KU LEUVEN

Causality-preservation capabilities in data replication methods: an overview

Yves-Cédric Bauwelinckx with
J. Dhaene, T. Verdonck & M. van den Heuvel
KU Leuven

0 Outline

1 Outline

1 Synthetic data

What?

- **Fake, generated data** made to **resemble** the **original, real data**

1 Synthetic data

What?

- **Fake, generated data** made to **resemble** the **original, real data**

Why?

- Rise in data driven methods for modeling
- But: limited data available due to **privacy and ethics** concerns
- No private information in synthetic data $\Rightarrow$ **data can be shared**

1 Synthetic data

What?

- **Fake, generated data** made to **resemble** the **original, real data**

Why?

- Rise in data driven methods for modeling
- But: limited data available due to **privacy and ethics** concerns
- No private information in synthetic data $\Rightarrow$ **data can be shared**

How?

- Machine learning: **generative models**
- Generative model **learns underlying distribution** from real data
- Sample from learned distribution to create synthetic data
- State-of-the-art methodology: GAN (Goodfellow, 2014)

1 GAN

- Generative Adversarial Network
- Gained popularity by generating realistic pictures

1 GAN

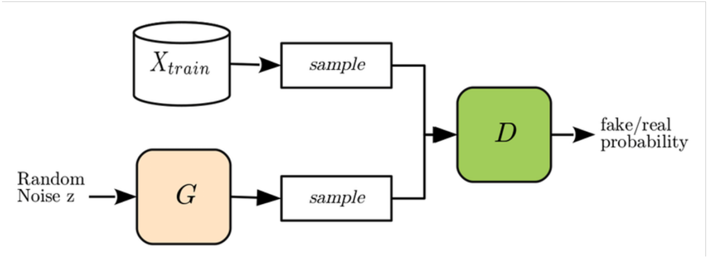
- Generative Adversarial Network
- Gained popularity by generating realistic pictures

1 GAN

- Generative Adversarial Network
- Gained popularity by generating realistic pictures

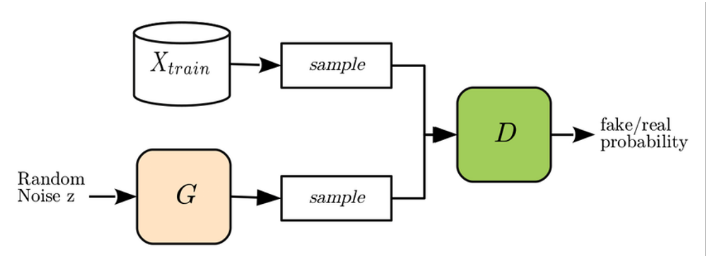
- Can also be used for tabular data (eg. insurance, finance)

1 GAN



- G(enerator) en D(iscriminator) $\Rightarrow$ 2 neural networks
- Generator and Discriminator compete against each other (adversarial)
- Goal: generator maps random noise to real data distribution

1 GAN

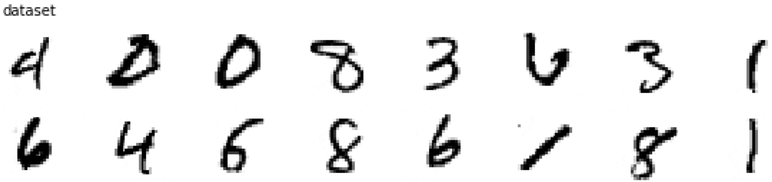


- **Generator generates** a random **sample** to "fool" Discriminator
- **Discriminator** tries to **distinguish real from generated samples**
- Discriminator gives feedback to Generator
- Generator performs better ⇒ Discriminator performs better ⇒ Generator performs better, etc.

1 GAN

- Understanding underlying distribution
- Generalizes concepts $\Rightarrow$ not copies from original dataset

Example for pictures of handwritten digits:



1 GAN

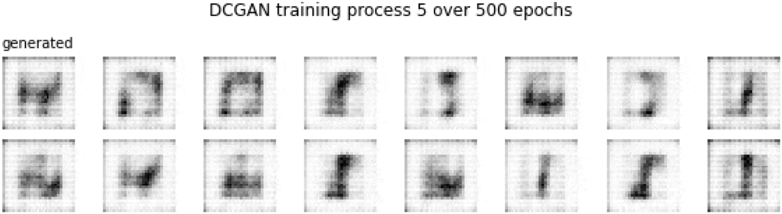
- Understanding underlying distribution
- Generalizes concepts $\Rightarrow$ not copies from original dataset

Example for pictures of handwritten digits:

1 GAN

- Understanding underlying distribution
- Generalizes concepts $\Rightarrow$ not copies from original dataset

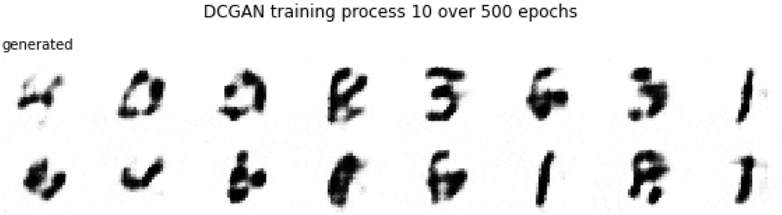
Example for pictures of handwritten digits:



1 GAN

- Understanding underlying distribution
- Generalizes concepts $\Rightarrow$ not copies from original dataset

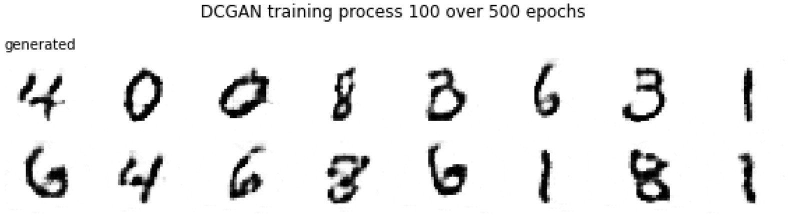
Example for pictures of handwritten digits:



1 GAN

- Understanding underlying distribution
- Generalizes concepts $\Rightarrow$ not copies from original dataset

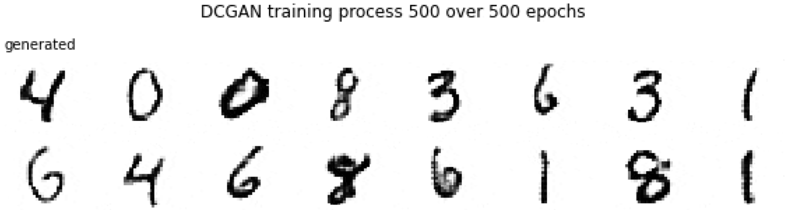
Example for pictures of handwritten digits:



1 GAN

- Understanding underlying distribution
- Generalizes concepts $\Rightarrow$ not copies from original dataset

Example for pictures of handwritten digits:

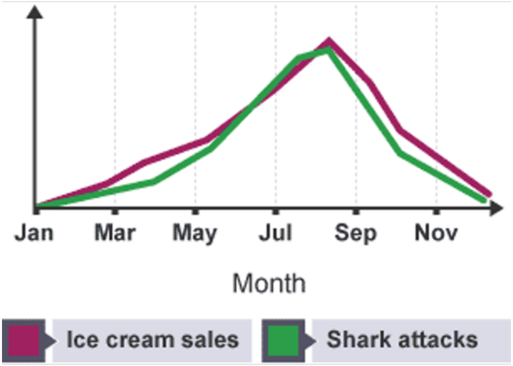


2 Outline

2 Causality

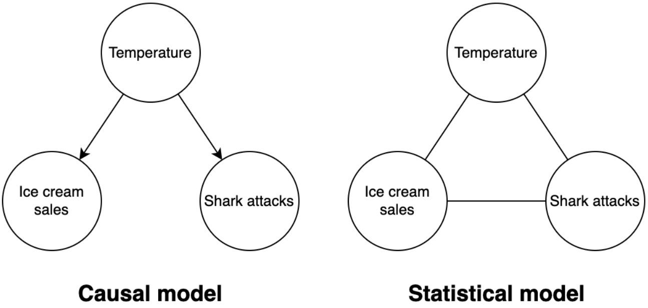
- Questions in business decisions: often causal
  - Observational: If I observe X, what will Y be?
  - Causal: If I do X, how will the outcome Y change?
- Rising interest for causality in insurance (eg. fairness, explainability)
  - Discrimination-free pricing (Lindholm et al., 2021; Araiza et al., 2022)
- GANs are good at replicating complex distributions
- Able to find correlations between variables
- **But: Correlation $\neq$ Causation**

2 Correlation $\neq$ Causation



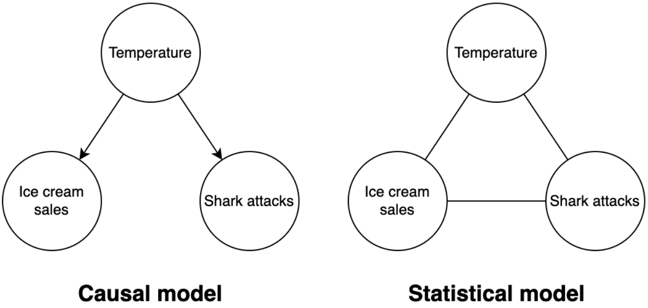
- High correlation between ice cream sales and shark attacks

2 Correlation $\neq$ Causation



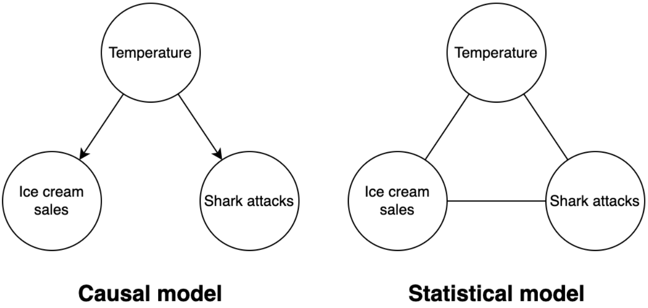
- Causal model allows to answer causal questions
  - How do we lessen the amount of shark attacks?

2 Correlation $\neq$ Causation



- Causal model allows to answer causal questions
  - How do we lessen the amount of shark attacks?
  - Causal: Lower temperature

2 Correlation $\neq$ Causation



- Causal model allows to answer causal questions
  - How do we lessen the amount of shark attacks?
  - Causal: Lower temperature
  - Statistical: stop selling ice cream?

2 Causality

GANs are capable of copying distributions, including correlations

But:
Correlation $\neq$ Causation

2 Causality

GANs are capable of copying distributions, including correlations

But:
Correlation $\neq$ Causation

Question:
How well are causal relations preserved in the synthetic data?

3 Outline

3 Experiment setup

- Create a dataset with known causal effects
- Train GAN with this dataset
- Sample synthetic data from GAN
- Perform analysis on both original and synthetic dataset
- Compare causal effects found from analysis
- Expectation:
  - Causal effects in original dataset $\approx$ original causal effects
  - Causal effects in synthetic dataset?
  - Difference between the two is due to GAN

3 Experiment setup

We create data with 3 different assumptions:

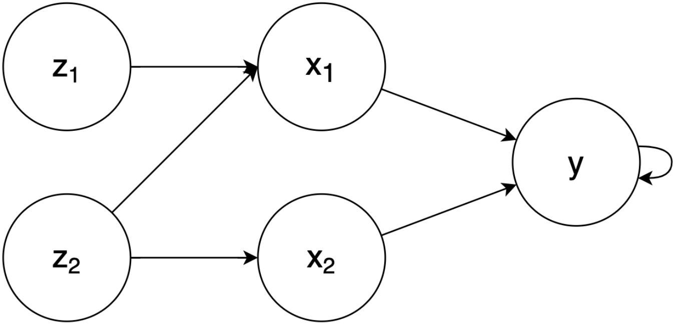
- Ordinary Least Squares
  - eg. $y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$
- Time Series (autoregressive)
  - eg. $y_t = \alpha y_{t-1} + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \epsilon_t$
- Full causal model
  - eg. $y = \beta_1 x_1 + \beta_2 x_2 + \epsilon_1$
  - $x_1 = \beta_3 z_1 + \beta_4 z_2 + \epsilon_2$
  - $x_2 = \beta_5 z_2 + \epsilon_3$

Data is made with certain α's and β's

Try to **recover them in the synthetic data**

3 Experiment setup

The dataset is made according to the following causal graph:



- $y_t = \alpha y_{t-1} + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \epsilon_{1,t}$
- $x_{1,t} = \beta_3 z_{1,t} + \beta_4 z_{2,t} + \epsilon_{2,t}$
- $x_{2,t} = \beta_5 z_{2,t} + \epsilon_{3,t}$

3 Experiment setup

3 variations of GAN:

- GAN
  - Train data: Table with data points
- TimeGAN
  - State-of-the-art GAN for time series
  - Train data: Table with data points in sequence
- CausalGAN
  - Causal graph is used for generator construction
  - Data generation follows causal graph ordering
  - Train data: Table with data points + causal graph

4 Outline

4 Results - Cross-sectional

	GAN	TimeGAN	CausalGAN
Ordinary Least Squares	**Good**	Good	Good
Time Series	-	-	-
Causal structure	-	-	-

- All GAN methods perform well at a cross-sectional level (OLS)
- Predictor variables are the causes
- Assumptions of OLS imply a single causal structure (predictor variables $\Rightarrow$ response variable)

4 Results - Time Series

	GAN	TimeGAN	CausalGAN
Ordinary Least Squares	Good	Good	Good
Time Series	/	**Shortcut**	/
Causal structure	-	-	-

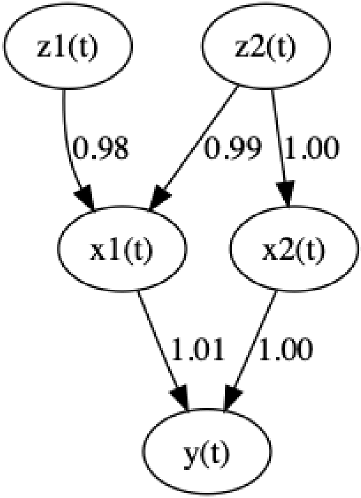
- GAN and CausalGAN are not able to produce time series
- TimeGAN fails at finding autocorrelation
- Real causal effects: $y_t = 0.5y_{t-1} + x_{1,t} + x_{2,t}$
- Found causal effects: $y_t = 2x_{1,t} + 2x_{2,t}$
- Found equation is approximation of original equation
- **TimeGAN** found a shortcut and did not keep causal relation

4 Results - Causal structure

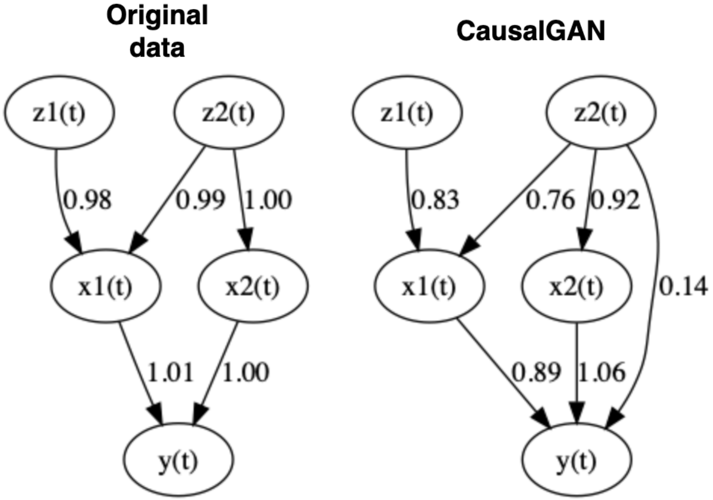
	GAN	TimeGAN	CausalGAN
Ordinary Least Squares	Good	Good	Good
Time Series	/	Shortcut	/
Causal structure	Bad	/	**OK**

- Attempt to reconstruct the original causal graph from data $\Rightarrow$ causal discovery
- Causal structure is lost in synthetic data from GAN
- Causal structure is mostly preserved by CausalGAN

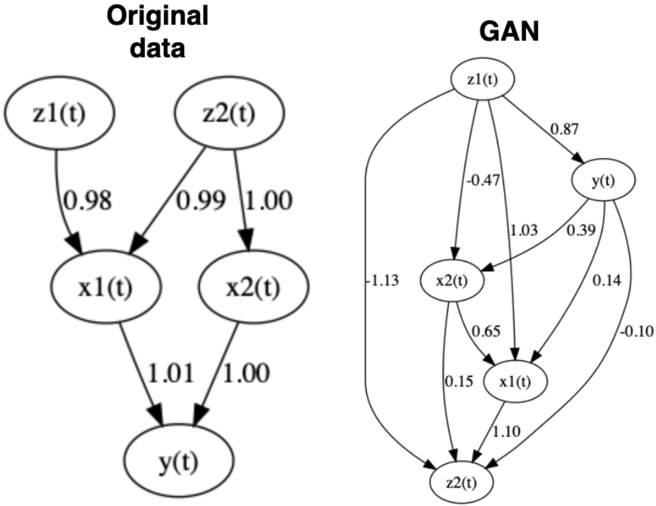
4 Causal graph - original data



4 Causal graph - CausalGAN data



4 Causal graph - GAN data



5 Outline

5 Conclusion

- Need for data for more models, but privacy concerns
- Synthetic data as a solution
- Only when assumptions are met that correlation does imply causation, causation is kept
- Generative models might simplify causal structures (shortcut)
- You should be careful about the capabilities of synthetic data asking causal questions

Thank you for your attention

5 Results - Cross-sectional

Model	Par.	**Real**	**GAN**	**TimeGAN**	**CausalGAN**
OLS	β_3	1.00 $\pm$ 0.005	1.02 $\pm$ 0.071	0.37 $\pm$ 0.432	0.98 $\pm$ 0.108
	β_4	1.00 $\pm$ 0.005	1.07 $\pm$ 0.127	1.22 $\pm$ 0.336	0.96 $\pm$ 0.102
	β_5	1.00 $\pm$ 0.005	1.01 $\pm$ 0.126	1.10 $\pm$ 0.017	1.00 $\pm$ 0.162

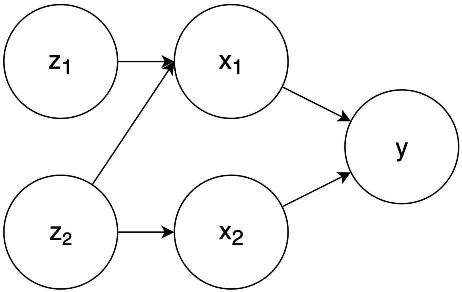
Model	Par.	**Real**	**GAN**	**TimeGAN**	**CausalGAN**
TS	α	0.50 $\pm$ 0.001	0.01 $\pm$ 0.002	0.02 $\pm$ 0.133	0.01 $\pm$ 0.006
	β_1	1.00 $\pm$ 0.004	1.00 $\pm$ 0.177	1.05 $\pm$ 1.523	0.96 $\pm$ 0.189
	β_2	1.00 $\pm$ 0.004	1.14 $\pm$ 0.168	0.87 $\pm$ 2.043	0.99 $\pm$ 0.203

5 Results - Time Series

- New experiment: variables are time series (have autocorr.)

Model	Parameter	**Real**	**TimeGAN**
OLS	β_3	1.00 $\pm$ 0.001	0.99 $\pm$ 0.024
	β_4	1.00 $\pm$ 0.001	1.00 $\pm$ 0.022
	β_5	1.00 $\pm$ 0.001	1.00 $\pm$ 0.002
TS	α	0.50 $\pm$ 0.001	-0.01 $\pm$ 0.021
	β_1	1.00 $\pm$ 0.002	2.07 $\pm$ 0.068
	β_2	1.00 $\pm$ 0.002	2.00 $\pm$ 0.155

5 Results - Causal structure



5 Results - Causal structure

Causal effect	**Real**	**CausalGAN**	**GAN**
$z_1 \to x_1$	1.00	0.93	1.03
$z_2 \to x_1$	1.01	0.80	1.07
$z_2 \to x_2$	0.99	0.83	0.16
$x_1 \to y$	1.02	1.04	0.14
$x_2 \to y$	1.01	1.00	0.39
$z_1 \to z_2$	0	0	-1.11
$z_1 \to x_2$	0	0	-0.47
$z_1 \to y$	0	0	0.86
$z_2 \to y$	0	0.14	-0.10
$x_2 \to x_1$	0	0.14	0.65