Berkeley School of Public Health

DAGs and probability

PHW250 B – Andrew Mertens

Link between DAGs and probability

- The **joint distribution** of the variables in the model is given by the product of the **conditional distributions** over all the "families" in the graph.

$$\underbrace{P(x_1, x_2, \ldots, x_n)}_{\text{Joint distribution}} = \prod_i \underbrace{P(x_i | pa_i)}_{\text{Conditional}}$$

- **Example:** the joint distribution for the DAG $\underline{X \rightarrow Y \rightarrow Z}$ is:

$$\underline{P(X = x, Y = y, Z = z)} = \underline{P(X = x)}\,\underline{P(Y = y | X = x)}\,\underline{P(Z = z | Y = y)}$$

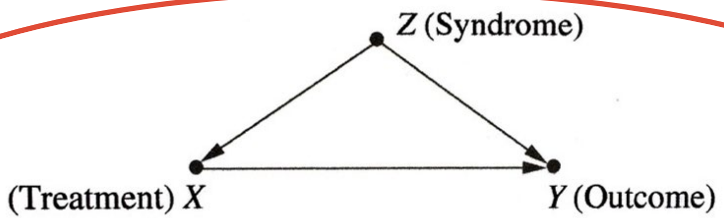
Pearl et al., *Causal Inference in Statistics: A Primer.* Wiley 2016.

Example: DAG

- A population of patients contains a fraction *r* of individuals with fatal syndrome *Z*
  - $Z=z_1$ represents patients with the syndrome
  - $Z=z_0$ represents patients without the syndrome
- The syndrome makes it difficult for them to take treatment *X*
  - $X=x_1$ represents patients who take the drug
  - $X=x_0$ represents patients who do not take the drug
- Their survival status is indicated by *Y*
  - $Y=y_1$ represents patients who die
  - $Y=y_0$ represents patients who survive

Z= random variable

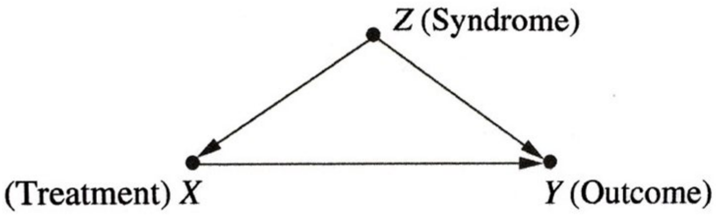
$Z_{1,0}$ = specific values



Pearl et al., *Causal Inference in Statistics: A Primer.* Wiley 2016.

Example: structural causal model

- $Z = f_Z(U_Z)$ No parents
- $X = f_X(Z, U_X)$
- $Y = f_Y(X, Z, U_Y)$



Pearl et al., *Causal Inference in Statistics: A Primer*. Wiley 2016.

Example: probabilities

$p \quad q \quad r$

Assume the following:

- Patients without the syndrome who do not take the drug die with probability p_1
  - $P(Y \mid Z = z_0, X = x_0) = p_1$
- Patients without the syndrome who take the drug die with probability p_2
  - $P(Y \mid Z = z_0, X = x_1) = p_2$
- Patients with the syndrome who do not take the drug die with probability p_3
  - $P(Y \mid Z = z_1, X = x_0) = p_3$
- Patients with the syndrome who take the drug die with probability p_4
  - $P(Y \mid Z = z_1, X = x_1) = p_4$
- Patients with the syndrome are more likely to avoid the drug with the following probabilities:
  - $P(X = x_1 \mid Z = z_0) = q_1$
  - $P(X = x_1 \mid Z = z_1) = q_2$
- The probability patients have the syndrome:
  - $P(Z = z_1) = r$

Pearl et al., *Causal Inference in Statistics: A Primer*. Wiley 2016.

Example: 2x2 table

- $P(Y \mid Z = z_0, X = x_0) = p_1 \quad = c_0 / (c_0 + d_0)$
- $P(Y \mid Z = z_0, X = x_1) = p_2 \quad = a_0 / (a_0 + b_0)$
- $P(Y \mid Z = z_1, X = x_0) = p_3 \quad = c_1 / (c_1 + d_1)$
- $P(Y \mid Z = z_1, X = x_1) = p_4 \quad = a_1 / (a_1 + b_1)$
- $P(X = x_1 \mid Z = z_0) = q_1 \quad = (a_0 + b_0) / N_0$
- $P(X = x_1 \mid Z = z_1) = q_2 \quad = (a_1 + b_1) / N_1$
- $P(Z = z_1) = r \quad = (a_1 + b_1 + c_1 + d_1) / N$

with

Z=z₁		
	$Y=y_1$	$Y=y_0$
$X=x_1$	a_1	b_1
$X=x_0$	c_1	d_1

without

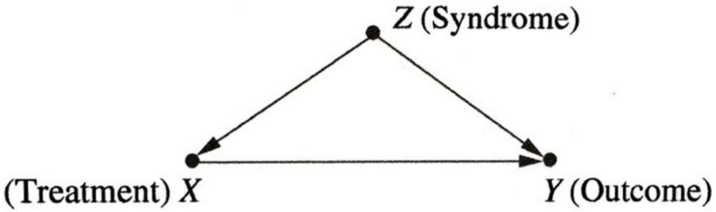
Z=z₀		
	$Y=y_1$	$Y=y_0$
$X=x_1$	a_0	b_0
$X=x_0$	c_0	d_0

Pearl et al., *Causal Inference in Statistics: A Primer*. Wiley 2016.

Example: computing joint probabilities

What is $P(X = x_1, Y = y_1, Z = z_1)$ (the probability that $X = x_1$, $Y = y_1$, and $Z = z_1$)?

- $P(Y \mid Z = z_0, X = x_0) = p_1$
- $P(Y \mid Z = z_0, X = x_1) = p_2$
- $P(Y \mid Z = z_1, X = x_0) = p_3$
- $P(Y \mid Z = z_1, X = x_1) = p_4$
- $P(X = x_1 \mid Z = z_0) = q_1$
- $P(X = x_1 \mid Z = z_1) = q_2$
- $P(Z = z_1) = r$



Pearl et al., *Causal Inference in Statistics: A Primer*. Wiley 2016.

Example: computing joint probabilities

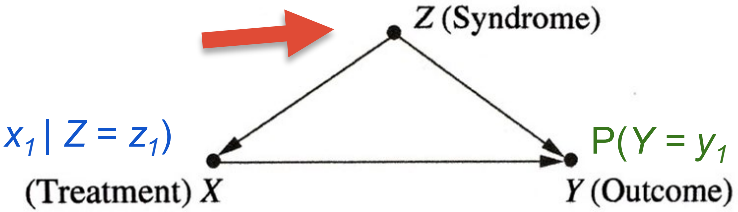
What is P($X = x_1$, $Y = y_1$, $Z = z_1$) (the probability that $X = x_1$, $Y = y_1$, and $Z = z_1$)?

$= P(Z = z_1)\ P(X = x_1 \mid Z = z_1)\ P(Y = y_1 \mid X = x_1, Z = z_1)$ (chain rule)

$= r * q_2 * p_4$

- $P(Y \mid Z = z_0, X = x_0) = p_1$
- $P(Y \mid Z = z_0, X = x_1) = p_2$
- $P(Y \mid Z = z_1, X = x_0) = p_3$
- $P(Y \mid Z = z_1, X = x_1) = p_4$
- $P(X = x_1 \mid Z = z_0) = q_1$
- $P(X = x_1 \mid Z = z_1) = q_2$
- $P(Z = z_1) = r$

$P(Z = z_1)$



$P(X = x_1 \mid Z = z_1)$

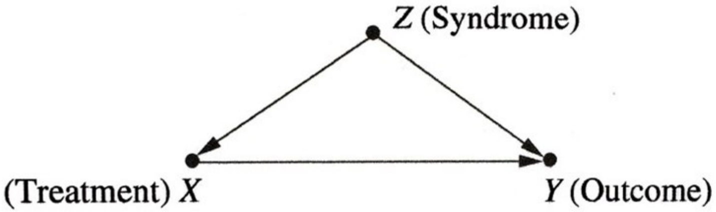
$P(Y = y_1 \mid X = x_1, Z = z_1)$

Pearl et al., *Causal Inference in Statistics: A Primer*. Wiley 2016.

Example: computing conditional probabilities

Calculate the difference in the probability of death among those with the syndrome comparing those who did and did not take the drug.

- $P(Y \mid Z = z_0, X = x_0) = p_1$
- $P(Y \mid Z = z_0, X = x_1) = p_2$
- $P(Y \mid Z = z_1, X = x_0) = p_3$
- $P(Y \mid Z = z_1, X = x_1) = p_4$
- $P(X = x_1 \mid Z = z_0) = q_1$
- $P(X = x_1 \mid Z = z_1) = q_2$
- $P(Z = z_1) = r$



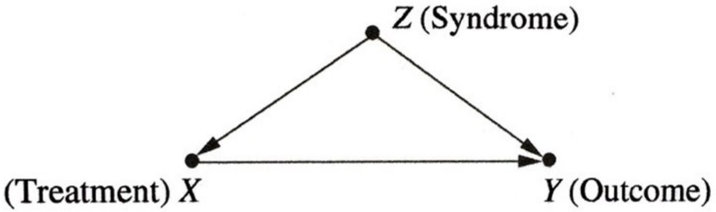
Pearl et al., *Causal Inference in Statistics: A Primer.* Wiley 2016.

Example: computing conditional probabilities

Calculate the difference in the probability of death among those with the syndrome comparing those who did and did not take the drug.

$P(Y = y_1 \mid X = x_1 , Z = z_1) - P(Y = y_1 \mid X = x_0, Z = z_1)$

- $P(Y \mid Z = z_0, X = x_0) = p_1$
- $P(Y \mid Z = z_0, X = x_1) = p_2$
- $P(Y \mid Z = z_1, X = x_0) = p_3$
- $P(Y \mid Z = z_1, X = x_1) = p_4$
- $P(X = x_1 \mid Z = z_0) = q_1$
- $P(X = x_1 \mid Z = z_1) = q_2$
- $P(Z = z_1) = r$



$p4 - p3$

Pearl et al., *Causal Inference in Statistics: A Primer.* Wiley 2016.

Summary of key points

- Structural causal models and DAGs directly translate into probabilities between variables in our data.
- Thus, DAGs and structural causal models encode our assumptions about the probability relationships (including dependence and independence of variables) in our data.