Title: Conditional Independence - Revisited

Speakers: Patrick Forre

Collection: Causal Inference & Quantum Foundations Workshop

Date: April 19, 2023 - 10:00 AM

URL: https://pirsa.org/23040116

Abstract: "Many relationships in causality, statistics or probability theory can be expressed as conditional independence relations between the occurring random variables. Since the invention of the notion of conditional independence one aim was to be able to also express such relationship between random and non-random variables, like the parameters of a stochastic model, the input variables of a probabilistic program or intervention variables in a causal model. Over time several different versions of such extended conditional independence notion have been proposed, each coming with their own advantages and disadvantages, oftentimes limited to certain subclasses of random variables like discrete variables or ones with densities.

In this talk we present another such notion of conditional independence, which can easily be expressed in measure-theoretic generality and even in categorical probability. We will study its expressivity, present its (convenient) properties, and relate it to other notions of conditional independence."

Conditional Independence - Revisited

Patrick Forré

AI4Science Lab, AMLab, Informatics Institute, University of Amsterdam, The Netherlands

Goal

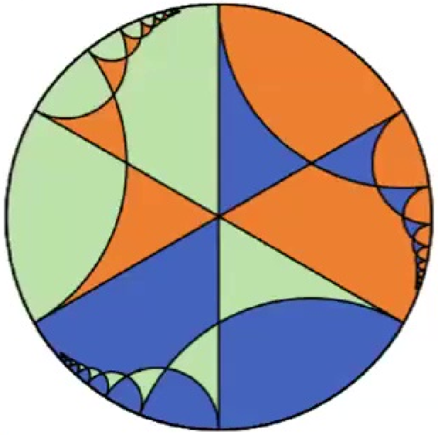
- Formalize a generalized notion of **Conditional Independence:**

$X \perp\!\!\!\perp Y \mid Z$ w.r.t. $P(X, Y, Z \mid T)$, that works with/for/in:

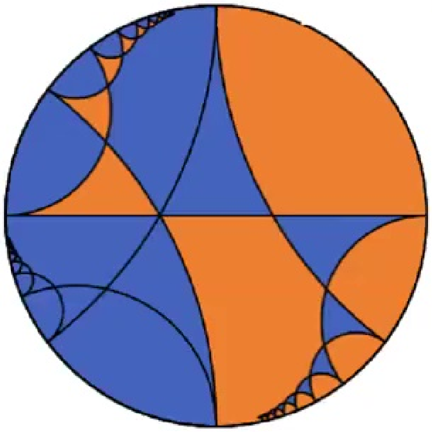
  - measure-theoretic generality (e.g. mix of discrete + continuous distributions),
  - transition probabilities / Markov kernels (i.e. in relative setting),
  - arbitrary combination of random (output) and non-random (input) variables,
  - causality, probabilistic graphical models (Global Markov Property),
  - statistical theory (express statistical concepts as CI relations),
  - meaningful (ir-)relevance rules, etc.

Why Measure Theory to do Probability Theory?

Why Measure Theory in the first place?



- The existence of the Lebesgue measure:
  - does not exist on whole power set $2^{\mathbb{R}}$.
  - but does exist on Borel σ-algebra $\mathscr{B}_{\mathbb{R}}$.



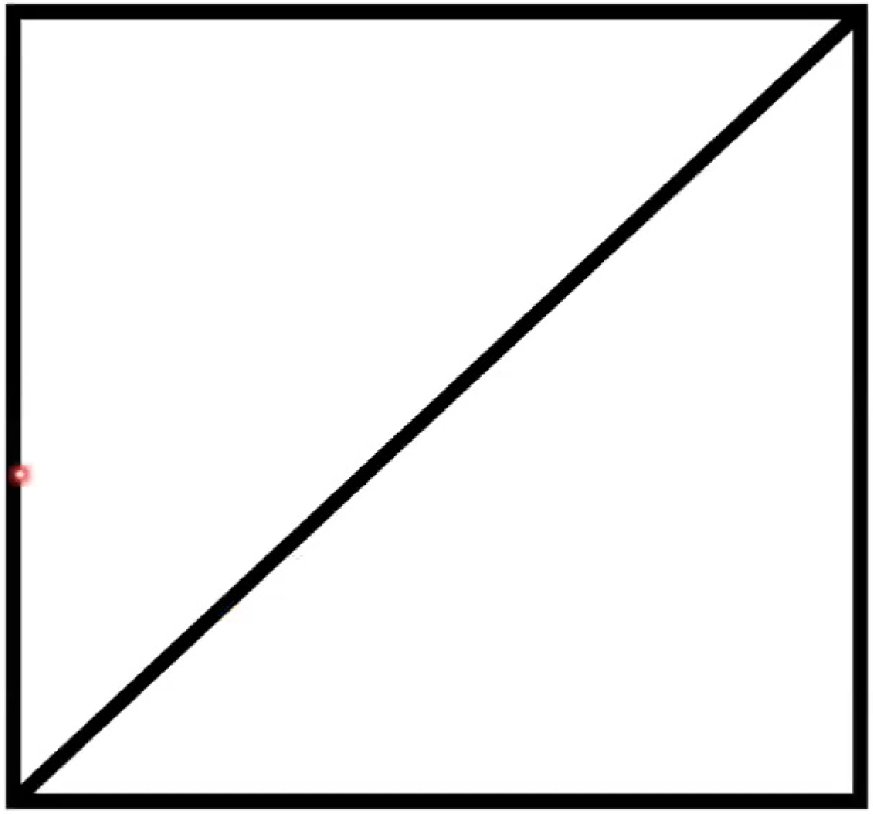
- To prevent set-theoretic paradoxa like Banach-Tarski:
  - the orange is both a third and a half of the Poincaré disk / hyperbolic plane.

- Stan Wagon. The Banach Tarski Paradox. CUP 1985. https://demonstrations.wolfram.com/TheBanachTarskiParadox/

Discrete and continuous distributions are not expressive enough

- The uniform distribution on the diagonal $\Delta \subseteq [0,1]^2$
  - is neither discrete nor absolute continuous w.r.t. λ^2.
  - so it can not be described with a probability mass function nor with a probability density w.r.t. λ^2.

- Also the Dirac delta peak δ_x has no density.



Conclusion:

Measure Theory is “safe” and expressive enough to do Probability Theory!

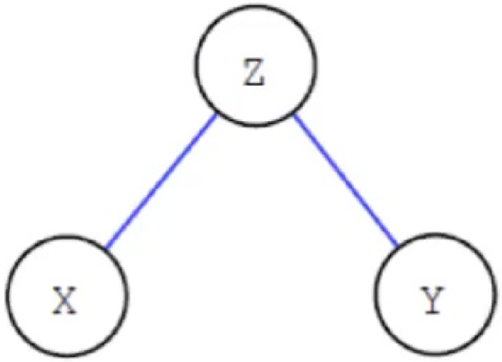
Independent Random Variables

- Let $(\mathscr{W}, P(W))$ be a probability space.
- Random variables: $X : \mathscr{W} \to \mathscr{X}$, $Y : \mathscr{W} \to \mathscr{Y}$, are **independent** if one of the following holds:
  - $P(X, Y) = P(X) \otimes P(Y)$,
  - $\forall A \in \mathscr{B}_{\mathscr{X}} \, . \, \mathbb{E}[\mathbb{1}_A(X) \,|\, Y] = P(X \in A)$ a.s.
  - $\forall B \in \mathscr{B}_{\mathscr{Y}} \, . \, \mathbb{E}[\mathbb{1}_B(Y) \,|\, X] = P(Y \in B)$ a.s.
- heuristically, if the outcome of Y does not provide information about the outcome of X.

8

Conditionally Independent Random Variables

- Random variable X is **independent** of Y **conditioned** on Z if one of the following holds:
  - $P(X, Y, Z) = P(X|Z) \otimes P(Y|Z) \otimes P(Z)$,
  - $P(X, Y|Z) = P(X|Z) \otimes P(Y|Z) \quad P(Z)$-a.s.
  - $\forall A \in \mathscr{B}_{\mathcal{X}} \,.\, \mathbb{E}[\mathbb{1}_A(X)\,|\,Y, Z] = \mathbb{E}[\mathbb{1}_A(X)\,|\,Z]$ a.s.
  - $\forall B \in \mathscr{B}_{\mathcal{Y}} \,.\, \mathbb{E}[\mathbb{1}_B(Y)\,|\,X, Z] = \mathbb{E}[\mathbb{1}_B(Y)\,|\,Z]$ a.s.



- I.e. if Y does not contain any additional information beyond the information already encoded in Z to predict the state of X.

Separoid Rules

- Redundancy: $X \precsim_P Z \implies X \perp\!\!\!\perp Y | Z,$
- Symmetry: $X \perp\!\!\!\perp Y | Z \implies Y \perp\!\!\!\perp X | Z,$
- Decomposition: $X \perp\!\!\!\perp Y, U | Z \implies X \perp\!\!\!\perp U | Z,$
- Weak Union: $X \perp\!\!\!\perp Y, U | Z \implies X \perp\!\!\!\perp Y | U, Z,$
- Contraction: $X \perp\!\!\!\perp U | Z \wedge X \perp\!\!\!\perp Y | U, Z$

$$\implies X \perp\!\!\!\perp Y, U | Z.$$

A.P. Dawid. *Separoids: A Mathematical Framework for Conditional Independence and Irrelevance*. Annals of Mathematics and Artificial Intelligence 32 (2001): 335-372.

Conditional Independence in Statistical Theory

If random variables X and Y are independent, given Z, we may write $X \perp\!\!\!\perp Y | Z$. The most fruitful intuitive interpretation of this statement is that the conditional distributions of X, given Y and Z, are in fact governed by the value of Z alone, further information about the value of Y being irrelevant. This intuitive property extends readily to statements such as $X \perp\!\!\!\perp \Theta | T$, in which X is a random variable with distributions governed by a parameter Θ, and T is (say) a function of X: a moment's reflection will show that this is just the requirement that T is *sufficient* for Θ based on data X. Similarly, $T \perp\!\!\!\perp \Theta$ (with the conditioning variable trivial) if and only if T is an *ancillary* statistic.

- However, the provided theory is of pure probabilistic nature.
- Non-random variables were implicitly turned into random variables.

- A.P. Dawid. *Conditional Independence in Statistical Theory*. Journal of the Royal Statistical Society: Series B (Methodological) 41.1 (1979): 1-15.
- A.P. Dawid. *Conditional Independence for Statistical Operations*. The Annals of Statistics 8.3 (1980): 598-617.

Motivation - Ancillary Statistics

- Consider statistical model $P(X|\Theta)$ and statistic $R: \mathscr{X} \to \mathscr{R}$.
- Definition: R is called an **ancillary statistic** of X if the distribution of R, i.e. $P(R|\Theta=\theta)$, does not depend on θ.
- We want to re-phrase this as a **conditional independence** relations:

$$R \perp\!\!\!\perp \Theta \quad [P(X|\Theta)],$$

Observations - Problems

- We want to **ancillarity** to be equivalent to:

$$R \perp\!\!\!\perp \Theta \quad [P(X \mid \Theta)].$$

- Note that the swapped relation: $\Theta \perp\!\!\!\perp R$ then would read:
  - The distribution of Θ, i.e. $P(\Theta \mid R) = ???$, is not dependent on R.
  - However, since there is no distribution $P(\Theta)$ provided, there is also no (conditional?) distribution $P(\Theta \mid R)$ to impose conditions on.
- $\Longrightarrow$ Such an extension of the notion of **conditional independence**, including **non-random variables** Θ, in such generality, needs to be assumed to be **asymmetric**!!!

Motivation - Sufficient Statistics

- Consider statistical model $P(X|\Theta)$ and statistic $S: \mathcal{X} \to \mathcal{S}$.
- Definition: S is called a **sufficient statistic** for $P(X|\Theta)$ if the conditional distribution $P(X|S=s, \Theta=\theta)$ does not depend on θ.
- We want to express this as a **conditional independence** relation:

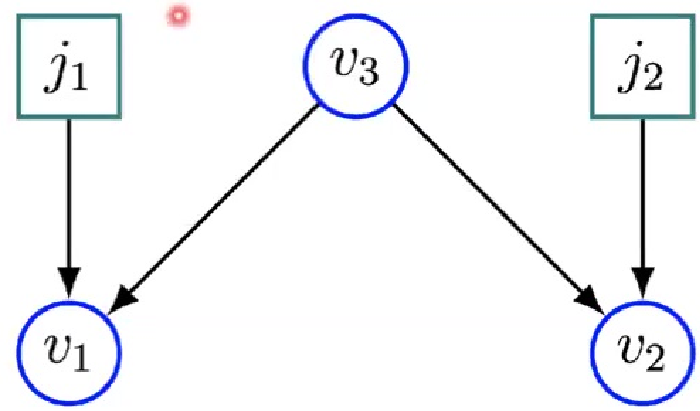
$$X \perp\!\!\!\perp \Theta \,|\, S \qquad [P(X|\Theta)].$$

- Again, note that Θ is a **non-random variable**, the relation between X and Θ is **asymmetric** and S still depends on X (explicitely) and on Θ (in distribution).

Ronald A. Fisher. *On the mathematical foundations of theoretical statistics*. Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character, 222.594-604 (1922): 309-368.

Motivation - Probabilistic Programs

- Consider computational graph of a probabilist program on the right:
- output variable: $O_2 := I_2 + O_3$ is independent of input variable I_1 given the input of I_2.
- we can reason about functional dependencies in probabilistic programs.

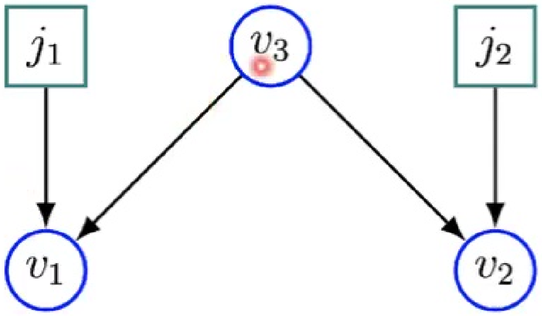


- we want to state this as:

$$O_2 \perp\!\!\!\perp I_1 \,|\, I_2$$

Causal Bayesian Networks with Input Variables

- A **causal Baysian network with input variables** consists of:
  - a **conditional directed acyclic graph** (CDAG): $G = (J, V, E)$,
  - a measurable space $\mathcal{X}_j$ for each $j \in J$,
  - a *standard* measurable space $\mathcal{X}_v$ for each $v \in V$,
  - a **transition probability**: $P_v(X_v \mid X_{\mathrm{Pa}^{G}(v)})$ for $v \in V$.



- The **joint transition probability** (product in a reverse topological ordering):

$$P(X_V \| X_J) := \bigotimes_{v \in V}^{<} P_v(X_v \mid X_{\mathrm{Pa}^{G}(v)}).$$

16

Motivation - Local Markov Property

- For the last node $w \in V$ w.r.t $<$ we get **factorization**:

$$P(X_V \| X_J) = P_w(X_w \,|\, X_{\mathrm{Pa}^G(w)}) \otimes \bigotimes_{v \in V \setminus \{w\}}^{<} P_v(X_v \,|\, X_{\mathrm{Pa}^G(v)})$$

$$= P_w(X_w \,|\, X_{\mathrm{Pa}^G(w)}) \otimes P(X_{V \setminus \{w\}} \| X_J),$$

joint = Markov kernel $\otimes$ marginal.

- We want to be able phrase this as a **conditional independence**:

$$X_w \perp\!\!\!\perp X_{V \setminus \{w\}} \,|\, X_{\mathrm{Pa}^G(w)} \qquad [P(X_V \| X_J)].$$

- Note that the middle Markov kernels does **not** depend on all of X_J !!!

Generalizing the Notion of Conditional Independence

Definition - Naive Extension of Conditional Independence

- Let $(P_t(W))_{t \in \mathscr{T}}$ be a family of probability distributions on $\mathscr{W}$.

- Say that X is **naively independent** of Y **conditioned** on Z if for every $t \in \mathscr{T}$ we have the conditional independence:

$$X \perp\!\!\!\perp Y | Z \quad [P_t(W)].$$

Problems - Naive Extension of Conditional Independence

- We cannot express ancillarity as for every $t \in \mathcal{T}$ the map T is constant $T = t$ under $P_t(W)$, so every variable X would satisfy:

$$X \perp\!\!\!\perp T \quad [P_t(W)]$$

- We are implicitely conditioning on the whole input T, so we cannot express (conditional) independence between input and output variables.

Definition - Extended Conditional Independence

- Consider a family of probability measures $(P_t(W))_{t \in \mathcal{T}}$ on $\mathcal{W}$.
- Consider measurable maps:
  - $X : \mathcal{W} \to \mathcal{X}, \quad Y : \mathcal{W} \to \mathcal{Y}, \quad Z : \mathcal{W} \to \mathcal{Z},$
  - $R : \mathcal{T} \to \mathcal{R}, \quad S : \mathcal{T} \to \mathcal{S}$ s.t. $(R, S) : \mathcal{T} \to \mathcal{R} \times \mathcal{S}$ injective.
- They display the **extended conditional independence**:

$$X \perp\!\!\!\perp (Y, R) \,|\, (Z, S) \qquad [(P_t(W))_{t \in \mathcal{T}}]_{\mathcal{E}}$$

- if for all $s \in S(\mathcal{T})$, all bounded measurable $h : \mathcal{X} \to \mathbb{R}$ there exists measurable map $g_{h,s} : \mathcal{Z} \to \mathbb{R}$ such that for all $t \in S^{-1}(s)$ we have:

$$\mathbb{E}_t[h(X) \,|\, Y, Z] = g_{h,s}(Z) \qquad P_t(W)\text{-a.s.}$$

P. Constantinou, A.P. Dawid. *Extended conditional independence and applications in causal inference*. The Annals of Statistics (2017): 2618-2653.

The Problems with "Extended Conditional Independence"

- variable R is meaningless:
  - definition of ECI does **not** dependent on R at all,
  - R only used to complement S,
  - however, $T := \mathrm{id}_{\mathcal{T}} : \mathcal{T} \to \mathcal{T}$ already complements every S,
- separation between stochastic variables X, Y, Z vs. non-stochastic variables R, S, rather than providing a unifying framework,
- arguably too technical definition,
- not enough separoid rules to be useful for graphical models, i.e. prove Global Markov Property for Causal Bayesian Networks with input variables,

P. Constantinou, A.P. Dawid. *Extended conditional independence and applications in causal inference*. The Annals of Statistics (2017): 2618-2653.

Fixing "Extended Conditional Independence"

- As discussed, with $T = \mathrm{id}_{\mathcal{T}}$ we always get:
  - $X \perp\!\!\!\perp (Y, R) \,|\, (Z, S) \iff X \perp\!\!\!\perp (Y, T) \,|\, (Z, S) \quad [(P_t(W))_{t\in\mathcal{T}}]_{\mathcal{E}}$
  - So, R can always be replaced by T, which then can be made implicit and removed from the notation:

$$X \perp\!\!\!\perp Y \,|\, (Z, S) \quad [(P_t(W))_{t\in\mathcal{T}}]_{\mathcal{E}}$$

- Express $(P_t(W))_{t\in\mathcal{T}}$ as Markov kernel $P(W \,|\, T)$,
- unify all variables to measurable maps:

$$X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}, \; Y : \mathcal{W} \times \mathcal{T} \to \mathcal{Y}, \; Z \leftrightarrow (Z, S) : \mathcal{W} \times \mathcal{T} \to \mathcal{Z}:$$

$$X \perp\!\!\!\perp Y \,|\, Z \quad [(P(W \,|\, T)]_{\mathcal{E}}$$

P. Constantinou, A.P. Dawid. *Extended conditional independence and applications in causal inference*. The Annals of Statistics (2017): 2618-2653.

Fixing "Extended Conditional Independence"

- Recall definition of: $X \perp\!\!\!\perp (Y, R) \mid (Z, S) \quad [(P_t(W))_{t \in \mathscr{T}}]_{\mathscr{E}}$:
  - $\forall s \in S(\mathscr{T})\ \forall h\ \exists g_{h,s}\ \forall t \in S^{-1}(s)$:

    $$\mathbb{E}_t[h(X) \mid Y, Z] = g_{h,s}(Z) \qquad P_t(W)\text{-a.s.}$$
- role of h is actually an indicator variable $\mathbb{1}_A$ with $A \in \mathscr{B}_{\mathscr{X}}$:
  - $\forall s \in S(\mathscr{T})\ \forall A \in \mathscr{B}_{\mathscr{X}}\ \exists g_{A,s}\ \forall t \in S^{-1}(s)$:

    $$\mathbb{E}_t[\mathbb{1}_A(X) \mid Y, Z] = g_{A,s}(Z) \qquad P_t(W)\text{-a.s.}$$
- axiom of choice now allows to reformulate:
  - $\exists G :\ \mathscr{B}_{\mathscr{X}} \times \mathscr{S} \times \mathscr{Z} \to \mathbb{R}, \quad \forall s \in S(\mathscr{T})\ \forall A \in \mathscr{B}_{\mathscr{X}}\ \forall t \in S^{-1}(s)$:

    $$P_t(X \in A \mid Y, Z) = G(A \mid s, Z) \qquad P_t(W)\text{-a.s.}$$
- finally, asking G to also have Markov kernel properties leads to following definition:

P. Constantinou, A.P. Dawid. *Extended conditional independence and applications in causal inference*. The Annals of Statistics (2017): 2618-2653.

Definition - Transitional Conditional Independence

- Consider measurable spaces $\mathscr{W}$, $\mathscr{T}$, and transition probability:

$$P(W|T) : \mathscr{T} - \to \mathscr{W}$$

- Consider measurable maps (or even transition probabilities):

$$X : \mathscr{W} \times \mathscr{T} - \to \mathscr{X}, \;\; Y : \mathscr{W} \times \mathscr{T} - \to \mathscr{Y}, \;\; Z : \mathscr{W} \times \mathscr{T} - \to \mathscr{Z}:$$

- We say that X is **independent** of Y **conditioned** on Z w.r.t. $P(W|T)$ if

<u>there exists</u> a transition probability: $Q(X|Z) : \mathscr{Z} - \to \mathscr{X}$, s.t.

$$P(X, Y, Z|T) = Q(X|Z) \otimes P(Y, Z|T).$$

- In symbols: $X \perp\!\!\!\perp Y | Z \quad [P(W|T)].$

Patrick Forré, *Transitional Conditional Independence*, 2021, https://arxiv.org/abs/2104.11547.

Recall: Operations on Markov Kernels

- We have the following operations on Markov kernels:
  - product: $P(X|Y,Z) \otimes Q(Y|Z)$,

$$(P(X|Y,Z) \otimes Q(Y|Z))(D|z) := \int P(X \in D^y | Y = y, Z = z)\, Q(Y \in dy | Z = z),$$

  - marginalization: $P(Y|Z) := P(X \in \mathscr{X}, Y|Z)$,
  - composition: $P(X|Y,Z) \circ Q(Y|Z) :=$ X-marginal of $P(X|Y,Z) \otimes Q(Y|Z)$,
  - push-forward: $P(W|T) : \mathscr{T} - \to \mathscr{W}$ along $X : \mathscr{W} \times \mathscr{T} \to \mathscr{X}$:

$$P(X|T) : \mathscr{T} - \to \mathscr{X}, \qquad P(X \in A | T = t) := P(W \in X_t^{-1}(A) | T = t),$$

  - conditioning needs extra attention …

Theorem: Disintegration of Markov Kernels

- Let $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{Z}$ be measurable spaces, where $\mathscr{X}$, $\mathscr{Y}$ are *standard*.
- Let $K(X, Y | Z) : \mathscr{Z} \to \mathscr{P}(\mathscr{X} \times \mathscr{Y})$ be a transition probability.
- Then there exists a **conditional transition probability**:

$$K(X | Y, Z) : \mathscr{Y} \times \mathscr{Z} \to \mathscr{P}(\mathscr{X})$$

such that:

$$K(X, Y | Z) = K(X | Y, Z) \otimes K(Y | Z).$$

- Furthermore, two such conditional transition probabilities $K(X | Y, Z)$ only differ by a measurable $K(Y | Z)$-null set of $\mathscr{Y} \times \mathscr{Z}$.

Remark - Recover Old Definitions

- **Transitional** implies **Extended Conditional Independence**:

$$X \perp\!\!\!\perp (Y,R)\,|\,(Z,S) \quad [P(W\,|\,T)]$$

$$\implies X \perp\!\!\!\perp (Y,R)\,|\,(Z,S) \quad [(P_t(W))_{t\in\mathcal{T}}]_{\mathcal{E}}$$

- For constant $T = \star$ we **recover the usual definition** on *standard* measurable spaces:

$$X \perp\!\!\!\perp Y\,|\,Z \quad [P(W\,|\,T)] \iff X \perp\!\!\!\perp Y\,|\,Z \quad \text{(usual def.)}$$

Statistical Theory - Ancillary Statistics

- Consider statistical model $P(X|\Theta)$ and statistic $R: \mathscr{X} \to \mathscr{R}$.
- Then the following statements are equivalent:
  - R is an ancillary statistic,
  - $R \perp\!\!\!\perp \Theta \quad [P(X|\Theta)]$,
  - There exists a probability distribution $Q(R)$ s.t.

$$P(R|\Theta) = Q(R).$$

Statistical Theory - Sufficient Statistics
A General Fisher-Neyman Factorization Theorem

- Consider statistical model $P(X \mid \Theta)$ and statistic $S: \mathscr{X} \to \mathscr{S}$.
- Then the following points are equivalent:
  - S is sufficient,
  - $X \perp\!\!\!\perp \Theta \mid S \quad [P(X \mid \Theta)]$,
  - There exists a factorization:

$$P(X, S \mid \Theta) = Q(X \mid S) \otimes P(S \mid \Theta).$$

Fisher-Neyman Factorization Theorem

- Consider statistical model $P(X \mid \Theta)$ and statistic $S : \mathscr{X} \to \mathscr{S}$.
- Assume: $P(X \mid \Theta)$ has density $p(x \mid \theta)$ w.r.t. σ-finite reference measure μ.
- Then the following are equivalent:
  - S is sufficient,
  - $X \perp\!\!\!\perp \Theta \mid S \quad [P(X \mid \Theta)]$,
  - There exists a factorization of the density:

$$p(x \mid \theta) = g(S(x) \mid \theta) \cdot h(x),$$

Basu's Theorem as a Rule for Conditional Independence

- Consider statistical model $P(X \mid \Theta)$ and
  - statistics: $S : \mathscr{X} \to \mathscr{S}, \quad R : \mathscr{X} \to \mathscr{R}$ such that:
    - R is ancillary: $R \perp\!\!\!\perp \Theta \qquad [P(X \mid \Theta)]$,
    - S is sufficient: $X \perp\!\!\!\perp \Theta \mid S \quad [P(X \mid \Theta)]$,
    - S is boundedly complete: for $g : \mathscr{S} \to \mathbb{R}$ bounded, measurable:
      - $\mathbb{E}[g(S) \mid \Theta] = 0 \implies g = 0 \quad P(S \mid \Theta)$-a.s.
- Then we have: $R \perp\!\!\!\perp \Theta, S \quad [P(X \mid \Theta)]$.

Statistical Theory - Propensity Score

- Consider transition probability $P(X|Z)$, put $E(z) := P(X|Z = z)$.
- Then $E: \mathscr{Z} \to \mathscr{P}(\mathscr{X}) =: \mathscr{E}$ is a measurable map.
- Let $S: \mathscr{Z} \to \mathscr{S}$ be another measurable map.
- Then we have the equivalence:

$$X \perp\!\!\!\perp Z | S \quad \Longleftrightarrow \quad E = g(S) \text{ for some measurable } g: \mathscr{S} \to \mathscr{E}.$$

- In particular: $X \perp\!\!\!\perp Z | E$.
- The **propensity score** function E captures **all information** about X in Z, and it is **minimal** among all $S: \mathscr{Z} \to \mathscr{S}$ w.r.t. that property.

P.R. Rosenbaum, D.B. Rubin. *The central role of the propensity score in observational studies for causal effects*. Biometrika 70.1 (1983): 41-55.

Posterior vs. Likelihood-based Inference

- Let $P(X \mid \Theta)$ be a statistical model
  - likelihood function: $L(\theta) := P(X \mid \Theta = \theta)$
  - family of priors: $P(\Theta \mid \Pi)$,
  - joint: $P(X, \Theta \mid \Pi) := P(X \mid \Theta) \otimes P(\Theta \mid \Pi)$,
  - marginal: $P(X \mid \Theta)$
  - posteriors: conditional $P(\Theta \mid X, \Pi)$ s.t.:

$$P(X, \Theta \mid \Pi) = P(\Theta \mid X, \Pi) \otimes P(X \mid \Pi)$$

  - posterior function: $Z(x, \pi) := P(\Theta \mid X = x, \Pi = \pi)$

34

Posterior vs. Likelihood-based Inference

- We get the following conditional independencies:
  - likelihood-based inference:
    - $X \perp\!\!\!\perp \Theta \mid L \quad [P(X \mid \Theta)]$
  - Bayesian posterior-based inference:
    - $\Theta \perp\!\!\!\perp X \mid Z \quad [P(X, \Theta \mid \Pi)]$
- => capture nuanced differences between both approaches with the asymmetry of transitional conditional independence.

Asymmetric Separoid Rules - Notation

- Consider transition probability: $P(W\,|\,T) : \mathscr{T} \dashrightarrow \mathscr{W}$.
- Let $T : \mathscr{W} \times \mathscr{T} \to \mathscr{T}$ the canonical projection map $T := \mathrm{pr}_{\mathscr{T}}$,
- and $\star : \mathscr{W} \times \mathscr{T} \to \{\,*\,\}$ be the constant map.
- Consider **"conditional" random variables**:

$$X : \mathscr{W} \times \mathscr{T} \dashrightarrow \mathscr{X}, \qquad Y : \mathscr{W} \times \mathscr{T} \dashrightarrow \mathscr{Y},$$
$$Z : \mathscr{W} \times \mathscr{T} \dashrightarrow \mathscr{Z}, \qquad U : \mathscr{W} \times \mathscr{T} \dashrightarrow \mathscr{U},$$

where $\mathscr{X}, \mathscr{Y}, \mathscr{Z}, \mathscr{U}$ are *standard* measurable spaces.

- We then get the following **(asymmetric) separoid rules**.

Asymmetric Separoid Rules - Special Rules

- Left Redundancy:

$$X \precsim_P Z \implies X \perp\!\!\!\perp Y | Z.$$

- T-Restricted Right Redundancy:

$$X \perp\!\!\!\perp \star \,|\, Z, T \quad \text{always holds.}$$

- T-Inverted Right Decomposition:

$$X \perp\!\!\!\perp Y | Z \implies X \perp\!\!\!\perp T, Y | Z.$$

Asymmetric Separoid Rules - Standard Rules

- Left Decomposition:

$$X, U \perp\!\!\!\perp Y | Z \implies U \perp\!\!\!\perp Y | Z.$$

- Right Decomposition:

$$X \perp\!\!\!\perp Y, U | Z \implies X \perp\!\!\!\perp U | Z.$$

- Left Weak Union:

$$X, U \perp\!\!\!\perp Y | Z \implies X \perp\!\!\!\perp Y | U, Z.$$

- Right Weak Union:

$$X \perp\!\!\!\perp Y, U | Z \implies X \perp\!\!\!\perp Y | U, Z.$$

Asymmetric Separoid Rules - Contraction Rules

- Left Contraction:

$$X \perp\!\!\!\perp Y \,|\, U, Z \;\wedge\; U \perp\!\!\!\perp Y \,|\, Z$$
$$\Longrightarrow\; X, U \perp\!\!\!\perp Y \,|\, Z.$$

- Right Contraction:

$$X \perp\!\!\!\perp Y \,|\, U, Z \;\wedge\; X \perp\!\!\!\perp U \,|\, Z$$
$$\Longrightarrow\; X \perp\!\!\!\perp Y, U \,|\, Z.$$

- Flipped Left Cross Contraction:

$$X \perp\!\!\!\perp Y \,|\, U, Z \;\wedge\; Y \perp\!\!\!\perp U \,|\, Z$$
$$\Longrightarrow\; Y \perp\!\!\!\perp X, U \,|\, Z.$$

- Right Cross Contraction:

$$X \perp\!\!\!\perp Y \,|\, U, Z \;\wedge\; U \perp\!\!\!\perp X \,|\, Z$$
$$\Longrightarrow\; X \perp\!\!\!\perp Y, U \,|\, Z.$$

Asymmetric Separoid Rules - Derived Rules

- Restricted Symmetry:

$$X \perp\!\!\!\perp Y | Z \wedge Y \perp\!\!\!\perp \star | Z \implies Y \perp\!\!\!\perp X | Z.$$

- Symmetry:

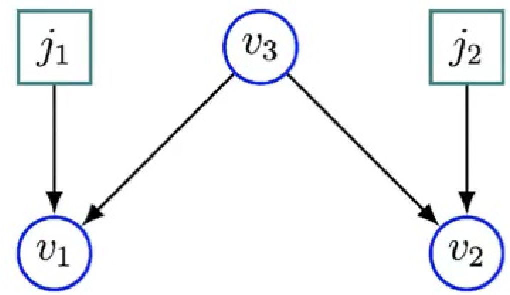
$$X \perp\!\!\!\perp Y | Z, T \implies Y \perp\!\!\!\perp X | Z, T.$$

Theorem: Global Markov Property

- Consider a **causal Bayesian network** with input variables J, observed variables V, latent variables U and marginal CADMG $G = (J, V, E, L)$.

- Then for every subsets: $A, B, C \subseteq V \cup J$

  (not necessarily disjoint) we have the implication:



$$A \perp' B \,|\, C \quad [G] \quad \Longrightarrow \quad X_A \perp\!\!\!\perp X_B \,|\, X_C \quad [P(X_V \| X_J)].$$

- Note: $A \perp' B \,|\, C :\Longleftrightarrow A \perp J \cup B \,|\, C.$

Remark - Use for Causality

- The **Global Markov Property** immediately implies validity of:
  - **Do-Calculus Rules**,
  - **Adjustment Criteria** (e.g. back-door criterion),
  - **ID-Algorithm** (for identification of causal effects)**,**

in **measure-theoretic generality**.

- Note: **Positivity**/absolute continuity **assumptions needed** to get well-defined Markov kernels (to take care of ambiguity up to null-sets of conditional Markov kernels).

Conditional Independence in Categorical Probability

- Let $\mathscr{C}$ be a Markov category (=abstraction of a category of spaces with transition probabilities / Markov kernels as morphisms).
- Definition: A morphism $K : \mathscr{T} \to \mathscr{X} \otimes \mathscr{Y} \otimes \mathscr{Z}$ in $\mathscr{C}$ displays

  the conditional independence: $X \perp\!\!\!\perp Y \,|\, Z,$

  iff there exists a factorization in $\mathscr{C}$:

$$K(X, Y, Z \,|\, T) = Q(X \,|\, Z) \otimes K(Y, Z \,|\, T).$$

- => leads to d-separation criterion for string diagrams!

T. Fritz, A. Klingler. *The d-Separation Criterion in Categorical Probability*. Journal of Machine Learning Research 24.46 (2023): 1-49.

Conclusion

- Simple definition of **Transitional Conditional Independence**:

  $$X \perp\!\!\!\perp Y \mid Z \quad :\Longleftrightarrow \quad \exists Q(X \mid Z). \quad P(X, Y, Z \mid T) = Q(X \mid Z) \otimes P(Y, Z \mid T).$$

- has **structural meaning** (in contrast to other definitions)
- works in **measure-theoretic generality**
  - e.g. combination of discrete and continuous distributions and more
- **unifying framework** allows reasoning between input, output, conditional random variables
- easily generalizes to **categorical probability theory**
- satisfies meaningful (ir-)relavance relations: left-/right versions of **(asymmetric) separoid axioms**
- "strong" definition: implies all previous definitions of (extended) conditional independence so far
- not too "strong: still get **Global Markov Property** for causal Bayesian networks with input variables
  - implies do-calculus rules, adjustment criteria, ID-algorithm
- can express typical notions in **statistical theory** equivalently as TCI relation
  - e.g. ancillarity, sufficiency, adequacy, propensity score, etc.