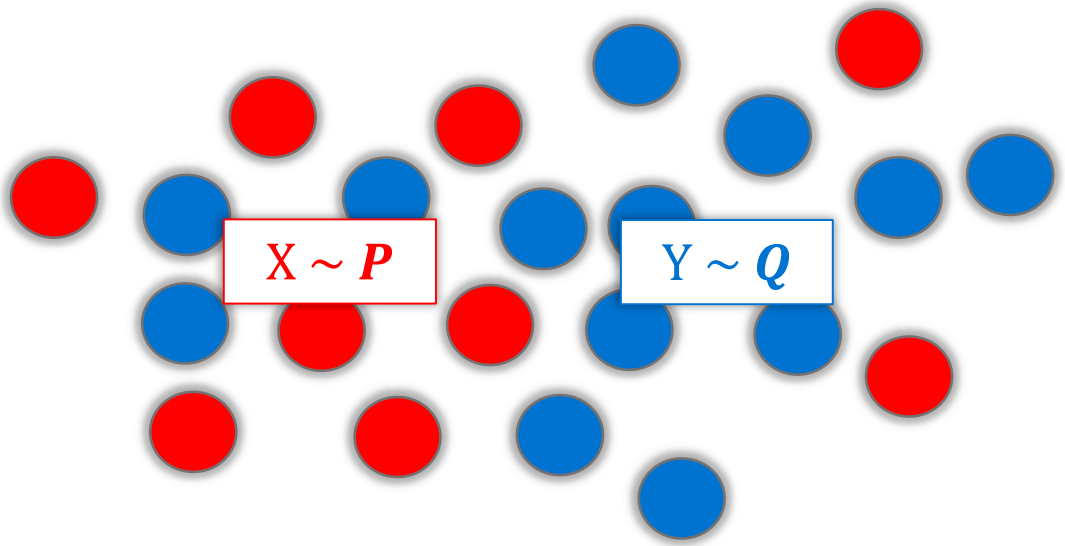
UNIVERSITY OF CAMBRIDGE

The Alan Turing Institute

Conditional Independence Testing using Adversarial Neural Networks

Alexis Bellot
Mihaela van der Schaar

From two-sample testing to independence testing



Two sample problem: Can we say whether $P = Q$?

Independence problem: Can we say whether $(X, Y) \perp (0, 1)$?

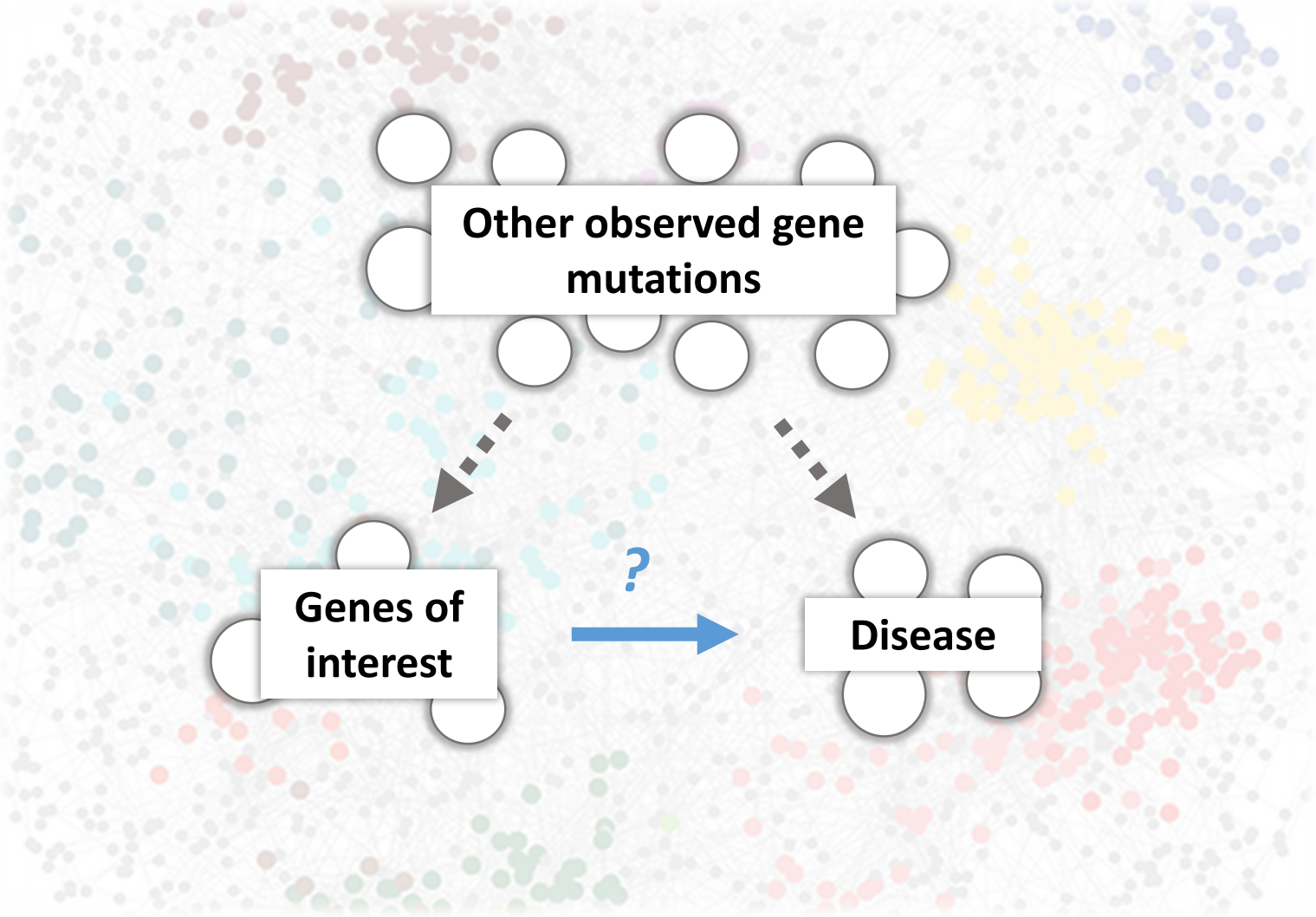
From independence to conditional independence

Two sample problem: Can we say whether $P = Q$?

Independence problem: Can we say whether $X \perp Y$?

Conditional Independence problem: Can we say whether $X \perp Y \mid Z$?

Can we find genes *directly* associated with disease?



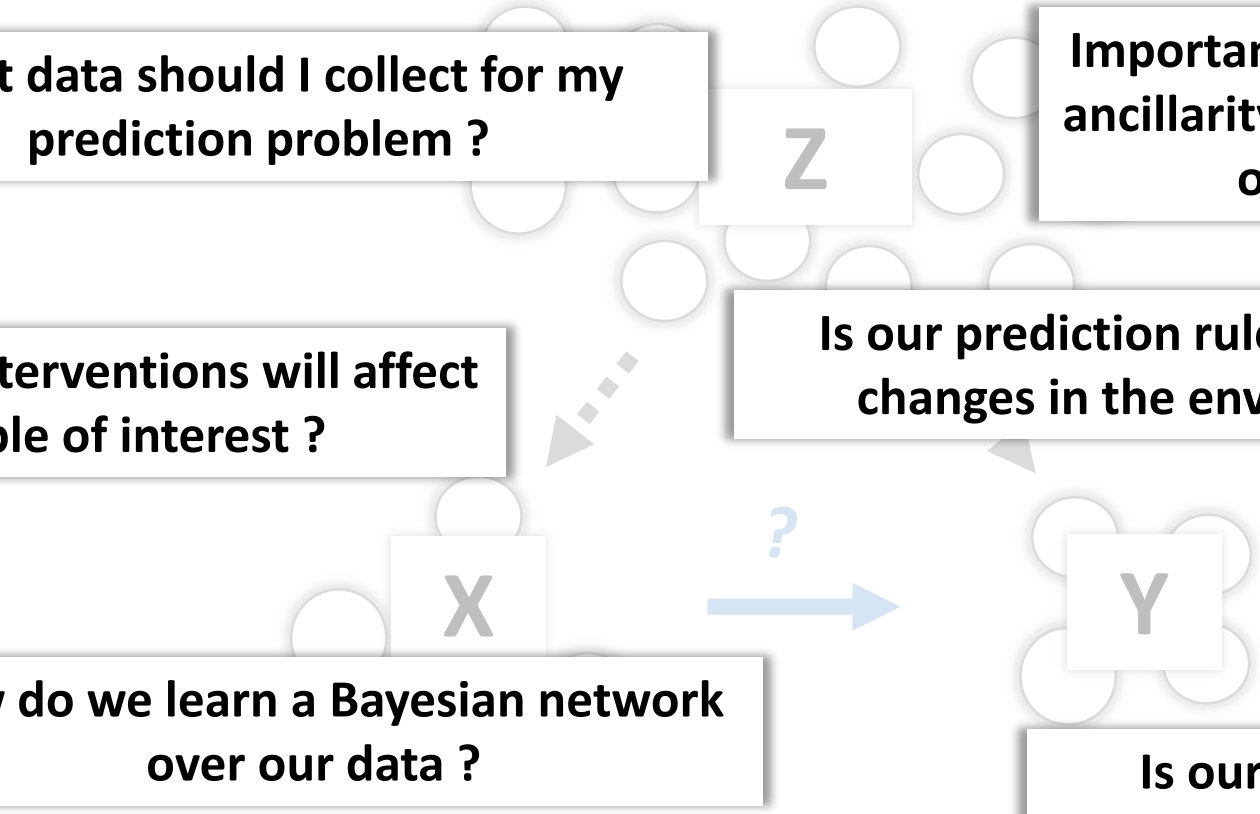
Many questions fit into this formalism

What data should I collect for my prediction problem ?

Important concepts of statistics (sufficiency, ancillarity …) can be regarded as expressions of conditional independence

Is it likely that interventions will affect the variable of interest ?

Is our prediction rule invariant to changes in the environment ?



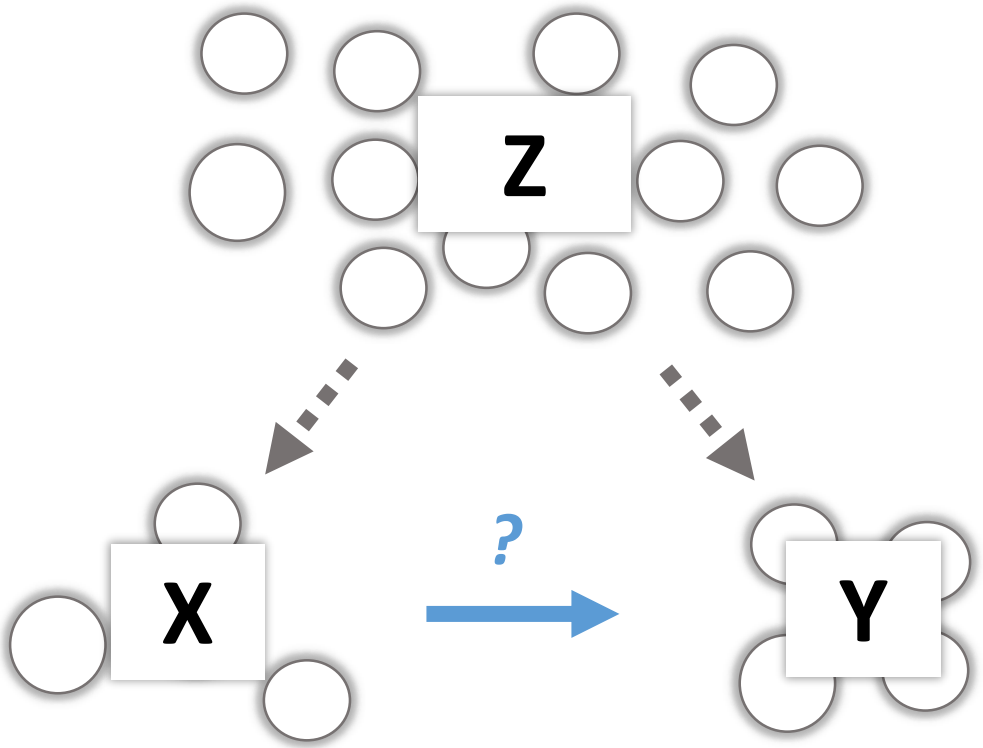
How do we learn a Bayesian network over our data ?

Is our prediction rule fair ? Are sensitive attributes influential ?

$$X \perp Y \mid Z \quad \text{iff} \quad p(Y \mid X, Z) = p(Y \mid Z)$$

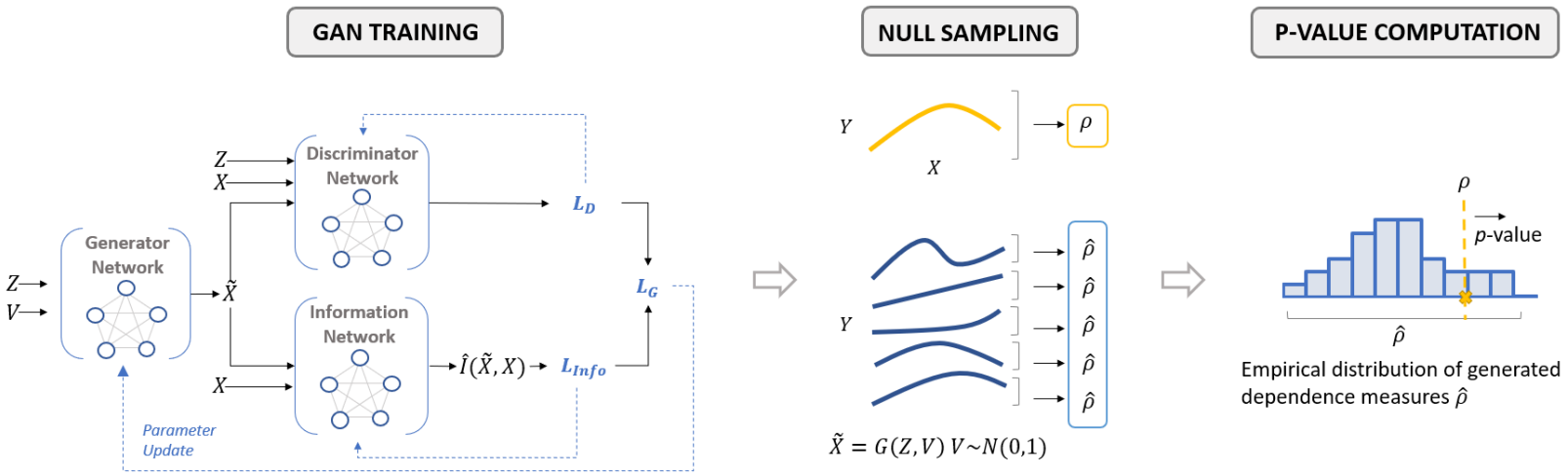
The intuition

- If we had access to $p(Y \mid Z)$
- Samples from this distribution breaks the direct dependency $X - Y$
- A comparison of the dependencies between *synthetic* and *observed* data should not reveal any differences under the null.



Why you should come see our poster

- We develop a modified GAN to sample from $\boldsymbol{p(Y \mid Z)}$ with high power
  - ➢ Better performance in high dimensions
- Provably valid testing
  - ➢ No assumptions on data distribution.
  - ➢ Non-asymptotic error bounds.



Join us for more details shortly

East Exhibition Hall B + C #75

The Alan Turing Institute