Causal Probabilistic Programing Without Tears

Eli Bingham[1], James Koppel[2], Alexander Lew[2], Robert Osazuwa Ness[3], Zenna Taveres[2], Sam Witty[4], Jeremy Zucker[5]

1. Broad Institute of MIT and Harvard, USA; 2. MIT, USA; 3. Microsoft Research, USA; 4. UMass Amherst, USA; 5. Pacific Northwest National Laboratory, USA

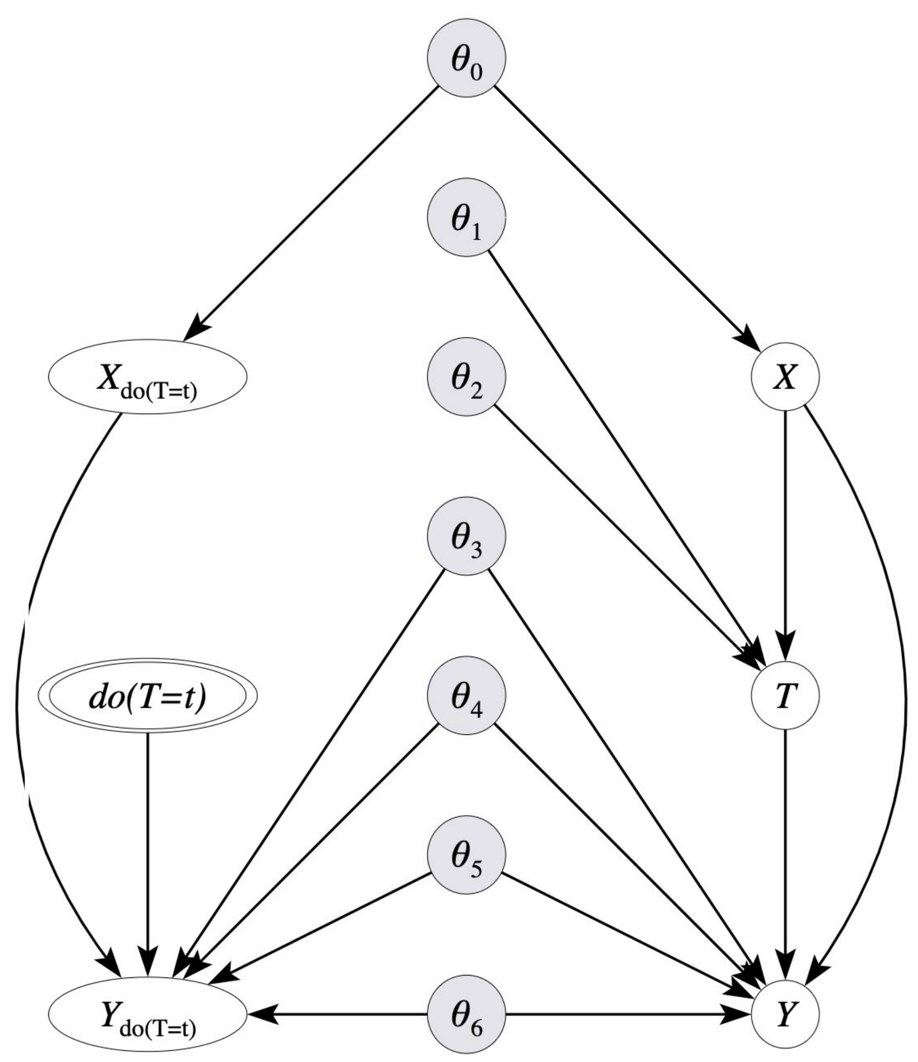
Abstract

We take an informal, example-driven tour of some of the underlying questions that causal inference researchers and practitioners seek to answer, and the causal assumptions that make it possible to derive these answers from data. Taken together with other recent literature, our examples point toward an emerging research agenda under which many conceptual and practical difficulties in applying causal inference methods could be alleviated by framing causal questions as source code transformations and standard probabilistic computations on causal models instantiated as programs in a modern generative probabilistic programming language.

Backdoor criterion:

A causal model over the observed covariates ***X*****,** the treatment ***T***, and the outcome ***Y*** [Pearl 2009]

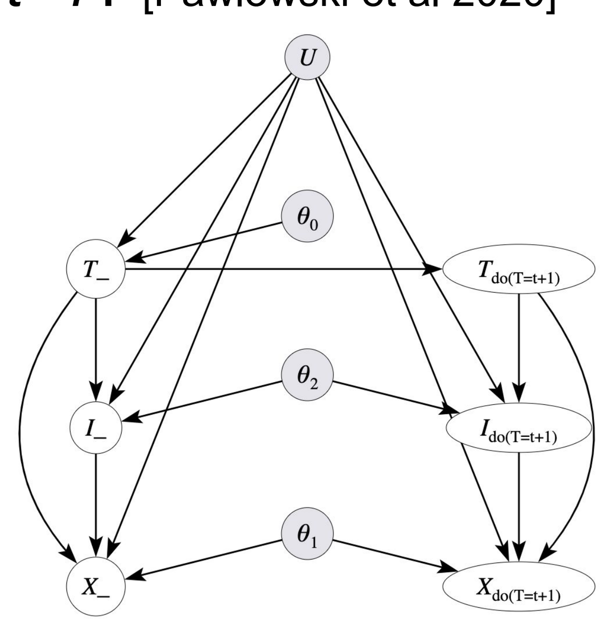
```
def causal_model(theta):
  X ~ bernoulli(theta[0])
  T ~ bernoulli(theta[X+1])
  Y ~ bernoulli(theta[T+2*X+3])
  return Y, T, X

def intervened_causal_model(theta, t):
  X ~ bernoulli(theta[0])
  T = t
  Y ~ bernoulli(theta[T+2*X+3])
  return Y

def joint_model():
  theta ~ ThetaPrior()
  for i in range(N):   # observed variables
    Y[i], T[i], X[i] ~ causal_model(theta)
  Y_treated ~ intervened_causal_model(theta,t=1)
  Y_untreated ~ intervened_causal_model(theta,t=0)
  return Y_treated - Y_untreated

ATE = Expectation(joint_model | [Y_obs, T_obs, X_obs])
```



Deep Structural Causal Models:

The following code models morphological transformations of MNIST, defining a causal generative model over digits *u*, containing endogenous variables to control the thickness t and intensity *i* of the image *x*. Given an observed image *x*, what would the image look like had *t* been *t+1*? [Pawlowski et al 2020]

```
def deep_structural_causal_model(theta):
  u ~ noise()

  t_ = f(u; theta[0])
  t = t_ + 1

  i_ = g(u; t_, theta[1])
  i = g(u; t, theta[1])

  x_ = h(u; i_, t_, theta[2])  # observed
  x = h(u; i, t, theta[2])
  return x
```

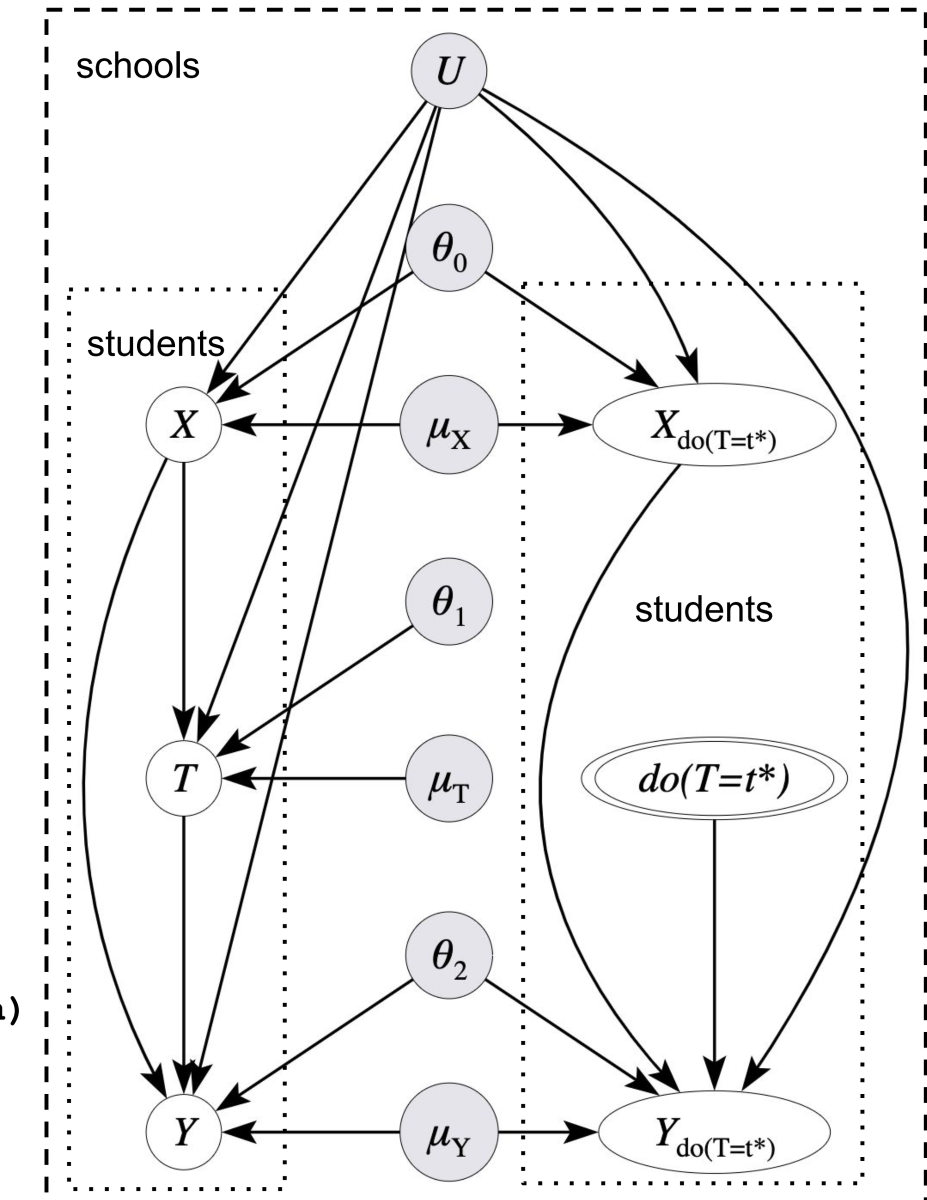


Individual Treatment Effects with Structured Latent Confounders:

Calculate the difference in educational outcome ***Y*** of a particular student *i* at school *o* with a particular intervention *t**:

$ITE^{(o,i)} = f_y(U^{(o)}, X^{(o,i)}, do(T^{(o,i)}=t^*)) - f_y(U^{(o)}, X^{(o,i)}, T^{(o,i)})$ [Witty et al 2020]

```
def instance_causal_model(f_x, f_t, f_y, U, theta):
  mu_X = f_x(U)
  X ~ normal(mu_X, theta[0])  # Generate observed covariates X, based on u

  mu_T =  f_t(U, X)
  T ~ normal(mu_T, theta[1])  # Generate observed treatment, based on u and x

  mu_Y = f_y(U, X, T)
  Y ~ normal(mu_Y, theta[2])  # Generate outcome as a function of u, x, and t
  return X, T, Y

def joint_model(n_schools, n_students, doT, theta):
  # Generate causal functions from a Gausssian process
  f_x ~ GP(m_x, k_x)
  f_t ~ GP(m_t, k_t)
  f_y ~ GP(m_y, k_y)

  for o in range(n_schools):
    U[o] ~ normal(0, I)  # Generate a school-level latent confounder
    for i in range(n_students):
      X[o,i], T[o,i], Y[o,i] ~ instance_causal_model(f_x, f_t, f_y, U[o],theta)
      ITE[o,i] = f_y(U[o], X[o,i], doT) - f_y(U[o], X[o,i], T[o,i])

  return ITE  # return array of all instance ITE values
```



A Research Program for Causal Probabilistic Programming

1. What program transformations and analyses might be necessary to cover a much larger fraction of the causal inference literature?
2. Can these transformations be formalized with efficient, model-agnostic implementations?
3. Can they be distilled into a core calculus of a small number of composable primitives?

References and Acknowledgements

1. Judea Pearl, 2009. "*Causality: Models, Reasoning and Inference (2nd Ed)*". Cambridge University Press, USA
2. Nick Pawlowski, Daniel C Castro, and Ben Glocker. 2020. Deep structural causal models for tractable counterfactual inference.*arXiv preprint arXiv:2006.06485*(2020)
3. Sam Witty, Kenta Takatsu, David Jensen, and Vikash Mansinghka. 2020. Causal inference using Gaussian processes with structured latent confounders. In *International Conference on Machine Learning*. PMLR, 10313–10323.

Jeremy Zucker's work was supported by the Data-Model Convergence Initiative, a component of the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830.