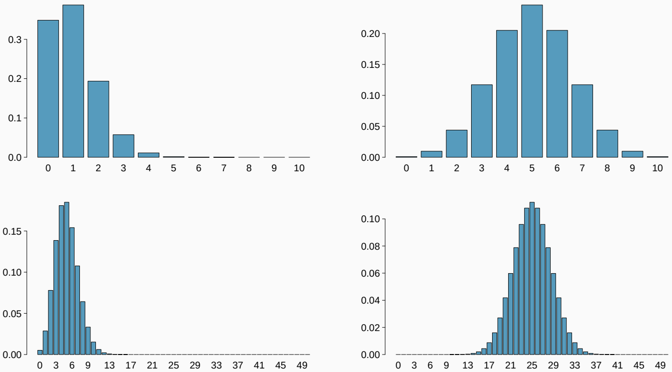
Normal approximation of Binomial distribution

If $X \sim \mathrm{Bin}(n, p)$, X is approximately normally distributed with mean np and variance $np(1-p)$,

$$X \overset{\text{approx.}}{\sim} \mathrm{N}(np, np(1-p)),$$

if both $np > 5$ and $n(1-p) > 5$.

Normal approximation



$n = 10$

$n = 50$

$p = 0.1$ $\quad$ $p = 0.5$

Bivariate distributions

Definition

Informal: A two-dimensional or bivariate random variable (X, Y) produces a pair of random numbers.

For discrete random variables we have the probability mass function

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

Here $f_{X,Y}(i,j) \geq 0$ and $\sum_{i,j} f_{X,Y}(i,j) = 1$.

Example

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

Y / X	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				

$\sum_{x=0}^{4} \sum_{y=0}^{4} f_{X,Y}(x, y) = 1$
$P(X = 0 \text{ and } Y = 1) = f_{X,Y}(0, 1) = 0.16$
$P(X = 2) = f_{XY}(2, 0) + f_{X,Y}(2, 1) + f_{XY}(2, 2) = 0.08$

Expected value

$$\mathsf{E}h(X,Y) = \sum_{\text{all } j} h(i,j) f_{X,Y}(i,j)$$

and called marginal densities / marginal p.m.f.'s.

For example:

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

with $h(i,j) = i + j$.

Expected number of children

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

$E[X+Y]$ is the expected number of boys + girls = children. So $h(i,j) = i + j$.

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				

$E[X+Y] = (0+0) \cdot 0.38 + (1+0) \cdot 0.17 + \ldots.$

Marginal distributions

Given a discrete random variable (X, Y) we the probability mass functions for X and Y are given by

$$f_X(i) = \sum_{\text{all } j} f_{X,Y}(i,j)$$

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

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 random variables

For bivariate continuous random variables we have a probability density function $f_{X,Y}(x, y)$ with properties

1. $f_{X,Y}(x, y) \geq 0$,
2. $\int \int f_{X,Y}(x, y) dx dy = 1$, and
3. $\mathsf{P}(a \leq X \leq b \text{ and } c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) \mathrm{d}x \mathrm{d}y$.

Marginal distributions

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

$$f_X(x) = \int f_{X,Y}(x, y)dy$$

$$f_Y(y) = \int f_{X,Y}(x, y)dx$$

Expected value

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

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

and

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

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

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

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

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,

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

- Note that $\mathrm{Cov}(X, X) = \mathsf{V}(X)$.
- $\mathrm{Cov}(X, Y)$ can be calculated as $\mathrm{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$.
- If X and Y are independent: $\mathrm{Cov}(X, Y) = 0$, $\mathsf{E}(XY) = \mathsf{E}(X)\mathsf{E}(Y)$.

Rules for covariance

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:

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

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

("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 (samvariation) of X and Y.
- It holds $-1 \leq \rho \leq 1$.
- X and Y are called uncorrelated if $\rho(X, Y) = 0$.

We have the following relationship between dependence and correlation:

- If X and Y are independent, they are also uncorrelated.
- If X and Y are uncorrelated, they do not need to be independent.

These relationships are natural because two random variables are independent if there is no co-variation at all, while they are not correlated if there is no *linear* co-variation.

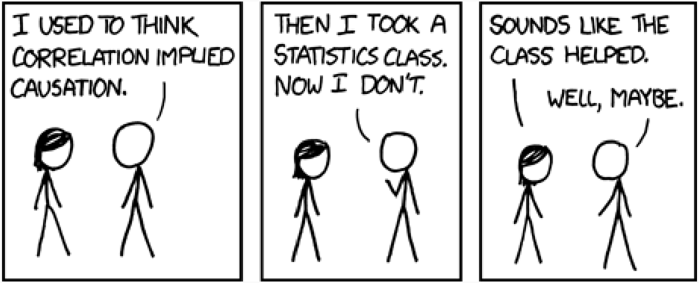
Correlation and causality

- Correlation does not say anything about causality!*
- Sometimes correlation can be explained by a third variable which was not measured.
- Days 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.

Causality

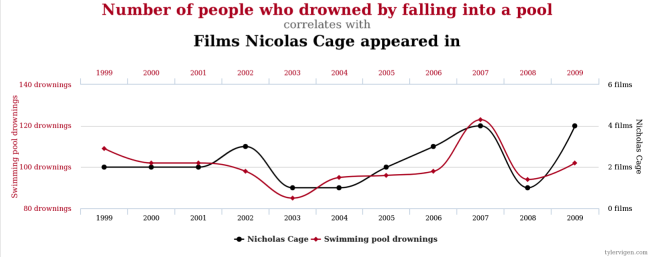
- If we want to know/predict what will change if we perform an action we need insight into causality.
- Will the number of drowning accidents change if we ban ice?
- There are many causal statements in the news!
    - "Do not skip breakfast if you want to reduce the risk of coronary heart disease"
-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

XKCD's take


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

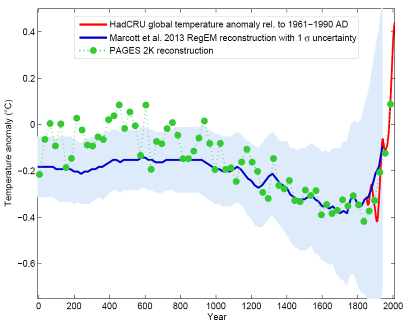
Thinking statistics

Spurious correlation



`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.