Transitional Conditional Independence

Forré, P.

Publication date
2021

Document Version
Submitted manuscript

License
Unspecified

Link to publication

Citation for published version (APA):
Forré, P. (2021). *Transitional Conditional Independence*. (v1 ed.) ArXiv.
https://arxiv.org/abs/2104.11547v1

Transitional Conditional Independence

Patrick Forré

AI4Science Lab, AMLab
Informatics Institute
University of Amsterdam
The Netherlands

We develope the framework of transitional conditional independence. For this we introduce transition probability spaces and transitional random variables. These constructions will generalize, strengthen and unify previous notions of (conditional) random variables and non-stochastic variables, (extended) stochastic conditional independence and some form of functional conditional independence. Transitional conditional independence is asymmetric in general and it will be shown that it satisfies all desired relevance relations in terms of left and right versions of the separoid rules, except symmetry, on standard, analytic and universal measurable spaces. As a preparation we prove a disintegration theorem for transition probabilities, i.e. the existence and essential uniqueness of (regular) conditional Markov kernels, on those spaces.
Transitional conditional independence will be able to express classical statistical concepts like sufficiency, adequacy and ancillarity.
As an application, we will then show how transitional conditional independence can be used to prove a directed global Markov property for causal graphical models that allow for non-stochastic input variables in strong generality. This will then also allow us to show the main rules of causal/do-calculus, relating observational and interventional distributions, in such measure theoretic generality.

2020 MSC: 62A99, 60A05.
Keywords: Extended conditional independence, asymmetric separoid axioms, conditional Markov kernels, causality, graphical models, global Markov property.

[0]p.d.forre@uva.nl

Contents

1. Introduction

Conditional independence nowadays is a widely used concept in statistics, probability theory and machine learning, e.g. see [Bis06, Mur12], especially in the areas of probabilistic graphical models and causality, see [DL93, Lau96, SGS00, Pea09, KF09, PJS17, Daw02, RS02, Ric03, ARS09, CMKR12, ER14, Eva16, Eva18, MMC20, BFPM21, GVP90] and many more. Already in its invention paper, see [Daw79a], a strong motivation for (further) developing conditional independence was the ability to express statistical concepts like sufficiency, adequacy and ancillarity, etc., in terms of conditional independence. For example, an ancillary statistic, see [Fis25, Bas64], is a function of the data that has the same probability distribution under any chosen model parameters. In the non-Bayesian setting the parameters of the model are not considered random variables, and thus stochastic conditional independence cannot express such concepts in its vanilla form. Since then the search for extensions of the definition of conditional independence started and several versions have been proposed and their characteristics been studied, e.g. see [Daw79a, Daw79b, Daw80, Daw98, Daw01a, GR01, CD17a, RERS17, FM20]. It is desirable that a definition of conditional independence satisfies many meaningful properties. Arguably, it should: (i) be related to functional and/or variation conditional independence and stochastic conditional independence, (ii) be able to express classical statistical concepts like sufficiency, adequacy, ancillarity, etc., (iii) work for large classes of measurable spaces, (iv) work for large classes of random variables and non-stochastic variables, or even combinations of those, (v) embrace and anticipate the inherent asymmetry in the dependency between the stochastic (output) and non-stochastic (input) parts of such variables, (vi) work for "conditional" (random) variables (which also need a proper definition first), (vii) work for large classes of (transition) probability distributions, e.g. for non-discrete variables that don't even have densities or are mixtures of discrete and absolute continuous probability distributions, etc., (viii) satisfy reasonable relevance relations and rules, e.g. as many of the separoid rules, see [Daw01a], as possible, (iix) give rise to meaningful factorizations of the used (transition) probability distributions, (ix) lead to global Markov properties for conditional probabilistic (causal) graphical models, (x) be as simple as possible.

All mentioned extensions of conditional independence, see [CD17a, RERS17, FM20], lack on some of those points. We will discuss this in Appendix L.

Contributions Instead of giving an ad hoc definition of extended conditional independence we go back to the roots of measure theoretic probability and first provide the proper framework of "conditional" versions of random variables, probability spaces, null sets, etc., which we will call *transitional random variables* and *transition probability spaces*, etc.. We prefer the word *transitional* over *conditional* to stress that it is related to transition probabilities (Markov kernels) and that there is no conditioning operation involved that would come from some joint probability space with a joint probability distribution. Also those construction will be defined for all points and not just up to some null sets.

Transitional probability spaces are products $\mathcal{W} \times \mathcal{T}$ of measurable spaces that contain the domain $\mathcal{T}$ and codomain $\mathcal{W}$ of a fixed transition probability/Markov kernel, which we will suggestively write similarly to conditional distributions: $\mathbf{P}(W|T)$ or $\mathbf{K}(W|T)$, etc.

Transitional random variables X are then measurable functions $X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}$ on such transition probability spaces. In fact, we will allow, more generally, for probabilistic maps $\mathbf{X} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$. Transitional random variables X can be thought of as "conditional" random variables, stochastic processes or random fields $(X_t)_{t\in\mathcal{T}}$ where we are not interested in their joint distribution like $\mathbf{P}((X_t)_{t\in\mathcal{T}})$, but in the functional dependence of their (push-forward) transition probability $\mathbf{P}(X_t|T = t)$ on their ("parameter") inputs t. They will already generalize and unify the notion of random variables and deterministic, non-stochastic variables and formalize "conditional" random variables.

We will then define that the transitional random variable X is *transitionally conditionally independent* from the transitional random variable Y given the transitional random variable Z w.r.t. $\mathbf{P}(W|T)$ if there exists a Markov kernel $\mathbf{Q}(X|Z)$ such that:

$$\mathbf{P}(X, Y, Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{P}(Y, Z|T),$$

where $\mathbf{P}(Y, Z|T)$ is the marginal Markov kernel of the joint push-forward Markov kernel $\mathbf{P}(X, Y, Z|T)$ of $\mathbf{P}(W|T)$ and where $\otimes$ denotes the product of Markov kernels. In symbols we will write:

$$X \underset{\mathbf{P}(W|T)}{\perp\!\!\!\perp} Y \mid Z.$$

This asymmetric notion of conditional independence will resolve all mentioned points above: It is arguably simple, comes with a meaningful factorization, can be defined for all measurable spaces and satisfies left and right versions of the *separoid rules*, see [Daw01a], except symmetry, as long as the transitional random variables have as codomains standard, analytic or just universal measurable spaces. Note that the full set of separoid rules was not possible to prove with the technically more involved definition of extended conditional independence from [CD17a] even for standard measurable spaces.

As a required step to prove the (asymmetric) *separoid rules* we will first need to prove the disintegration of transition probabilities/Markov kernels $\mathbf{K}(X, Y|T)$ in two transitional random variables that have as codomains standard, analytic or universally measurable spaces. In other words, we will show that there exists a *(regular) conditional Markov kernel* $\mathbf{K}(X|Y, T)$ such that:

$$\mathbf{K}(X, Y|T) = \mathbf{K}(X|Y, T) \otimes \mathbf{K}(Y|T),$$

where $\mathbf{K}(Y|T)$ is the marginal Markov kernel of $\mathbf{K}(X, Y|T)$ and $\otimes$ denotes the product of Markov kernels. The difficulty is to arrive at a *conditional Markov kernel* that is a probability measure in X for each value of Y and T and that, at the same time, is jointly measurable in (Y, T), and not just measurable in one variable when the other variable is fixed. This is the reason that we need to restrict ourselves to measurable spaces that come with some topological underpinning and built-in countability properties, like the mentioned standard, analytic and universal measurable spaces.

One can also argue that one can push the disintegration of Markov kernels (a bit, but) not much further than universal measurable spaces. Already the theory of regular conditional probability distributions was shown to quickly run into the foundations and axiomization of set theory, e.g. if the continuum hypothesis holds or not, see [Fad85,Fre15]. Also one of the most general disintegration result for probability distributions requires (countably) compact approximating classes, see [Pac78, Fre15] 452I. Furthermore, to achive measurablity usually a countability assumption for the used σ-algebra is needed, like countably generated and/or countably separated, see [Fad85, BD75, BRN63, Fre15] 452. This shows that if one does not want to specialize to specific probability measures or make strong topological assumptions one cannot go much further than universal measurable spaces for the disintegration of probability measures let alone Markov kernels.

Transitional conditional independence will imply the usual notion of conditional independence for random variables (via the corner case where $\mathcal{T} = \{*\}$, the one-point space), but also the other notions of extended conditional independence, see [CD17a, RERS17, FM20].

Transitional conditional independence can express the statistical concepts of ancillary, sufficient and adequate statistic, see [Fis22, Fis25], $S(X)$ for statistical model $\mathbf{P}(W|\Theta)$ and transitional random variables X (and Y) via:

1. Ancillarity: $S(X) \underset{\mathbf{P}(W|\Theta)}{\perp\!\!\!\perp} \Theta$.
2. Sufficiency: $X \underset{\mathbf{P}(W|\Theta)}{\perp\!\!\!\perp} \Theta \,|\, S(X)$.
3. Adequacy: $X \underset{\mathbf{P}(W|\Theta)}{\perp\!\!\!\perp} \Theta, Y \,|\, S(X)$.

Transitional conditional independence can also encode deterministic functional relations. For example, if F is a function on a product space, $F : \mathcal{T}_1 \times \mathcal{T}_2 \to \mathcal{F}$, with $(t_1, t_2) \mapsto F(t_1, t_2)$. Then F is a function of t_1 alone, if and only if:

$$F \underset{\mathbf{P}(W|T_1,T_2)}{\perp\!\!\!\perp} T_2 \,|\, T_1,$$

where T_i are the canonical projections onto factors $\mathcal{T}_i$, $i = 1, 2$, and F is viewed as transitional random variable on $\mathcal{W} \times \mathcal{T}_1 \times \mathcal{T}_2$.

As an application of *transitional conditional independence* and its separoid rules we show that "conditional"/transitional graphical models like causal Bayesian networks or structural causal models (with latent and input variables) will follow a *directed global Markov property*, i.e. they entail transitional conditional independence relations that are graphically encoded via σ-separation, a generalization of d-separation and m-separation, see [Pea09,Ric03,FM17,FM18,FM20]. The proof relies on the fact that both σ-separation and transitional conditional independence follow analogous asymmetric separoid rules.

We will formalize the notion of asymmetric separoids and study when we recover the usual notion of (symmetric) separoids.

Finally, we can use the global Markov property and apply it to an extended causal Bayesian network that also contains soft intervention variables to gain a graphical criterion about when one can remove hard interventions or replace them with conditioning operations, see [Pea09]. Transitional conditional independence then automatically presents us with meaningful factorizations and Markov kernels. Such rules are of importance for the identification of causal effects, see [Pea09]. In practice such rules can give us guidelines about when one can replace costly randomized control trials with observational studies and also how to estimate the corresponding causal effects.

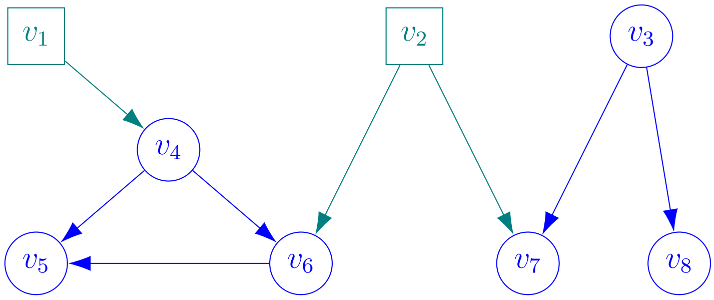


Figure 1: Causal Bayesian network with input nodes v_1, v_2 and output nodes $v_3, \dots, v_8$. The graph allows us to read off the transitional conditional independencies: $X_7 \perp\!\!\!\perp X_1 \,|\, X_2$ and $X_7 \perp\!\!\!\perp X_1, X_5 \,|\, X_2, X_4, X_6$, etc.

Overview In Section 2 we will develop transitional probability theory built on the notion of Markov kernels/transition probabilities. We introduce the notions of transition probability spaces, transitional random variables and null-sets, etc. We also go over typical constructions for Markov kernels like marginalization, product, composition, etc. Our main Theorem of this section will be concerned with the existence of (regular) conditional Markov kernels.

In Section 3 we will define transitional conditional independence for general transitional random variables. We then demonstrate its meaning in the two corner cases: random variables and deterministic maps. Our main result of this section will be to show that transitional conditional independence satisfies all left and right versions of the separoid rules.

In Section 4 we will show how transitional conditional independence can express classical statistical concepts like ancillarity, sufficiency, adequacy, etc. We also demonstrate what it can say about propensity scores, likelihoods and Bayesian statistics.

In Section 5 we will review some graph theory, which is needed for the section after. The main result will be the introduction of the notion of σ-separation and that it also satisfies all the asymmetric separoid rules in total analogy to transitional conditional independence.

In Section 6, as a the most striking application of transitional conditional independence, we will introduce the notion of causal Bayesian networks that allow for (non-stochastic) input variables. The main theorem will be that such causal Bayesian networks will satisfy a directed global Markov property, relating its graphical structure to transitional conditional independence relations. Since the notion of transitional conditional independence is stronger than any other notion of (extended) conditional independence, our result is thus the strongest form of a global Markov property for such graphical models proven so far, while using the same assumptions.

Finally, in Section 7 we will discuss our findings and also indicate how to extend all concepts from this paper to universal measurable spaces, which generalize standard and analytic measurable spaces, where all the proofs can be found in the Appendix.

All of the proofs of the above can be found in the corresponding Appendices.

Of own interest is Appendix A, where we introduce and formalize the notion of τ-κ-separoids, an asymmetric version of separoids. In two theorems we will show how τ-κ-separoids can be constructed from symmetric ones and vice versa, demonstrating the close relationship between these two concepts.

In Appendix K we will extend the theory of causal Bayesian networks and relax its definition to also allow for latent veriables. Then we introduce hard interventions, soft interventions and marginization for them. We also show how to extend the global Markov property to more general probabilistic graphical models. Finally, we demonstrate how the global Markov property can be used to get a measure theoretic clean proof of the 3 rules of do-calculus and the most important adjustment formulas.

In Appendix L we have a more detailed comparison of transitional conditional independence to other notions of conditional independence. In particular, we compare to "the" extendend conditional independence, to symmetric extendend conditional independence, to extendend conditional independence based on families of probability distributions and also to variation conditional independence.

Notations We will use curly letters like $\mathcal{W}$, $\mathcal{T}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ to indicate *measurable spaces*. We implicitly assume that they are endowed with a fixed *σ-algebra*, which we will denote by $\mathcal{B}_\mathcal{W}$, $\mathcal{B}_\mathcal{T}$, etc., if needed. If we say that $D \subseteq \mathcal{W}$ is a *measurable subset* we will mean $D \in \mathcal{B}_\mathcal{W}$. We will, unless stated otherwise, always assume that *topological spaces* like $\mathbb{R}^D$, $[0,1]$, $\bar{\mathbb{R}} := [-\infty, +\infty]$, etc., are endowed with their *Borel σ-algebra*. Similarly, we will assume that *product spaces* like $\mathcal{W} \times \mathcal{T}$ carry the *product σ-algebra*. For the *space of probability measures* $\mathcal{P}(\mathcal{W})$ on $\mathcal{W}$ we will use the smallest σ-algebra such that all evaluations maps $j_D : \mathcal{P}(\mathcal{W}) \to [0,1]$ given by $j_D(\mathbf{P}) := \mathbf{P}(D)$ for $D \in \mathcal{B}_\mathcal{W}$ are measurable. Maps will usually be denoted by capital letters X, Y, Z in correspondence to their codomains $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, resp. We use bold font letters $\mathbf{K}$, $\mathbf{P}$, $\mathbf{X}$, etc., to indicate probability distributions or Markov kernels. Later we will use $\mathbf{G}$ to denote graphs. If we say that $X : \mathcal{W} \to \mathcal{X}$ is a *measurable map* we implicitely assume that $\mathcal{W}$ and $\mathcal{X}$ are measurable spaces and that $\mathcal{B}_\mathcal{W} \supseteq f^*\mathcal{B}_\mathcal{X} := \{f^{-1}(A) \,|\, A \in \mathcal{B}_\mathcal{X}\}$. A *standard measurable space* (standard Borel space) is a measurable space $\mathcal{X}$ whose measurable sets $B \in \mathcal{B}_\mathcal{X}$ are the Borel sets of a complete separable metric d on $\mathcal{X}$, e.g. $\mathbb{R}^D$ or $\mathbb{Z}$. We call $\mathcal{X}$ (or $\mathcal{B}_\mathcal{X}$) *countably generated* if $\mathcal{B}_\mathcal{X} = \sigma(\mathcal{E})$ for a countable subset $\mathcal{E} \subseteq \mathcal{B}_\mathcal{X}$. We will also treat analytic and universal measurable spaces. For simplicity of presentation we will treat only (standard) measurable spaces in the main paper and refer for the corresponding but technically more demanding results for analytic and universal measurable spaces to the Appendix B.

2. Transitional Probability Theory

2.1. Transition Probabilities/Markov Kernels

Here we will review the notion of transitional probabilities, also known as Markov kernels. We mainly introduce our suggestive notations, which make the later theory more intuitive.

Definition 2.1 (Markov kernel)**.** *Let $\mathcal{T}$, $\mathcal{W}$ be measurable spaces. A* Markov kernel *or* transition probability *from $\mathcal{T}$ to $\mathcal{W}$ is - per definition - a map:*

$$\mathbf{K} : \mathcal{B}_{\mathcal{W}} \times \mathcal{T} \to [0,1], \quad (D,t) \mapsto \mathbf{K}(D|t),$$

such that:

1. *For each $t \in \mathcal{T}$ the mapping:*

$$\mathcal{B}_{\mathcal{W}} \to [0,1], \quad D \mapsto \mathbf{K}(D|t)$$

   is a probability measure (i.e. normalized and countably additive).

2. *For each $D \in \mathcal{B}_{\mathcal{W}}$ the mapping:*

$$\mathcal{T} \to [0,1], \quad t \mapsto \mathbf{K}(D|t)$$

   is measurable.

Notation 2.2 (Markov kernel)**.** *We will most of the time use the dashed arrow $\dashrightarrow$ to $\mathcal{W}$ instead of a usual arrow $\to$ on other spaces to indicate Markov kernels:*

$$\mathbf{K} : \mathcal{T} \dashrightarrow \mathcal{W}, \quad (D,t) \mapsto \mathbf{K}(D|t).$$

Furthermore, we will often use suggestive notations as follows:

$$\mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{W}, \quad (D,t) \mapsto \mathbf{K}(W \in D|T = t) := \mathbf{K}(D|t).$$

We also use the following notations. For fixed $D \in \mathcal{B}_{\mathcal{W}}$ the map:

$$\mathbf{K}(W \in D|T) : \mathcal{T} \to [0,1], \quad t \mapsto \mathbf{K}(W \in D|T = t),$$

and for fixed $t \in \mathcal{T}$ the map:

$$\mathbf{K}(W|T = t) : \mathcal{B}_{\mathcal{W}} \to [0,1], \quad D \mapsto \mathbf{K}(W \in D|T = t).$$

We might also use the same notation as above to represent the Markov kernel as a measurable probabilistic map:

$$\mathbf{K}(W|T) : \mathcal{T} \to \mathcal{P}(\mathcal{W}), \quad t \mapsto \mathbf{K}(W|T = t).$$

Here W and T are considered suggestive symbols only, but one could give W the meaning of the (identity or) projection map $\mathrm{pr}_{\mathcal{W}}$ onto $\mathcal{W}$. From the point on we also have a map T mapping to $\mathcal{T}$ the notation becomes ambiguous: $\mathbf{K}(W|T)$ could also mean $\mathbf{K}(W|T)$ where we plugged in T for t in "$T = t$", similar to conditional expectations $\mathbb{E}[W|T]$, but the meaning should become clear from the context.

Remark 2.3 (Markov kernels generalize probability distributions)**.**

1. *Every probability distribution $\mathbf{P}(W) \in \mathcal{P}(\mathcal{W})$ can be considered as a* constant Markov kernel *from $\mathcal{T}$ to $\mathcal{W}$ via:*

$$\mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{W}, \quad (D, t) \mapsto \mathbf{K}(W \in D|T = t) := \mathbf{P}(W \in D).$$

2. *Every Markov kernel from the one-point space: $\mathcal{T} = \circledast := \{*\}$ to $\mathcal{W}$:*

$$\mathbf{K}(W|T) : \circledast \dashrightarrow \mathcal{W}, \quad (D, *) \mapsto \mathbf{K}(W \in D|T = *),$$

defines a unique probability distribution $\mathbf{P}(W) \in \mathcal{P}(\mathcal{W})$ given via:

$$\mathbf{P}(W \in D) := \mathbf{K}(W \in D|T = *).$$

So we can identify probability distributions on $\mathcal{W}$ with Markov kernels $\circledast \dashrightarrow \mathcal{W}$.

Remark 2.4 (Markov kernels generalize deterministic maps)**.** *Consider a measurable map $\tilde{X} : \mathcal{T} \to \mathcal{X}$. Then we can turn $\tilde{X}$ into a Markov kernel $\boldsymbol{\delta}_{\tilde{X}}(X|T)$ via:*

$$\boldsymbol{\delta}_{\tilde{X}}(X|T) : \mathcal{T} \dashrightarrow \mathcal{X}, \quad (A, t) \mapsto \boldsymbol{\delta}_{\tilde{X}}(X \in A|T = t) := \mathbb{1}_A(\tilde{X}(t)),$$

which puts 100% We will often also use the notation without the dummy variable X: $\boldsymbol{\delta}(\tilde{X}|T) := \boldsymbol{\delta}_{\tilde{X}}(X|T)$.

2.2. Constructing Transition Probabilities from Others

In the following we will demonstrate how one can construct new Markov kernels from others. The constructions include marginalization, product, composition, push-forward. Later an own subsection is dedicated to conditioning. Note that the measurability of those constructions is either clear or can be proven using Dynkin's π-λ theorem, see [Kle14] Thm. 1.19, also see [Bog07] Thm. 10.7.2.

Definition 2.5 (Marginalizing Markov kernels)**.** *Let*

$$\mathbf{K}(X, Y|T) : \mathcal{T} \dashrightarrow \mathcal{X} \times \mathcal{Y}$$

be a Markov kernel in two variables. We can then define the marginal Markov kernels *as follows:*

$$\mathbf{K}(X|T) : \mathcal{T} \dashrightarrow \mathcal{X}, \quad (A, t) \mapsto \mathbf{K}(X \in A, Y \in \mathcal{Y}|T = t),$$

and:

$$\mathbf{K}(Y|T) : \mathcal{T} \dashrightarrow \mathcal{Y}, \quad (B, t) \mapsto \mathbf{K}(X \in \mathcal{X}, Y \in B|T = t).$$

Definition 2.6 (Product of Markov kernels)**.** *Consider two Markov kernels:*

$$\mathbf{Q}(Z|Y, W, T) : \mathcal{Y} \times \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Z}, \qquad \mathbf{K}(W, U|T, X) : \mathcal{T} \times \mathcal{X} \dashrightarrow \mathcal{W} \times \mathcal{U}.$$

Then we define the product Markov kernel*:*

$$\mathbf{Q}(Z|Y, W, T) \otimes \mathbf{K}(W, U|T, X) : \mathcal{Y} \times \mathcal{T} \times \mathcal{X} \dashrightarrow \mathcal{Z} \times \mathcal{W} \times \mathcal{U},$$

using measurable sets $E \subseteq \mathcal{Z} \times \mathcal{W} \times \mathcal{U}$ *via:* $(E, (y, t, x)) \mapsto$

$$\int\int \mathbb{1}_E(z, w, u)\, \mathbf{Q}(Z \in dz|Y = y, W = w, T = t)\, \mathbf{K}((W, U) \in d(w, u)|T = t, X = x),$$

where the inner integration is over $z \in \mathcal{Z}$ *and the outer integration over* $(w, u) \in \mathcal{W} \times \mathcal{U}$*.*

Definition 2.7 (Composition of Markov kernels)**.** *Consider two Markov kernels:*

$$\mathbf{Q}(Z|Y, W, T) : \mathcal{Y} \times \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Z}, \qquad \mathbf{K}(W, U|T, X) : \mathcal{T} \times \mathcal{X} \dashrightarrow \mathcal{W} \times \mathcal{U}.$$

Then we define their composition*:*

$$\mathbf{Q}(Z|Y, W, T) \circ \mathbf{K}(W, U|T, X) : \mathcal{Y} \times \mathcal{T} \times \mathcal{X} \dashrightarrow \mathcal{Z},$$

using measurable sets $C \subseteq \mathcal{Z}$ *via:* $(C, (y, t, x)) \mapsto$

$$\int \mathbf{Q}(Z \in C|Y = y, W = w, T = t)\, \mathbf{K}(W \in dw|T = t, X = x).$$

Note that we implicitely marginalized U *out, i.e. in the composition we integrate over all variables (here:* W *and* U*) from the right hand Markov kernel.*

Remark 2.8.

1. *It is clear from the Definitions 2.7, 2.6 and 2.5 that the composition:*

$$\mathbf{Q}(Z|Y, W, T) \circ \mathbf{K}(W, U|T, X)$$

is the Z*-marginal of the product:*

$$\mathbf{Q}(Z|Y, W, T) \otimes \mathbf{K}(W, U|T, X).$$

2. *Both, products and compositions, are each associative, but clearly not commutative in general.*

3. *If the left Markov kernel $\mathbf{Q}(Z|Y,T)$ has no dependence in the second arguments w.r.t. to a first argument of the right Markov kernel $\mathbf{K}(W,U|T,X)$, i.e. no W in the above terms, then they commute by Fubini's theorem:*

$$\mathbf{Q}(Z|Y,T) \otimes \mathbf{K}(W,U|T,X) = \mathbf{K}(W,U|T,X) \otimes \mathbf{Q}(Z|Y,T).$$

Remark 2.9 (Composition of deterministic Markov kernels)**.** *Consider measurable maps:*

$$\tilde{X} : \mathcal{T} \to \mathcal{X}, \qquad \tilde{Z} : \mathcal{X} \to \mathcal{Z},$$

and their composition $\tilde{Z} \circ \tilde{X}$. Then the composition of the corresponding Markov kernels satisfies:

$$\boldsymbol{\delta}_{\tilde{Z}\circ\tilde{X}}(Z|T) = \boldsymbol{\delta}_{\tilde{Z}}(Z|X) \circ \boldsymbol{\delta}_{\tilde{X}}(X|T),$$

where $\boldsymbol{\delta}_{\tilde{Z}}(Z \in C|X = x) := \mathbb{1}_C(\tilde{Z}(x))$ and $\boldsymbol{\delta}_{\tilde{X}}(X \in A|T = t) := \mathbb{1}_A(\tilde{X}(t))$.
So the composition of Markov kernels extends the composition of maps.

Definition 2.10 (Push-forward Markov kernel w.r.t. measurable maps)**.** *Consider a Markov kernel $\mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{W}$ and a measurable map: $X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}$. Then we define the* push-forward Markov kernel*:*

$$X_*\mathbf{K}(W|T) := \mathbf{K}(X(W,T)|T) := \mathbf{K}(X|T) : \mathcal{T} \dashrightarrow \mathcal{X},$$

of $\mathbf{K}(W|T)$ w.r.t. X via: $(A,t) \mapsto$

$$\mathbf{K}(X \in A|T = t) := \mathbf{K}(W \in X_t^{-1}(A)|T = t),$$

where:

$$X_t^{-1}(A) = X^{-1}(A)_t := \{w \in \mathcal{W} \,|\, X(w,t) \in A\}.$$

Remark 2.11. *We can also write push-forwards as compositions:*

$$\mathbf{K}(X|T) = \boldsymbol{\delta}(X|W,T) \circ \mathbf{K}(W|T),$$

where: $\boldsymbol{\delta}(X \in A|W = w, T = t) := \mathbb{1}_A(X(w,t)) = \mathbb{1}_{X^{-1}(A)}(w,t)$. In this sense compositions of Markov kernels generalize push-forward Markov kernels.

Definition 2.12 (Push-forward Markov kernel w.r.t. another Markov kernel)**.** *Consider Markov kernels $\mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{W}$ and $\mathbf{X}(X|W,T) : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$. Then we define the* push-forward Markov kernel *as the composition:*

$$\mathbf{K}(X|T) := \mathbf{X}(X|W,T) \circ \mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{X}.$$

Remark 2.13. *Any Markov kernel $\mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{W}$ can always be extended to include the canonical projection map $T = \mathrm{pr}_{\mathcal{T}} : \mathcal{W} \times \mathcal{T} \to \mathcal{T}$ via:*

$$\mathbf{K}(W,T|T) : \mathcal{T} \dashrightarrow \mathcal{W} \times \mathcal{T}, \quad (E,t) \mapsto$$

$$\mathbf{K}((W,T) \in E|T = t) = \mathbf{K}(W \in E_t|T = t),$$

where $E_t = \{w \in \mathcal{W} \,|\, (w,t) \in E\}$. Using Definition 2.6, we can also write this as:

$$\mathbf{K}(W,T|T) = \mathbf{K}(W|T) \otimes \boldsymbol{\delta}(T|T) = \boldsymbol{\delta}(T|T) \otimes \mathbf{K}(W|T),$$

where $\boldsymbol{\delta}(T \in D|T = t) := \mathbb{1}_D(t)$ for measurable $D \subseteq \mathcal{T}$ and $t \in \mathcal{T}$.

2.3. Null Sets w.r.t. Transition Probabilities

Definition 2.14 (Null sets w.r.t. transition probabilities). *Let* $\mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{W}$ *be a transition probability. A subset* $M \subseteq \mathcal{W} \times \mathcal{T}$ *will be called a* $\mathbf{K}(W|T)$-null set *if every section/fibre* $M_t := \{w \in \mathcal{W} \,|\, (w,t) \in M\}$ *is a* $\mathbf{K}(W|T=t)$*-null set, i.e. there exist measurable* $N_t \subseteq W$ *with* $M_t \subseteq N_t$ *and* $\mathbf{K}(W \in N_t|T=t) = 0$*, for every* $t \in \mathcal{T}$*.*

We are usually interested in measurable null sets. The notion of null sets w.r.t. transition probabilities generalizes the notion of null sets in probability spaces, which can be recovered by taking $\mathcal{T} = \{*\}$, the one-point space.

2.4. Transition Probability Spaces

We will now give the definition of a transition probability space, which will generalize the notion of probability spaces.

Definition 2.15 (Transition probability space). *Consider measurable spaces* $\mathcal{T}$ *and* $\mathcal{W}$ *and a Markov kernel/transition probability:*

$$\mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{W}, \quad (D,t) \mapsto \mathbf{K}(W \in D|T=t).$$

We then call the tuple: $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ *a* transition probability space*. It naturally comes with the canonical projection map:*

$$T : \mathcal{W} \times \mathcal{T} \to \mathcal{T}, \quad T(w,t) := t,$$

and the Markov kernel: $\mathbf{K}(W,T|T) := \mathbf{K}(W|T) \otimes \boldsymbol{\delta}(T|T)$*, which then satisfies* $\mathbf{K}(T|T) = \boldsymbol{\delta}(T|T)$*.*

The notion of transition probability space generalizes the notion of probability spaces, which can be recovered by taking $\mathcal{T} = *$, the one-point space.

2.5. Transitional Random Variables

In this subsection we will introduce the notion of transitional random variables, which will generalize the usual notion of random variables, formalizes what one could call “conditional” random variables. Furthermore, we start from a bit more general point of view as we not only allow for (deterministic) measurable maps, but also for stochastic maps, which again will be formalized as Markov kernels.

Definition 2.16 (Transitional random variables). *If* $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ *is a transition probability space then a* transitional random variable *is a Markov kernel:*

$$\mathbf{X} = \mathbf{X}(X|W,T) : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$$

to any other measurable space $\mathcal{X}$*. It will come with its push-forward Markov kernel:*

$$\mathbf{K}(X|T) := \mathbf{X}(X|W,T) \circ \mathbf{K}(W|T).$$

Remark 2.17. *1. If $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ is a transition probability space then any measurable map $X : \mathcal{W}\times\mathcal{T} \to \mathcal{X}$ induces a transitional random variable $\boldsymbol{\delta}(X|W,T)$ given by:*

$$\boldsymbol{\delta}(X \in A|W = w, T = t) := \mathbb{1}_A(X(w,t)).$$

By slight abuse of notation we will call X itself a transitional random variable as well, by actually referring to $\boldsymbol{\delta}(X|W,T)$. Transitional random variables of this form will be of the main focus in the following.

2. *A transitional random variable $X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}$ can be considered as a family of random variables measurably parameterized by $t \in \mathcal{T}$. For $t \in \mathcal{T}$ we have the measurable maps:*

$$X_t : \mathcal{W} \to \mathcal{X}, \quad w \mapsto X_t(w) := X(w,t),$$

each of which can be considered a random variable on the probability space $(\mathcal{W}, \mathbf{K}(W|T = t))$. Note that in this setting we are not modelling the joint distribution of $(X_t)_{t\in\mathcal{T}}$, but rather how the individual distribution of X_t depends on and varies with $t \in \mathcal{T}$.

3. *Note that by going from transition probability space $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ to the one $(\mathcal{X} \times \mathcal{T}, \mathbf{K}(X|T))$ for a transitional random variable $\mathbf{X}(X|W,T)$ the projection map:*

$$X : \mathcal{X} \times \mathcal{T} \to \mathcal{X}, \quad (x,t) \mapsto x,$$

can be considered a transitional random variable of the form $\boldsymbol{\delta}(X|X,T)$. So with only slight loss of generality one can replace a general transitional random variable $\mathbf{X}$ by one of the form $\boldsymbol{\delta}(X|W,T)$. More will be said in Theorem 4.5.

4. *The notion of transitional random variables generalizes the notion of random variables and formalizes what one could call a* (probabilistic) "conditional" random variable. *Note that the Markov kernel can be given without any conditioning operation.*

5. *Transitional random variables can model* probabilistic programs. *For each user chosen input $T = t$ a random input $w \sim \mathbf{K}(W|T = t)$ is drawn. Then the input (w,t) is presented to the probabilistic program $\mathbf{X}$ and an output is sampled $x_1 \sim \mathbf{X}(X|W = w, T = t)$. Using Markov kernels to represent transitional random variables allows for random noise inside the program that generates the output x_1. So even when presented with the same input (w,t) again another output $x_2 \neq x_1$ might be drawn. So $\mathbf{X}(X|W = w, T = t)$ models the output distribution for fixed input (w,t). Certainly, if one has no insight into the input sampling proceedure $\mathbf{K}(W|T)$ one might only be interested in the push-forward: $\mathbf{K}(X|T)$, which directly describes the output distribution for each user chosen input $T = t$.*

6. *If we want to model a deterministic variable with no stochasticity we could consider transitional random variables of the form $\boldsymbol{\delta}_\varphi(X|T)$ that do not depend on the W-argument.*

Example 2.18 (Special transitional random variables of importance)**.** *Let $(\mathcal{W}\times\mathcal{T}, \mathbf{K}(W|T))$ be a transitional probability space. Then we denote by:*

1. *T the* canonical projection *onto $\mathcal{T}$:*

$$T := \mathrm{pr}_{\mathcal{T}} : \mathcal{W} \times \mathcal{T} \to \mathcal{T}, \quad (w,t) \mapsto T(w,t) := t.$$

We also put:

$$\mathbf{T} = \mathbf{T}(T|W,T) = \boldsymbol{\delta}(T|W,T) : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{T}, \quad (D,(w,t)) \mapsto \mathbb{1}_D(t).$$

2. *$*$ the* constant transitional random variable*:*

$$* : \mathcal{W} \times \mathcal{T} \to \circledast,, \quad (w,t) \mapsto *,$$

where $\circledast := \{*\}$ is the one-point space. We also use the same symbol $*$ to denote the Markov kernel:

$$\boldsymbol{\delta}_* : \mathcal{W} \times \mathcal{T} \dashrightarrow \circledast, \quad (D,*) \mapsto \mathbb{1}_D(*).$$

2.6. Ordering the Class of Transitional Random Variables

We now introduce several comparison relations between transitional random variables. All them model to some degree that one variable $\mathbf{X}$ is a (deterministic or stochastic) measurable function of another one $\mathbf{Y}$ (up to some form of null set).

Notation 2.19. *Let $(\mathcal{W}\times\mathcal{T}, \mathbf{K}(W|T))$ be a transitional probability space and $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ be transitional random variables with joint Markov kernel:*

$$\mathbf{K}(X,Y,Z|T) := (\mathbf{X}(X|W,T) \otimes \mathbf{Y}(Y|W,T) \otimes \mathbf{Z}(Z|W,T)) \circ \mathbf{K}(W|T).$$

We put:

1. *$X \precsim Y$, for transitional random variables of the form $\mathbf{X} = \boldsymbol{\delta}(X|W,T)$, $\mathbf{Y} = \boldsymbol{\delta}(Y|W,T)$, if there exists a measurable map $\varphi : \mathcal{Y} \to \mathcal{X}$ such that $X = \varphi \circ Y$.*
2. *$\mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y}$ if there exists a measurable map $\varphi : \mathcal{Y} \to \mathcal{X}$ such that:*

$$\mathbf{K}(X,Y|T) = \boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y|T),$$

where $\mathbf{K}(Y|T)$ is the marginal of $\mathbf{K}(X,Y|T)$.

3. *$\mathbf{X} \precsim^*_{\mathbf{K}} \mathbf{Y}$ if there exists a Markov kernel $\mathbf{Q}(X|Y) : \mathcal{Y} \dashrightarrow \mathcal{X}$ such that:*

$$\mathbf{K}(X,Y|T) = \mathbf{Q}(X|Y) \otimes \mathbf{K}(Y|T).$$

Remark 2.20. *1. We have the implications:*

$$X \precsim Y \implies \mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y} \implies \mathbf{X} \precsim^*_{\mathbf{K}} \mathbf{Y}.$$

2. *The relation* $\lesssim_{\mathbf{K}}$ *will be the most crucial one in the following.*

3. *Note that for general* $\mathbf{X}$ *we do not even have reflexivity:* $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{X}$*. Indeed, e.g. the distribution of the product of two standard normal distributions does not look like the graph of a function.*

4. *We also do not have anti-symmetry, i.e. we can not conclude from:* $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{X}$ *that then* $\mathbf{X} = \mathbf{Y}$ *holds, since such variables might differ on some null-set.*

5. *We can fix the anti-symmetry by going over to almost-sure anti-symmetry, i.e. by defining:*

$$\mathbf{X} \approx_{\mathbf{K}} \mathbf{Y} :\iff \mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{X}.$$

The relation $\lesssim_{\mathbf{K}}$ satisfies the following rules, from which "product extension" is the most important one. The lack of this rule for $\lesssim^*_{\mathbf{K}}$ is the reason we will stick to $\lesssim_{\mathbf{K}}$ later on.

Theorem 2.21. *Let* $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ *be a transitional probability space and* $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, $\mathbf{U}$ *be transitional random variables. The relatation* $\lesssim_{\mathbf{K}}$ *satisfies the following rules:*

1. *Almost-sure anti-symmetry:* $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{X} \implies: \mathbf{X} \approx_{\mathbf{K}} \mathbf{Y}$.

2. *Transitivity:* $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \implies \mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z}$.

3. *Bottom element:* $\boldsymbol{\delta}_* \lesssim_{\mathbf{K}} \mathbf{X}$.

4. *Product stays bounded:* $(\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z}) \wedge (\mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z}) \implies \mathbf{X} \otimes \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z}$.

5. *Product extension:* $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \implies \mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \otimes \mathbf{Z}$.

6. *Product compatibility:* $(\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z}) \wedge (\mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{U}) \implies \mathbf{X} \otimes \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \otimes \mathbf{U}$.

Furthermore, the relation $\lesssim_{\mathbf{K}}$ *turns the sub-class of all transitional random variables on* $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ *of the form* $\mathbf{X} = \boldsymbol{\delta}(X|W,T)$*, where* $X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}$ *is a measurable map, and where* $\mathcal{X}$ *may vary, into a* join-semi-lattice *up to almost-sure anti-symmetry with join* $\otimes$ *and bottom element* $\boldsymbol{\delta}_*$*.*

Note that the mentioned sub-class might not be a set, but all the properties of join-semi-lattice can be proven. The proofs will be given in Appendix D and Theorem D.10.

2.7. Disintegration of Transition Probabilities

In this subsection we prove the existence of (regular) conditional Markov kernels. Since we will factorize a joint Markov kernel into a marginal part and a conditional part such procedures are also called disintegration. First, we will talk about the essential uniqueness of such factorizations and then existence. For proofs see Appendix C.

Definition 2.22 (Conditional Markov kernels)**.** *Consider a Markov kernel*

$$\mathbf{K}(X, Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{X} \times \mathcal{Y}$$

and its marginal $\mathbf{K}(Y|Z)$*. A* (regular) conditional Markov kernel of $\mathbf{K}(X, Y|Z)$ conditioned on Y given Z *is a Markov kernel:*

$$\mathbf{K}(X|Y, Z) : \mathcal{Y} \times \mathcal{Z} \dashrightarrow \mathcal{X}$$

such that:

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

Lemma 2.23 (Essential uniqueness of conditional Markov kernels)**.** *If we have Markov kernels:*

$$\mathbf{P}(X, Y, Z),\ \mathbf{Q}(X|Y, Z) : \mathcal{Y} \times \mathcal{Z} \dashrightarrow \mathcal{X}, \qquad \text{and} \qquad \mathbf{K}(Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{Y},$$

between any measurable spaces $\mathcal{X}$*,* $\mathcal{Y}$*,* $\mathcal{Z}$ *such that:*

$$\mathbf{P}(X|Y, Z) \otimes \mathbf{K}(Y|Z) = \mathbf{Q}(X|Y, Z) \otimes \mathbf{K}(Y|Z),$$

then for every $A \in \mathcal{B}_{\mathcal{X}}$ *the set:*

$$N_A := \{(y, z) \in \mathcal{Y} \times \mathcal{Z} \mid \mathbf{P}(X \in A|Y = y, Z = z) \neq \mathbf{Q}(X \in A|Y = y, Z = z)\}$$

is a measurable $\mathbf{K}(Y|Z)$*-null set.*
If, furthermore, $\mathcal{B}_{\mathcal{X}}$ *is countably generated then also* $N := \bigcup_{A \in \mathcal{B}_{\mathcal{X}}} N_A$ *is a measurable* $\mathbf{K}(Y|Z)$*-null set.*

Theorem 2.24 (Existence of conditional Markov kernels)**.** *Let* $\mathbf{K}(X, Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{X} \times \mathcal{Y}$ *be a Markov kernel. Furthermore, assume one of the following:*

1. $\mathcal{X}$ *is standard and* $\mathcal{Y}$ *is countably generated.*
2. $X \lesssim_{\mathbf{K}} (Y, Z)$*.*
3. $Y \lesssim_{\mathbf{K}} Z$*.*

Then $X \lesssim^*_{\mathbf{K}} (Y, Z)$*, i.e. there exists a (regular) conditional Markov kernel* $\mathbf{K}(X|Y, Z)$ *of* $\mathbf{K}(X, Y|Z)$*.*

Remark 2.25. *In Appendix C we will prove an analogous result with the weaker assumption of* $\mathcal{X}$ *being only an analytic or even just a universal measurable space. The price one has to pay is that the conditional Markov kernel will come with a weaker measurability property. It will only be measurable w.r.t. the* σ*-algebra generated by analytic subsets, universally measurable subsets, resp.*

The following corrollary is a well known result for probability measures:

Corollary 2.26 (Conditional probability distributions). *Let X and Y be random variables on probability space $(\mathcal{W}, \mathbf{P}(W))$ with standard measurable spaces $\mathcal{X}$, $\mathcal{Y}$, resp., as codomains. Then there always exist* regular[1] conditional probability distributions *$\mathbf{P}(X|Y)$ and $\mathbf{P}(Y|X)$ satisfying:*

$$\mathbf{P}(X,Y) = \mathbf{P}(X|Y) \otimes \mathbf{P}(Y), \qquad \mathbf{P}(X,Y) = \mathbf{P}(Y|X) \otimes \mathbf{P}(X).$$

Furthermore, these conditional probability distributions are essentially unique in the strong sense of Lemma 2.23.

3. Transitional Conditional Independence

3.1. Definition of Transitional Conditional Independence

In this section we will introduce the notion of transitional conditional independence for transitional random variables. It generalizes prior notions of (extended) conditional independence, see [Daw79a, Daw80, Daw01a, CD17a, RERS17, FM20], and it unifies stochastic conditional independence and some form of functional conditional independence. A comparison with other notions of (extended) conditional independence from the literature will be done in Appendix L.

Definition 3.1 (Transitional conditional independence). *Let $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ be a transition probability space with Markov kernel:*

$$\mathbf{K}(W|T) : \mathcal{T} \dashrightarrow \mathcal{W}.$$

Consider transitional random variables: $\mathbf{X} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$ and $\mathbf{Y} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Y}$ and $\mathbf{Z} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Z}$. The joint push-forward Markov kernel is then given by:

$$\mathbf{K}(X,Y,Z|T) := (\mathbf{X}(X|W,T) \otimes \mathbf{Y}(Y|W,T) \otimes \mathbf{Z}(Z|W,T)) \circ \mathbf{K}(W|T).$$

We say that $\mathbf{X}$ is (transitionally) independent of $\mathbf{Y}$ conditioned on $\mathbf{Z}$ w.r.t. $\mathbf{K} = \mathbf{K}(W|T)$, *in symbols:*

$$\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z},$$

if there exists *a Markov kernel:*

$$\mathbf{Q}(X|Z) : \mathcal{Z} \dashrightarrow \mathcal{X},$$

such that:

$$\mathbf{K}(X,Y,Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{K}(Y,Z|T), \tag{1}$$

[1] The word "regular" refers to the fact that the conditional probabilities are Markov kernels as defined above.

where $\mathbf{K}(Y, Z|T)$ *is the marginal of* $\mathbf{K}(X, Y, Z|T)$.
We use the following notations for the following special case:

$$\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \qquad :\Longleftrightarrow \qquad \mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \boldsymbol{\delta}_*.$$

For transitional random variables of the forms $\mathbf{X} = \boldsymbol{\delta}(X|W,T)$, $\mathbf{Y} = \boldsymbol{\delta}(Y|W,T)$, $\mathbf{Z} = \boldsymbol{\delta}(Z|W,T)$, *where* $X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}$, $Y : \mathcal{W} \times \mathcal{T} \to \mathcal{Y}$, $Z : \mathcal{W} \times \mathcal{T} \to \mathcal{Z}$, *etc., are measurable maps, we might also just write* X, Y, Z, *instead of* $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, *in those relations* $\perp\!\!\!\perp$. *E.g. we would write:*

$$X \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} Y \,|\, Z \qquad :\Longleftrightarrow \qquad \boldsymbol{\delta}(X|W,T) \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \boldsymbol{\delta}(Y|W,T) \,|\, \boldsymbol{\delta}(Z|W,T).$$

Several remarks are in order.

Remark 3.2. *If a candidate* $\mathbf{Q}(X|Z)$ *is found then for the Equation 1 to hold it is sufficient to check that for all* $t \in \mathcal{T}$ *and all measurable* $A \subseteq \mathcal{X}$, $B \subseteq \mathcal{Y}$, $C \subseteq \mathcal{Z}$ *one has that:*

$$\mathbf{K}(X \in A, Y \in B, Z \in C|T = t) = \int_C \int_B \mathbf{Q}(X \in A|Z = z)\, \mathbf{K}(Y \in dy, Z \in dz|T = t).$$

Remark 3.3 (Essential uniqueness)**.** *The Markov kernel* $\mathbf{Q}(X|Z)$ *appearing in the conditional independence* $\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}(W|T)} \mathbf{Y} \,|\, \mathbf{Z}$ *in Definition 3.1 is then a version of a conditional Markov kernel* $\mathbf{K}(X|Y, Z, T)$ *and is thus essentially unique (up to* $\mathbf{K}(Z|T)$*-null set) in the sense of Lemma 2.23.*

Notation 3.4. *The Markov kernel* $\mathbf{Q}(X|Z)$ *appearing in the conditional independence* $\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}(W|T)} \mathbf{Y} \,|\, \mathbf{Z}$ *is essentially unique as remarked in 3.3 and we can suggestively write it as:*

$$\mathbf{K}(X|\not{Y}, Z, \not{T}) := \mathbf{K}(X|\not{T}, \not{Y}, Z) := \mathbf{Q}(X|Z),$$

or similarly with crossed variables in different order. So we have in case of $\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}(W|T)} \mathbf{Y} \,|\, \mathbf{Z}$:

$$\mathbf{K}(X, Y, Z|T) = \mathbf{K}(X|\not{Y}, Z, \not{T}) \otimes \mathbf{K}(Y, Z|T).$$

This notation indicates that $\mathbf{K}(X|\not{Y}, Z, \not{T})$ *is a version of the conditional Markov kernel* $\mathbf{K}(X|Y, Z, T)$, *but does not (directly) depend on the arguments of* Y *and* T.

Remark 3.5 (Conditional independence includes conditional independence from **T**)**.** *It is easy to see from the Definition 3.1 that we have the equivalence:*

$$\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \quad \Longleftrightarrow \quad \mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{T} \otimes \mathbf{Y} \,|\, \mathbf{Z}.$$

The heuristic of why we automatically declare independence from T *as well is that we do* not *require* Z *to be in some sense "orthogonal" to or "functionally independent" of* T, *like other notion of conditional independence require.* Z *can be a direct function of* T. *In this way, i.e. by carefully exploiting the interplay between the second and third argument of* $\perp\!\!\!\perp_{\mathbf{K}}$, *we can re-introduce dependence on* T *through* Z. *This is what makes the (asymmetric) notion of transitional conditional independence so flexible.*

Remark 3.6 (Existence of conditional Markov kernels expressed as conditional independence)**.** *Let* $\mathbf{X}$, $\mathbf{Y}$ *be transitional random variables on transition probability space* $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$*. Then we can express the existence of a conditional Markov kernel* $\mathbf{K}(X|Y,T)$ *of* $\mathbf{K}(X,Y|T)$ *equivalently in one of the following equivalent statements:*

$$\mathbf{X} \precsim^*_{\mathbf{K}} \mathbf{Y} \otimes \mathbf{T} \qquad \iff \qquad \mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \boldsymbol{\delta}_* \,|\, \mathbf{Y} \otimes \mathbf{T} \qquad \iff \qquad \mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{T} \,|\, \mathbf{Y} \otimes \mathbf{T}.$$

Note that for standard measurable space $\mathcal{X}$ *and countably generated* $\mathcal{Y}$ *the above statements always hold by Theorem 2.24.*

3.2. Transitional Conditional Independence for Random Variables

Remark 3.7 (Transitional conditional independence for random variables)**.** *If we translate* transitional conditional independence *to random variables* X, Y, Z *on a probability space* $(\mathcal{W}, \mathbf{P}(W))$*, i.e. taking* $\mathcal{T} = \{*\}$*, then we arrive at:*

$$X \underset{\mathbf{P}(W)}{\perp\!\!\!\perp} Y \,|\, Z \qquad \iff \qquad \exists \mathbf{Q}(X|Z) : \mathbf{P}(X,Y,Z) = \mathbf{Q}(X|Z) \otimes \mathbf{P}(Y,Z). \tag{2}$$

Such a $\mathbf{Q}(X|Z)$ *would then clearly be a regular version of both,* $\mathbf{P}(X|Y,Z)$ *and* $\mathbf{P}(X|Z)$*. It thus directly implies what we will call* weak conditional independence*:*

$$X \overset{\omega}{\underset{\mathbf{P}(W)}{\perp\!\!\!\perp}} Y \,|\, Z \quad :\iff \quad \forall A \in \mathcal{B}_{\mathcal{X}} : \mathbb{E}[\mathbb{1}_A(X)|Y,Z] = \mathbb{E}[\mathbb{1}_A(X)|Z] \quad \mathbf{P}(W)\text{-}a.s., \tag{3}$$

which makes use of the conditional expectations for each A*, which exist for all measurable spaces, in contrast to regular conditional probability distributions. Both notions of conditional independence can be defined for all measurable spaces. Transitional conditional independence incorporates the existence of such a* $\mathbf{P}(X|Z)$ *and a factorization of the joint directly into its definition. Certainly, if a regular version of* $\mathbf{P}(X|Z)$ *does not even exist the variables are declared (transitionally) conditionally dependent. But in case* $\mathbf{P}(X|Z)$ *exists, e.g. for standard measurable* $\mathcal{X}$ *and* $\mathcal{Z}$ *by Theorem 2.24, then both notions of conditional independence are equivalent. So the choice of which notion to pick depends on how much meaning one finds in the extistence of such a regular version* $\mathbf{P}(X|Z)$ *and a factorization. In the applications to causal graphical models, where one wants to connect and work with many different subsystems, the existence of such conditional Markov kernels is crucial, because otherwise those subsystems might not even be well-defined.*

Even though, one might argue that asking to check for the existence of a regular versions of $\mathbf{P}(X|Z)$ *seems like an unnessary burden, from the point on we prove how the existence of such Markov kernels can be inherited through the (asymmetric)* separoid rules*, see Theorem 3.11, or can be guaranteed just through graphical criteria, see the* global Markov property *in Theorem 6.3, it will turn out to be very usuful to get such regular conditional Markov kernels (almost) for free.*

3.3. Transitional Conditional Independence for Deterministic Variables

We now demonstrate how transitional conditional independence behaves when applied to the other corner case of deterministic functions that contain no stochasticity.

Theorem 3.8 (Transitional conditional independence for deterministic variables)**.** *Let $F : \mathcal{T} \to \mathcal{F}$ and $H : \mathcal{T} \to \mathcal{H}$ be measurable maps and $\mathcal{F}$ a standard measurable space. We now consider them as (deterministic) transitional random variables on the transition probability space $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$. Let $Y : \mathcal{W} \times \mathcal{T} \to \mathcal{Y}$ be another transitional random variable.*
Then the following statements are equivalent:

1. $F \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} Y \,|\, H$.
2. *There exists a measurable function $\varphi : \mathcal{H} \to \mathcal{F}$ such that $F = \varphi \circ H$.*

Remark 3.9. 1. *Note that the second statement is independent of Y.*

2. *The first direction can we weakened by only assuming $\mathcal{B}_{\mathcal{F}}$ to separate the points of $\mathcal{F}$ to get a map $\varphi : H(\mathcal{T}) \to \mathcal{F}$ such that $F = \varphi \circ H$.*
3. *Theorem 3.8 shows how transitional conditional independence can express certain functional conditional (in)dependences. It also shows its (restricted) relation to variation conditional independence, see [CD17a, CD17b].*
4. *The full equivalence in Theorem 3.8 for standard $\mathcal{F}$ needs Kuratowski's extension theorem for standard measurable spaces (see [Kec95] 12.2). Versions for analytic and universal measurable spaces can be found in Appendix B.8. The proof of Theorem 3.8 can be found in the Appendix L.3 in Theorem L.7.*

Example 3.10. *If, for example, $\mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2$ and $T_i : \mathcal{W} \times \mathcal{T}_1 \times \mathcal{T}_2 \to \mathcal{T}_i$ the canonical projection onto $\mathcal{T}_i$, then F is a function in two variables (t_1, t_2). We then have:*

$$F \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} T_1 \,|\, T_2,$$

if and only if F - as a function - is only dependent on the argument t_2 (and not on t_1).

3.4. Separoid Rules for Transitional Conditional Independence

In the following we will list all the left and right versions of the separoid rules (see [Daw01a] for the symmetric versions or Appendix A) that hold for transitional conditional independence. Note that almost all of these work for all measurable spaces. Some of the rules, especially Left Weak Union, require the codomains of the transitional random variables to be standard or countably generated. This is required for the existence of a (regular) conditional Markov kernel, see Theorem 2.24. For those rules one can weaken the assumptions to analytic or even universal measurable spaces if one is willing to accept Markov kernels that are only universally measurable, see Theorem C.9 and Theorem C.10 in Appendix C.3.

Formally we will show that the class of transitional random variables with standard (or universal) measurable spaces as codomains together with transitional conditional independence will form what we will call a T-$*$-*separoid*, see Appendix A Definition A.3, an asymmetric notion of separoid. Note that these rules could not been proven in this amplitude for other versions of extendend conditional independence, see [CD17a].

The proofs for these separoid rules for transitional random variables will be given in Appendix E.

Theorem 3.11 (Separoid rules for transitional conditional independence)**.** *Consider a transition probability space* $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ *and transitional random variables* $\mathbf{X} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$ *and* $\mathbf{Y} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Y}$ *and* $\mathbf{Z} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Z}$ *and* $\mathbf{U} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{U}$. *Then the ternary relation* $\perp\!\!\!\perp = \perp\!\!\!\perp_{\mathbf{K}(W|T)}$ *satisfies the following rules:*

a) *Left Redundancy E.3:*

$$\mathbf{X} \precsim_{\mathbf{K}} \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}.$$

b) $\mathbf{T}$-*Restricted Right Redundancy E.4 (for $\mathcal{X}$ standard, $\mathcal{Z}$ countably generated)*[2]*:*

$$\mathbf{X} \perp\!\!\!\perp \boldsymbol{\delta}_* | \mathbf{Z} \otimes \mathbf{T} \text{ always holds.}$$

c) *Left Decomposition E.5:*

$$\mathbf{X} \otimes \mathbf{U} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z} \implies \mathbf{U} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}.$$

d) *Right Decomposition E.6:*

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{U} \,|\, \mathbf{Z}.$$

e) $\mathbf{T}$-*Inverted Right Decomposition E.7:*

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{T} \otimes \mathbf{Y} \,|\, \mathbf{Z}.$$

f) *Left Weak Union E.8 (for $\mathcal{X}$ standard, $\mathcal{U}$ countably generated)*[2]*:*

$$\mathbf{X} \otimes \mathbf{U} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}.$$

g) *Right Weak Union E.11:*

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}.$$

[2](Only) T-Restricted Right Redundancy, Left Weak Union and T-Restricted Symmetry need the existence of conditional Markov kernels. That is the reason we assume standard and countably generated measurable spaces there. If one is only interested in universal measurability one can instead assume the weaker assumption of universal and universally countably generated measurable spaces, resp.

h) *Left Contraction* E.12:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \wedge (\mathbf{U} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}) \implies \mathbf{X} \otimes \mathbf{U} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}.$$

i) *Right Contraction* E.13:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \wedge (\mathbf{X} \perp\!\!\!\perp \mathbf{U} \,|\, \mathbf{Z}) \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z}.$$

j) *Right Cross Contraction* E.14:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \wedge (\mathbf{U} \perp\!\!\!\perp \mathbf{X} \,|\, \mathbf{Z}) \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z}.$$

k) *Flipped Left Cross Contraction* E.15:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \wedge (\mathbf{Y} \perp\!\!\!\perp \mathbf{U} \,|\, \mathbf{Z}) \implies \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \otimes \mathbf{U} \,|\, \mathbf{Z}.$$

Remark 3.12. *In particular, we have the equivalence:*

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z}) \quad \iff \quad (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \quad \wedge \quad (\mathbf{X} \perp\!\!\!\perp \mathbf{U} \,|\, \mathbf{Z}),$$

and for standard measurable spaces:

$$(\mathbf{X} \otimes \mathbf{U} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}) \quad \iff \quad (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \quad \wedge \quad (\mathbf{U} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}).$$

Corollary 3.13 (Symmetry)**.** *Let the setting be like in Theorem 3.11. We then have:*

l) *Restricted Symmetry* E.19:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}) \wedge (\mathbf{Y} \perp\!\!\!\perp \boldsymbol{\delta}_* \,|\, \mathbf{Z}) \implies \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \,|\, \mathbf{Z}.$$

m) **T**-*Restricted Symmetry* E.20 *(for $\mathcal{Y}$ standard, $\mathcal{Z}$ countably generated)*[2]:

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z} \otimes \mathbf{T} \implies \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \,|\, \mathbf{Z} \otimes \mathbf{T}.$$

n) *Symmetry* E.21 *(for $\mathcal{Y}$ standard, $\mathcal{Z}$ countably generated, $\mathcal{T} = \{*\}$)*[2]:

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z} \implies \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \,|\, \mathbf{Z}.$$

Corollary 3.14. *Let the setting be like in Theorem 3.11. We then have:*

o) *Inverted Left Decomposition* E.22:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}) \wedge (\mathbf{U} \precsim_{\mathbf{K}} \mathbf{X} \otimes \mathbf{Z}) \implies \mathbf{X} \otimes \mathbf{U} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}.$$

p) **T**-*Extended Inverted Right Decomposition* E.23:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}) \wedge (\mathbf{U} \precsim_{\mathbf{K}} \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{Z}) \implies \mathbf{X} \perp\!\!\!\perp \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z}.$$

q) *Equivalent Exchange* E.24:

$$(\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{Z}) \wedge (\mathbf{Z} \approx_{\mathbf{K}} \mathbf{Z}') \implies \mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{Z}'.$$

r) *Full Equivalent Exchange* E.25*: If* $(\mathbf{X}' \approx_{\mathbf{K}} \mathbf{X}) \wedge (\mathbf{Y}' \approx_{\mathbf{K}} \mathbf{Y}) \wedge (\mathbf{Z}' \approx_{\mathbf{K}} \mathbf{Z})$ *then:*

$$\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{Z} \qquad \iff \qquad \mathbf{X}' \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y}' \,|\, \mathbf{Z}'.$$

4. Applications to Statistical Theory

In the following we will collect some illustrative applications of transitional conditional independence.

4.1. Ancillarity, Sufficiency, Adequacy

In this subsection we want to relate the concepts of acillarity, sufficiency and adequacy, see [Fis22,Fis25,Bas59,Bas64,Daw75,GRF10], to transitional conditional independence.

Example 4.1 (Certain statistics expressed as conditional independence)**.** *Let* $\mathbf{P}(W|\Theta)$ *be a statistical model, considered as a Markov kernel* $\mathcal{F} \dashrightarrow \mathcal{W}$. *Let* X *and* Y *be two transitional random variables w.r.t.* $\mathbf{P}(W|\Theta)$. *A* statistic *of* X *is a measurable map* $S : \mathcal{X} \to \mathcal{S}$, *which we consider as the transitional random variable* $S \precsim X$ *given via:*

$$S : \mathcal{W} \times \mathcal{F} \to \mathcal{S}, \quad (w, \theta) \mapsto S(X(w, \theta)).$$

1. Ancillarity. *S is an* ancillary statistic *of* X *w.r.t.* Θ *if and only if:*

$$S \underset{\mathbf{P}(W|\Theta)}{\perp\!\!\!\perp} \Theta.$$

   This means that every parameter $\Theta = \theta$ *induces the same distribution for* S*:*

$$\mathbf{P}(S|\Theta = \theta) = \mathbf{P}(S|\not{\emptyset}).$$

2. Sufficiency. *S is a* sufficient statistic *of* X *w.r.t.* Θ *if and only if:*

$$X \underset{\mathbf{P}(W|\Theta)}{\perp\!\!\!\perp} \Theta \,|\, S.$$

   This means that there is a Markov kernel $\mathbf{P}(X|S, \not{\Theta})$, *not dependent on* Θ, *such that:*

$$\mathbf{P}(X, S|\Theta) = \mathbf{P}(X|S, \not{\Theta}) \otimes \mathbf{P}(S|\Theta).$$

   So X *only "interacts" with the parameters* Θ *through* S.

3. Adequacy. *S is an* adequate statistic *of* X *for* Y *w.r.t.* Θ *if and only if:*

$$X \underset{\mathbf{P}(W|\Theta)}{\perp\!\!\!\perp} \Theta, Y \,|\, S.$$

   This means we have a factorization:

$$\mathbf{P}(X, Y, S|\Theta) = \mathbf{P}(X|\not{Y}, S, \not{\Theta}) \otimes \mathbf{P}(Y, S|\Theta),$$

   for some Markov kernel $\mathbf{P}(X|\not{Y}, S, \not{\Theta})$, *only dependent on* S. *This means that all information of* X *about the (parameters and/or) labels* Y *are fully captured already by* S.

Now we want to show that the classical Fisher-Neyman factorization criterion for sufficiency (see [Fis22, Ney36, HS49, Bur61]) is in line with the reformulation of sufficiency as a transitional conditional independence. The proof is given in Theorem F.1 in Appendix F.

Theorem 4.2 (Fisher-Neyman)**.** *Consider a statistical model, witten as a Markov kernel, $\mathbf{P}(X|\Theta)$. Assume that we can write it in the following form:*

$$\mathbf{P}(X \in A|\Theta = \theta) = \int_A h(x) \cdot g(S(x); \theta)\, d\boldsymbol{\mu}(x),$$

where X takes values in a standard measurable space $\mathcal{X}$, $S : \mathcal{X} \to \mathcal{S}$ is measurable, $\mathcal{S}$ countably generated, $\boldsymbol{\mu}$ is a measure on $\mathcal{X}$, $g, h \geqslant 0$ are measurable and h is $\boldsymbol{\mu}$-integrable (e.g. exponential families are of such form).
Then S is a sufficient statistic of X for $\mathbf{P}(X|\Theta)$: $\quad X \underset{\mathbf{P}(X|\Theta)}{\perp\!\!\!\perp} \Theta \,|\, S$.

This is one direction of the Fisher-Neyman factorization theorem for sufficiency. Our definition of transitional conditional independence generalizes the factorization theorem to Markov kernels (per definition) without the necessity of densities and/or reference measures.

4.2. Invariant Reductions

Example 4.3 (Invariant reduction)**.** *Let $\mathbf{P}(X|\Theta)$ be a statistical Model, given as a Markov kernel. Assume that we are only interested in a certain quantity of the parameters $\Gamma = \Gamma(\Theta)$, considered as a measurable function in Θ. For the estimation of Γ we then might only need parts of the information encoded in the data X. An* invariant reduction *of $\mathbf{P}(X|\Theta)$* w.r.t. *Γ, see [HWG65], is then a measurable function $U : \mathcal{X} \to \mathcal{U}$ such that $\mathbf{P}(U(X)|\Theta)$ only depends on Γ. We can rephrase this as the transitional conditional independence:*

$$U \underset{\mathbf{P}(X|\Theta)}{\perp\!\!\!\perp} \Theta \,|\, \Gamma,$$

and the occuring Markov kernel $\mathbf{P}(U|\Gamma, \not{\Theta})$ would give the correct model to further work with.

4.3. Reparameterizing Transitional Random Variables

We want to generalize two somewhat related folklore results from random variables to transitional random variables:

1.) Since the paper [Dar53] it was developed that for a real-values random variable X that has a continuous cumulative distribution function F and quantile function $R = F^{-1}$ that $E := F(X)$ is uniformly distributed on $[0, 1]$ and $R(E) = X$ a.s.

2.) It is known that for random variables X and Z with a well-behaved joint distribution $\mathbf{P}(X, Z)$ there exists a random variable E that is independent of Z and a measurable map g such that $X = g(E, Z)$ a.s.

To establish such results for transitional random variables we will use the following constructions.

Definition 4.4. *Let X be a transitional random variable with values in $\mathcal{X} = \bar{\mathbb{R}} = [-\infty, +\infty]$ on a transition probability space $(\mathcal{W} \times \mathcal{Z}, \mathbf{K}(W|Z))$. We then define the* interpolated transitional cumulative distribution function (itcdf) *of X as:*

$$F(x; u|z) := \mathbf{K}(X < x|Z = z) + u \cdot \mathbf{K}(X = x|Z = z),$$

with $u \in [0, 1]$, and the transitional quantile function (tqf) *of X as:*

$$R(e|z) := \inf \left\{ \tilde{x} \in \bar{\mathbb{R}} \,|\, F(\tilde{x}; 1|z) \geqslant e \right\},$$

for $e \in [0, 1]$.

Theorem 4.5. *Let $(\mathcal{W} \times \mathcal{Z}, \mathbf{K}(W|Z))$ be any transition probability space and X be a transitional random variable with values in a standard measurable space $\mathcal{X}$ and $\iota : \mathcal{X} \hookrightarrow \bar{\mathbb{R}}$ a fixed embedding onto a Borel subset of $\bar{\mathbb{R}}$ (i.e. w.l.o.g. $\mathcal{X} = \bar{\mathbb{R}}$). Let $\mathbf{K}(U)$ be the uniform distribution on $\mathcal{U} := [0, 1]$. We put $\bar{\mathcal{W}} = \mathcal{U} \times \mathcal{W}$, $\bar{W} = (U, W)$ and $\mathbf{K}(\bar{W}|Z) = \mathbf{K}(U) \otimes \mathbf{K}(W|Z)$. We then consider the transitional random variables X, U, Z, E on the transition probability space $\left(\bar{\mathcal{W}} \times \mathcal{Z}, \mathbf{K}(\bar{W}|Z)\right)$ where:*

$$E := F(X; U|Z) : \bar{\mathcal{W}} \times \mathcal{Z} \to \mathcal{E} := [0, 1],$$

and F is the itcdf of X from 4.4. Then we have the transitional independence:

$$E \underset{\mathbf{K}(\bar{W}|Z)}{\perp\!\!\!\perp} Z, \qquad \text{with} \qquad \mathbf{K}(E|\cancel{Z}) \text{ the uniform distribution on } \mathcal{E} = [0, 1],$$

and:

$$X = R(E|Z) \quad \mathbf{K}(\bar{W}|Z)\text{-a.s.},$$

where R is the tqf of X from 4.4.

The proof of this theorem can be found in the Appendix G in Theorem G.4.

4.4. Propensity Score

For a random variable X and binary random variable $Y \in \{0, 1\}$ the *propensity score* is $e(x) := \mathbf{P}(Y = 1|X = x)$. It is the "smallest" statistic of X such that $Y \perp\!\!\!\perp X \,|\, e(X)$ (see [RR83]). This is one of the core concepts of causal inference using the potential outcome formulation. We now claim that the above can be generalized to arbitrary (non-binary) Y with Markov kernel $\mathbf{P}(Y|X)$, even when no distribution for X is specified.

Theorem 4.6 (Propensity score)**.** *Let $\mathbf{P}(Y|X)$ be a Markov kernel. We define the* propensity *of $x \in \mathcal{X}$ w.r.t. $\mathbf{P}(Y|X)$ as:*

$$E(x) := \mathbf{P}(Y|X = x) \;\in\; \mathcal{P}(\mathcal{Y}).$$

Then $E : \mathcal{X} \to \mathcal{E} := \mathcal{P}(\mathcal{Y})$ *is measurable,* $E \precsim X$ *and we have:*

$$Y \underset{\mathbf{P}(Y|X)}{\perp\!\!\!\perp} X \,|\, E.$$

Furthermore, if $S : \mathcal{X} \to \mathcal{S}$ *is another measurable map* ($S \precsim X$) *with:*

$$Y \underset{\mathbf{P}(Y|X)}{\perp\!\!\!\perp} X \,|\, S,$$

then:

$$E \precsim S \precsim X.$$

So E *is in this sense the smallest statistic of* X *such that the above conditional independence holds.*

The proof is given in Theorem F.2 in Appendix F.

4.5. A Weak Likelihood Principle

The same Theorem F.2 applied to a statistical model $\mathbf{P}(X|\Theta)$ can give a weak form of the *likelihood principle*. For the history of the likehood principle and discussions see [SBC+62, Fis22, Hac65, Edw74, Edw92, Roy97, Bir62, Jay03, May14, Eva13, Gan15].

Theorem 4.7 (A weak likelihood principle)**.** *Let* $\mathbf{P}(X|\Theta)$ *be a statistical model. Define the* likelihood function *as:* $L(\theta) := \mathbf{P}(X|\Theta = \theta) \in \mathcal{P}(\mathcal{X})$*. We then have the transitional conditional independence:*

$$X \underset{\mathbf{P}(X|\Theta)}{\perp\!\!\!\perp} \Theta \,|\, L.$$

Furthermore, any other measurable map S *of the parameters* Θ *with* $X \perp\!\!\!\perp_{\mathbf{P}(X|\Theta)} \Theta \,|\, S$ *satisfies:*

$$L \precsim S \precsim \Theta.$$

In this sense the likelihood captures all information of the parameters Θ about the data X and it does so most efficiently.

4.6. Bayesian Statistics

Let $\mathbf{P}(X|\Theta)$ be a statistical model (for simplicity between standard measurable spaces) and $\mathbf{P}(\Theta|\Pi)$ be a prior with hyperparameters $\Pi = \pi$. Then by the standard Bayesian setting we have a joint (transition) probability distribution:

$$\mathbf{P}(X, \Theta|\Pi) := \mathbf{P}(X|\Theta) \otimes \mathbf{P}(\Theta|\Pi).$$

A conditional Markov kernel gives us the posterior (transitional) probability distributions:

$$\mathbf{P}(\Theta|X, \Pi),$$

which is unique up to $\mathbf{P}(X|\Pi)$-null set. We now define the transitional random variable:

$$Z(x, \pi) := \mathbf{P}(\Theta|X = x, \Pi = \pi),$$

which gives us a joint (transition) probability distribution: $\mathbf{P}(X, \Theta, Z|\Pi)$.

The following result then formalizes the basic idea that the posterior (given via Z) most efficiently incorporates all information from the data (X) about the state of the parameters (Θ) as soon as a prior (Π) is specified.

Theorem 4.8 (Bayesian statistics)**.** *With the above notations we have the conditional independence:*

$$\Theta \underset{\mathbf{P}(X,\Theta|\Pi)}{\perp\!\!\!\perp} X \mid Z.$$

Furthermore, if S is another deterministic measurable function in (X, Π) such that:

$$\Theta \underset{\mathbf{P}(X,\Theta|\Pi)}{\perp\!\!\!\perp} X \mid S,$$

then:

$$Z \precsim S \quad \mathbf{P}(X|\Pi)\text{-}a.s.$$

The proof of this theorem is given in Theorem F.3 in Appendix F.

Theorem 4.9 (A weak likelihood principle for Bayesian statistics)**.** *We also have the transitional conditional independence with $L(\theta) := \mathbf{P}(X|\Theta = \theta)$:*

$$X \underset{\mathbf{P}(X,\Theta|\Pi)}{\perp\!\!\!\perp} \Theta, \Pi \mid L.$$

Furthermore, if $\mathcal{X}$ is countably generated, then any other measurable map S in Θ with

$$X \underset{\mathbf{P}(X,\Theta|\Pi)}{\perp\!\!\!\perp} \Theta, \Pi \mid S$$

satisfies:

$$L \precsim S \precsim \Theta \quad \mathbf{P}(\Theta|\Pi)\text{-}a.s.$$

The proof of this theorem is given in Theorem F.4 in Appendix F.

5. Graph Theory

In this section we introduce a few graph theoretic notions that are required to apply transitional conditional independence to graphical models in the next section. Of importance, also on its own, is the notion of *σ-separation*, a generalization of d/m/m*-separation, see [Pea09, Ric03, FM17, FM18, FM20], to graphs that allow for cycles and input nodes. We define σ-separation in such a way that it forms a *J-$\emptyset$-separoid*, see Appendix A Definition A.3. The reason is that we want to match those separoid rules to the ones for transitional conditional independence in Theorem 3.11.

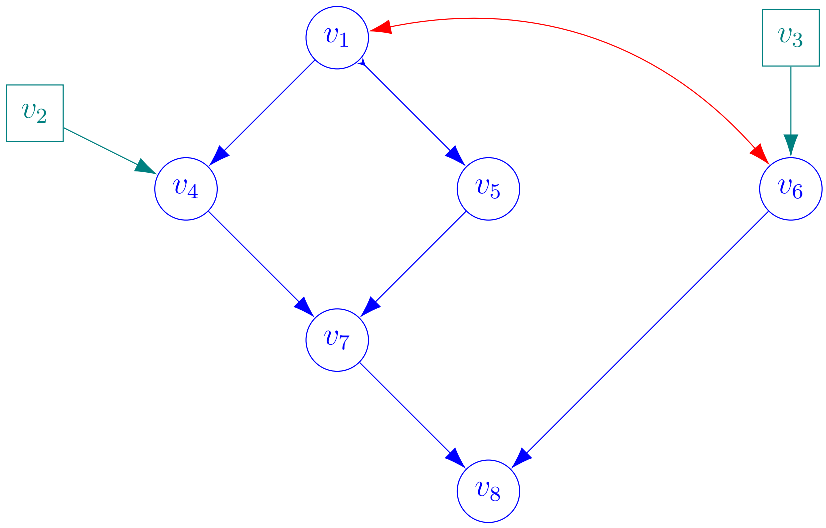


Figure 2: Conditional Acyclic Directed Mixed Graph (CADMG).

5.1. Conditional Directed Mixed Graphs (CDMGs)

Definition 5.1 (Conditional directed mixed graphs (CDMGs))**.** *A* conditional directed mixed graph (CDMG) $\mathbf{G} = (J, V, E, L)$ *consists of two (disjoint) sets of vertices/nodes: the set of* input nodes J*, the set of* output nodes V*, and two (disjoint) sets of edges:*

1. $E \subseteq \{w \rightarrowtail v \,|\, w \in J \cup V, v \in V\}$*, the set of* directed edges,
2. $L \subseteq \{v_1 \leftrightarrow v_2 \,|\, v_1, v_2 \in V, v_1 \neq v_2\}$*, the set of* bi-directed edges,

   with: $v_1 \leftrightarrow v_2 \in L \implies v_2 \leftrightarrow v_1 \in L$.

So - per definition - there won't be any arrow heads pointing to input nodes $j \in J$*.*

We drop "mixed" from the definition if $L = \emptyset$ *(CDG), and "conditional" if* $J = \emptyset$ *(DMG) or both (DG).*

Notation 5.2. *For a CDMG we will suggestively write:* $\mathbf{G}(V|\,\mathrm{do}(J)) := (J, V, E, L) = \mathbf{G}$*, where the sets of edges* E *and* L *are implicit. We will also write:* $v \in \mathbf{G}$ *to mean* $v \in J \cup V$*;* $v_1 \leftarrowtail v_2 \in \mathbf{G}$ *to mean* $v_2 \rightarrowtail v_1 \in E$*;* $v_1 \leftrightarrow v_2 \in \mathbf{G}$ *to mean* $v_1 \leftrightarrow v_2 \in L$*;* $v_1 \mathrel{*\!\!\rightarrow} v_2 \in \mathbf{G}$ *to mean* $v_1 \leftrightarrow v_2 \in \mathbf{G} \,\vee\, v_1 \rightarrowtail v_2 \in \mathbf{G}$*; etc.*

Definition 5.3 (Walks)**.** *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG and* $v, w \in \mathbf{G}$*.*

1. *A* walk *from* v *to* w *in* $\mathbf{G}$ *is a finite sequence of nodes and edges*

$$v = v_0 \mathrel{*\!\!-\!\!*} v_1 \mathrel{*\!\!-\!\!*} \cdots \mathrel{*\!\!-\!\!*} v_{n-1} \mathrel{*\!\!-\!\!*} v_n = w$$

   in $\mathbf{G}$ *for some* $n \geqslant 0$*, i.e. such that for every* $k = 1, \ldots, n$ *we have that* $v_{k-1} \mathrel{*\!\!-\!\!*} v_k \in \mathbf{G}$*, and with* $v_0 = v$ *and* $v_n = w$*.*

In the definition of walk the appearance of the same nodes several times is allowed. Also the trivial walk *consisting of a single node* $v_0 \in \mathbf{G}$ *is allowed as well (if* $v = w$*).*

2. *A* directed walk *from* v *to* w *in* $\mathbf{G}$ *is of the form:*

$$v = v_0 \rightarrowtail v_1 \rightarrowtail \cdots \rightarrowtail v_{n-1} \rightarrowtail v_n = w,$$

for some $n \geqslant 0$*, where all arrow heads point in the direction of* w *and there are no arrow heads pointing back.*

Definition 5.4. *For a CDMG* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *and* $v \in \mathbf{G}$ *we define the sets:*

1. Parents*:* $\mathrm{Pa}^{\mathbf{G}}(v) := \{w \in \mathbf{G} \,|\, w \rightarrowtail v \in \mathbf{G}\}$,
2. Children*:* $\mathrm{Ch}^{\mathbf{G}}(v) := \{w \in \mathbf{G} \,|\, v \rightarrowtail w \in \mathbf{G}\}$,
3. Ancestors*:* $\mathrm{Anc}^{\mathbf{G}}(v) := \{w \in \mathbf{G} \,|\, \exists$ *directed walk in* $\mathbf{G} : w \rightarrowtail \cdots \rightarrowtail v\}$,
4. Descendents*:* $\mathrm{Desc}^{\mathbf{G}}(v) := \{w \in \mathbf{G} \,|\, \exists$ *directed walk in* $\mathbf{G} : v \rightarrowtail \cdots \rightarrowtail w\}$,
5. Strongly connected component*:* $\mathrm{Sc}^{\mathbf{G}}(v) := \mathrm{Anc}^{\mathbf{G}}(v) \cap \mathrm{Desc}^{\mathbf{G}}(v) \ni v$.

We extend these notions to sets $A \subseteq J \cup V$ *by taking the union, e.g.* $\mathrm{Anc}^{\mathbf{G}}(A) = \bigcup_{v \in A} \mathrm{Anc}^{\mathbf{G}}(v)$.

Definition 5.5 (Acyclicity)**.** *A CDMG* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *is called* acyclic *if every directed walk from any node* $v \in \mathbf{G}$ *to itself is trivial, i.e.* $\mathrm{Sc}^{\mathbf{G}}(v) = \{v\}$ *and* $\mathrm{Pa}^{\mathbf{G}}(v) \not\ni v$ *for all* $v \in \mathbf{G}$*.*

Definition 5.6 (Topological order)**.** *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG. A* topological order *of* $\mathbf{G}$ *is a total order* $<$ *of* $J \cup V$ *such that for all* $v, w \in \mathbf{G}$*:*

$$v \in \mathrm{Pa}^{\mathbf{G}}(w) \quad \Longrightarrow \quad v < w.$$

Remark 5.7. *A CDMG* $\mathbf{G}(V|\,\mathrm{do}(J))$ *is acyclic iff it has a topological order.*

5.2. Sigma-Separation in CDMGs

Here we will generalize the notion of σ-separation, a generalization of d/m/m*-separation, see [Pea09, Ric03, FM17, FM18, FM20], to graphs that allow for cycles and input nodes.

Definition 5.8 (σ-blocked walks)**.** *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG and* $C \subseteq J \cup V$ *a subset of nodes and* π *a walk in* $\mathbf{G}(V|\,\mathrm{do}(J))$*:* $\pi = (v_0 \ast\!\!-\!\!\ast \cdots \ast\!\!-\!\!\ast v_n)$.

1. *We say that the walk* π *is* C*-*σ-blocked *or* σ-blocked by C *if either:*
   a) $v_0 \in C$ *or* $v_n \in C$ *or:*
   b) *there are two adjacent edges in* π *of one of the following forms:*

$$\begin{array}{lllll} \textit{left chain:} & v_{k-1} \leftarrow v_k \leftarrow\!\!* \, v_{k+1} & \textit{with} & v_k \in C \wedge v_k \notin \mathrm{Sc}^{\mathbf{G}}(v_{k-1}), \\ \textit{right chain:} & v_{k-1} *\!\!\rightarrow v_k \rightarrow v_{k+1} & \textit{with} & v_k \in C \wedge v_k \notin \mathrm{Sc}^{\mathbf{G}}(v_{k+1}), \\ \textit{fork:} & v_{k-1} \leftarrow v_k \rightarrow v_{k+1} & \textit{with} & v_k \in C \wedge v_k \notin \mathrm{Sc}^{\mathbf{G}}(v_{k-1}) \cap \mathrm{Sc}^{\mathbf{G}}(v_{k+1}), \\ \textit{collider:} & v_{k-1} *\!\!\rightarrow v_k \leftarrow\!\!* \, v_{k+1} & \textit{with} & v_k \notin C. \end{array}$$

2. *We say that the walk* π *is* C-σ-open *if it is not* C-σ*-blocked.*

Definition 5.9 (σ-separation)**.** *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG and* $A, B, C \subseteq J \cup V$ *(not necessarily disjoint) subset of nodes. We then say that:*

1. A is σ-separated from B given C in $\mathbf{G}$, *in symbols:* $A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B\,|\,C$,

   if every walk from a node in A *to a node in* $J \cup B$[3] *is* σ*-blocked by* C*.*

2. *Otherwise we write:* $A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\not\perp}} B\,|\,C.$

3. *Special case:* $A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \quad :\Longleftrightarrow \quad A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B|\emptyset.$

Remark 5.10. *If* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *is acyclic then* $\mathrm{Sc}^{\mathbf{G}}(v) = \{v\}$ *and* $v \notin \mathrm{Pa}^{\mathbf{G}}(v)$ *for all* $v \in \mathbf{G}$ *and the additional conditions in "*σ*-blocked" for the non-colliders are of the form* $v_k \notin \mathrm{Sc}^{\mathbf{G}}(v_{k\pm1}) = \{v_{k\pm1}\}$*. These are then automatically satisfied, because the node* $v_k \in \mathrm{Pa}^{\mathbf{G}}(v_{k\pm1}) \not\ni v_{k\pm1}$ *for the relevant other node* $v_{k\pm1}$ *and thus* $v_k \neq v_{k\pm1}$*. So in the acyclic case the additional conditions involving* $\mathrm{Sc}^{\mathbf{G}}(v_{k\pm1})$ *can be dropped. This shows that in the acyclic case* σ*-separation is equivalent to d/m/m*-separation, see [Pea09, Ric03]. It turns out that in the non-acyclic case* σ*-separation is the better concept, see [FM17, FM18, FM20].*

5.3. Separoid Rules for Sigma-Separation

Here we collect the formal rules that σ-separation satisfies. Note that these rules match the rules of transitional conditional independence in Theorem 3.11. The proofs will be given in Appendix I. We formally show that the subsets of $J \cup V$ of a CDMG $\mathbf{G}(V|\,\mathrm{do}(J))$ together with the relations $=$, $\subseteq$, $\perp^{\sigma}$, operation $\cup$ and element $\emptyset$ form a J-$\emptyset$-separoid, see Appendix A Definition A.3.

Theorem 5.11 (Separoid rules for σ-separation)**.** *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG and* $A, B, C, D \subseteq J \cup V$ *subset of nodes. Then the ternary relation* $\perp = \perp^{\sigma}_{\mathbf{G}(V|\,\mathrm{do}(J))}$ *satisfy the following rules:*

a) *Left Redundancy I.1:*

   $A \subseteq C \implies A \perp B\,|\,C.$

[3] Note the inclusion of J here. This is done to get similar asymmetric separoid rules to transitional conditional independence in order to have a more "nice" looking global Markov property later on.

b) *J-Restricted Right Redundancy I.2:*

$A \perp \emptyset \,|\, C \cup J$ *always holds.*

c) *Left Decomposition I.3:*

$A \cup D \perp B \,|\, C \implies D \perp B \,|\, C.$

d) *Right Decomposition I.4:*

$A \perp B \cup D \,|\, C \implies A \perp D \,|\, C.$

e) *J-Inverted Right Decomposition I.5:*

$A \perp B \,|\, C \implies A \perp J \cup B \,|\, C.$

f) *Left Weak Union I.6:*

$A \cup D \perp B \,|\, C \implies A \perp B \,|\, D \cup C.$

g) *Right Weak Union I.7:*

$A \perp B \cup D \,|\, C \implies A \perp B \,|\, D \cup C.$

h) *Left Contraction I.8:*

$(A \perp B \,|\, D \cup C) \wedge (D \perp B \,|\, C) \implies A \cup D \perp B \,|\, C.$

i) *Right Contraction I.9:*

$(A \perp B \,|\, D \cup C) \wedge (A \perp D \,|\, C) \implies A \perp B \cup D \,|\, C.$

j) *Right Cross Contraction I.10:*

$(A \perp B \,|\, D \cup C) \wedge (D \perp A \,|\, C) \implies A \perp B \cup D \,|\, C.$

k) *Flipped Left Cross Contraction I.11:*

$(A \perp B \,|\, D \cup C) \wedge (B \perp D \,|\, C) \implies B \perp A \cup D \,|\, C.$

Remark 5.12. *In particular, we have the equivalences:*

$$(A \perp B \cup D \,|\, C) \iff (A \perp B \,|\, D \cup C) \wedge (A \perp D \,|\, C),$$
$$(A \cup D \perp B \,|\, C) \iff (A \perp B \,|\, D \cup C) \wedge (D \perp B \,|\, C).$$

Remark 5.13 (Symmetry)**.** *Let the assumptions be like in 5.11. We also have the following rules:*

l) *Restricted Symmetry I.16:*

$(A \perp B \,|\, C) \wedge (B \perp \emptyset \,|\, C) \implies B \perp A \,|\, C.$

m) *J-Restricted Symmetry I.17:*

$$A \perp B \mid C \cup J \implies B \perp A \mid C \cup J.$$

n) *Symmetry I.18: If $J = \emptyset$ then:*

$$A \perp B \mid C \implies B \perp A \mid C.$$

Lemma 5.14 (More separoid like rules)**.** *Let the assumptions be like in 5.11. We also have the following rules:*

o) *Left Composition I.12:*

$$(A \perp B \mid C) \wedge (D \perp B \mid C) \implies A \cup D \perp B \mid C.$$

p) *Right Composition I.13:*

$$(A \perp B \mid C) \wedge (A \perp D \mid C) \implies A \perp B \cup D \mid C.$$

q) *Left Intersection I.14: If $A \cap D = \emptyset$ then:*

$$(A \perp B \mid D \cup C) \wedge (D \perp B \mid A \cup C) \implies A \cup D \perp B \mid C.$$

r) *Right Intersection I.15: If $B \cap D = \emptyset$ then:*

$$(A \perp B \mid D \cup C) \wedge (A \perp D \mid B \cup C) \implies A \perp B \cup D \mid C.$$

s) *More Redundancies I.19:*

$$A \perp B \mid C \iff (A \setminus C) \perp (B \setminus C) \mid C \iff A \cup C \perp J \cup B \cup C \mid C.$$

6. Applications to Graphical Models

6.1. Causal Bayesian Networks

We now introduce a definition of causal Bayesian networks that allows for non-stochastic input variables.

Definition 6.1 (Causal Bayesian network)**.** *A* causal Bayesian network (CBN) $\mathbf{M}$ *consists of:*

1. *a (finite) conditional directed acyclic graph (CDAG):* $\mathbf{G} = \mathbf{G}(V \mid \mathrm{do}(J))$,
2. *input variables X_j, $j \in J$, and (stochastic) output variables X_v, $v \in V$,*
3. *a measurable space $\mathcal{X}_v$ for every $v \in J \dot\cup V$, where $\mathcal{X}_v$ is standard[4] if $v \in V$,*

[4] We could also work under the weaker assumption of universal measurable spaces if we would replace all mentionings of measurability with the weaker notion of universal measurability.

4. *a Markov kernel, suggestively written as:* $\mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\operatorname{Pa}^{\mathbf{G}}(v)}\right)\right)$*:*

$$\begin{aligned} \mathcal{X}_{\operatorname{Pa}^{\mathbf{G}}(v)} &\dashrightarrow \mathcal{X}_v, \\ (A, x_{\operatorname{Pa}^{\mathbf{G}}(v)}) &\mapsto \mathbf{P}_v\left(X_v \in A \middle| \operatorname{do}\left(X_{\operatorname{Pa}^{\mathbf{G}}(v)} = x_{\operatorname{Pa}^{\mathbf{G}}(v)}\right)\right), \end{aligned}$$

for every $v \in V$*, where we write for* $D \subseteq J \dot\cup V$*:*

$$\begin{aligned} \mathcal{X}_D &:= \prod_{v \in D} \mathcal{X}_v, & \mathcal{X}_\emptyset &:= \circledast = \{*\}, \\ X_D &:= (X_v)_{v \in D}, & X_\emptyset &:= *, \\ x_D &:= (x_v)_{v \in D}, & x_\emptyset &:= *. \end{aligned}$$

By abuse of notation, we denote the causal Bayesian network as:

$$\mathbf{M}(V|\operatorname{do}(J)) = \left(\mathbf{G}(V|\operatorname{do}(J)), \left(\mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\operatorname{Pa}^{\mathbf{G}}(v)}\right)\right)\right)_{v \in V}\right).$$

Definition 6.2. *Any CBN* $\mathbf{M} = \mathbf{M}(V|\operatorname{do}(J))$ *comes with its* joint Markov kernel*:*

$$\mathbf{P}(X_V|\operatorname{do}(X_J)) : \mathcal{X}_J \dashrightarrow \mathcal{X}_V,$$

given by:

$$\mathbf{P}(X_V|\operatorname{do}(X_J)) := \bigotimes_{v \in V}^{>} \mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\operatorname{Pa}^{\mathbf{G}}(v)}\right)\right),$$

where the product $\otimes^{>}$ *is taken in reverse order of a fixed topological order* $<$*, i.e. children appear only on the left of their parents in the product. Note that by Remark 2.8 about associativity and (restricted) commutativity of the product the joint Markov kernels does actually not depend on the topological order.*

6.2. Global Markov Property for Causal Bayesian Networks

We now turn to probably the most striking application of transitional conditional independence: the *global Markov property* for causal Bayesian networks that allow for (non-stochastic) input variables. The global Markov property relates the graphical structure $\mathbf{G}(V|\operatorname{do}(J))$ to transitional conditional independence relations between the corresponding transitional random variables X_A. So checking the graph for σ-separation relations will then automatically imply the existence of a regular conditional Markov kernel that does not depend on the specified variables, which can be stochastic or not.

This will be the first time the global Markov property will be proven in this generality of measure theoretic probability, in the presence of input variables and with such a strong notion of conditional independence. Direct generalizations to causal Bayesian networks or *structural causal models* that allow for latent confounders, cycles, selection bias and (non-stochastic) input variables are immidiate, see [FM17, FM18, FM20, Ric03, Eva16, Eva18, RERS17], etc., because all those cases can be reduced to a CBN in one or the other form and only the graphical separation criteria need to be adjusted. In Appendix K we will show how the global Markov property together with transitional conditional independence gives us a measure theoretically clean proof of the main rules of *do-calculus*, see [Pea09], in this generality.

The proof of the global Markov property follows similar arguments as used in [LDLL90, Ver93, Ric03, FM17, FM18, RERS17], namely chaining the separoid rules for conditional independence (see Theorem 3.11) and the ones for $d/m/\sigma$-separation (see Theorem 5.11) together in an inductive way. The main difference here is that we never rely on the Symmetry property but instead use the left and right versions of the separoid rules separately. Note, again, that the validity of those separoid rules in this vast generality is a non-trival result, see Theorem 3.11, and were only known for corner cases for other notions of extended conditional independence, rendering those less useful for the applications to graphical models. The full proof of the global Markov property is given in Appendix J.

Theorem 6.3 (Global Markov property for causal Bayesian networks)**.** *Consider a causal Bayesian network* $\mathbf{M}(V|\,\mathrm{do}(J))$ *with graph* $\mathbf{G}(V|\,\mathrm{do}(J))$ *and joint Markov kernel:* $\mathbf{P}(X_V|\,\mathrm{do}(X_J))$*. Then for all* $A, B, C \subseteq J \dot\cup V$ *(not-necessarily disjoint) we have the implication:*

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \qquad \Longrightarrow \qquad X_A \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C.$$

Recall that we have - per definition - an implicit dependence on J*,* X_J*, resp., in the second argument on each side.*

Remark 6.4. *The global Markov property says that already checking the graphical criterion* $A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C$ *is enough to get the existence of a Markov kernel, suggestively written as:* $\mathbf{P}\left(X_A|\cancel{X_B}, X_{C\cap V}, \mathrm{do}(X_{C\cap J}), \cancel{\mathrm{do}(X_J)}\right)$*, such that:*

$$\mathbf{P}(X_A, X_B, X_C|\,\mathrm{do}(X_J)) = \mathbf{P}\left(X_A|\cancel{X_B}, X_{C\cap V}, \mathrm{do}(X_{C\cap J}), \cancel{\mathrm{do}(X_J)}\right) \otimes \mathbf{P}(X_B, X_C|\,\mathrm{do}(X_J)).$$

The full power of the global Markov property is unleashed when applied to the causal Bayesian network that arises from augmentation with further soft intervention variables and hard interventions, etc. It will then automatically relate many of the marginal conditional interventional Markov kernels of extended and sub-CBNs to each other, see Appendix K for a glimpse of the possibilities.

Example 6.5. *Let* $A \subseteq \mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be an ancestral subset, i.e.* $A = \bigcup_{v\in A} \mathrm{Anc}^{\mathbf{G}}(v)$*. Then we have:* $(V \cap A) \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} (J \setminus A) \,|\, J \cap A$*, and thus, by the global Markov property, a Markov kernel* $\mathbf{P}(X_{V\cap A}|\,\mathrm{do}(X_{J\cap A}))$ *such that:*

$$\mathbf{P}(X_A|\,\mathrm{do}(X_J)) = \mathbf{P}(X_{V\cap A}|\,\mathrm{do}(X_{J\cap A})) \otimes \mathbf{P}(X_{J\cap A}|\,\mathrm{do}(X_J)).$$

So for ancestral subsets we can only work with input variables from $J \cap A$ *and ignore the ones from* $J\setminus A$*, which is in correspondence with our expectations about ancestral causal relations.*

7. Discussion

7.1. Extensions to Universal Measurable Spaces

Remark 7.1. *We presented the theory of the main paper is terms of general measurable spaces whereever it was possible. Some results needed more restrictive assumptions like standard measurable spaces. These restrictions can be weakend if one replaces standard with universal measurable spaces and measurability with the weaker notion of universal measurability. All theory has been developed for this more general setting in the Appendix B. If $\mathcal{X}$ is a measurable space we denote by $\mathcal{X}_\bullet = (\mathcal{X}, (\mathcal{B}_\mathcal{X})_\bullet)$ its* universal completion, *where $(\mathcal{B}_\mathcal{X})_\bullet$ is the intersection of all completions, see Definition B.2. A map $f : \mathcal{X} \to \mathcal{Y}$ is* universally measurable *if $f : \mathcal{X}_\bullet \to \mathcal{Y}$ is measurable. We then advocate the use of transition probability spaces $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$, with universally measurable $\mathbf{K}(W|T) : \mathcal{T}_\bullet \dashrightarrow \mathcal{W}$ and the following class of transitional random variables:*

$$\mathcal{C} := \{X : (\mathcal{W} \times \mathcal{T})_\bullet \to \mathcal{X} \text{ universally measurable} \,|\, \mathcal{X} \text{ universal measurable space}\},$$

where a universal measurable space, *see Corollary B.30, is - up to universal completion - countably generated, countably separated with perfect probability measures. The class $\mathcal{C}$ together with $\precsim_{\mathbf{K},\bullet}$*[5] *and $X \vee Y := (X, Y)$ forms a* join-semi-lattice *up to $\approx_{\mathbf{K},\bullet}$-anti-symmetry, see Appendix D, and together with $\perp\!\!\!\perp_{\mathbf{K},\bullet}$*[6] *a T-$*$-separoid, see Appendix E. Furthermore, all $\bullet$-regular conditional Markov kernels $\mathbf{K}(X|Y,Z) : (\mathcal{Y} \times \mathcal{Z})_\bullet \dashrightarrow \mathcal{X}$ exist for $\mathbf{K}(X,Y|Z)$, when $\mathcal{X}, \mathcal{Y}$ are universal measurable spaces, see Theorem C.10. This is our contribution to make the theory applicable in as much generality as possible.*

7.2. Comparison to Other Notions of Extended Conditional Independence

Remark 7.2. *There can at least 4 other notions of extended conditional independence be found in the literature, these are from [CD17a, RERS17, FM20] plus variation conditional independence. We elaborate in more details in Appendix L. In overview, it can be said that variation conditional independence, a non-probabilistic, set-theoretic notion of conditional independence, only shares formal similarities with the other notions of extended conditional independence. The correspondence would be to replace the space of probability measures $\mathcal{P}(\mathcal{X})$ with the power set $2^\mathcal{X}$, see Example L.6.*
All other 3 notions of (stochastic) extended conditional independence, see [CD17a, RERS17, FM20], are weaker than transitional conditional independence, i.e. whenever they were defined transitional conditional independence implies the other extended conditional independence.
In contrast to [CD17a] transitional conditional independence satisfies all (asymmetric)

[5] We define: $X \precsim_{\mathbf{K},\bullet} Y$ if there is a univerally measurable $\varphi : \mathcal{Y}_\bullet \to \mathcal{X}$ such that $\mathbf{K}(X,Y|T) = \boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y|T)$.

[6] We define: $X \perp\!\!\!\perp_{\mathbf{K},\bullet} Y \,|\, Z$ if there exists a universally measurable Markov kernel $\mathbf{Q}(X|Z) : \mathcal{Z}_\bullet \dashrightarrow \mathcal{X}$ such that $\mathbf{K}(X,Y,Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{K}(Y,Z|T)$.

separoid axioms, where two of them need standard or universal measurable spaces, see Theorem 3.11, making it suitable for the use in graphical models, see the global Markov property, Theorem 6.3.
In contrast to [RERS17] transitional conditional independence is an asymmetric notion, directly propagating the asymmetry between "input and output variables", making it able to equivalently express classical statistical concepts like ancillarity, sufficiency, adequacy, etc. Furthermore, it was possible to develop the theory of transitional conditional independence in a much more general context of measurable spaces, transition probability spaces and transitional random variables.
In contrast to [FM20] transitional conditional independence provides one with the existence of certain Markov kernels and factorizations. This gives a much stronger conclusion in the global Markov property in graphical models, while starting from the same assumptions, see Theorem 6.3.
More will be said in Appendix L.

7.3. Conclusion

We developed the theory of transition probability spaces, transitional random variables and transitional conditional indpendence. These concepts are most well behaved if the underlying spaces have similar properties to standard measurable spaces. To extend the theory as far as possible we studied, developed and generalized the theory of univeral measurable spaces. Furthermore, we proved the disintegration of transition probabilities, i.e. the existence of regular conditional Markov kernels on standard, analytic and universal measurable spaces.
Transitional conditional independence was defined as an asymmetric notion of (ir)relevance relations. We developed the theory of asymmetric separoids and showed that transitional conditional independence and the graphical notion of σ-separation satisfy all those asymmetric separoid rules. We then showed how to relate those notions in graphical models and proved a global Markov property for those. We also showed in the Appendix how the global Markov property implies the validity of the 3 main rules of the causal do-calculus in measure theoretic generality and indicated how to apply it to further causal concepts. We then compared transitional conditional independence to other notions of extended conditional independence and showed that it is stronger than all of those.
We also showed that transitional conditional independence can express classical statistical concepts like ancillarity, sufficiency, adequacy and invariant reductions, etc. We also demonstrated what it can say about Bayesian statistics, the likelihood principle, propensity scores, etc.
Finally, we want to stress the simplicity of the definition of *transitional conditional independence*:

$$X \underset{\mathbf{K}}{\perp\!\!\!\perp} Y \mid Z \qquad :\Longleftrightarrow \qquad \exists \mathbf{Q}(X|Z): \quad \mathbf{K}(X,Y,Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{K}(Y,Z|T).$$

Acknowledgements

The author wants to express his gratitude towards Joris M. Mooij for many inspiring discussions and his constant support. This work was partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 639466).

References

[ARS09] R. Ayesha Ali, Thomas S. Richardson, and Peter Spirtes, *Markov equivalence for ancestral graphs*, The Annals of Statistics **37** (2009), no. 5B, 2808–2837.

[Bas59] Debabrata Basu, *The Family of Ancillary Statistics*, Sankhyā (1959), 247–256.

[Bas64] ______, *Recovery of Ancillary Information*, Sankhyā, Series A (1964), no. 26, 3–16.

[BD75] David Blackwell and Lester E. Dubins, *On Existence and Non-Existence of Proper, Regular, Conditional Distributions*, The Annals of Probability (1975), 741–752.

[BFPM21] Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M. Mooij, *Foundations of Structural Causal Models with Cycles and Latent Variables*, https://arxiv.org/abs/1611.06221, accepted to The Annals of Statistics (2021).

[Bir62] Allan Birnbaum, *On the Foundations of Statistical Inference*, Journal of the American Statistical Association **57** (1962), no. 298, 269–306.

[Bis06] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[Bog07] Vladimir I. Bogachev, *Measure Theory*, vol. 1+2, Springer, 2007.

[BRN63] David Blackwell and Czesław Ryll-Nardzewski, *Non-Existence of Everywhere Proper Conditional Distributions*, The Annals of Mathematical Statistics **34** (1963), no. 1, 223–225.

[BTP14] Elias Bareinboim, Jin Tian, and Judea Pearl, *Recovering from Selection Bias in Causal and Statistical Inference*, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 28, 2014.

[Bur61] D. L. Burkholder, *Sufficiency in the undominated case*, The Annals of Mathematical Statistics (1961), 1191–1200.

[CB17] Juan Correa and Elias Bareinboim, *Causal Effect Identification by Adjustment under Confounding and Selection Biases*, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, 2017.

[CD17a] Panayiota Constantinou and A. Philip Dawid, *Extended Conditional Independence and Applications in Causal Inference*, The Annals of Statistics (2017), 2618–2653.

[CD17b] ______, *Supplement to "Extended Conditional Independence and Applications in Causal Inference"*, The Annals of Statistics (2017), DOI:10.1214/16-AOS1537SUPP.

[CMKR12] Diego Colombo, Marloes H. Maathuis, Markus Kalisch, and Thomas S. Richardson, *Learning high-dimensional directed acyclic graphs with latent and selection variables*, The Annals of Statistics (2012), 294–321.

[Dar53] George Darmois, *Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire*, Revue de l'Institut international de statistique (1953), 2–8.

[Daw75] A. Philip Dawid, *On the Concepts of Sufficiency and Ancillarity in the Presence of Nuisance Parameters*, Journal of the Royal Statistical Society: Series B (Methodological) **37** (1975), no. 2, 248–258.

[Daw79a] ______, *Conditional Independence in Statistical Theory*, Journal of the Royal Statistical Society: Series B (Methodological) **41** (1979), no. 1, 1–15.

[Daw79b] ______, *Some Misleading Arguments Involving Conditional Independence*, Journal of the Royal Statistical Society: Series B (Methodological) **41** (1979), no. 2, 249–252.

[Daw80] ______, *Conditional Independence for Statistical Operations*, The Annals of Statistics (1980), 598–617.

[Daw98] ______, *Conditional Independence*, Encyclopedia of statistical sciences, update **2** (1998), 146–153.

[Daw01a] ______, *Separoids: a Mathematical Framework for Conditional Independence and Irrelevance.*, Annals of Mathematics and Artificial Intelligence **32** (2001), no. 1-4, 335–372.

[Daw01b] ______, *Some Variations on Variation Independence*, International Workshop on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, 2001, pp. 83–86.

[Daw02] ______, *Influence diagrams for causal modelling and inference*, International Statistical Review **70** (2002), 161–189.

[DL93] A. Philip Dawid and Steffen L. Lauritzen, *Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models*, The Annals of Statistics (1993), 1272–1317.

[Edw74] Anthony W. F. Edwards, *The History of Likelihood*, International Statistical Review/Revue Internationale de Statistique (1974), 9–15.

[Edw92] ———, *Likelihood*, 2nd ed., John Hopkins University Press, Baltimore, 1992.

[ER14] Robin J. Evans and Thomas S. Richardson, *Markovian Acyclic Directed Mixed Graphs for Discrete Data*, The Annals of Statistics (2014), 1452–1482.

[Eva13] Michael Evans, *What does the proof of Birnbaum's theorem prove?*, Electronic Journal of Statistics **7** (2013), 2645–2655.

[Eva16] Robin J. Evans, *Graphs for Margins of Bayesian Networks*, Scandinavian Journal of Statistics **43** (2016), no. 3, 625–648.

[Eva18] ———, *Margins of discrete Bayesian networks*, The Annals of Statistics **46** (2018), no. 6A, 2623–2656.

[Fad85] Arnold M. Faden, *The Existence of Regular Conditional Probabilities: Necessary and Sufficient Conditions*, The Annals of Probability **13** (1985), no. 1, 288–298.

[Fis22] Ronald Aylmer 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** (1922), no. 594-604, 309–368.

[Fis25] ———, *Theory of Statistical Estimation*, Mathematical Proceedings of the Cambridge Philosophical Society, vol. 22, Cambridge University Press, 1925, pp. 700–725.

[FM17] Patrick Forré and Joris M. Mooij, *Markov Properties for Graphical Models with Cycles and Latent Variables*, https://arxiv.org/abs/1710.08775 (2017).

[FM18] ———, *Constraint-based Causal Discovery for Non-linear Structural Causal Models with Cycles and Latent Confounders*, Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-2018), 2018.

[FM20] ———, *Causal Calculus in the Presence of Cycles, Latent Confounders and Selection Bias*, Proceedings of the 35th Annual Conference on Uncertainty in Artificial Intelligence (UAI-2019), vol. 115, Proceedings of Machine Learning Reasearch (PMLR), 2020, pp. 71–80.

[Fre15] David H. Fremlin, *Measure Theory*, vol. 1-6, Torres Fremlin, 2000-2015, https://www1.essex.ac.uk/maths/people/fremlin/mt.htm.

[Gan15] Greg Gandenberger, *A new proof of the likelihood principle*, British Journal for the Philosophy of Science **66** (2015), no. 3, 475–503.

[GP95] David Galles and Judea Pearl, *Testing Identifiability of Causal Effects*, Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-1995), Morgan Kaufmann Publishers Inc., 1995, pp. 185–195.

[GR01] Richard D. Gill and James M. Robins, *Causal Inference for Complex Longitudinal Data: The Continuous Case*, The Annals of Statistics (2001), 1785–1811.

[GRF10] Malay Ghosh, N. Reid, and D. A. S. Fraser, *Ancillary Statistics: A Review*, Statistica Sinica (2010), 1309–1332.

[GVP90] Dan Geiger, Thomas Verma, and Judea Pearl, *Identifying independence in Bayesian networks*, Networks **20** (1990), no. 5, 507–534.

[Hac65] Ian Hacking, *Logic of Statistical Inference*, Cambridge University Press, 1965.

[HK99] Petr Holický and Ondřej F. K. Kalenda, *Descriptive properties of spaces of measures*, Bulletin of the Polish Academy of Sciences. Mathematics **47** (1999), no. 1, 37–51.

[HS49] Paul R. Halmos and Leonard J. Savage, *Application of the Radon-Nikodym theorem to the theory of sufficient statistics*, The Annals of Mathematical Statistics **20** (1949), no. 2, 225–241.

[HV06] Yimin Huang and Marco Valtorta, *Pearl's Calculus of Intervention is Complete*, Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI-2006), AUAI Press, 2006, pp. 217–224.

[HV08] ______, *On the Completeness of an Identifiability Algorithm for Semi-Markovian Models*, Annals of Mathematics and Artificial Intelligence **54** (2008), no. 4, 363–408.

[HWG65] William Jackson Hall, Robert A. Wijsman, and Jayanta K. Ghosh, *The Relationship between Sufficiency and Invariance with Applications in Sequential Analysis*, The Annals of Mathematical Statistics **36** (1965), no. 2, 575–614.

[Jay03] Edwin T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

[Kec95] Alexander S. Kechris, *Classical Descriptive Set Theory*, Graduate Texts in Mathematics, vol. 156, Springer-Verlag, New York, 1995.

[KF09] Daphne Koller and Nir Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.

[Kle14] Achim Klenke, *Probability Theory - A Comprehensive Course*, 2nd ed., Universitext, Springer, London, 2014.

[Lau96] Steffen L. Lauritzen, *Graphical Models*, vol. 17, Clarendon Press, 1996.

[LDLL90] S. L. Lauritzen, A. P. Dawid, B. N. Larsen, and H.-G. Leimer, *Independence properties of directed Markov fields*, Networks **20** (1990), no. 5, 491–505.

[Mah75] Dorothy Maharam, *On Smoothing Compact Measure Spaces by Multiplication*, Transactions of the American Mathematical Society **204** (1975), 1–39.

[May14] Deborah G. Mayo, *On the Birnbaum argument for the strong likelihood principle*, Statistical Science (2014), 227–239.

[MMC20] Joris M. Mooij, Sara Magliacane, and Tom Claassen, *Joint Causal Inference from Multiple Contexts*, Journal of Machine Learning Research **21** (2020), no. 99, 1–108.

[Mur12] Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

[Ney36] Jerzy Neyman, *Su un teorema concernente le cosiddette statistiche sufficienti*, Istituto Italiano degli Attuari, 1936.

[Pac78] Jan K. Pachl, *Disintegration and Compact Measures*, Mathematica Scandinavica (1978), 157–168.

[Pea93a] Judea Pearl, *Aspects of graphical models connected with causality*, In Proceedings of the 49th Session of the International Statistical Institute, 1993, pp. 391–401.

[Pea93b] ______, *Comment: Graphical Models, Causality, And Intervention*, Statistical Science **8** (1993), no. 3, 266–269.

[Pea09] ______, *Causality: Models, Reasoning, and Inference*, 2nd ed., Cambridge University Press, 2009.

[PJS17] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf, *Elements of Causal Inference: Foundation and Learning Algorithms*, MIT Press, 2017.

[PP85] Judea Pearl and Azaria Paz, *Graphoids: A Graph-based Logic for Reasoning about Relevance Relations*, University of California (Los Angeles). Computer Science Department, 1985.

[PP14] ______, *Confounding Equivalence in Causal Inference*, Journal of Causal Inference **2** (2014), no. 1.

[PTKM15] Emilija Perković, Johannes Textor, Markus Kalisch, and Marloes H. Maathuis, *A Complete Generalized Adjustment Criterion*, Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence (UAI-2015), AUAI Press, 2015, pp. 682–691.

[RERS17] Thomas S. Richardson, Robin J. Evans, James M. Robins, and Ilya Shpitser, *Nested Markov Properties for Acyclic Directed Mixed Graphs*, https://arxiv.org/abs/1701.06686 (2017).

[Res77] Paul Ressel, *Some Continuity and Measurability Results on Spaces of Measures*, Mathematica Scandinavica **40** (1977), no. 1, 69–78.

[Ric03] Thomas S. Richardson, *Markov Properties for Acyclic Directed Mixed Graphs*, Scandinavian Journal of Statistics **30** (2003), no. 1, 145–157.

[Roy97] Richard Royall, *Statistical Evidence: a Likelihood Paradigm*, vol. 71, CRC press, 1997.

[RR83] Paul R. Rosenbaum and Donald B. Rubin, *The central role of the propensity score in observational studies for causal effects*, Biometrika **70** (1983), 41–55.

[RS02] Thomas S. Richardson and Peter Spirtes, *Ancestral graph Markov models*, The Annals of Statistics **30** (2002), no. 4, 962–1030.

[SBC+62] Leonard J. Savage, George Barnard, Jerome Cornfield, Irwin Bross, I. J. Good, D. V. Lindley, C. W. Clunies-Ross, John W. Pratt, Howard Levene, Thomas Goldman, et al., *On the Foundations of Statistical Inference: Discussion*, Journal of the American Statistical Association **57** (1962), no. 298, 307–326.

[Sch74] Laurent Schwartz, *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*, Oxford University Press, Bombay, 1974.

[SGS00] Peter Spirtes, Clark Glymour, and Richard Scheines, *Causation, Prediction, and Search*, 2nd ed., MIT Press, Cambridge, MA, 2000.

[SP06] Ilya Shpitser and Judea Pearl, *Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models*, Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2, AAAI Press, 2006, pp. 1219–1226.

[Spo94] Wolfgang Spohn, *On the Properties of Conditional Independence*, Patrick Suppes: Scientific Philosopher, Springer, 1994, pp. 173–196.

[SVR10] Ilya Shpitser, Tyler J. VanderWeele, and James M. Robins, *On the Validity of Covariate Adjustment for Estimating Causal Effects*, Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI-2010), AUAI Press, 2010, pp. 527–536.

[Tia02] Jin Tian, *Studies in causal reasoning and learning*, Ph.D. thesis, University of California, Los Angeles, 2002.

[Tia04] ______, *Identifying Conditional Causal Effects*, Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI-2004), AUAI Press, 2004, pp. 561–568.

[TP02] Jin Tian and Judea Pearl, *A General Identification Condition for Causal Effects*, Eighteenth National Conference on Artificial Intelligence, American Association for Artificial Intelligence, 2002, pp. 567–573.

[Ver93] Tom S. Verma, *Graphical Aspects of Causal Models*, Tech. Report R-191, Computer Science Department, University of California, Los Angeles, 1993.

Appendices

A. Symmetric Separoids and Asymmetric Separoids

This section aims at formalizing a notion of "asymmetric separoids". Separoids were introduced in [Daw01a] as a way to express (ir)relevance relations. Similar ideas and the notion of graphoids were also investigated in [PP85, Spo94]. "Asymmetric separoids" will naturally arise from symmetric ones when one fixes parts of one, say the second argument and/or the third argument. Depending on their behaviour in the second and third argument we will call such a ternary relation $\perp\!\!\!\perp$ a *τ-κ-separoid*. Our notion of *transitional conditional independence*, see Definition 3.1, will form an T-$*$-separoid, see Appendices D and E. Also σ-separation, an extension of d-separation, to graphs that allow for (cycles and) external "input" nodes will form a J-$\emptyset$-separoid, see Section 5 and Appendix I. The global Markov property for the probabilistic graphical models, which relates the conditional independence structure of the graph in terms of σ-separation to transitional conditional independence of the corresponding transitional random variables can then be interpreted as a τ-κ-separoid homomorphism. The global Markov property will be presented in Section 6.2 and proof is given in Appendix J.

Since the separoid axioms from [Daw01a] are not logically independent, i.e. parts of the rules can be deduced from the others, we have to investigate which of the rules generalize in which form. This is the reason we give a slightly different definition of separoids than it was presented in [Daw01a]. We also do not require the order on the space to fully form a join-semi-lattice. To make the distinction to [Daw01a] clearer we will stick to the names *symmetric separoid* and *τ-κ-separoid* in this paper. We then study the relation between symmetric separoids and τ-κ-separoid. The main findings are Theorem A.11 and Theorem A.12. The first, A.11, states, as discussed above, that if we start with a symmetric separoid and fix parts of the second and/or third argument we will arrive at a τ-κ-separoid. The other one, A.12, shows that if we symmetrize an τ-κ-separoid with the logical "or" we will get a symmetric separoid.

One can certainly argue which of the separoid rules should be included in the notion of τ-κ-separoid and which not. The guiding principle of our definition was that it should: (i) generalize the notion of (symmetric) separoids and recover it, (ii) be strong enough to prove the global Markov property, see Theorem 6.3 and Appendix J.1, (iii) be weak enough such that we can verify these rules for both, transitional conditional independence, see D and E, and σ-separation in graphs with input nodes, see 5 and I, (iv) be adjusted in such a way that we can prove the two Theorems, A.11 and A.12, mentioned above to get a close relationship between symmetric separoids and τ-κ-separoids. This is, for instance, the reason we included 4 Contraction rules, which are all needed to prove Theorem A.12. These then also give us Restricted Symmetry 1. The obstruction to become symmetric then gives a clear motivation for τ-Restricted Right Redundancy 2, allowing $\varnothing$-κ-separoids to become symmetric separoids. A look at the proof of the global Markov property in Appendix J.1 shows that all the other separoid rules are needed, thus motivating the minimal set of rules that we require.

Remark A.1. *In the following let Ω be a class that is endowed with an equivalence relation $\cong$, another binary relation $\ll$, a binary operation $\vee$ and a ternary relation $\perp\!\!\!\perp$ and an element $\varnothing \in \Omega$ (so called* bottom element*). We will assume without further indication that they will satisfy the following rules for all $\alpha, \beta, \gamma, \dots \in \Omega$:*

1. $(\alpha \cong \alpha') \wedge (\beta \cong \beta') \implies (\alpha \vee \beta) \cong (\alpha' \vee \beta')$.
2. $(\alpha \vee \beta) \cong (\beta \vee \alpha)$.
3. $((\alpha \vee \beta) \vee \gamma) \cong (\alpha \vee (\beta \vee \gamma))$.
4. $(\alpha \ll \beta) \wedge (\alpha \cong \alpha') \wedge (\beta \cong \beta') \implies (\alpha' \ll \beta')$.
5. $(\alpha \ll \beta) \implies (\alpha \ll (\beta \vee \gamma))$.
6. $(\alpha \perp\!\!\!\perp \beta \,|\, \gamma) \wedge (\alpha \cong \alpha') \wedge (\beta \cong \beta') \wedge (\gamma \cong \gamma') \implies (\alpha' \perp\!\!\!\perp \beta' \,|\, \gamma')$.
7. $\varnothing \ll \alpha$.
8. $\varnothing \vee \alpha \cong \alpha$.

From this it is clear that if $\varnothing'$ is another bottom element we have: $\varnothing \cong \varnothing'$ and $\varnothing \ll \varnothing' \ll \varnothing$.

Definition A.2 (Symmetric separoid)**.** *We call $(\Omega, \cong, \ll, \vee, \perp\!\!\!\perp, \varnothing)$ a* symmetric separoid *if the following rules hold for all $\alpha, \beta, \gamma, \lambda \in \Omega$:*

1. *(Unrestricted) Symmetry:*

   $\alpha \perp\!\!\!\perp \beta \,|\, \gamma \implies \beta \perp\!\!\!\perp \alpha \,|\, \gamma$.

2. *(Left) Redundancy:*

$\alpha \lll \gamma \implies \alpha \perp\!\!\!\perp \beta \,|\, \gamma.$

3. *(Left) Decomposition:*

   $\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma \implies \lambda \perp\!\!\!\perp \beta \,|\, \gamma.$

4. *(Left) Weak Union:*

   $\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma \implies \alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma.$

5. *(Left) Contraction:*

   $(\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \wedge (\lambda \perp\!\!\!\perp \beta \,|\, \gamma) \implies \alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma.$

Definition A.3 (τ-κ-separoid)**.** *Let $\tau, \kappa \in \Omega$ be fixed. We call $(\Omega, \cong, \lll, \vee, \perp\!\!\!\perp, \varnothing)$ a* τ-κ-separoid *if the following rules hold for all $\alpha, \beta, \gamma, \lambda \in \Omega$:*

1. *κ-Extended Left Redundancy:*

   $\alpha \lll \kappa \vee \gamma \implies \alpha \perp\!\!\!\perp \beta \,|\, \gamma.$

2. *τ-Restricted Right Redundancy:*

   $\alpha \perp\!\!\!\perp \varnothing \,|\, \gamma \vee \tau.$

3. *τ-Inverted Right Decomposition:*

   $\alpha \perp\!\!\!\perp \beta \,|\, \gamma \implies \alpha \perp\!\!\!\perp \tau \vee \beta \,|\, \gamma.$

4. *Left Decomposition:*

   $\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma \implies \lambda \perp\!\!\!\perp \beta \,|\, \gamma.$

5. *Right Decomposition:*

   $\alpha \perp\!\!\!\perp \beta \vee \lambda \,|\, \gamma \implies \alpha \perp\!\!\!\perp \lambda \,|\, \gamma.$

6. *Left Weak Union:*

   $\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma \implies \alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma.$

7. *Right Weak Union:*

   $\alpha \perp\!\!\!\perp \beta \vee \lambda \,|\, \gamma \implies \alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma.$

8. *Left Contraction:*

   $(\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \wedge (\lambda \perp\!\!\!\perp \beta \,|\, \gamma) \implies \alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma.$

9. *Right Contraction:*

   $(\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \wedge (\alpha \perp\!\!\!\perp \lambda \,|\, \gamma) \implies \alpha \perp\!\!\!\perp \beta \vee \lambda \,|\, \gamma.$

10. *Right Cross Contraction:*

    $$(\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \wedge (\lambda \perp\!\!\!\perp \alpha \,|\, \gamma) \implies \alpha \perp\!\!\!\perp \beta \vee \lambda \,|\, \gamma.$$

11. *Flipped Left Cross Contraction:*

    $$(\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \wedge (\beta \perp\!\!\!\perp \lambda \,|\, \gamma) \implies \beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma.$$

Remark A.4. 1. *Every symmetric separoid is a $\varnothing$-$\varnothing$-separoid.*

2. *Every $\varnothing$-κ-separoid is a symmetric separoid by Lemma A.10 below.*

3. *Let $(\Omega, \cong, \ll, \vee, \perp\!\!\!\perp, \varnothing)$ be a τ-κ-separoid. Then $(\Omega, \cong, \ll_\kappa, \vee, \perp\!\!\!\perp, \varnothing)$ is a τ-$\varnothing$-separoid, where $\ll_\kappa$ is given via:*

   $$\alpha \ll_\kappa \beta :\iff \alpha \ll \kappa \vee \beta.$$

4. *So every $\varnothing$-κ-separoid is a (symmetric) $\varnothing$-$\varnothing$-separoid using $\ll_\kappa$.*

In the following we will, by abuse of notation, refer to the whole structure $(\Omega, \cong, \ll, \vee, \perp\!\!\!\perp, \varnothing)$ as Ω.

Lemma A.5 (κ-Extended Inverted Left Decomposition)**.** *Let Ω be a τ-κ-separoid then we have the implication:*

$$(\alpha \perp\!\!\!\perp \beta \,|\, \gamma) \wedge (\lambda \ll \alpha \vee \kappa \vee \gamma) \implies \alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma.$$

Proof. We have the implications:

$$\lambda \ll \alpha \vee \kappa \vee \gamma \xRightarrow{\kappa\text{-Extended Left Redundancy } 1} \lambda \perp\!\!\!\perp \beta \,|\, \alpha \vee \gamma$$
$$\xRightarrow{(\alpha \perp\!\!\!\perp \beta \,|\, \gamma) \wedge \text{ Left Contraction } 8} \alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma.$$

□

Lemma A.6 (τ-κ-Extended Inverted Right Decomposition)**.** *Let Ω be a τ-κ-separoid then we have the implication:*

$$(\alpha \perp\!\!\!\perp \beta \,|\, \gamma) \wedge (\lambda \ll \tau \vee \beta \vee \kappa \vee \gamma) \implies \alpha \perp\!\!\!\perp \tau \vee \beta \vee \lambda \,|\, \gamma.$$

Proof. We have the implications:

$$\alpha \perp\!\!\!\perp \beta \,|\, \gamma \xRightarrow{\tau\text{-Inverted Right Decomposition } 3} \alpha \perp\!\!\!\perp \tau \vee \beta \,|\, \gamma,$$
$$\lambda \ll \tau \vee \beta \vee \kappa \vee \gamma \xRightarrow{\kappa\text{-Extended Left Redundancy } 1} \lambda \perp\!\!\!\perp \alpha \,|\, \tau \vee \beta \vee \gamma,$$
$$\text{both together} \xRightarrow{\text{Flipped Left Cross Contraction } 11} \alpha \perp\!\!\!\perp \tau \vee \beta \vee \lambda \,|\, \gamma.$$

□

Lemma A.7 (κ-Equivalent Exchange). *Let Ω be a τ-κ-separoid then we have the implication:*

$$(\alpha \perp\!\!\!\perp \beta \,|\, \gamma) \wedge (\gamma \ll \kappa \vee \gamma') \wedge (\gamma' \ll \kappa \vee \gamma) \implies \alpha \perp\!\!\!\perp \beta \,|\, \gamma'.$$

Proof. We have the implications:

$$\begin{aligned}
\gamma' \ll \kappa \vee \gamma \quad &\implies && \gamma' \ll \tau \vee \beta \vee \kappa \vee \gamma, \\
\alpha \perp\!\!\!\perp \beta \,|\, \gamma \quad &\xRightarrow{\tau\text{-}\kappa\text{-Ext. Inv. Right Decomposition A.6}} && \alpha \perp\!\!\!\perp \tau \vee \beta \vee \gamma' \,|\, \gamma \\
&\xRightarrow{\text{Right Decomposition 5}} && \alpha \perp\!\!\!\perp \beta \vee \gamma' \,|\, \gamma \\
&\xRightarrow{\text{Right Weak Union 7}} && \alpha \perp\!\!\!\perp \beta \,|\, \gamma' \vee \gamma, && \text{(a)} \\
\gamma \ll \kappa \vee \gamma' \quad &\xRightarrow{\kappa\text{–Extended Left Redundancy 1}} && \gamma \perp\!\!\!\perp \beta \,|\, \gamma', && \text{(b)} \\
\text{(a)} \wedge \text{(b)} \quad &\xRightarrow{\text{Left Contraction 8}} && \alpha \vee \gamma \perp\!\!\!\perp \beta \,|\, \gamma' \\
&\xRightarrow{\text{Left Decomposition 4}} && \alpha \perp\!\!\!\perp \beta \,|\, \gamma'.
\end{aligned}$$

□

Remark A.8. *Note that if we have $\gamma \ll \gamma' \ll \gamma$ then we also have the condition:*

$$\gamma \ll \kappa \vee \gamma' \quad \wedge \quad \gamma' \ll \kappa \vee \gamma,$$

since we have the rule: $\alpha \ll \beta \implies \alpha \ll \lambda \vee \beta$, as stated in the beginning.

Lemma A.9 (Full κ-Equivalent Exchange). *Let Ω be a τ-κ-separoid. If $\alpha' \ll \kappa \vee \alpha$ and $\beta' \ll \kappa \vee \beta$ and $\gamma \ll \kappa \vee \gamma'$ and $\gamma' \ll \kappa \vee \gamma$ then we have:*

$$\alpha \perp\!\!\!\perp \beta \,|\, \gamma \qquad \implies \qquad \alpha' \perp\!\!\!\perp \beta' \,|\, \gamma'.$$

Proof. We have the implications:

$$\begin{aligned}
\alpha' \ll \kappa \vee \alpha \quad &\implies && \alpha' \ll \alpha \vee \kappa \vee \gamma \\
&\xRightarrow{\kappa\text{-Ext. Inv. Left Decomposition A.5}} && \alpha \vee \alpha' \perp\!\!\!\perp \beta \,|\, \gamma \\
&\xRightarrow{\text{Left Decomposition 4}} && \alpha' \perp\!\!\!\perp \beta \,|\, \gamma, \\
\beta' \ll \kappa \vee \beta \quad &\implies && \beta' \ll \tau \vee \beta \vee \kappa \vee \gamma \\
&\xRightarrow{\tau\text{-}\kappa\text{–Ext. Inv. Right Decomposition A.6}} && \alpha' \perp\!\!\!\perp \tau \vee \beta \vee \beta' \,|\, \gamma \\
&\xRightarrow{\text{Right Decomposition 5}} && \alpha' \perp\!\!\!\perp \beta' \,|\, \gamma, \\
(\gamma \ll \kappa \vee \gamma') \wedge (\gamma' \ll \kappa \vee \gamma) \quad &\xRightarrow{\kappa\text{-Equivalent Exchange A.7}} && \alpha' \perp\!\!\!\perp \beta' \,|\, \gamma'.
\end{aligned}$$

□

Lemma A.10 (Symmetry)**.** *Let Ω be a τ-κ-separoid then we have the implication:*

1. *Restricted Symmetry:*

   $$(\alpha \perp\!\!\!\perp \beta \,|\, \gamma) \wedge (\beta \perp\!\!\!\perp \varnothing \,|\, \gamma) \implies \beta \perp\!\!\!\perp \alpha \,|\, \gamma.$$

2. *τ-Restricted Symmetry:*

   $$\alpha \perp\!\!\!\perp \beta \,|\, \gamma \vee \tau \implies \beta \perp\!\!\!\perp \alpha \,|\, \gamma \vee \tau.$$

3. *Symmetry: If $\tau \cong \varnothing$ then:*

   $$\alpha \perp\!\!\!\perp \beta \,|\, \gamma \implies \beta \perp\!\!\!\perp \alpha \,|\, \gamma.$$

Proof. Restricted Symmetry 1 follows from Flipped Left Cross Contraction 11 with $\lambda = \varnothing$. τ-Restricted Symmetry 2 follows by replacing γ with $\gamma \vee \tau$ and using τ-Restricted Right Redundancy 2. Symmetry follows from τ-Restricted Symmetry 2 with $\tau \cong \varnothing$. □

The following will be the first main theorem of this section.

Theorem A.11. *Let $(\Omega, \cong, \ll, \vee, \perp\!\!\!\perp, \varnothing)$ be a τ_1-κ_1-separoid and $\tau_2, \kappa_2 \in \Omega$ further elements such that $\tau_2 \ll \tau_2$. For $\alpha, \beta, \gamma, \lambda \in \Omega$ we define the ternary relation:*

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma :\iff \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma.$$

Then $(\Omega, \cong, \ll, \vee, \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp}, \varnothing)$ is a $(\tau_1 \vee \tau_2)$-$(\kappa_1 \vee \kappa_2)$-separoid.

Proof. We put $\tau := \tau_1 \vee \tau_2$ and $\kappa := \kappa_1 \vee \kappa_2$.

1. κ-Extended Left Redundancy:

   $$\alpha \ll \kappa \vee \gamma \implies \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma.$$

   Proof:

   $$\begin{aligned} \alpha \ll \kappa \vee \gamma \quad &\implies \quad \alpha \ll \kappa_1 \vee (\kappa_2 \vee \gamma), \\ &\xrightarrow{\kappa_1\text{-Extended Left Redundancy } 1} \quad \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma, \\ &\implies \quad \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma. \end{aligned}$$

2. τ-Restricted Right Redundancy:

   $$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \varnothing \,|\, \gamma \vee \tau.$$

Proof:

$$\xrightarrow{\tau_1\text{-Restricted Right Redundancy 2}} \quad \alpha \perp\!\!\!\perp \varnothing \,|\, \kappa_2 \vee \gamma \vee \tau_2 \vee \tau_1,$$

$$\tau_2 \ll \tau_2 \quad \Longrightarrow \quad \tau_2 \ll \kappa_1 \vee \kappa_2 \vee \gamma \vee \tau_2 \vee \tau_1,$$

$$\text{both} \quad \xrightarrow{\tau_1\text{-}\kappa_1\text{-Ext. Inv. Right Decomposition A.6}} \quad \alpha \perp\!\!\!\perp \tau_2 \vee \varnothing \,|\, \kappa_2 \vee \gamma \vee \tau_2 \vee \tau_1$$

$$\Longrightarrow \quad \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \varnothing \,|\, \gamma \vee \tau.$$

3. τ-Inverted Right Decomposition:

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma \Longrightarrow \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \tau \vee \beta \,|\, \gamma.$$

Proof:

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma \quad \Longrightarrow \quad \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma$$

$$\xrightarrow{\tau_1\text{-Inverted Right Decomposition 3}} \quad \alpha \perp\!\!\!\perp \tau_2 \vee \tau_1 \vee \beta \,|\, \kappa_2 \vee \gamma,$$

$$\tau_2 \ll \tau_2 \quad \Longrightarrow \quad \tau_2 \ll \tau_2 \vee \tau_2 \vee \tau_1 \vee \beta \vee \kappa_2 \vee \kappa_2 \vee \gamma,$$

$$\xrightarrow{\tau_1\text{-}\kappa_1\text{-Ext. Inv. Right Decomposition A.6}} \quad \alpha \perp\!\!\!\perp \tau_2 \vee \tau_1 \vee \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma$$

$$\Longrightarrow \quad \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \tau \vee \beta \,|\, \gamma.$$

4. Left Decomposition:

$$\alpha \vee \lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma \Longrightarrow \lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma.$$

Proof:

$$\alpha \vee \lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma \quad \Longrightarrow \quad \alpha \vee \lambda \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma,$$

$$\xrightarrow{\text{Left Decomposition 4}} \quad \lambda \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma$$

$$\Longrightarrow \quad \lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma.$$

5. Right Decomposition:

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \vee \lambda \,|\, \gamma \Longrightarrow \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \lambda \,|\, \gamma.$$

Proof:

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \vee \lambda \,|\, \gamma \quad \Longrightarrow \quad \alpha \perp\!\!\!\perp \tau_2 \vee \beta \vee \lambda \,|\, \kappa_2 \vee \gamma$$

$$\xrightarrow{\text{Right Decomposition 5}} \quad \alpha \perp\!\!\!\perp \tau_2 \vee \lambda \,|\, \kappa_2 \vee \gamma$$

$$\Longrightarrow \quad \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \lambda \,|\, \gamma.$$

6. Left Weak Union:

$$\alpha \vee \lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma \implies \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma.$$

Proof:

$$\begin{aligned}
\alpha \vee \lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma \quad &\implies \quad \alpha \vee \lambda \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma \\
&\xRightarrow{\text{Left Weak Union 6}} \quad \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \lambda \vee \gamma \\
&\implies \quad \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma.
\end{aligned}$$

7. Right Weak Union:

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \vee \lambda \,|\, \gamma \implies \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma.$$

Proof:

$$\begin{aligned}
\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \vee \lambda \,|\, \gamma \quad &\implies \quad \alpha \perp\!\!\!\perp \tau_2 \vee \beta \vee \lambda \,|\, \kappa_2 \vee \gamma \\
&\xRightarrow{\text{Right Weak Union 7}} \quad \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \lambda \vee \gamma \\
&\implies \quad \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma.
\end{aligned}$$

8. Left Contraction:

$$(\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma) \wedge (\lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma) \implies \alpha \vee \lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma.$$

Proof:

$$\begin{aligned}
\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma \quad &\implies \quad \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \lambda \vee \gamma, \\
\lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma \quad &\implies \quad \lambda \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma, \\
\text{both together} \quad &\xRightarrow{\text{Left Contraction 8}} \quad \alpha \vee \lambda \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \gamma \\
&\implies \quad \alpha \vee \lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \gamma.
\end{aligned}$$

9. Right Contraction:

$$(\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma) \wedge (\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \lambda \,|\, \gamma) \implies \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \vee \lambda \,|\, \gamma.$$

Proof:

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma \implies \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \lambda \vee \gamma,$$

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \lambda \,|\, \gamma \implies \alpha \perp\!\!\!\perp \tau_2 \vee \lambda \,|\, \kappa_2 \vee \gamma$$

$$\xRightarrow{\text{Right Decomposition 5}} \alpha \perp\!\!\!\perp \lambda \,|\, \kappa_2 \vee \gamma,$$

$$\text{both together} \xRightarrow{\text{Right Contraction 9}} \alpha \perp\!\!\!\perp \tau_2 \vee \beta \vee \lambda \,|\, \kappa_2 \vee \gamma$$

$$\implies \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \vee \lambda \,|\, \gamma.$$

10. Right Cross Contraction:

$$(\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma) \wedge (\lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \alpha \,|\, \gamma) \implies \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \vee \lambda \,|\, \gamma.$$

Proof:

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma \implies \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \lambda \vee \gamma,$$

$$\lambda \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \alpha \,|\, \gamma \implies \lambda \perp\!\!\!\perp \tau_2 \vee \alpha \,|\, \kappa_2 \vee \gamma$$

$$\xRightarrow{\text{Right Decomposition 5}} \lambda \perp\!\!\!\perp \alpha \,|\, \kappa_2 \vee \gamma,$$

$$\text{both together} \xRightarrow{\text{Right Cross Contraction 10}} \alpha \perp\!\!\!\perp \tau_2 \vee \beta \vee \lambda \,|\, \kappa_2 \vee \gamma$$

$$\implies \alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \vee \lambda \,|\, \gamma.$$

11. Flipped Left Cross Contraction:

$$(\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma) \wedge (\beta \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \lambda \,|\, \gamma) \implies \beta \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \alpha \vee \lambda \,|\, \gamma.$$

Proof:

$$\alpha \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma \implies \alpha \perp\!\!\!\perp \tau_2 \vee \beta \,|\, \kappa_2 \vee \lambda \vee \gamma$$

$$\xRightarrow{\text{Right Weak Union 7}} \alpha \perp\!\!\!\perp \beta \,|\, \kappa_2 \vee \tau_2 \vee \lambda \vee \gamma,$$

$$\beta \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \lambda \,|\, \gamma \implies \beta \perp\!\!\!\perp \tau_2 \vee \lambda \,|\, \kappa_2 \vee \gamma,$$

$$A \xRightarrow{\text{Flipped Left Cross Contraction 11}} \beta \perp\!\!\!\perp \tau_2 \vee \alpha \vee \lambda \,|\, \kappa_2 \vee \gamma,$$

$$\implies \beta \underset{(\tau_2|\kappa_2)}{\perp\!\!\!\perp} \alpha \vee \lambda \,|\, \gamma.$$

□

We now directly present the second main theorem of this paper, showing the close relation between symmetric and τ-κ-separoids.

Theorem A.12. *Let $(\Omega, \cong, \ll, \vee, \perp\!\!\!\perp, \varnothing)$ be a τ-κ-separoid. We then define for $\alpha, \beta, \gamma \in \Omega$ the symmetrized ternary relation:*

$$\alpha \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma \qquad :\iff \qquad (\alpha \perp\!\!\!\perp \beta \,|\, \gamma) \quad \vee \quad (\beta \perp\!\!\!\perp \alpha \,|\, \gamma)\,,$$

where with $\vee$ in the middle of the right we mean the logical OR. We further define:

$$\alpha \ll_\kappa \beta \qquad :\iff \qquad \alpha \ll \kappa \vee \beta.$$

Then $(\Omega, \cong, \ll_\kappa, \vee, \overset{\vee}{\perp\!\!\!\perp}, \varnothing)$ is a (symmetric) $\varnothing$-$\varnothing$-separoid.

Proof. 1. (Unrestricted) Symmetry:

$$\alpha \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma \implies \beta \overset{\vee}{\perp\!\!\!\perp} \alpha \,|\, \gamma.$$

Proof. Clear by construction.

2. (Left) Redundancy:

$$\alpha \ll_\kappa \gamma \implies \alpha \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma.$$

Proof.

$$\begin{aligned}
\alpha \ll_\kappa \gamma \quad &\implies && \alpha \ll \kappa \vee \gamma,\\
&\xRightarrow{\kappa\text{-Extended Left Redundancy 1}} && \alpha \perp\!\!\!\perp \beta \,|\, \gamma,\\
&\implies && (\alpha \perp\!\!\!\perp \beta \,|\, \gamma) \vee (\beta \perp\!\!\!\perp \alpha \,|\, \gamma)\\
&\implies && \alpha \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma.
\end{aligned}$$

3. (Left) Decomposition:

$$\alpha \vee \lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma \implies \lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma$$

Proof.

$$\begin{aligned}
\alpha \vee \lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma \quad &\implies && (\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma) \vee (\beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma)\,,\\
\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma \quad &\xRightarrow{\text{Left Decomposition 4}} && \lambda \perp\!\!\!\perp \beta \,|\, \gamma,\\
\beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma \quad &\xRightarrow{\text{Right Decomposition 5}} && \beta \perp\!\!\!\perp \lambda \,|\, \gamma,\\
\text{both together} \quad &\implies && (\lambda \perp\!\!\!\perp \beta \,|\, \gamma) \vee (\beta \perp\!\!\!\perp \lambda \,|\, \gamma)\\
&\implies && \lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma.
\end{aligned}$$

4. (Left) Weak Union:

$$\alpha \vee \lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma \implies \alpha \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma.$$

Proof.

$$\begin{aligned}
\alpha \vee \lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma &\implies && (\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma) \vee (\beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma)\,, \\
\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma &\xRightarrow{\text{Left Weak Union 6}} && \alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma, \\
\beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma &\xRightarrow{\text{Right Weak Union 7}} && \beta \perp\!\!\!\perp \alpha \,|\, \lambda \vee \gamma, \\
\text{both together} &\implies && (\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \vee (\beta \perp\!\!\!\perp \alpha \,|\, \lambda \vee \gamma) \\
&\implies && \alpha \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma.
\end{aligned}$$

5. (Left) Contraction:

$$(\alpha \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma) \wedge (\lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma) \implies \alpha \vee \lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma.$$

Proof.

$$\begin{aligned}
(\alpha \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \lambda \vee \gamma) \wedge (\lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma) &\implies && ((\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \vee (\beta \perp\!\!\!\perp \alpha \,|\, \lambda \vee \gamma)) \\
& && \wedge ((\lambda \perp\!\!\!\perp \beta \,|\, \gamma) \wedge (\beta \perp\!\!\!\perp \lambda \,|\, \gamma))\,, \\
(\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \wedge (\lambda \perp\!\!\!\perp \beta \,|\, \gamma) &\xRightarrow{\text{Left Contr. 8}} && \alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma, \\
(\alpha \perp\!\!\!\perp \beta \,|\, \lambda \vee \gamma) \wedge (\beta \perp\!\!\!\perp \lambda \,|\, \gamma) &\xRightarrow{\text{FL Cr. Contr. 11}} && \beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma, \\
(\beta \perp\!\!\!\perp \alpha \,|\, \lambda \vee \gamma) \wedge (\lambda \perp\!\!\!\perp \beta \,|\, \gamma) &\xRightarrow{\text{R. Cr. Contr. 10}} && \beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma, \\
(\beta \perp\!\!\!\perp \alpha \,|\, \lambda \vee \gamma) \wedge (\beta \perp\!\!\!\perp \lambda \,|\, \gamma) &\xRightarrow{\text{Right Contr. 9}} && \beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma, \\
\text{all together} &\implies && (\alpha \vee \lambda \perp\!\!\!\perp \beta \,|\, \gamma) \vee (\beta \perp\!\!\!\perp \alpha \vee \lambda \,|\, \gamma)\,, \\
&\implies && \alpha \vee \lambda \overset{\vee}{\perp\!\!\!\perp} \beta \,|\, \gamma.
\end{aligned}$$

□

B. Measure Theory

In this section we review the necessary measure theory that is needed to prove the disintegration theorem for Markov kernels in more generality. Since already the existence of disintegrations or regular conditional probability distributions is in general dependent on the existence of countably compact approximating classes (see Pachl's theorem in [Fre15] 452I and remarks 452J, 452K) we need to dive into topological measure theory and descriptive set theory. To short-cut this endeavor (i.e. to skip Souslin-F, K-analytic, analytic spaces, etc.) here we directly define Souslin's $\mathcal{S}$-operator on the level of measurable spaces (without the need for complicated combinatorics and topological considerations in the statements) by only defining it on σ-algebras and cite or rephrase the relevant results from [Kec95], [Bog07], [Fre15], etc. directly for measurable spaces. We also introduce the notion of (or generalize from separable metric spaces to) "universal measurable spaces", a measure space analogon to Radon spaces (plus some countability conditions). Universal measurable spaces will have the a property that allows for general disintegrations. We can not expect to get a general existence result for (regular) conditional Markov kernels in spaces far beyond these universal measurable spaces, but we will get stronger versions for the more restrictive standard and analytic measurable spaces.

B.1. Completion and Universal Completion of a Sigma-Algebra

Definition B.1 (Completion of a σ-algebra)**.** *Let $\mathcal{X}$ be a measurable space and $\mathcal{P}(\mathcal{X})$ the space of probability measures on $\mathcal{X}$ or, more precisely, on its σ-algebra $\mathcal{B}_\mathcal{X}$.*

1. *For $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$ consider the set of* $\boldsymbol{\mu}$-null sets *of $\mathcal{X}$:*

$$\mathcal{I}_{\boldsymbol{\mu}} := \{M \subseteq \mathcal{X} \,|\, \exists N \in \mathcal{B}_\mathcal{X},\, M \subseteq N,\, \boldsymbol{\mu}(N) = 0\}.$$

   $\mathcal{I}_{\boldsymbol{\mu}}$ is a σ-ideal, i.e. it contains $\varnothing$ and is closed under subsets and countable unions.

2. *The* $\boldsymbol{\mu}$-completion *of $\mathcal{B}_\mathcal{X}$ is the following σ-algebra:*

$$(\mathcal{B}_\mathcal{X})_{\boldsymbol{\mu}} := \{A \triangle M \,|\, A \in \mathcal{B}_\mathcal{X}, M \in \mathcal{I}_{\boldsymbol{\mu}}\}.$$

   We use the notation: $A \triangle M := (A \cup M) \backslash (A \cap M)$.

   Note that $\boldsymbol{\mu}$ can uniquely be extended to $(\mathcal{B}_\mathcal{X})_{\boldsymbol{\mu}}$.

3. *We will write $\mathcal{X}_{\boldsymbol{\mu}}$ to refer to $\mathcal{X}$ equipped with the σ-algebra $(\mathcal{B}_\mathcal{X})_{\boldsymbol{\mu}}$.*

Definition B.2 (Universal completion of a σ-algebra)**.** *Let $\mathcal{X}$ be a measurable space.*

1. *The* universal completion *of the σ-algebra $\mathcal{B}_\mathcal{X}$ is the following σ-algebra:*

$$(\mathcal{B}_\mathcal{X})_\bullet := \bigcap_{\boldsymbol{\nu} \in \mathcal{P}(\mathcal{X})} (\mathcal{B}_\mathcal{X})_{\boldsymbol{\nu}},$$

   i.e. the intersection of all $\boldsymbol{\nu}$-completions of $\mathcal{B}_\mathcal{X}$.

2. *The sets $A \in (\mathcal{B}_\mathcal{X})_\bullet$ will be called the* universally measurable subsets of $\mathcal{X}$.

3. *We will write $\mathcal{X}_\bullet$ to refer to $\mathcal{X}$ equipped with the σ-algebra $(\mathcal{B}_\mathcal{X})_\bullet$.*

4. *We call $\mathcal{X}$, $\mathcal{B}_\mathcal{X}$, resp.,* universally complete *if $\mathcal{B}_\mathcal{X} = (\mathcal{B}_\mathcal{X})_\bullet$.*

Remark B.3. *For a measurable space $\mathcal{X}$ and $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$ we clearly have the following inclusions:*

$$\mathcal{B}_{\mathcal{X}} \subseteq (\mathcal{B}_{\mathcal{X}})_{\bullet} \subseteq (\mathcal{B}_{\mathcal{X}})_{\boldsymbol{\mu}}.$$

Notation B.4. *Let $\mathcal{X}$ be a set and $\mathcal{E} \subseteq 2^{\mathcal{X}}$ a non-empty set of subsets of $\mathcal{X}$. We then write:*

$$(\mathcal{E})_{\bullet} := (\sigma(\mathcal{E}))_{\bullet} := \bigcap_{\boldsymbol{\nu} \in \mathcal{P}(\sigma(\mathcal{E}))} (\sigma(\mathcal{E}))_{\boldsymbol{\nu}},$$

for the smallest universally complete σ-algebra containing $\mathcal{E}$.

Definition B.5. *Let $f : \mathcal{X} \to \mathcal{Y}$ be a (not necessarily measurable) map between measurable spaces. f will be called* universally measurable *if it is $(\mathcal{B}_{\mathcal{X}})_{\bullet}$-$(\mathcal{B}_{\mathcal{Y}})_{\bullet}$-measurable.*

Lemma B.6. *Let $\mathcal{X}$ and $\mathcal{Y}$ be measurable spaces and $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$. Let $f : \mathcal{X} \to \mathcal{Y}$ be a (not necessarily measurable) map. We then have:*

1. *If f is $\mathcal{B}_{\mathcal{X}}$-$\mathcal{B}_{\mathcal{Y}}$-measurable then also $(\mathcal{B}_{\mathcal{X}})_{\boldsymbol{\mu}}$-$(\mathcal{B}_{\mathcal{Y}})_{f_*\boldsymbol{\mu}}$-measurable, where $f_*\boldsymbol{\mu}$ is the push-forward measure of $\boldsymbol{\mu}$.*

2. *If f is $\mathcal{B}_{\mathcal{X}}$-$\mathcal{B}_{\mathcal{Y}}$-measurable then also $(\mathcal{B}_{\mathcal{X}})_{\bullet}$-$(\mathcal{B}_{\mathcal{Y}})_{\bullet}$-measurable.*

3. *f is $(\mathcal{B}_{\mathcal{X}})_{\bullet}$-$\mathcal{B}_{\mathcal{Y}}$-measurable if and only if it is $(\mathcal{B}_{\mathcal{X}})_{\bullet}$-$(\mathcal{B}_{\mathcal{Y}})_{\bullet}$-measurable.*

4. *If f is $(\mathcal{B}_{\mathcal{X}})_{\bullet}$-$(\mathcal{B}_{\mathcal{Y}})_{\bullet}$-measurable then also $(\mathcal{B}_{\mathcal{X}})_{\boldsymbol{\mu}}$-$(\mathcal{B}_{\mathcal{Y}})_{f_*\boldsymbol{\mu}}$-measurable.*

Proof. 1.) If $B \in (\mathcal{B}_{\mathcal{Y}})_{f_*\boldsymbol{\mu}}$ then $B = C \triangle M$ with $C \in \mathcal{B}_{\mathcal{Y}}$ and $M \subseteq N$ with $\boldsymbol{\mu}(f^{-1}(N)) = 0$ for some $N \in \mathcal{B}_{\mathcal{Y}}$. Then $f^{-1}(B) = f^{-1}(C) \triangle f^{-1}(M)$ with $f^{-1}(B) \in \mathcal{B}_{\mathcal{X}}$, $f^{-1}(M) \subseteq f^{-1}(N) \in \mathcal{B}_{\mathcal{X}}$ and $\boldsymbol{\mu}(f^{-1}(N)) = 0$. So $f^{-1}(B) \in (\mathcal{B}_{\mathcal{X}})_{\boldsymbol{\mu}}$.
2.) This follows from the previous point by noting that:

$$(\mathcal{B}_{\mathcal{Y}})_{\bullet} = \bigcap_{\boldsymbol{\nu} \in \mathcal{P}(\mathcal{Y})} (\mathcal{B}_{\mathcal{Y}})_{\boldsymbol{\nu}} \subseteq \bigcap_{\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})} (\mathcal{B}_{\mathcal{Y}})_{f_*\boldsymbol{\mu}}.$$

3.) Clear from point 2.
4.) Same as point 1. with $(\mathcal{B}_{\mathcal{X}})_{\bullet} \subseteq (\mathcal{B}_{\mathcal{X}})_{\boldsymbol{\mu}}$. ☐

B.2. Analytic Subsets of a Measurable Space

Definition B.7 (Analytic subsets, see [Fre15] 421H, 423E)**.** *Let $\mathcal{X}$ be a measurable space and $\mathbb{R}$ be equipped with its Borel σ-algebra.*

1. *A subset $A \subseteq \mathcal{X}$ will be called* $\mathcal{B}_{\mathcal{X}}$-analytic, Souslin-$\mathcal{B}_{\mathcal{X}}$ *or just* analytic *(if the context is clear) if A is the projection of a measurable set $C \in \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathcal{X}}$ of the product space $\mathbb{R} \times \mathcal{X}$:*

$$A = \mathrm{pr}_{\mathcal{X}}(C) := \{x \in \mathcal{X} \,|\, \exists u \in \mathbb{R} : (u, x) \in C\}.$$

2. *The set of* analytic subsets *of* $\mathcal{X}$ *is then abbreviated by:*

$$\mathcal{S}(\mathcal{B}_\mathcal{X}) := \{\mathrm{pr}_\mathcal{X}(C) \,|\, C \in \mathcal{B}_\mathbb{R} \otimes \mathcal{B}_\mathcal{X}\}.$$

   Note that $\mathcal{B}_\mathcal{X} \subseteq \mathcal{S}(\mathcal{B}_\mathcal{X})$ *and that* $\mathcal{S}(\mathcal{B}_\mathcal{X})$ *is closed under countable unions and countable intersections, but in general NOT w.r.t. complements (see [Fre15] 421E).*

3. *We denote the* σ*-algebra generated by the analytic subsets by:* $\sigma\mathcal{S}(\mathcal{B}_\mathcal{X})$.

4. *We will write* $\mathcal{X}_\mathcal{S}$ *to refer to* $\mathcal{X}$ *equipped with the* σ*-algebra* $\sigma\mathcal{S}(\mathcal{B}_\mathcal{X})$.

Lemma B.8 (See [Fre15] 421C (c)+(d))**.** *Let* $f : (\mathcal{X}, \mathcal{B}_\mathcal{X}) \to (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ *be a measurable map between any measurable spaces.*

1. *If* $B \in \mathcal{S}(\mathcal{B}_\mathcal{Y})$ *then* $f^{-1}(B) \in \mathcal{S}(\mathcal{B}_\mathcal{X})$.

2. *If* f *is surjective,* $A \in \mathcal{S}(f^*\mathcal{B}_\mathcal{Y})$ *then* $f(A) \in \mathcal{S}(\mathcal{B}_\mathcal{Y})$.

Note that we use: $f^*\mathcal{B}_\mathcal{Y} := \{f^{-1}(B) \subseteq \mathcal{X} \,|\, B \in \mathcal{B}_\mathcal{Y}\}$.

Lemma B.9 (See [Fre15] 431A, 421D)**.** *For any measurable space* $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ *and probability measure* $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$ *we have the inclusion of* σ*-algebras:*

$$\mathcal{B}_\mathcal{X} \subseteq \sigma\mathcal{S}(\mathcal{B}_\mathcal{X}) \subseteq (\mathcal{B}_\mathcal{X})_\bullet \subseteq (\mathcal{B}_\mathcal{X})_{\boldsymbol{\mu}}.$$

Furthermore, we have the equalities:

$$\mathcal{S}((\mathcal{B}_\mathcal{X})_\bullet) = (\mathcal{B}_\mathcal{X})_\bullet, \qquad \text{and} \qquad \mathcal{S}((\mathcal{B}_\mathcal{X})_{\boldsymbol{\mu}}) = (\mathcal{B}_\mathcal{X})_{\boldsymbol{\mu}}.$$

Lemma B.10. *Let* $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ *be a a measurable space. Then the identity maps:*

$$\mathrm{id}: \ (\mathcal{X}, (\mathcal{B}_\mathcal{X})_\bullet) \to (\mathcal{X}, \sigma\mathcal{S}(\mathcal{B}_\mathcal{X})) \to (\mathcal{X}, \mathcal{B}_\mathcal{X}),$$

are measurable. They induce measurable bijective restriction maps:

$$\mathrm{res}: \ \mathcal{P}(\mathcal{X}_\bullet) \to \mathcal{P}(\mathcal{X}_\mathcal{S}) \to \mathcal{P}(\mathcal{X}), \qquad \boldsymbol{\mu} \mapsto \boldsymbol{\mu}_|,$$

where we endow the spaces of probability measures each with the smallest σ*-algebra such that the evaluation maps:*

$$j_A : \ \mathcal{P}(\mathcal{X}, \mathcal{B}) \to [0,1], \quad j_A(\boldsymbol{\nu}) := \boldsymbol{\nu}(A),$$

are measurable for all $A \in \mathcal{B}$*, where we use* $\mathcal{B} = (\mathcal{B}_\mathcal{X})_\bullet$*,* $\mathcal{B} = \sigma\mathcal{S}(\mathcal{B}_\mathcal{X})$*,* $\mathcal{B} = \mathcal{B}_\mathcal{X}$*, resp. Note that the inverses of the bijective maps might not be measurable in general.*

Lemma B.11. *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be measurable spaces,* $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$ *and* $f : \mathcal{X} \to \mathcal{Y}$ *be a map.*

1. *If* f *is* $\mathcal{B}_\mathcal{X}$*-*$\mathcal{B}_\mathcal{Y}$*-measurable then also* $\sigma\mathcal{S}(\mathcal{B}_\mathcal{X})$*-*$\sigma\mathcal{S}(\mathcal{B}_\mathcal{Y})$*-measurable.*

2. *If* f *is* $\sigma\mathcal{S}(\mathcal{B}_\mathcal{X})$*-*$\sigma\mathcal{S}(\mathcal{B}_\mathcal{Y})$*-measurable then also* $(\mathcal{B}_\mathcal{X})_\bullet$*-*$(\mathcal{B}_\mathcal{Y})_\bullet$*-measurable.*

3. *If* f *is* $(\mathcal{B}_\mathcal{X})_\bullet$*-*$(\mathcal{B}_\mathcal{Y})_\bullet$*-measurable then also* $(\mathcal{B}_\mathcal{X})_{\boldsymbol{\mu}}$*-*$(\mathcal{B}_\mathcal{Y})_{f_*\boldsymbol{\mu}}$*-measurable.*

B.3. Countably Generated and Countably Separated Sigma-Algebras

Definition B.12 ((Universally) countably generated σ-algebras). *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be a measurable space. Then $\mathcal{X}$ or $\mathcal{B}_\mathcal{X}$ are called:*

1. countably generated *if there exists a countable $\mathcal{E} \subseteq \mathcal{B}_\mathcal{X}$ such that $\sigma(\mathcal{E}) = \mathcal{B}_\mathcal{X}$.*

2. universally countably generated *if there exists a countable $\mathcal{E} \subseteq (\mathcal{B}_\mathcal{X})_\bullet$ such that $(\mathcal{E})_\bullet = (\mathcal{B}_\mathcal{X})_\bullet$.*

Clearly, the first implies the second.

Theorem B.13 (See [Bog07] Thm. 6.5.5). *Let $\mathcal{X}$ be a measurable space. The following are equivalent:*

1. *$\mathcal{X}$ is countably generated.*

2. *There exists a measurable map $f : \mathcal{X} \to [0,1]$ such that $\mathcal{B}_\mathcal{X} = f^*\mathcal{B}_{[0,1]}$.*

Here $\mathcal{B}_{[0,1]}$ is the Borel σ-algebra of $[0,1]$. We denote $f^*\mathcal{B}_\mathcal{Y} := \{f^{-1}(C) \,|\, C \in \mathcal{B}_\mathcal{Y}\}$.

Proof. If $\mathcal{B}_\mathcal{X} = \sigma(\{A_n \,|\, n \in \mathbb{N}\})$ then:

$$f := \sum_{n\in\mathbb{N}} 3^{-n} \cdot \mathbb{1}_{A_n} : \quad \mathcal{X} \to [0,1]$$

does the trick. The reverse is clear. □

Corollary B.14. *Let $\mathcal{X}$ be a measurable space. The following are equivalent:*

1. *$\mathcal{X}$ is universally countably generated.*

2. *There exists a universally measurable map $f : \mathcal{X} \to [0,1]$ such that $(\mathcal{B}_\mathcal{X})_\bullet = (f^*\mathcal{B}_{[0,1]})_\bullet$.*

Proof. Assume that $\mathcal{X}$ is universally countably generated. Let $(\mathcal{B}_\mathcal{X})_\bullet = (\mathcal{E})_\bullet$ with countable $\mathcal{E} \subseteq (\mathcal{B}_\mathcal{X})_\bullet$. Then by Theorem B.13 there exists a measurable map:

$$f : (\mathcal{X}, \sigma(\mathcal{E})) \to ([0,1], \mathcal{B}_{[0,1]}),$$

such that $f^*\mathcal{B}_{[0,1]} = \sigma(\mathcal{E}) \subseteq (\mathcal{B}_\mathcal{X})_\bullet$. It thus follows that: $(\mathcal{B}_\mathcal{X})_\bullet = (\sigma(\mathcal{E}))_\bullet = (f^*\mathcal{B}_{[0,1]})_\bullet$. This shows the claim.
For the reverse implication now assume that $(\mathcal{B}_\mathcal{X})_\bullet = (f^*\mathcal{B}_{[0,1]})_\bullet$. Since $\mathcal{B}_{[0,1]}$ is countably generated so is $f^*\mathcal{B}_{[0,1]}$ and we get the claim. □

This result inspires another definition:

Definition B.15 (Universally generated σ-algebras). *A measurable space $\mathcal{X}$ will be called* universally generated *if there exists a universally measurable map $f : \mathcal{X} \to [0,1]$ such that $(\mathcal{B}_\mathcal{X})_\bullet = \big(f^*(\mathcal{B}_{[0,1]})_\bullet\big)_\bullet$.*

Definition B.16 ((Universally) countably separated σ-algebras)**.** *Let* $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ *be a measurable space. Then* $\mathcal{X}$ *or* $\mathcal{B}_\mathcal{X}$ *are called:*

1. countably separated *if there exists a countable* $\mathcal{E} \subseteq \mathcal{B}_\mathcal{X}$ *that separates the points of* $\mathcal{X}$*, i.e. for all* $x_1, x_2 \in \mathcal{X}$ *with* $x_1 \neq x_2$ *there exists* $A \in \mathcal{E}$ *such that* $\#(A \cap \{x_1, x_2\}) = 1$.

2. universally countably separated *if there exists a countable* $\mathcal{E} \subseteq (\mathcal{B}_\mathcal{X})_\bullet$ *that separates the points of* $\mathcal{X}$.

Clearly, the first implies the second.

Theorem B.17 (See [Bog07] Thm. 6.5.7)**.** *Let* $\mathcal{X}$ *be a measurable space and* $\Delta_\mathcal{X} := \{(x, x) \in \mathcal{X} \times \mathcal{X} \,|\, x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{X}$ *its diagonal. The following are equivalent:*

1. $\mathcal{X}$ *is countably separated.*

2. *There exists an injective measurable map* $f : \mathcal{X} \to [0, 1]$.

3. $\Delta_\mathcal{X} \in \mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{X}$.

Proof. If $\mathcal{E} = \{A_n \,|\, n \in \mathbb{N}\} \subseteq \mathcal{B}_\mathcal{X}$ separates the points of $\mathcal{X}$ then:

$$f := \sum_{n \in \mathbb{N}} 3^{-n} \cdot \mathbb{1}_{A_n} : \quad \mathcal{X} \to [0, 1]$$

is injective and measurable. For the rest see [Bog07] Thm. 6.5.7. □

Corollary B.18. *Let* $\mathcal{X}$ *be a measurable space. The following are equivalent:*

1. $\mathcal{X}$ *is universally countably separated.*

2. *There exists an injective universally measurable map* $f : \mathcal{X} \to [0, 1]$.

3. $\Delta_\mathcal{X} \in (\mathcal{B}_\mathcal{X})_\bullet \otimes (\mathcal{B}_\mathcal{X})_\bullet$.

Proof. This follows directly from Theorem B.17 and Definition B.16. For this note that we have the equivalence between $(\mathcal{B}_\mathcal{X})_\bullet$-$\mathcal{B}_\mathcal{Y}$-measurability and $(\mathcal{B}_\mathcal{X})_\bullet$-$(\mathcal{B}_\mathcal{Y})_\bullet$-measurability. □

This result inspires another definition:

Definition B.19 (Universally separated σ-algebras)**.** *A measurable space* $\mathcal{X}$ *will be called* universally separated *if* $\Delta_\mathcal{X} \in (\mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{X})_\bullet$.

Lemma B.20. *Let* $\mathcal{X}$ *be a measurable space.*

1. *If* $\mathcal{X}$ *is countably separated then we have* $\{x\} \in \mathcal{B}_\mathcal{X}$ *for all* $x \in \mathcal{X}$.

2. *If* $\mathcal{X}$ *is universally (countably) separated then we have* $\{x\} \in (\mathcal{B}_\mathcal{X})_\bullet$ *for all* $x \in \mathcal{X}$.

Proof. Let $\Delta_{\mathcal{X}} := \{(x,x) \in \mathcal{X} \times \mathcal{X} \,|\, x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{X}$ be the diagonal of $\mathcal{X}$. Fix $x_0 \in \mathcal{X}$. For any σ-algebra $\mathcal{B}$ on $\mathcal{X}$ the identity map $\mathrm{id} : \mathcal{X} \to \mathcal{X}$ and the constant map $c_{x_0} : \mathcal{X} \to \mathcal{X}$ with $c_{x_0}(x) := x_0$ are $\mathcal{B}$-$\mathcal{B}$-measurable. This implies that the map:

$$\iota_{x_0} : \mathcal{X} \to \mathcal{X} \times \mathcal{X}, \quad x \mapsto (x, x_0),$$

is $\mathcal{B}$-$(\mathcal{B} \otimes \mathcal{B})$-measurable and thus $(\mathcal{B})_\bullet$-$(\mathcal{B} \otimes \mathcal{B})_\bullet$-measurable. This implies that $\{x_0\} = \iota_{x_0}^{-1}(\Delta_{\mathcal{X}}) \in \mathcal{B}$ if $\Delta_{\mathcal{X}} \in \mathcal{B} \otimes \mathcal{B}$ and $\{x_0\} = \iota_{x_0}^{-1}(\Delta_{\mathcal{X}}) \in (\mathcal{B})_\bullet$ if $\Delta_{\mathcal{X}} \in (\mathcal{B} \otimes \mathcal{B})_\bullet$.
Each of the cases is implied by the assumptions by Theorem B.17, Corollary B.18 and Definition B.19. □

Corollary B.21. *Let $\mathcal{X}$ be a measurable space. The following are equivalent:*

1. *$\mathcal{X}$ is countably generated and countably separated.*
2. *There exists an injective measurable map $f : \mathcal{X} \to [0,1]$ such that $\mathcal{B}_{\mathcal{X}} = f^*\mathcal{B}_{[0,1]}$.*

Proof. If $\mathcal{B}_{\mathcal{X}} = \sigma(\mathcal{E})$ and $\mathcal{F} \subseteq \mathcal{B}_{\mathcal{X}}$ separates the points of $\mathcal{X}$ with countable $\mathcal{E}$, $\mathcal{F}$ then with $\mathcal{E} \cup \mathcal{F} = \{A_n \,|\, n \in \mathbb{N}\} \subseteq \mathcal{B}_{\mathcal{X}}$: and

$$f := \sum_{n \in \mathbb{N}} 3^{-n} \cdot \mathbb{1}_{A_n} : \quad \mathcal{X} \to [0,1]$$

we get the claim. The reverse is clear. □

Corollary B.22. *Let $\mathcal{X}$ be a measurable space. The following are equivalent:*

1. *$\mathcal{X}$ is universally countably generated and universally countably separated.*
2. *There exists an injective universally measurable map $f : \mathcal{X} \to [0,1]$ such that $(\mathcal{B}_{\mathcal{X}})_\bullet = (f^*\mathcal{B}_{[0,1]})_\bullet$.*

Proof. Assume the first. If $(\mathcal{B}_{\mathcal{X}})_\bullet = (\mathcal{E})_\bullet$ and $\mathcal{F}$ separates the points of $\mathcal{X}$ with countable $\mathcal{E}, \mathcal{F} \subseteq (\mathcal{B}_{\mathcal{X}})_\bullet$ then put $\mathcal{G} := \mathcal{E} \cup \mathcal{F} = \{A_n \,|\, n \in \mathbb{N}\}$. Again using:

$$f := \sum_{n \in \mathbb{N}} 3^{-n} \cdot \mathbb{1}_{A_n} : \quad \mathcal{X} \to [0,1],$$

gives us an injective $\sigma(\mathcal{G})$-$\mathcal{B}_{[0,1]}$-measurable map f with $f^*\mathcal{B}_{[0,1]} = \sigma(\mathcal{G}) \subseteq (\mathcal{B}_{\mathcal{X}})_\bullet$. From this it follow that $(\mathcal{B}_{\mathcal{X}})_\bullet = (\sigma(\mathcal{G}))_\bullet = (f^*\mathcal{B}_{[0,1]})_\bullet$. This shows the claim.
The reverse direction is clear from Corollaries B.18 and B.14. □

B.4. Standard, Analytic and Universal Measurable Spaces

Definition B.23. *Let $([0,1], \mathcal{B}_{[0,1]})$ be the unit interval with its Borel σ-algebra. A measurable space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ will be called:*

1. standard (Borel) *if $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}) \cong (\mathcal{Y}, \mathcal{B}_{[0,1]|\mathcal{Y}})$ with a $\mathcal{Y} \in \mathcal{B}_{[0,1]}$,*

2. analytic *if* $(\mathcal{X}, \mathcal{B}_\mathcal{X}) \cong (\mathcal{Y}, \mathcal{B}_{[0,1]|\mathcal{Y}})$ *with a* $\mathcal{Y} \in \mathcal{S}(\mathcal{B}_{[0,1]})$,

3. countably perfect *if* $(\mathcal{X}, \mathcal{B}_\mathcal{X}) \cong (\mathcal{Y}, \mathcal{B}_{[0,1]|\mathcal{Y}})$ *with a* $\mathcal{Y} \in (\mathcal{B}_{[0,1]})_\bullet$.

4. universal *if* $(\mathcal{X}, (\mathcal{B}_\mathcal{X})_\bullet) \cong (\mathcal{Y}, (\mathcal{B}_{[0,1]|\mathcal{Y}})_\bullet)$ *with a* $\mathcal{Y} \in (\mathcal{B}_{[0,1]})_\bullet$.

Here "$\cong$" means "measurably isomorphic" (i.e. using a measurable map with a measurable inverse) and we will in the following write $\mathcal{B}_{|\mathcal{X}}$ for the induced Borel (sub-)σ-algebra $\mathcal{B}_{[0,1]|\mathcal{Y}}$ in $(\mathcal{B}_\mathcal{X})_\bullet$ (for universal measurable spaces). In the first cases of course: $\mathcal{B}_\mathcal{X} = \mathcal{B}_{|\mathcal{X}}$.

Remark B.24. *1. The inclusion $\mathcal{B}_{[0,1]} \subseteq \mathcal{S}(\mathcal{B}_{[0,1]}) \subseteq (\mathcal{B}_{[0,1]})_\bullet$ shows that every standard measurable space is also an analytic measurable space and every analytic measurable space is countably perfect and (the universal completion of) any countably perfect measurable space is a universal measurable space.*

2. Note that in the definition of universal measurable spaces we have $(\mathcal{B}_\mathcal{X})_\bullet = (\mathcal{B}_{|\mathcal{X}})_\bullet$, but we don't need to have $\mathcal{B}_\mathcal{X} = \mathcal{B}_{|\mathcal{X}}$. Since the definition only depends on the universal completion we, nevertheless, can always replace $\mathcal{B}_\mathcal{X}$ with $\mathcal{B}_{|\mathcal{X}}$, when needed and which usually has the better properties.

Lemma B.25. *1. Any countable product of standard, analytic, countably perfect or universal measurable spaces, resp., is of the same type.*

2. If $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ is a standard, analytic, countably perfect or universal measurable space then $\mathcal{B}_{|\mathcal{X}}$ is countably generated and countably separated (cp. [Bog07] 6.5.8).

Proof. This directly follows from the Borel isomorphism $[0,1]^\mathbb{N} \cong [0,1]$ (see [Fre15] 424C). ☐

Example B.26 (Standard measurable spaces, see [Bog07] 6.8.8, [Fre15] 424C)**.** *Any* Polish space $\mathcal{X}$, *i.e. a separable completely metrizable space (e.g. $\mathbb{R}^k$, $[0,1]$, $\mathbb{N}^\mathbb{N}$, any topological manifold, any countable simplicial complex, etc.), is a standard measurable space with its Borel σ-algebra $\mathcal{B}_\mathcal{X}$.*

Example B.27 (Analytic measurable spaces, see [Bog07] 6.7.4)**.** *Any* Souslin space $\mathcal{X}$, *i.e. a Hausdorff space that is the continuous image of a Polish space, is an analytic measurable space with its Borel σ-algebra $\mathcal{B}_\mathcal{X}$.*

B.5. Perfect Measures and Perfect Measurable Spaces

Definition B.28 (Perfect measures and perfect measurable spaces)**.** *Let $\mathcal{X}$ be a measurable space. Then we call:*

1. $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$ perfect *if for all measurable maps:*

$$f: \mathcal{X} \to \mathbb{R},$$

we have that:

$$f_*((\mathcal{B}_\mathcal{X})_{\boldsymbol{\mu}}) = (\mathcal{B}_\mathbb{R})_{f_*\boldsymbol{\mu}}.$$

In particular: $f(\mathcal{X}) \in (\mathcal{B}_\mathbb{R})_{f_*\boldsymbol{\mu}}$.

2. $\mathcal{X}$ perfect *if every $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$ is perfect.*

Note that we define the push-forward σ-algebra of $\mathcal{B}$ along a map $f : \mathcal{X} \to \mathcal{Y}$ as:

$$f_*\mathcal{B} := \left\{C \subseteq \mathcal{Y} \mid f^{-1}(C) \in \mathcal{B}\right\}.$$

Theorem B.29 (See [Bog07] 7.5.7 (i), 7.5.7 (iii), 6.5.8, 7.5.6 (iv))**.** *Let $\mathcal{X}$ be a measurable space. Then the following statements are equivalent:*

1. *$\mathcal{X}$ is countably perfect.*
2. *$\mathcal{X}$ is countably generated, countably separated and perfect.*
3. *$\mathcal{X}$ is countably generated, countably separated and every probability measure $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$ has a countably compact approximating class (see [Fre15] 451A or [Bog07] §1.12(ii) for precise definitions).*
4. *$\mathcal{X}$ is countably generated, countably separated and there exists a (countably) compact class that approximates every probability measure $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$.*

Corollary B.30. *Let $\mathcal{X}$ be a measurable space. Then the following are equivalent:*

1. *$\mathcal{X}$ is a universal measurable space.*
2. *$\mathcal{X}$ is universally countably generated, univerally countably separated and perfect.*
3. *$\mathcal{X}$ is universally countably generated, universally countably separated and every probability measure $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{B}_\mathcal{X})$ has a countably compact approximating class.*
4. *$\mathcal{X}$ is universally countably generated, universally countably separated and there exists a (countably) compact class that approximates every probability measure $\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X})$.*

Example B.31.
1. *Every universally measurable subset $\mathcal{X} \in (\mathcal{B}_\mathcal{Y})_\bullet$ of an analytic or standard measurable space $\mathcal{Y}$ is a countably perfect space when equipped with the subspace σ-algebra of measurable sets: $\mathcal{B}_{\mathcal{Y}|\mathcal{X}}$.*
2. *Every universally measurable subset $\mathcal{X} \in (\mathcal{B}_\mathcal{Y})_\bullet$ of an analytic or standard measurable space $\mathcal{Y}$ is a countably perfect space when equipped with the subspace σ-algebra of universal measurable sets: $((\mathcal{B}_\mathcal{Y})_\bullet)_{|\mathcal{X}}$.*

Example B.32. *Consider a* Radon space *$\mathcal{X}$, i.e. a Hausdorff space where every Borel probability measure is inner regular w.r.t. the (closed) compact subsets. Let $\mathcal{B}_\mathcal{X}$ be the Borel σ-algebra. A* countable network *for the topology is a countable set $\mathcal{E} \subseteq 2^\mathcal{X}$ such that every open set O can be written as:*

$$O = \bigcup_{\substack{A \in \mathcal{E} \\ A \subseteq O}} A.$$

Then we have the following:

1. *If $\mathcal{X}$ has a countable network $\mathcal{E} \subseteq \mathcal{B}_\mathcal{X}$ then $\mathcal{X}$ is a countably perfect measurable space.*
2. *If $\mathcal{X}$ has a countable network $\mathcal{E} \subseteq (\mathcal{B}_\mathcal{X})_\bullet$ then $\mathcal{X}$ is a universal measurable space.*

Proof. The first statement is similar to the second.
The second statement follows from B.30. For this let $\mathcal{T}_\mathcal{X}$ be the set of open sets. By assumption every open set $A \in \mathcal{T}_\mathcal{X}$ is the (countable) union of elements of $\mathcal{E} \subseteq (\mathcal{B}_\mathcal{X})_\bullet$. So we have: $\mathcal{T}_\mathcal{X} \subseteq \sigma(\mathcal{E})$ and thus:

$$\mathcal{B}_\mathcal{X} = \sigma(\mathcal{T}_\mathcal{X}) \subseteq \sigma(\mathcal{E}) \subseteq (\mathcal{B}_\mathcal{X})_\bullet.$$

Since $\mathcal{X}$ is Hausdorff $\mathcal{T}_\mathcal{X}$ separates points, so $\mathcal{E}$ as well. Furthermore, on Radon spaces every Borel measure is Radon. And in general every Radon measure is perfect (see [Bog07] 7.5.10(i)). Or, alternatively, the compact sets build a compact approximating class for every Borel measure. So $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ is a universal measurable space.
Also see [Fre15] 451M for a bit more general situation. □

B.6. The Space of Probability Measures

Remark B.33 (The σ-algebra on the space of probability measures)**.** *Recall that for any measurable space $\mathcal{X}$ we endow the space of probability measures $\mathcal{P}(\mathcal{X})$ with the smallest σ-algebra $\mathcal{B}_{\mathcal{P}(\mathcal{X})}$ such that all the evaluation maps:*

$$j_A : \mathcal{P}(\mathcal{X}) \to [0,1], \qquad j_A(\boldsymbol{\nu}) := \boldsymbol{\nu}(A),$$

are measurable for $A \in \mathcal{B}_\mathcal{X}$.

Lemma B.34. *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be a measurable space and $\mathcal{B}_\mathcal{X} = \sigma(\mathcal{E})$ with a set $\mathcal{E}$ that is closed under pairwise intersections. Then we have:*

$$\mathcal{B}_{\mathcal{P}(\mathcal{X})} = \sigma\left(j_A^{-1}((\alpha,1]) \,\middle|\, A \in \mathcal{E}, \alpha \in \mathbb{Q}\right).$$

Note that:

$$j_A^{-1}((\alpha,1]) = \{\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X}) \,|\, \boldsymbol{\mu}(A) > \alpha\}.$$

Proof. Let $\mathcal{B}_\mathcal{E} := \sigma\left(\left\{j_A^*\mathcal{B}_{[0,1]} \,\middle|\, A \in \mathcal{E}\right\}\right)$. The system:

$$\mathcal{D} := \left\{A \in \mathcal{B}_\mathcal{X} \,\middle|\, j_A = (\boldsymbol{\mu} \mapsto \boldsymbol{\mu}(A)) \text{ is } \mathcal{B}_\mathcal{E}\text{-}\mathcal{B}_{[0,1]}\text{-measurable}\right\}$$

is a Dynkin system that contains the π-system $\mathcal{E}$. So $\mathcal{B}_\mathcal{X} = \sigma(\mathcal{E}) \subseteq \mathcal{D}$ by Dynkin's lemma (see [Kle14] Thm. 1.19). So all j_A are $\mathcal{B}_\mathcal{E}$-$\mathcal{B}_{[0,1]}$-measurable. So by definition of $\mathcal{B}_{\mathcal{P}(\mathcal{X})}$ as the smallest σ-algebra for which all j_A are measurable we get:

$$\sigma\left(\left\{j_A^*\mathcal{B}_{[0,1]} \,\middle|\, A \in \mathcal{B}_\mathcal{X}\right\}\right) = \mathcal{B}_{\mathcal{P}(\mathcal{X})} \subseteq \mathcal{B}_\mathcal{E} = \sigma\left(\left\{j_A^*\mathcal{B}_{[0,1]} \,\middle|\, A \in \mathcal{E}\right\}\right).$$

Thus equality: $\mathcal{B}_{\mathcal{P}(\mathcal{X})} = \mathcal{B}_\mathcal{E}$. Since $\mathcal{B}_{[0,1]} = \sigma(\mathcal{F})$ with the countable π-system $\mathcal{F} := \{(\alpha,1] \,|\, \alpha \in \mathbb{Q}\}$ we get:

$$\mathcal{B}_{\mathcal{P}(\mathcal{X})} = \sigma\left(j_A^{-1}((\alpha,1]) \,\middle|\, A \in \mathcal{E}, \alpha \in \mathbb{Q}\right).$$

□

Lemma B.35. *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be measurable spaces.*

1. *If $f : \mathcal{X} \to \mathcal{Y}$ is measurable then the induced map:*

$$f_* : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y}), \quad \boldsymbol{\mu} \mapsto f_*\boldsymbol{\mu} = (B \mapsto \boldsymbol{\mu}(f^{-1}(B)))$$

   is measurable as well.

2. *The map:*

$$\delta : \mathcal{X} \to \mathcal{P}(\mathcal{X}), \quad x \mapsto \boldsymbol{\delta}_x,$$

   is measurable and $\delta^*\mathcal{B}_{\mathcal{P}(\mathcal{B}_\mathcal{X})} = \mathcal{B}_\mathcal{X}$. δ is injective iff $\mathcal{B}_\mathcal{X}$ separates points.

3. *The map:*

$$\begin{array}{ccc} \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) & \to & \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \\ (\boldsymbol{\mu}, \boldsymbol{\nu}) & \mapsto & \boldsymbol{\mu} \otimes \boldsymbol{\nu} \end{array}$$

   is measurable.

4. *If $g : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is measurable then the map:*

$$\begin{array}{ccl} \mathcal{P}(\mathcal{X}) \times \mathcal{Y} & \to & \mathcal{P}(\mathcal{Z}), \\ (\boldsymbol{\mu}, y) & \mapsto & g^y_*\boldsymbol{\mu} := g_*(\boldsymbol{\mu} \otimes \boldsymbol{\delta}_y) \\ & = & (C \mapsto \boldsymbol{\mu}(\{x \in \mathcal{X} \,|\, g(x, y) \in C\})) \end{array}$$

   is measurable as well.

Proof. 1.) For every $B \in \mathcal{B}_\mathcal{Y}$ we have:

$$(j_B \circ f_*)(\boldsymbol{\mu}) = \boldsymbol{\mu}(f^{-1}(B)) = j_{f^{-1}(B)}(\boldsymbol{\mu}).$$

Since the latter is measurable in $\boldsymbol{\mu}$ by definition the former is as well. Since this holds for all $B \in \mathcal{B}_\mathcal{Y}$ the measurability of f_* is shown.
2.) Let $A \in \mathcal{B}_\mathcal{X}$ then we have:

$$x \mapsto j_A \circ \boldsymbol{\delta}_x = \mathbb{1}_A(x)$$

is measurable in x. So $x \mapsto \boldsymbol{\delta}_x$ is measurable.
3.) For $A \in \mathcal{B}_\mathcal{X}$ and $B \in \mathcal{B}_\mathcal{Y}$ the composition of maps:

$$\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \xrightarrow{j_A \times j_B} [0,1] \times [0,1] \xrightarrow{\cdot} [0,1], \quad (\boldsymbol{\mu}, \boldsymbol{\nu}) \mapsto \boldsymbol{\mu}(A) \cdot \boldsymbol{\nu}(B),$$

is clearly measurable. Since this map agrees with the following map:

$$\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \xrightarrow{j_{A \times B}} [0,1], \quad (\boldsymbol{\mu}, \boldsymbol{\nu}) \mapsto \boldsymbol{\mu}(A) \cdot \boldsymbol{\nu}(B),$$

we have the measurability of the latter map as well. Now let:

$$\mathcal{D} := \{D \in \mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y} \,|\, (\boldsymbol{\mu}, \boldsymbol{\nu}) \mapsto (\boldsymbol{\mu} \otimes \boldsymbol{\nu})(D) \text{ is measurable}\}.$$

By aboves arguments we see that $\mathcal{D}$ contains the system $\mathcal{E} := \{A \times B \,|\, A \in \mathcal{B}_\mathcal{X}, B \in \mathcal{B}_\mathcal{Y}\}$, which is stable under finite intersections. Since $\boldsymbol{\mu} \otimes \boldsymbol{\nu}$ are probability measures it is easily verified that with $D \in \mathcal{D}$ also $D^c \in \mathcal{D}$ and with $D_n \in \mathcal{D}$, $n \in \mathbb{N}$, pairwise disjoint also $\bigcup_{n\in\mathbb{N}} D_n \in \mathcal{D}$. So $\mathcal{D}$ is a Dynkin system. By Dynkin's lemma (see [Kle14] Thm. 1.19) we see that:

$$\mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y} = \sigma(\mathcal{E}) \subseteq \mathcal{D}.$$

This means that for every $D \in \mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y}$ the composition of maps:

$$\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \xrightarrow{j_D} [0,1], \quad (\boldsymbol{\mu}, \boldsymbol{\nu}) \mapsto (\boldsymbol{\mu} \otimes \boldsymbol{\nu})(D),$$

is measurable. So the map

$$\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \mathcal{P}(\mathcal{X} \times \mathcal{Y}), \quad (\boldsymbol{\mu}, \boldsymbol{\nu}) \mapsto \boldsymbol{\mu} \otimes \boldsymbol{\nu},$$

is measurable by definition of the induced σ-algebras.
4.) If $g : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and $C \in \mathcal{B}_\mathcal{Z}$ then by the previous points we get the measurability of the maps:

$$\mathcal{P}(\mathcal{X}) \times \mathcal{Y} \xrightarrow{\mathrm{id}\times\boldsymbol{\delta}} \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \xrightarrow{\otimes} \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \xrightarrow{g_*} \mathcal{P}(\mathcal{Z}) \xrightarrow{j_C} [0,1].$$

It maps $(\boldsymbol{\mu}, y) \mapsto (g_*(\boldsymbol{\mu} \otimes \boldsymbol{\delta}_y))(C)$. Furthermore, we have:

$$g^{-1}(C)^y = \{x \in \mathcal{X} \,|\, g(x,y) \in C\} = (g^y)^{-1}(C).$$

Furthermore, for $D \in \mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y}$ we have:

$$(\boldsymbol{\mu} \otimes \boldsymbol{\delta}_y)(D) = \boldsymbol{\mu}(D^y),$$

which again by Dynkin's lemma is enough to test on sets $D = A \times B$, where it is obvious:

$$(\boldsymbol{\mu} \otimes \boldsymbol{\delta}_y)(A \times B) = \boldsymbol{\mu}(A) \cdot \mathbb{1}_B(y) = \boldsymbol{\mu}((A \times B)^y).$$

Together we get:

$$j_C(g^y_*\boldsymbol{\mu}) = (g^y_*\boldsymbol{\mu})(C) = \boldsymbol{\mu}(g^{-1}(C)^y) = (\boldsymbol{\mu} \otimes \boldsymbol{\delta}_y)(g^{-1}(C)) = j_C(g_*(\boldsymbol{\mu} \otimes \boldsymbol{\delta}_y)).$$

Since the latter is measurable in $(\boldsymbol{\mu}, y)$ by aboves arguments also the former is. Since this holds for all $C \in \mathcal{B}_\mathcal{Z}$ we have shown the measurability of the claimed map: $(\boldsymbol{\mu}, y) \mapsto g^y_*\boldsymbol{\mu}$. ☐

Lemma B.36. *If $\mathcal{X}$ is countably generated then $\mathcal{P}(\mathcal{X})$ is countably generated and countably separated.*

Proof. If $\mathcal{B}_\mathcal{X} = \sigma(\mathcal{E})$ with countable $\mathcal{E}$, then the algebra generated by $\mathcal{E}$ is countable and closed under finite intersections, so w.l.o.g. $\mathcal{E}$ an algebra. Then $\mathcal{E} \times \mathbb{Q}$ is countable and we have by Lemma B.34:

$$\mathcal{B}_{\mathcal{P}(\mathcal{X})} = \sigma\left(\left\{ j_A^{-1}((\alpha, 1]) \,\middle|\, A \in \mathcal{E}, \alpha \in \mathbb{Q} \right\}\right).$$

This shows that $\mathcal{B}_{\mathcal{P}(\mathcal{X})}$ is countably generated. Now consider $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{P}(\mathcal{X})$ with $\boldsymbol{\mu} \neq \boldsymbol{\nu}$. Assume that $\boldsymbol{\mu}(A) = \boldsymbol{\nu}(A)$ for all $A \in \mathcal{E}$. Then define:

$$\mathcal{D} := \{A \in \mathcal{B}_{\mathcal{X}} \mid \boldsymbol{\mu}(A) = \boldsymbol{\nu}(A)\}.$$

Then $\mathcal{D}$ is a Dynkin system that contains the π-system $\mathcal{E}$. So $\mathcal{B}_{\mathcal{X}} = \sigma(\mathcal{E}) \subseteq \mathcal{D}$ by Dynkin's lemma (see [Kle14] Thm. 1.19). But this means $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ agree on all $A \in \mathcal{B}_{\mathcal{X}}$ and are thus equal, in contradiction to the assumption. So there must be an $A \in \mathcal{E}$ such that $\boldsymbol{\nu}(A) \neq \boldsymbol{\mu}(A)$. So there exists an $\alpha \in \mathbb{Q}$ with either $\boldsymbol{\nu}(A) > \alpha > \boldsymbol{\mu}(A)$ or $\boldsymbol{\nu}(A) < \alpha < \boldsymbol{\mu}(A)$. In the first case $\boldsymbol{\nu} \in j_A^{-1}((\alpha, 1])$ and $\boldsymbol{\mu} \notin j_A^{-1}((\alpha, 1])$ and in the second case $\boldsymbol{\nu} \notin j_A^{-1}((\alpha, 1])$ and $\boldsymbol{\mu} \in j_A^{-1}((\alpha, 1])$. So the above system also separates points of $\mathcal{P}(\mathcal{X})$. This shows the claim. □

Theorem B.37 (See [Res77] Thm. 3). *Let $\mathcal{X}$ be an analytic measurable space.*

1. *If $A \in \sigma\mathcal{S}(\mathcal{B}_{\mathcal{X}})$ then j_A is $\sigma\mathcal{S}(\mathcal{B}_{\mathcal{P}(\mathcal{X})})$-$\mathcal{B}_{[0,1]}$-measurable.*
2. *If $A \in \mathcal{S}(\mathcal{B}_{\mathcal{X}})$ then $j_A^{-1}((\alpha, 1]) \in \mathcal{S}(\mathcal{B}_{\mathcal{P}(\mathcal{X})})$ for every $\alpha \in \mathbb{R}$.*

Theorem B.38 (See [Res77] Thm. 4). *Let $\mathcal{X}$ be any measurable space.*

1. *If $A \in (\mathcal{B}_{\mathcal{X}})_\bullet$ then j_A is $(\mathcal{B}_{\mathcal{P}(\mathcal{X})})_\bullet$-$\mathcal{B}_{[0,1]}$-measurable.*
2. *We have the inclusions: $\mathcal{B}_{\mathcal{P}(\mathcal{X})} \subseteq \mathcal{B}_{\mathcal{P}(\mathcal{X}_\bullet)} \subseteq (\mathcal{B}_{\mathcal{P}(\mathcal{X})})_\bullet = (\mathcal{B}_{\mathcal{P}(\mathcal{X}_\bullet)})_\bullet$.*

Proof. The first statement is due to [Res77] Thm. 4. We reproduce the proof here. Let $A \in (\mathcal{B}_{\mathcal{X}})_\bullet$. For the claim we need to show that $j_A^{-1}((\alpha, 1]) \in (\mathcal{B}_{\mathcal{P}(\mathcal{X})})_\bullet$ for all $\alpha \in \mathbb{R}$. So we fix $\alpha \in \mathbb{R}$ and $\boldsymbol{\rho} \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$. It is thus enough to show that $j_A^{-1}((\alpha, 1]) \in (\mathcal{B}_{\mathcal{P}(\mathcal{X})})_{\boldsymbol{\rho}}$. Now define $\boldsymbol{\nu} \in \mathcal{P}(\mathcal{X})$ for $B \in \mathcal{B}_{\mathcal{X}}$ as follows:

$$\boldsymbol{\nu}(B) := \int_{\mathcal{P}(\mathcal{X})} \mu(B)\, d\boldsymbol{\rho}(\mu) = \int_{\mathcal{P}(\mathcal{X})} j_B(\mu)\, d\boldsymbol{\rho}(\mu).$$

Since $A \in (\mathcal{B}_{\mathcal{X}})_\bullet \subseteq (\mathcal{B}_{\mathcal{X}})_{\boldsymbol{\nu}}$ there exist: $B_1, B_2 \in \mathcal{B}_{\mathcal{X}}$ such that:

$$B_1 \subseteq A \subseteq B_2, \qquad \boldsymbol{\nu}(B_2 \backslash B_2) = 0.$$

The last condition shows:

$$0 = \boldsymbol{\nu}(B_2 \backslash B_1) = \int j_{B_2 \backslash B_1}(\mu)\, \boldsymbol{\rho}(\mu).$$

Since $j_{B_2 \backslash B_1}(\mu) \geqslant 0$ and $j_{B_2 \backslash B_1} : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ is measurable we get that:

$$C := j_{B_2 \backslash B_1}^{-1}((0, 1]) = \{\mu \in \mathcal{P}(\mathcal{X}) \mid j_{B_2 \backslash B_1}(\mu) > 0\} \in \mathcal{B}_{\mathcal{P}(\mathcal{X})} \qquad \text{and} \qquad \boldsymbol{\rho}(C) = 0.$$

We now have the inclusions:

$$j_{B_1}^{-1}((\alpha, 1]) \subseteq j_A^{-1}((\alpha, 1]) \subseteq j_{B_2}^{-1}((\alpha, 1]),$$

with $j_{B_i}^{-1}((\alpha,1]) \in \mathcal{B}_{\mathcal{P}(\mathcal{X})}$, $i = 1, 2$, and:

$$j_{B_2}^{-1}((\alpha,1]) \backslash j_{B_1}^{-1}((\alpha,1]) \subseteq j_{B_2 \backslash B_1}^{-1}((0,1]) = C.$$

Indeed, if $\mu(B_2) > \alpha$ and $\mu(B_1) \leqslant \alpha$, then:

$$\mu(B_2 \backslash B_1) = \mu(B_2) - \mu(B_1) > \alpha - \alpha = 0.$$

Since we have $\boldsymbol{\rho}(C) = 0$ we also have:

$$0 \leqslant \boldsymbol{\rho}\left(j_{B_2}^{-1}((\alpha,1]) \backslash j_{B_1}^{-1}((\alpha,1])\right) \leqslant \boldsymbol{\rho}(C) = 0.$$

This shows that $j_A^{-1}((\alpha,1]) \in (\mathcal{B}_{\mathcal{P}(\mathcal{X})})_{\boldsymbol{\rho}}$, which shows the first claim as α and $\boldsymbol{\rho}$ were arbitrary.
Now consider the second statement.
The inclusion $\mathcal{B}_{\mathcal{P}(\mathcal{X})} \subseteq \mathcal{B}_{\mathcal{P}(\mathcal{X}_\bullet)}$ is clear since $\mathcal{B}_\mathcal{X} \subseteq (\mathcal{B}_\mathcal{X})_\bullet$.
The inclusion $\mathcal{B}_{\mathcal{P}(\mathcal{X}_\bullet)} \subseteq (\mathcal{B}_{\mathcal{P}(\mathcal{X})})_\bullet$ follows from the first point, since $\mathcal{B}_{\mathcal{P}(\mathcal{X}_\bullet)}$ is per definition the smallest σ-algebra such that all j_A for $A \in (\mathcal{B}_\mathcal{X})_\bullet$ are measurable.
The equality $(\mathcal{B}_{\mathcal{P}(\mathcal{X})})_\bullet = (\mathcal{B}_{\mathcal{P}(\mathcal{X}_\bullet)})_\bullet$ follows from the sandwich:

$$\mathcal{B}_{\mathcal{P}(\mathcal{X})} \subseteq \mathcal{B}_{\mathcal{P}(\mathcal{X}_\bullet)} \subseteq (\mathcal{B}_{\mathcal{P}(\mathcal{X})})_\bullet.$$

□

Corollary B.39. *Let $\mathcal{X}$ be universally countably generated then $\mathcal{P}(\mathcal{X})$ is universally countably generated and universally countably separated.*

Proof. By assumption we have a countably generated σ-algebra $\mathcal{E} \subseteq (\mathcal{B}_\mathcal{X})_\bullet$ such $(\mathcal{E})_\bullet = (\mathcal{B}_\mathcal{X})_\bullet$. By Lemma B.36 we have that $\mathcal{P}(\mathcal{X}, \mathcal{E})$ is countably generated and countably separated. If we now apply Theorem *B*.38 twice, once to $\mathcal{B}_\mathcal{X}$ and once to $\mathcal{E}$, we get the equalities:

$$(\mathcal{B}_{\mathcal{P}(\mathcal{X},\mathcal{E})})_\bullet = (\mathcal{B}_{\mathcal{P}(\mathcal{X},(\mathcal{E})_\bullet)})_\bullet = (\mathcal{B}_{\mathcal{P}(\mathcal{X},(\mathcal{B}_\mathcal{X})_\bullet)})_\bullet = (\mathcal{B}_{\mathcal{P}(\mathcal{X},\mathcal{B}_\mathcal{X})})_\bullet.$$

So $\mathcal{B}_{\mathcal{P}(\mathcal{X},\mathcal{E})}$ is a countably generated and countably separated σ-algebra that universally generates $(\mathcal{B}_{\mathcal{P}(\mathcal{X},\mathcal{B}_\mathcal{X})})_\bullet$. This shows the claim. □

Theorem B.40 (See [Sch74] Appendix §5 Thm. 7+8)**.** *Let $\mathcal{X}$ be a standard measurable space, an analytic measurable space, resp., then the space of probability measures $\mathcal{P}(\mathcal{X})$ is also a standard measurable space, an analytic measurable space, resp., (in the σ-algebra from this section).*

Lemma B.41. *Let $f : \mathcal{X} \to \mathcal{Y}$ be a measurable map between measurable spaces with $\mathcal{B}_\mathcal{X} = f^*\mathcal{B}_\mathcal{Y}$. Then the map:*

$$f_* : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y}), \quad \boldsymbol{\mu} \mapsto f_*\boldsymbol{\mu},$$

is injective, measurable and satisfies: $\mathcal{B}_{\mathcal{P}(\mathcal{X})} = (f_*)^*\mathcal{B}_{\mathcal{P}(\mathcal{Y})}$.

Proof. f_* is measurable since for any $B \in \mathcal{B}_\mathcal{Y}$ we have:

$$j_B(f_*\boldsymbol{\mu}) = j_{f^{-1}(B)}(\boldsymbol{\mu}),$$

and $j_{f^{-1}(B)}$ is measurable since $f^{-1}(B) \in \mathcal{B}_\mathcal{X}$.
f_* is injective. Indeed, if $f_*\boldsymbol{\mu} = f_*\boldsymbol{\nu}$ and $A \in \mathcal{B}_\mathcal{X} = f^*\mathcal{B}_\mathcal{Y}$ is of the form $A = f^{-1}(B)$ with $B \in \mathcal{B}_\mathcal{Y}$ then:

$$\boldsymbol{\mu}(A) = \boldsymbol{\mu}(f^{-1}(B)) = (f_*\boldsymbol{\mu})(B) = (f_*\boldsymbol{\nu})(B) = \boldsymbol{\nu}(f^{-1}(B)) = \boldsymbol{\nu}(A),$$

which implies $\boldsymbol{\mu} = \boldsymbol{\nu}$.
The measurability of f_* shows: $\mathcal{B}_{\mathcal{P}(\mathcal{X})} \supseteq (f_*)^*\mathcal{B}_{\mathcal{P}(\mathcal{Y})}$.
For the reverse, consider generators: $j_A^{-1}((\alpha, 1]) \in \mathcal{B}_{\mathcal{P}(\mathcal{X})}$ for $\alpha \in \mathbb{R}$ and $A = f^{-1}(B) \in \mathcal{B}_\mathcal{X} = f^*\mathcal{B}_\mathcal{Y}$. We then get:

$$\begin{aligned}
j_A^{-1}((\alpha, 1]) &= \{\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X}) \,|\, \boldsymbol{\mu}(A) > \alpha\} \\
&= \left\{\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X}) \,|\, \boldsymbol{\mu}(f^{-1}(B)) > \alpha\right\} \\
&= \{\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X}) \,|\, (f_*\boldsymbol{\mu})(B) > \alpha\} \\
&= \{\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X}) \,|\, j_B((f_*\boldsymbol{\mu})) \in (\alpha, 1]\} \\
&= \left\{\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X}) \,|\, (f_*\boldsymbol{\mu}) \in j_B^{-1}((\alpha, 1])\right\} \\
&= \left\{\boldsymbol{\mu} \in \mathcal{P}(\mathcal{X}) \,|\, \boldsymbol{\mu} \in (f_*)^{-1}(j_B^{-1}((\alpha, 1]))\right\} \\
&= (f_*)^{-1}(j_B^{-1}((\alpha, 1])) \\
&\in (f_*)^*\mathcal{B}_{\mathcal{P}(\mathcal{Y})}.
\end{aligned}$$

This implies:

$$\mathcal{B}_{\mathcal{P}(\mathcal{X})} = \sigma\left(\left\{j_A^{-1}((\alpha, 1]) \,|\, \alpha \in \mathbb{R}, A \in \mathcal{B}_{\mathcal{P}(\mathcal{X})}\right\}\right) \subseteq (f_*)^*\mathcal{B}_{\mathcal{P}(\mathcal{Y})}.$$

□

Theorem B.42. *Let $\mathcal{Y}$ be an analytic measurable space and $\mathcal{X} \in \mathcal{B}_\mathcal{Y}$, $\mathcal{X} \in \mathcal{S}(\mathcal{B}_\mathcal{Y})$, resp. Then the inclusion map $\iota : (\mathcal{X}, \mathcal{B}_{\mathcal{Y}|\mathcal{X}}) \hookrightarrow (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ induces a canonical measurable isomorphism:*

$$\iota_* : \mathcal{P}(\mathcal{B}_{\mathcal{Y}|\mathcal{X}}) \cong \{\nu \in \mathcal{P}(\mathcal{B}_\mathcal{Y}) \,|\, \nu(\mathcal{X}) = 1\} \subseteq \mathcal{P}(\mathcal{B}_\mathcal{Y}).$$

Furthermore, $\mathcal{P}(\mathcal{B}_{\mathcal{Y}|\mathcal{X}}) \in \mathcal{B}_{\mathcal{P}(\mathcal{B}_\mathcal{Y})}$, $\mathcal{P}(\mathcal{B}_{\mathcal{Y}|\mathcal{X}}) \in \mathcal{S}(\mathcal{B}_{\mathcal{P}(\mathcal{B}_\mathcal{Y})})$, resp.

Proof. W.l.o.g. assume $\mathcal{B}_{|\mathcal{Y}} = \mathcal{B}_\mathcal{Y}$. The bijection can be found in [Fre15] 437Q, 434F(c). The additional claims follow from [HK99], [Bog07] 8.10.28. □

Lemma B.43. *Let $f : \mathcal{X} \to \mathcal{Y}$ be a measurable injective map between countably perfect measurable spaces with $\mathcal{B}_\mathcal{X} = f^*\mathcal{B}_\mathcal{Y}$. Then $f(\mathcal{X}) \in (\mathcal{B}_\mathcal{Y})_\bullet$ and f_* induces an identification:*

$$f_* : \quad \mathcal{P}(\mathcal{X}) \cong \{\boldsymbol{\nu} \in \mathcal{P}(\mathcal{Y}) \,|\, \boldsymbol{\nu}(f(\mathcal{X})) = 1\},$$

where the latter lies in $(\mathcal{B}_{\mathcal{P}(\mathcal{Y})})_\bullet$.

Proof. Since by definition we can embed $\mathcal{Y}$ into $[0,1]$ we can for the first statement w.l.o.g. assume $\mathcal{Y} = [0,1]$. Since $\mathcal{X}$ is perfect by [Bog07] Thm. 7.5.7 (i) we get the claim: $f(\mathcal{X}) \in (\mathcal{B}_\mathcal{Y})_\bullet$.
Since $f(\mathcal{X}) \in (\mathcal{B}_\mathcal{Y})_\bullet$ we can evaluate at $f(\mathcal{X})$. By Theorem B.38 the evaluation map $j_{f(\mathcal{X})}$ is $(\mathcal{B}_{\mathcal{P}(\mathcal{Y})})_\bullet$-$\mathcal{B}_{[0,1]}$-measurable. So we get that:

$$\mathcal{P}_1 := \{\boldsymbol{\nu} \in \mathcal{P}(\mathcal{Y}) \,|\, \boldsymbol{\nu}(f(\mathcal{X})) = 1\} = j_{f(\mathcal{X})}^{-1}(1) \in (\mathcal{B}_{\mathcal{P}(\mathcal{Y})})_\bullet.$$

By Lemma B.41 we know that f_* is injective, measurable and satisfies: $\mathcal{B}_{\mathcal{P}(\mathcal{X})} = (f_*)^*\mathcal{B}_{\mathcal{P}(\mathcal{Y})}$. It is clear that $f_*\mathcal{P}(\mathcal{X}) \subseteq \mathcal{P}_1$.
If now $\boldsymbol{\nu} \in \mathcal{P}_1$ then we define for $A \in \mathcal{B}_\mathcal{X} = f^*\mathcal{B}_\mathcal{Y}$:

$$\boldsymbol{\mu}(A) := \boldsymbol{\nu}(B \cap f(\mathcal{X}))$$

for $A = f^{-1}(B)$, $B \in \mathcal{B}_\mathcal{Y}$.
$\boldsymbol{\mu}$ is well-defined. If $B_1, B_2 \in \mathcal{B}_\mathcal{Y}$ with $f^{-1}(B_1) = A = f^{-1}(B_2)$ then, because of the injectivity of f, we get: $B_1 \cap f(\mathcal{X}) = B_2 \cap f(\mathcal{X})$. Also note that $B \cap f(\mathcal{X}) \in (\mathcal{B}_\mathcal{Y})_\bullet$ can be measured by $\boldsymbol{\nu}$.
It is clear that $\boldsymbol{\mu}$ is a probability measure and satisfies $f_*\boldsymbol{\mu} = \boldsymbol{\nu}$.

$$(f_*\boldsymbol{\mu})(B) = \boldsymbol{\mu}(f^{-1}(B)) = \boldsymbol{\nu}(B \cap f(\mathcal{X})) = \boldsymbol{\nu}(B \cap f(\mathcal{X})) + \underbrace{\boldsymbol{\nu}(B \cap f(\mathcal{X})^\mathrm{c})}_{=0} = \boldsymbol{\nu}(B).$$

□

Corollary B.44. *Let $\mathcal{X}$ be a countably perfect measurable space then the space of probability measures $\mathcal{P}(\mathcal{X})$ is also a countably perfect measurable space.*

Proof. $\mathcal{X}$ by definition can be embedded as a universally measurable set into $\mathcal{Y} = [0,1]$. By Lemma B.43 we get that $\mathcal{P}(\mathcal{X})$ is measurably isomorphic to a universally measurable subset of $\mathcal{P}(\mathcal{Y})$. By Theorem B.40 $\mathcal{P}(\mathcal{Y})$ is a standard measurable space. A universally measurable subset of standard measurable space with the induced σ-algebra is countably perfect (see [Bog07] Thm. 7.5.7 (i)). □

Corollary B.45. *Let $\mathcal{X}$ be a universal measurable space then the space of probability measures $\mathcal{P}(\mathcal{X})$ is a universal measurable space.*

Proof. By definition there is a countably generated and countably separated σ-algebra $\mathcal{B}_{|\mathcal{X}} \subseteq (\mathcal{B}_\mathcal{X})_\bullet$ such that $(\mathcal{X}, \mathcal{B}_{|\mathcal{X}})$ is countably perfect and $(\mathcal{B}_{|\mathcal{X}})_\bullet = (\mathcal{B}_\mathcal{X})_\bullet$. By Corollary B.44 the space $\mathcal{P}(\mathcal{X}, \mathcal{B}_{|\mathcal{X}})$ is countably perfect. By Theorem B.38 we also have the equalities:

$$(\mathcal{B}_{\mathcal{P}(\mathcal{X},\mathcal{B}_{|\mathcal{X}})})_\bullet = (\mathcal{B}_{\mathcal{P}(\mathcal{X},(\mathcal{B}_{|\mathcal{X}})_\bullet)})_\bullet = (\mathcal{B}_{\mathcal{P}(\mathcal{X},(\mathcal{B}_\mathcal{X})_\bullet)})_\bullet = (\mathcal{B}_{\mathcal{P}(\mathcal{X},\mathcal{B}_\mathcal{X})})_\bullet.$$

So $(\mathcal{B}_{\mathcal{P}(\mathcal{X},\mathcal{B}_\mathcal{X})})_\bullet$ is the universal completion of the σ-algebra of a countably perfect space. This shows that $\mathcal{P}(\mathcal{X})$ is a universal measurable space. □

B.7. Measurable Selection Theorem

Lemma B.46 (See [Fre15] 423G)**.** *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be an analytic measurable space and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ a countably separated measurable space (e.g. another analytic measurable space) and let $f : (\mathcal{X}, \mathcal{B}_\mathcal{X}) \to (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be a measurable map. We then have:*

$$A \in \mathcal{S}(\mathcal{B}_\mathcal{X}) \implies f(A) \in \mathcal{S}(\mathcal{B}_\mathcal{Y}).$$

We also have that the graph $\Gamma_f \in \mathcal{S}(\mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y})$.

Proof. See [Fre15] 423G for analytic measurable spaces. If $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ is countably separated we find an injective measurable map $\iota : (\mathcal{Y}, \mathcal{B}_\mathcal{Y}) \hookrightarrow ([0,1], \mathcal{B}_{[0,1]})$ by [Bog07] Thm. 6.5.7. Then by the previous argument $\iota(f(A)) \in \mathcal{S}(\mathcal{B}_{[0,1]})$ and because of the injectivity $f(A) = \iota^{-1}(\iota(f(A))) \in \mathcal{S}(\mathcal{B}_\mathcal{Y})$. ☐

Theorem B.47 (Measurable selection theorem, [Bog07] Thm. 6.9.12)**.** *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be an analytic measurable space and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be any measurable space. Let $D \in \mathcal{S}(\mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y})$. Then $\mathrm{pr}_\mathcal{Y}(D) \in \mathcal{S}(\mathcal{B}_\mathcal{Y})$ and there exists a $\sigma\mathcal{S}(\mathcal{B}_{\mathcal{Y}|\mathrm{pr}_\mathcal{Y}(D)})$-$\mathcal{B}_\mathcal{X}$-measurable map:*

$$g : \quad \mathrm{pr}_\mathcal{Y}(D) \to \mathcal{X},$$

such that $(g(y), y) \in D \subseteq \mathcal{X} \times \mathcal{Y}$ for every $y \in \mathrm{pr}_\mathcal{Y}(D)$.

Corollary B.48 ($\sigma\mathcal{S}(\mathcal{B})$-measurable right inverse)**.** *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be an analytic measurable space and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ a countably separated measurable space (e.g. another analytic measurable space) and let $f : (\mathcal{X}, \mathcal{B}_\mathcal{X}) \to (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be a measurable map. Then there exists a $\sigma\mathcal{S}(\mathcal{B}_{\mathcal{Y}|f(\mathcal{X})})$-$\mathcal{B}_\mathcal{X}$-measurable map: $g : f(\mathcal{X}) \to \mathcal{X}$ such that $f(g(y)) = y$ for all $y \in f(\mathcal{X})$.*

Proof. This follows directly from the measurable selection theorem B.47 with $D = \Gamma_f \in \mathcal{S}(\mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y})$ (and lemma B.46). ☐

The following corollary highlights the power of working in the category of analytic measurable spaces.

Corollary B.49 (Maps between analytic measurable spaces)**.** *Let $f : (\mathcal{X}, \mathcal{B}_\mathcal{X}) \to (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be a measurable map between analytic measurable spaces. Then $f(\mathcal{X}) \in \mathcal{S}(\mathcal{B}_\mathcal{Y})$, $(f(\mathcal{X}), \mathcal{B}_{\mathcal{Y}|f(\mathcal{X})})$ is an analytic measurable space and f factors into a surjective and an injective measurable map:*

$$f : (\mathcal{X}, \mathcal{B}_\mathcal{X}) \twoheadrightarrow (f(\mathcal{X}), \mathcal{B}_{\mathcal{Y}|f(\mathcal{X})}) \hookrightarrow (\mathcal{Y}, \mathcal{B}_\mathcal{Y}).$$

Furthermore, the induced maps also give a factorization into a surjective and an injective measurable map of analytic measurable spaces (w.r.t. induced σ-algebras):

$$f_* : (\mathcal{P}(\mathcal{B}_\mathcal{X}), \mathcal{B}_{\mathcal{P}(\mathcal{B}_\mathcal{X})}) \twoheadrightarrow (\mathcal{P}(\mathcal{B}_{\mathcal{Y}|f(\mathcal{X})}), \mathcal{B}_{\mathcal{P}(\mathcal{B}_{\mathcal{Y}|f(\mathcal{X})})}) \hookrightarrow (\mathcal{P}(\mathcal{B}_\mathcal{Y}), \mathcal{B}_{\mathcal{P}(\mathcal{B}_\mathcal{Y})})$$

with $\mathcal{P}(\mathcal{B}_{\mathcal{Y}|f(\mathcal{X})}) \in \mathcal{S}\left(\mathcal{B}_{\mathcal{P}(\mathcal{B}_\mathcal{Y})}\right)$. We get the identification of analytic measurable spaces:

$$\mathcal{P}(\mathcal{B}_{\mathcal{Y}|f(\mathcal{X})}) \quad = \quad \{\boldsymbol{\nu} \in \mathcal{P}(\mathcal{B}_\mathcal{Y}) \,|\, \boldsymbol{\nu}(\mathcal{Y} \setminus f(\mathcal{X})) = 0\}.$$

This shows the equality of sets:

$$\{f_*\boldsymbol{\mu} \,|\, \boldsymbol{\mu} \in \mathcal{P}(\mathcal{B}_\mathcal{X})\} = \{\boldsymbol{\nu} \in \mathcal{P}(\mathcal{B}_\mathcal{Y}) \,|\, \boldsymbol{\nu}(\mathcal{Y} \setminus f(\mathcal{X})) = 0\}.$$

B.8. Measurable Extension Theorems

Theorem B.50 (Kuratowski extension theorem for standard measurable spaces, see [Kec95] 12.2)**.** *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be any measurable space, $\mathcal{W} \subseteq \mathcal{X}$ any subset endowed with the subspace σ-algebra $\mathcal{B}_{\mathcal{X}|\mathcal{W}}$ and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be a standard measurable space. Let f : $(\mathcal{W}, \mathcal{B}_{\mathcal{X}|\mathcal{W}}) \to (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be a measurable map. Then there exists a measurable map:*

$$F : (\mathcal{X}, \mathcal{B}_\mathcal{X}) \to (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$$

such that the restriction to $\mathcal{W}$ equals f, i.e. $F_{|\mathcal{W}} = f$.
In short: There exists F such that the following diagram commutes:

$$\begin{array}{ccc} (\mathcal{W}, \mathcal{B}_{\mathcal{X}|\mathcal{W}}) & \xrightarrow{f} & (\mathcal{Y}, \mathcal{B}_\mathcal{Y}) \\ \big\downarrow & \overset{\exists F}{\nearrow} & \\ (\mathcal{X}, \mathcal{B}_\mathcal{X}). & & \end{array}$$

Corollary B.51 (Extensions of Kuratowski's extension theorem)**.** *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be any measurable space, $\mathcal{W} \subseteq \mathcal{X}$ any subset endowed with the subspace σ-algebra $\mathcal{B}_{\mathcal{X}|\mathcal{W}}$ and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be an analytic measurable space, a countably perfect measurable space, resp. Let $f : (\mathcal{W}, \mathcal{B}_{\mathcal{X}|\mathcal{W}}) \to (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be a measurable map. Then there exists a measurable map:*

$$F : (\mathcal{X}, \sigma\mathcal{S}(\mathcal{B}_\mathcal{X})) \to (\mathcal{Y}, \sigma\mathcal{S}(\mathcal{B}_\mathcal{Y})),$$

a measurable map:

$$F : (\mathcal{X}, (\mathcal{B}_\mathcal{X})_\bullet) \to (\mathcal{Y}, (\mathcal{B}_\mathcal{Y})_\bullet),$$

resp., such that the restriction to $\mathcal{W}$ equals f, i.e. $F_{|\mathcal{W}} = f$.

Proof. Since $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ is analytic, countably perfect, resp., we have a a standard measurable space $\mathcal{Z} = [0, 1]$ and a $\mathcal{B}_\mathcal{Y}$-$\mathcal{B}_\mathcal{Z}$-measurable injective map $j : \mathcal{Y} \hookrightarrow \mathcal{Z}$ with $\mathcal{B}_\mathcal{Y} = j^*\mathcal{B}_\mathcal{Z}$ and $j(\mathcal{Y}) \in \sigma\mathcal{S}(\mathcal{B}_\mathcal{Z})$, $j(\mathcal{Y}) \in (\mathcal{B}_\mathcal{Z})_\bullet$, resp.
By Kuratowski's extension theorem B.50 we now have a measurable extension of $j \circ f$:

$$G : (\mathcal{X}, \mathcal{B}_\mathcal{X}) \to (\mathcal{Z}, \mathcal{B}_\mathcal{Z}).$$

So the following outer diagram commutes:

$$\begin{array}{ccc} (\mathcal{W}, \mathcal{B}_{\mathcal{X}|\mathcal{W}}) & \xrightarrow{f} & (\mathcal{Y}, \mathcal{B}_\mathcal{Y}) \\ \big\downarrow \iota & \overset{\exists ?F}{\nearrow} & \big\downarrow j \\ (\mathcal{X}, \mathcal{B}_\mathcal{X}) & \xrightarrow{\exists G} & (\mathcal{Z}, \mathcal{B}_\mathcal{Z}). \end{array}$$

Now let $\mathcal{C} := G^{-1}(j(\mathcal{Y}))$ and $y_0 \in \mathcal{Y}$ be an arbitrary point. Then $\mathcal{C} \in \sigma\mathcal{S}(\mathcal{B}_\mathcal{X})$, $\mathcal{C} \in (\mathcal{B}_\mathcal{X})_\bullet$, resp., and we define:

$$\begin{array}{rccc} F : & \mathcal{X} & \to & \mathcal{Y}, \\ F_{|\mathcal{C}} : & \mathcal{C} & \to & \mathcal{Y}, \\ & x & \mapsto & j^{-1}(G(x)), \\ F_{|\mathcal{C}^\complement} : & \mathcal{C}^\complement & \to & \mathcal{Y}, \\ & x & \mapsto & y_0. \end{array}$$

For $B \subseteq \mathcal{Y}$ we have:

$$F^{-1}(B) = \left(F_{|\mathcal{C}}^{-1}(B) \cap \mathcal{C}\right) \cup \left(F_{|\mathcal{C}^c}^{-1}(B) \cap \mathcal{C}^c\right).$$

It is thus clear that F is $\sigma\mathcal{S}(\mathcal{B}_\mathcal{X})$-$\sigma\mathcal{S}(\mathcal{B}_\mathcal{Y})$-measurable, F is $(\mathcal{B}_\mathcal{X})_\bullet$-$(\mathcal{B}_\mathcal{Y})_\bullet$-measurable, resp.
Since $G(w) = j(f(w)) \in j(\mathcal{Y})$ we have that: $w \in \mathcal{C}$ and $F(w) = j^{-1}(G(w)) = f(w)$. □

Corollary B.52 (Kuratowski's extension theorem for universal measurable spaces)**.** *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be any measurable space, $\mathcal{W} \subseteq \mathcal{X}$ any subset and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be a universal measurable space. Let*

$$f : (\mathcal{W}, (\mathcal{B}_\mathcal{X})_{\bullet,|\mathcal{W}}) \to (\mathcal{Y}, \mathcal{B}_\mathcal{Y})$$

be a measurable map. Then there exists a universally measurable map:

$$F : (\mathcal{X}, (\mathcal{B}_\mathcal{X})_\bullet) \to (\mathcal{Y}, (\mathcal{B}_\mathcal{Y})_\bullet)$$

such that the restriction to $\mathcal{W}$ equals f, i.e. $F_{|\mathcal{W}} = f$.

Proof. Since $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ is universal we have an universal embedding into a standard measurable space $(\mathcal{Z}, \mathcal{B}_\mathcal{Z})$ (in fact $[0, 1]$), i.e. an $(\mathcal{B}_\mathcal{Y})_\bullet$-$(\mathcal{B}_\mathcal{Z})_\bullet$-measurable injective map $j : \mathcal{Y} \hookrightarrow \mathcal{Z}$ with $(\mathcal{B}_\mathcal{Y})_\bullet = j^*(\mathcal{B}_\mathcal{Z})_\bullet$ and $j(\mathcal{Y}) \in (\mathcal{B}_\mathcal{Z})_\bullet$.
Let $\mathcal{B}_{|\mathcal{Y}} = j^*\mathcal{B}_\mathcal{Z}$. We consider $f^*\mathcal{B}_{|\mathcal{Y}} \subseteq (\mathcal{B}_\mathcal{X})_{\bullet,|\mathcal{W}}$. Then we can pick for every $A \in f^*\mathcal{B}_{|\mathcal{Y}}$ a $\hat{A} \in (\mathcal{B}_\mathcal{X})_\bullet$ such that:

$$A = \hat{A} \cap \mathcal{W}.$$

We then consider the sub-σ-algebra:

$$\hat{\mathcal{B}}_\mathcal{X} = \sigma\left(\left\{\hat{A} \,|\, A \in f^*\mathcal{B}_{|\mathcal{Y}}\right\}\right) \subseteq (\mathcal{B}_\mathcal{X})_\bullet.$$

It is clear that $\hat{\mathcal{B}}_{\mathcal{X}|\mathcal{W}} = f^*\mathcal{B}_{|\mathcal{Y}}$ and that $j \circ f$ is $(f^*\mathcal{B}_{|\mathcal{Y}})$-$\mathcal{B}_\mathcal{Z}$-measurable.
By Kuratowski's extension theorem B.50 we now have a measurable extension of $j \circ f$: $(\mathcal{W}, f^*\mathcal{B}_{|\mathcal{Y}}) \to (\mathcal{Z}, \mathcal{B}_\mathcal{Z})$ to $(\mathcal{B}_\mathcal{X}, \hat{\mathcal{B}}_\mathcal{X})$:

$$G : \; (\mathcal{X}, \hat{\mathcal{B}}_\mathcal{X}) \to (\mathcal{Z}, \mathcal{B}_\mathcal{X}).$$

So the following diagram commutes:

$$\begin{array}{ccc}
(\mathcal{W}, f^*\mathcal{B}_{|\mathcal{Y}}) & \xrightarrow{\;f\;} & (\mathcal{Y}, \mathcal{B}_{|\mathcal{Y}}) \\
\big\downarrow \iota & & \big\downarrow j \\
(\mathcal{X}, \hat{\mathcal{B}}_\mathcal{X}) & \xrightarrow{\;\exists G\;} & (\mathcal{Z}, \mathcal{B}_\mathcal{Z}).
\end{array}$$

Now let $\mathcal{C} := G^{-1}(j(\mathcal{Y}))$ and $y_0 \in \mathcal{Y}$ be an arbitrary point. Then $\mathcal{C} \in (\hat{\mathcal{B}}_\mathcal{X})_\bullet \subseteq (\mathcal{B}_\mathcal{X})_\bullet$ and we define:

$$\begin{array}{rccc}
\hat{G} : & \mathcal{X} & \to & \mathcal{Y}, \\
\hat{G}_{|\mathcal{C}} = G_{|\mathcal{C}} : & \mathcal{C} & \to & \mathcal{Z}, \\
 & x & \mapsto & G(x), \\
\hat{G}_{|\mathcal{C}^c} : & \mathcal{C}^c & \to & \mathcal{Z}, \\
 & x & \mapsto & j(y_0).
\end{array}$$

For $C \subseteq \mathcal{Z}$ we have:

$$\hat{G}^{-1}(C) = \left(G_{|\mathcal{C}}^{-1}(C) \cap \mathcal{C}\right) \cup \underbrace{\left(\hat{G}_{|\mathcal{C}^c}^{-1}(C) \cap \mathcal{C}^c\right)}_{=\emptyset \text{ or } \mathcal{C}^c}.$$

It is thus clear that $\hat{G}$ is $(\mathcal{B}_\mathcal{X})_\bullet$-$(\mathcal{B}_\mathcal{Z})_\bullet$-measurable. Furthermore, $G(\mathcal{X}) \subseteq j(\mathcal{Y})$, where j is injective. So we can define $F := j^{-1} \circ G$. Then F is $(\mathcal{B}_\mathcal{X})_\bullet$-$(\mathcal{B}_\mathcal{Y})_\bullet$-measurable because $j \circ F = \hat{G}$ is $(\mathcal{B}_\mathcal{X})_\bullet$-$(\mathcal{B}_\mathcal{Z})_\bullet$-measurable and: $(\mathcal{B}_\mathcal{Y})_\bullet = j^*(\mathcal{B}_\mathcal{Z})_\bullet$.
Finally, for $w \in \mathcal{W}$ we have $j(f(w)) \in j(\mathcal{Y})$, so:

$$F(w) = j^{-1} \circ G(w) = j^{-1}(j(f(w))) = f(w).$$

□

C. Proofs - Disintegration of Transition Probabilities

C.1. Definition of Regular Conditional Markov Kernels

Definition C.1 (Regular conditional Markov kernel). *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$, $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$, $(\mathcal{Z}, \mathcal{B}_\mathcal{Z})$ be measurable spaces and:*

$$\mathbf{K}(X, Y|Z) : (\mathcal{Z}, \mathcal{B}_\mathcal{Z}) \dashrightarrow (\mathcal{X} \times \mathcal{Y}, \mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y}),$$

be a Markov kernel in two variables, and

$$\mathbf{K}(Y|Z) : (\mathcal{Z}, \mathcal{B}_\mathcal{Z}) \dashrightarrow (\mathcal{Y}, \mathcal{B}_\mathcal{Y}), \quad (B, z) \mapsto \mathbf{K}(X \in \mathcal{X}, Y \in B|Z = z),$$

the marginal Markov kernel. A regular ($\sigma\mathcal{S}$-regular, $\bullet$-regular, resp.) conditional Markov kernel *of $\mathbf{K}(X, Y|Z)$ conditioned on Y given Z is a Markov kernel:*

$$\begin{array}{llll} \mathbf{K}(X|Y, Z) & : & (\mathcal{Y} \times \mathcal{Z}, \quad \mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z}) & \dashrightarrow (\mathcal{X}, \mathcal{B}_\mathcal{X}), \\ \mathbf{K}(X|Y, Z) & : & (\mathcal{Y} \times \mathcal{Z}, \quad \sigma\mathcal{S}(\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z})) & \dashrightarrow (\mathcal{X}, \mathcal{B}_\mathcal{X}), \\ \mathbf{K}(X|Y, Z) & : & (\mathcal{Y} \times \mathcal{Z}, \quad (\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z})_\bullet) & \dashrightarrow (\mathcal{X}, \mathcal{B}_\mathcal{X}), \end{array}$$

resp., such that:

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

C.2. Essential Uniqueness of Regular Conditional Markov Kernels

Lemma C.2 (Essential uniqueness). *Consider Markov kernels:*

$$\mathbf{P}(X|Y, Z), \mathbf{Q}(X|Y, Z) : \mathcal{Y} \times \mathcal{Z} \dashrightarrow \mathcal{X},$$

and

$$\mathbf{K}(Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{Y}$$

with any measurable spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ such that:

$$\mathbf{P}(X|Y,Z) \otimes \mathbf{K}(Y|Z) = \mathbf{Q}(X|Y,Z) \otimes \mathbf{K}(Y|Z).$$

We then have the following statements.

1. *For every $A \in \mathcal{B}_{\mathcal{X}}$ the set:*

$$N_A := \{(y,z) \in \mathcal{Y} \times \mathcal{Z} \,|\, \mathbf{P}(X \in A|Y=y, Z=z) \neq \mathbf{Q}(X \in A|Y=y, Z=z)\}$$

   is a $\mathbf{K}(Y|Z)$-null set with $N_A \in \mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}}$.

2. *If $\mathcal{B}_{\mathcal{X}}$ is countably generated then $N := \bigcup_{A \in \mathcal{B}_{\mathcal{X}}} N_A$ is a $\mathbf{K}(Y|Z)$-null set with $N \in \mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}}$.*

3. *If $(\mathcal{B}_{\mathcal{X}})_\bullet$ is universally countably generated then $\tilde{N} := \bigcup_{A \in (\mathcal{B}_{\mathcal{X}})_\bullet} N_A$ is a $\mathbf{K}(Y|Z)$-null set with $\tilde{N} \in (\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}})_\bullet$.*

Analogous results hold for $\sigma\mathcal{S}(\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}})$-measurable or $(\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}})_\bullet$-measurable $\mathbf{P}$ and $\mathbf{Q}$.

Proof. For $A \subseteq (\mathcal{B}_{\mathcal{X}})_\bullet$ we put $N_A^> := h^{-1}(D^>)$, where $D^> := \{(x_1,x_2) \,|\, x_1, x_2 \in \mathbb{R}, x_1 > x_2\} \in \mathcal{B}_{\mathbb{R}^2}$ and h is given as the following composition of maps:

$$h: \mathcal{Y} \times \mathcal{Z} \xrightarrow{\mathbf{P} \times \mathbf{Q}} \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \xrightarrow{j_A \times j_A} \mathbb{R} \times \mathbb{R}.$$

We define $N_A^< := h^{-1}(D^<)$ similarly. Then $N_A = N_A^> \cup N_A^<$.
1.) If $A \in \mathcal{B}_{\mathcal{X}}$ then the evaluation map j_A is measurable by definition of the σ-algebra on $\mathcal{P}(\mathcal{X})$. So $N_A^>, N_A^< \in \mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}}$. By assumption we have:

$$\mathbf{P}(X|Y,Z) \otimes \mathbf{K}(Y|Z) = \mathbf{Q}(X|Y,Z) \otimes \mathbf{K}(Y|Z).$$

Evaluating boths sides on $A \times N_A^>$ gives thus the same value. So their difference equals 0:

$$0 = \int \mathbb{1}_{N_A^>}(y,z) \cdot (\mathbf{P}(X \in A|Y=y, Z=z) - \mathbf{Q}(X \in A|Y=y, Z=z)) \; \mathbf{K}(Y \in dy|Z=z),$$

where the integrand is:

$$\mathbb{1}_{N_A^>}(y,z) \cdot (\mathbf{P}(X \in A|Y=y, Z=z) - \mathbf{Q}(X \in A|Y=y, Z=z)) \geqslant 0.$$

This implies that: $N_A^>$ must be a $\mathbf{K}(Y|Z)$-null set in $\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}}$. By symmetry we get that also $N_A^<$ is a $\mathbf{K}(Y|Z)$-null set in $\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}}$ and thus $N_A = N_A^< \cup N_A^>$ is a $\mathbf{K}(Y|Z)$-null set in $\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}}$ as well.
2.) If now $\mathcal{B}_{\mathcal{X}}$ is countably generated then $\mathcal{B}_{\mathcal{X}} = \sigma(\mathcal{A})$ with a countable set $\mathcal{A}$ that is closed under finite intersections. One then puts $M := \bigcup_{A \in \mathcal{A}} N_A$, which is, as countable union of measurable $\mathbf{K}(Y|Z)$-null sets, a measurable $\mathbf{K}(Y|Z)$-null set. Then one can define:

$$\mathcal{D} := \{A \in (\mathcal{B}_{\mathcal{X}})_\bullet \,|\, \forall (y,z) \in M^{\mathsf{c}} : \mathbf{P}(X \in A|Y=y, Z=z) = \mathbf{Q}(X \in A|Y=y, Z=z)\}.$$

One easily sees that $\mathcal{D}$ is closed under complements, countable disjoint unions and contains $\mathcal{X} \in \mathcal{D}$. This shows that $\mathcal{D}$ is a Dynkin system (aka λ-system). Furthermore, we have: $\mathcal{A} \subseteq \mathcal{D}$ and that $\mathcal{A}$ is closed under finite intersections. By Dynkin's lemma we get that:

$$\mathcal{B}_{\mathcal{X}} = \sigma(\mathcal{A}) \subseteq \mathcal{D}.$$

If now $(y, z) \in N = \bigcup_{A \in \mathcal{B}_{\mathcal{X}}} N_A$ then there is an $A \in \mathcal{B}_{\mathcal{X}}$ such that:

$$\mathbf{P}(X \in A | Y = y, Z = z) \neq \mathbf{Q}(X \in A | Y = y, Z = z).$$

This implies $(y, z) \in M$ since $A \in \mathcal{D}$ (otherwise we had equality above). Since this holds for all (y, z) we get:

$$N = \bigcup_{A \in \mathcal{B}_{\mathcal{X}}} N_A \subseteq M = \bigcup_{A \in \mathcal{A}} N_A \subseteq N,$$

thus equality. This shows that $N = M$ is a measurable $\mathbf{K}(Y|Z)$-null set.
3.) If now $(\mathcal{B}_{\mathcal{X}})_{\bullet}$ is universally countably generated then $(\mathcal{B}_{\mathcal{X}})_{\bullet} = (\mathcal{A})_{\bullet}$ with a countable set $\mathcal{A} \subseteq (\mathcal{B}_{\mathcal{X}})_{\bullet}$ that is closed under finite intersections. For $A \in (\mathcal{B}_{\mathcal{X}})_{\bullet}$ the map j_A is universally measurable by Theorem B.38. This implies that $N_A \in (\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}})_{\bullet}$ and $M := \bigcup_{A \in \mathcal{A}} N_A \in (\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}})_{\bullet}$ are $\mathbf{K}(Y|Z)$-null sets. By the same arguments as above we get $\sigma(\mathcal{A}) \subseteq \mathcal{D}$. This shows that for $(y, z) \in M^{\mathsf{c}}$ we have that:

$$\forall A \in \sigma(\mathcal{A}) : \quad \mathbf{P}(X \in A | Y = y, Z = z) = \mathbf{Q}(X \in A | Y = y, Z = z).$$

Since every probability measure extends uniquely to the universal completion we even get for those points $(y, z) \in M^{\mathsf{c}}$:

$$\forall A \in (\mathcal{A})_{\bullet} : \quad \mathbf{P}(X \in A | Y = y, Z = z) = \mathbf{Q}(X \in A | Y = y, Z = z).$$

This then shows that even: $(\mathcal{B}_{\mathcal{X}})_{\bullet} = (\mathcal{A})_{\bullet} \subseteq \mathcal{D}$. As before we get that:

$$\tilde{N} = \bigcup_{A \in (\mathcal{B}_{\mathcal{X}})_{\bullet}} N_A = M \in (\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}})_{\bullet},$$

and thus that $\tilde{N}$ a $\mathbf{K}(Y|Z)$-null set in $(\mathcal{B}_{\mathcal{Y}} \otimes \mathcal{B}_{\mathcal{Z}})_{\bullet}$. □

C.3. Existence of Regular Conditional Markov Kernels

Remark C.3 (Existence of conditional Markov kernels)**.** *If* $\mathbf{K}(X, Y|Z)$ *is a Markov kernel then we want* $\mathbf{K}(X|Y, Z)$ *such that:*

$$\mathbf{K}(X, Y|Z) = \mathbf{K}(X|Y, Z) \otimes \mathbf{K}(Y|Z)$$

holds. The heuristic here is to find something like a Radon-Nikodym derivative:

$$\mathbf{K}(X \in A | Y = y, Z = z) = \frac{\mathbf{K}(X \in A, Y \in dy | Z = z)}{\mathbf{K}(Y \in dy | Z = z)}(y),$$

in a way that it is still a probability measure in X and jointly measurable in (y, z).
To achieve measurability from the start we will make use of Besicovitch density theorem C.4:

$$\frac{\mathbf{K}(X \in A, Y \in dy|Z = z)}{\mathbf{K}(Y \in dy|Z = z)}(y) = \lim_{\varepsilon \to 0} \frac{\mathbf{K}(X \in A, Y \in B_\varepsilon(y)|Z = z)}{\mathbf{K}(Y \in B_\varepsilon(y)|Z = z)},$$

and the fact that limits of countably many measurable functions stay measurable. To make sure we get a probability measure out we need to evaluate on sufficiently many sets A, which we will restrict to sets of the form $A = [-\infty, x]$. We will adjust the construction above on null-sets to make sure it will be a probability measure everywhere.
We will show for $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ (or $\bar{\mathbb{R}}$ or $[0, 1]$, etc.) and $\mathbf{K}(Y|Z)$-almost-all $(y, z) \in \mathcal{Y} \times \mathcal{Z}$ and all $x \in \mathcal{X}$ that we have that:

$$\mathbf{K}(X \leqslant x|Y = y, Z = z) = \inf_{m \in \mathbb{N}} \limsup_{n \in \mathbb{N}} \frac{\mathbf{K}(X \leqslant \lceil x \rceil_m, Y \in [y - \frac{1}{n}, y + \frac{1}{n}]|Z = z)}{\mathbf{K}(Y \in [y - \frac{1}{n}, y + \frac{1}{n}]|Z = z)},$$

where $x < \lceil x \rceil_m := \frac{\lfloor mx+1 \rfloor}{m} \leqslant x + \frac{1}{m}$ for $m \in \mathbb{N}$. On the remaining points (y, z), which lie inside the $\mathbf{K}(Y|Z)$-null set, we can take a somewhat arbitrary choice, e.g. we can put:

$$\mathbf{K}(X \leqslant x|Y = y, Z = z) := \mathbf{K}(X \leqslant x|Z = z).$$

Theorem C.4 (Besicovitch density theorem, [Fre15] 472D)**.** *Let $\boldsymbol{\mu}$ be a Radon measure on $\mathbb{R}^D$ (e.g. any finite or probability measure) and $f : \mathbb{R}^D \to \bar{\mathbb{R}} = [-\infty, +\infty]$ be any (locally) $\boldsymbol{\mu}$-integrable function. Then we have for $\boldsymbol{\mu}$-almost-all $x \in \mathbb{R}^D$:*

1. $\displaystyle\lim_{\varepsilon \to 0} \frac{1}{\boldsymbol{\mu}(B_\varepsilon(x))} \int_{B_\varepsilon(x)} f(z)\, \boldsymbol{\mu}(dz) = f(x).$

2. $\displaystyle\lim_{\varepsilon \to 0} \frac{1}{\boldsymbol{\mu}(B_\varepsilon(x))} \int_{B_\varepsilon(x)} |f(z) - f(x)|\, \boldsymbol{\mu}(dz) = 0.$

Here $B_\varepsilon(x)$ denote the closed balls of radius $\varepsilon > 0$ centered at x (in Euclidean norm). The above, in particular, holds for the density $f = \frac{d\boldsymbol{\nu}}{d\boldsymbol{\mu}}$ of another measure $\boldsymbol{\nu}$ w.r.t. $\boldsymbol{\mu}$:

$$\lim_{\varepsilon \to 0} \frac{\boldsymbol{\nu}(B_\varepsilon(x))}{\boldsymbol{\mu}(B_\varepsilon(x))} = \frac{d\boldsymbol{\nu}}{d\boldsymbol{\mu}}(x),$$

for $\boldsymbol{\mu}$-almost-all $x \in \mathbb{R}^D$.

Lemma C.5 (Also see [Mah75] §3.2)**.** *Let $\mathcal{Y} := \mathbb{R}^D$ endowed with its Borel σ-algebra and $\mathcal{Z}$ be any measurable space. Let*

$$F : \mathcal{Y} \times \mathcal{Z} \to [0, 1], \quad (y, z) \mapsto F(y, z)$$

be a function such that:

1. *For each fixed $y \in \mathcal{Y}$ the function:*

$$\mathcal{Z} \to [0, 1], \quad z \mapsto F(y, z)$$

   is measurable.

2. *For each fixed $z \in \mathcal{Z}$ the function:*

$$\mathcal{Y} \to [0,1], \quad y \mapsto F(y,z)$$

is:

   a) *monotone non-decreasing (i.e. $F(y,z) \leqslant F(y',z)$ if $y_d \leqslant y'_d$ for all $d = 1,\dots,D$), and:*
   b) *continuous from above (i.e. $F(y^{(n)},z) \to F(y,z)$ for any sequence $y^{(n)} \to y$ of elements $y^{(n)} \in \mathcal{Y}$ with $y_d^{(n)} \geqslant y_d^{(n+1)} \geqslant y_d$ for all $d = 1,\dots,D$.)*

Then F is (jointly) $(\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z})$-$\mathcal{B}_{[0,1]}$-measurable.
The same conclusion holds if we replace "continuity from above" with "continuity from below".

Proof. Define for $m \in \mathbb{N}$ and $x \in \mathbb{R}$: $\lceil x \rceil_m := \frac{\lfloor m\cdot x+1 \rfloor}{m}$. Then $\lceil x \rceil_m \in \mathbb{Q}$ and:

$$\lceil x \rceil_m - \frac{1}{m} \leqslant x < \lceil x \rceil_m \leqslant x + \frac{1}{m}.$$

For $y \in \mathcal{Y}$ we define component-wise:

$$\lceil y \rceil_m := \left(\lceil y_1 \rceil_m, \dots, \lceil y_D \rceil_m\right).$$

Note that $y \mapsto \lceil y \rceil_m$ is measurable and $\lceil y \rceil_m$ converges to y from above for $m \to \infty$. Now define:

$$\begin{aligned} F_m(y,z) &:= F(\lceil y \rceil_m, z) \\ &= \sum_{k \in \mathbb{Z}^D} F\left(\frac{k}{m}, z\right) \cdot \prod_{d=1}^{D} \mathbb{1}_{\left[\frac{k_d-1}{m}, \frac{k_d}{m}\right)}(y_d). \end{aligned}$$

Since F is measurable in z for fixed $y = \frac{k}{m}$ the functions F_m are jointly measurable for every $m \in \mathbb{N}$. Since F is monotoneous non-decreasing and continuous from the above we have for fixed (y,z):

$$F_m(y,z) = F(\lceil y \rceil_m, z) \longrightarrow F(y,z),$$

for $m \to \infty$. This means that F is the countable limit of (jointly) measurable functions and thus (jointly) measurable itself.
The same construction can be used for the "continuous from below" case by replacing right open intervals with left open intervals. □

If we repeat the same inductively and also with continuity from below we get the following result:

Lemma C.6. *Let $\mathcal{Z}$ be any measurable space and $\mathcal{X} = \mathcal{Y} = \bar{\mathbb{R}} := [-\infty, +\infty]$ (or $\mathbb{R}$ or $[0,1]$) and:*

$$\mathbf{K}(X, Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{X} \times \mathcal{Y}$$

be a Markov kernel. Define for $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$:

$$F_0(x, y|z) := \mathbf{K}(X \leqslant x, Y < y|Z = z),$$
$$F_1(x, y|z) := \mathbf{K}(X \leqslant x, Y \leqslant y|Z = z).$$

Then the functions:

$$F_0,\, F_1 : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to [0,1],$$

are both (jointly) measurable.

Proof. The measurability of F_1 directly follows from Lemma C.5 with (x, y) for the first argument and z for the second argument.
The measurability of F_0 follows inductively. For fixed x the map $(y, z) \mapsto F_0(x, y|z)$ is measurable in z for fixed y and monotone non-decreasing and continuous from below in y for fixed z. Now Lemma C.5 implies that still for fixed x the map $(y, z) \mapsto F_0(x, y|z)$ is measurable. Since for fixed (y, z) the map $x \mapsto F_0(x, y|z)$ is also monotone non-decreasing and continuous from above another application of Lemma C.5 gives us the joint measurability of $(x, y, z) \mapsto F_0(x, y|z)$. □

Proposition C.7 (Existence of conditional Markov kernels for unit intervals)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be the unit interval $[0,1]$ endowed with its Borel σ-algebra and $\mathcal{Z}$ any measurable space. Let*

$$\mathbf{K}(X, Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{X} \times \mathcal{Y},$$

be a Markov kernel in two variables. Then a (regular) conditional Markov kernel conditioned on Y given Z:

$$\mathbf{K}(X|Y, Z) : \mathcal{Y} \times \mathcal{Z} \dashrightarrow \mathcal{X},$$

exists.

Proof. For $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$ we put:

$$F_0(x, y|z) := \mathbf{K}(X \leqslant x, Y < y|Z = z),$$
$$F_1(x, y|z) := \mathbf{K}(X \leqslant x, Y \leqslant y|Z = z).$$

Then the functions:

$$F_0,\, F_1 : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to [0,1],$$

are both (jointly) measurable by C.6.
For $n \in \mathbb{N}$ we further want to define the following functions, anticipating the use of the Besocovitch density theorem C.4 for the closed balls/intervals $B_{1/n}(y) = [y - \frac{1}{n}, y + \frac{1}{n}]$):

$$F_n(x|y, z) := \frac{\mathbf{K}(X \leqslant x, Y \in [y - \frac{1}{n}, y + \frac{1}{n}]|Z = z)}{\mathbf{K}(Y \in [y - \frac{1}{n}, y + \frac{1}{n}]|Z = z)},$$

as long as we do not divide by 0. More formally, using the functions from above, we then actually define:

$$\begin{array}{rcll} F_n(x|y,z) & := & \frac{F_1(x,y+\frac{1}{n}|z)-F_0(x,y-\frac{1}{n}|z)}{F_1(1,y+\frac{1}{n}|z)-F_0(1,y-\frac{1}{n}|z)} & \cdot \mathbb{1}_{F_1(1,y+\frac{1}{n}|z)>F_0(1,y-\frac{1}{n}|z)} \\ & & +F_1(x,1|z) & \cdot \mathbb{1}_{F_1(1,y+\frac{1}{n}|z)=F_0(1,y-\frac{1}{n}|z)}. \end{array}$$

Since F_0, F_1 from above are measurable so are all the F_n, $n \in \mathbb{N}$. Since measurability is preserved under $\inf_{n\in\mathbb{N}}$ and $\sup_{n\in\mathbb{N}}$ also the following maps are jointly measurable:

$$\begin{array}{rcl} \overline{F}(x|y,z) & := & \limsup_{n\to\infty} F_n(x|y,z), \\ \underline{F}(x|y,z) & := & \liminf_{n\to\infty} F_n(x|y,z). \end{array}$$

It follows that the set $S \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ of all points (x,y,z) where either $\overline{F}(x|y,z) \neq \underline{F}(x|y,z)$ or $\overline{F}(x|y,z) \notin \mathbb{R}_{\geqslant 0}$ or $\underline{F}(x|y,z) \notin \mathbb{R}_{\geqslant 0}$ lies in $\mathcal{B}_\mathcal{X}\otimes\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z}$. Now the Besicovitch density theorem C.4 shows that for fixed $x \in \mathcal{X}$ and $z \in \mathcal{Z}$ the limits from above agree and are finite $\mathbf{K}(Y|Z=z)$-almost everywhere. In other words:

$$\mathbf{K}(Y \in S_{(x,z)}|Z=z) = 0,$$

with $S_{(x,z)} := \{y \in \mathcal{Y} \,|\, (x,y,z) \in S\}$. Furthermore, the Besicovitch density theorem C.4 says that for every $y \in S^{\complement}_{(x,z)}$ we have:

$$\overline{F}(x|y,z) = \underline{F}(x|y,z) = \frac{\mathbf{K}(X \leqslant x, Y \in dy|Z=z)}{\mathbf{K}(Y \in dy|Z=z)}(y),$$

in other words that both left hand sides are versions of the Radon-Nikodym derivative in y (for fixed x and z).
By the defining property of Radon-Nikodym derivative we get that for fixed x and z and every $B \in \mathcal{B}_\mathcal{Y}$ we have the equation:

$$\mathbf{K}(X \leqslant x, Y \in B|Z=z) = \int \mathbb{1}_B(y) \cdot \overline{F}(x|y,z)\, \mathbf{K}(Y \in dy|Z=z).$$

As a next step we want to modify $\overline{F}(x|y,z)$ such that it becomes a cumulative distribution function in x, i.e. it corresponds to a probability distribution on $\mathcal{X}$. For this define $\mathcal{X}_\mathbb{Q} := \mathcal{X} \cap \mathbb{Q}$, which is countable and dense in $\mathcal{X}$. First note that $F_n(1|y,z) = 1$ for all (y,z), which implies that:

$$\overline{F}(1|y,z) = \underline{F}(1|y,z) = 1$$

for all $(y,z) \in \mathcal{Y} \times \mathcal{Z}$.
Then define $S' := \bigcup_{x\in\mathcal{X}_\mathbb{Q}} S_x$ with $S_x := \{(y,z) \,|\, (x,y,z) \in S\}$.

$$\mathbf{K}(Y \in S'_z|Z=z) = \mathbf{K}(Y \in \bigcup_{x\in\mathcal{X}_\mathbb{Q}} S_{(x,z)}|Z=z) = 0$$

then shows that S' is a $\mathbf{K}(Y|Z)$-zero set in $\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z}$.
For every pair $x_1 < x_2$ in $\mathcal{X}_\mathbb{Q}$ consider:

$$E_{(x_1,x_2)} := \{(y,z) \,|\, \overline{F}(x_1|y,z) > \overline{F}(x_2|y,z)\} \in \mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z}.$$

Since we have the equations:

$$\begin{array}{cl} & \int \mathbb{1}_{E_{(x_1,x_2),z}}(y) \cdot \overline{F}(x_1|y,z)\,\mathbf{K}(Y \in dy|Z=z) \\ = & \mathbf{K}(X \leqslant x_1, Y \in E_{(x_1,x_2),z}|Z=z) \\ \overset{x_1<x_2}{\leqslant} & \mathbf{K}(X \leqslant x_2, Y \in E_{(x_1,x_2),z}|Z=z) \\ = & \int \mathbb{1}_{E_{(x_1,x_2),z}}(y) \cdot \overline{F}(x_2|y,z)\,\mathbf{K}(Y \in dy|Z=z) \\ \overset{\overline{F}(x_2|y,z)<\overline{F}(x_1|y,z)}{\leqslant} & \int \mathbb{1}_{E_{(x_1,x_2),z}}(y) \cdot \overline{F}(x_1|y,z)\,\mathbf{K}(Y \in dy|Z=z) \end{array}$$

we necessarily have $\mathbf{K}(Y \in E_{(x_1,x_2),z}|Z=z) = 0$ for every $z \in \mathcal{Z}$. Then $E := \bigcup_{x_1<x_2 \in \mathcal{X}_\mathbb{Q}} E_{(x_1,x_2)} \cup S'$ is also a $\mathbf{K}(Y|Z)$-zero set in $\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z}$. So for every $(y,z) \in E^\mathsf{c}$ and every $x_1, x_2 \in \mathcal{X}_\mathbb{Q}$, $x_1 \leqslant x_2$, we get:

$$\underline{F}(x_1|y,z) = \overline{F}(x_1|y,z) \leqslant \overline{F}(x_2|y,z) = \underline{F}(x_2|y,z).$$

Now for $x \in \mathcal{X}_\mathbb{Q}$ put:

$$D_x := \{(y,z)\,|\,\overline{F}(x|y,z) < \inf_{n\in\mathbb{N}} \overline{F}(x+1/n|y,z)\} \in \mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z}.$$

By the dominated convergence theorem (see [Kle14] Cor. 6.26) we get:

$$\begin{array}{cl} & \int \mathbb{1}_{D_{x,z}}(y) \cdot \overline{F}(x|y,z)\,\mathbf{K}(Y \in dy|Z=z) \\ \overset{D_x}{\leqslant} & \int \mathbb{1}_{D_{x,z}}(y) \cdot \inf_{n\in\mathbb{N}} \overline{F}(x+\frac{1}{n}|y,z)\,\mathbf{K}(Y \in dy|Z=z) \\ = & \inf_{n\in\mathbb{N}} \int \mathbb{1}_{D_{x,z}}(y) \cdot \overline{F}(x+\frac{1}{n}|y,z)\,\mathbf{K}(Y \in dy|Z=z) \\ = & \inf_{n\in\mathbb{N}} \mathbf{K}(X \leqslant x+\frac{1}{n}, Y \in D_{x,z}|Z=z) \\ = & \mathbf{K}(X \leqslant x, Y \in D_{x,z}|Z=z) \\ = & \int \mathbb{1}_{D_{x,z}}(y) \cdot \overline{F}(x|y,z)\,\mathbf{K}(Y \in dy|Z=z). \end{array}$$

This shows that $\mathbf{K}(Y \in D_{x,z}|Z=z) = 0$ for all $z \in \mathcal{Z}$. So $D := \bigcup_{x\in\mathcal{X}_\mathbb{Q}} D_x \cup E$ is again a $\mathbf{K}(Y|Z)$-zero set in $\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z}$.
So for all $(x,y,z) \in \mathcal{X}_\mathbb{Q} \times D^\mathsf{c}$ the following is then well-defined:

$$F_+(x|y,z) := \overline{F}(x|y,z) = \underline{F}(x|y,z).$$

When restricted to $\mathcal{X}_\mathbb{Q} \times D^\mathsf{c}$ the function F_+ is thus well-defined, jointly measurable and monotone non-decreasing and continuous from above in x and with $F_+(1|x,z) = 1$. We now aim to extend F_+ to $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$.
For $x \in \mathcal{X} = [0,1]$ and $n \in \mathbb{N}$ put $\lceil x \rceil_n := \min(1, \lfloor nx+1 \rfloor/n)$. Then $\lceil x \rceil_n \in [0,1] \cap \mathbb{Q} = \mathcal{X}_\mathbb{Q}$. The map $x \mapsto \lceil x \rceil_n$ is measurable and for $x \in [0,1)$ we have:

$$x < \lceil x \rceil_n \leqslant x + \frac{1}{n}.$$

So $\lceil 1 \rceil_n = 1$ and $\lceil x \rceil_n \in \mathcal{X}_\mathbb{Q}$ converges to $x \in \mathcal{X}$, $x \neq 1$, from above for $n \to \infty$.
We then define for all $(x,y,z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$:

$$F_+(x|y,z) := \inf_{n\in\mathbb{N}} \left\{\overline{F}(\lceil x \rceil_n|y,z)\right\} \cdot \mathbb{1}_{D^\mathsf{c}}(y,z) + F_1(x,1|z) \cdot \mathbb{1}_D(y,z).$$

It is clear that F_+ is again jointly measurable and agrees with the F_+, $\overline{F}$ and $\underline{F}$ from before on $\mathcal{X}_\mathbb{Q} \times D^c$ by construction. As a monotone approximation from above it is clearly continuous from above, montone non-decreasing and satifies $F_+(1|y,z) = 1$ for all (y,z). So for fixed (y,z) now $F_+(\cdot|y,z)$ corresponds to a probability distribution $\mathbf{K}(X|Y=y, Z=z)$ on $\mathcal{B}_\mathcal{X}$, uniquely given by the defining relations on sets $[0,x]$:

$$F_+(x|y,z) =: \mathbf{K}(X \leqslant x|Y=y, Z=z)$$

for all $x \in \mathcal{X}$. We now show that $(y,z) \mapsto \mathbf{K}(X \in A|Y=y, Z=z)$ is measurable for all $A \in \mathcal{B}_\mathcal{X}$. For this define:

$$\mathcal{D} := \{A \in \mathcal{B}_\mathcal{X} \,|\, (y,z) \mapsto \mathbf{K}(X \in A|Y=y, Z=z) \text{ is } (\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z})\text{-}\mathcal{B}_\mathbb{R}\text{-measurable}\}.$$

Since $\mathbf{K}(X|Y=y, Z=z)$ is a probability measure in X the system $\mathcal{D}$ is closed under countable disjoint unions and complements and contains $\mathcal{X} = [0,1]$. So $\mathcal{D}$ is a Dynkin system. Since the map $(y,z) \mapsto \mathbf{K}(X \leqslant x|Y=y, Z=z) = F_+(x|y,z)$ is measurable in (y,z) for all $x \in \mathcal{X}_\mathbb{Q}$ we have $\mathcal{E} := \{[0,x] \,|\, x \in \mathcal{X}_\mathbb{Q}\} \subseteq \mathcal{D}$. Since $\mathcal{E}$ is closed under finite intersections Dynkin's lemma (see [Kle14] Thm. 1.19) gives $\mathcal{B}_\mathcal{X} = \sigma(\mathcal{E}) \subseteq \mathcal{D}$ and $\mathbf{K}(X|Y,Z)$ is shown to be measurable in (y,z) for all $A \in \mathcal{B}_\mathcal{X}$. So $\mathbf{K}(X|Y,Z)$ is a Markov kernel.
Finally, we need to show that for all $A \in \mathcal{B}_\mathcal{X}$, $B \in \mathcal{B}_\mathcal{Y}$ and $z \in \mathcal{Z}$:

$$\mathbf{K}(X \in A, Y \in B|Z=z) = \int \mathbb{1}_B(y) \cdot \mathbf{K}(X \in A|Y=y, Z=z)\, \mathbf{K}(Y \in dy|Z=z)$$

holds. Since for any $B \in \mathcal{B}_\mathcal{Y}$ and $A = [0,x]$ for $x \in \mathcal{X}_\mathbb{Q}$ we have:

$$\mathbb{1}_B(y) \cdot \mathbf{K}(X \leqslant x|Y=y, Z=z) = \mathbb{1}_B(y) \cdot \overline{F}(x|y,z)$$

up to the $\mathbf{K}(Y|Z=z)$-zero set D_z we already have aboves equation on a generating set that is stable under finite intersections. Again Dynkin's lemma (see [Kle14] Thm. 1.19) shows that the equality from above holds for all sets $A \in \mathcal{B}_\mathcal{X}$, $B \in \mathcal{B}_\mathcal{Y}$ and every $z \in \mathcal{Z}$. This concludes the proof. □

Proposition C.8. *Let $\mathcal{X} = [0,1]$ and $\mathcal{Y}$ be countably generated and $\mathcal{Z}$ be any measurable space. Let*

$$\mathbf{K}(X,Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{X} \times \mathcal{Y},$$

be a Markov kernel in two variables. Then a regular conditional Markov kernel conditioned on Y given Z:

$$\mathbf{K}(X|Y,Z) : \mathcal{Y} \times \mathcal{Z} \dashrightarrow \mathcal{X},$$

exists.

Proof. Since $\mathcal{B}_\mathcal{Y}$ is countably generated by Theorem B.13 (or see [Bog07] Thm. 6.5.5) we can find a measurable map $\psi : (\mathcal{Y}, \mathcal{B}_\mathcal{Y}) \to ([0,1], \mathcal{B}_{[0,1]}) =: (\mathcal{Y}', \mathcal{B}_{\mathcal{Y}'})$ such that

$\mathcal{B}_\mathcal{Y} = \{\psi^{-1}(B') \,|\, B' \in \mathcal{B}_{\mathcal{Y}'}\}$.
Now consider the push-forward Markov kernel $\mathbf{K}(X, Y'|Z) := \mathbf{K}(X, \psi(Y)|Z)$:

$$\mathbf{K}(X, Y'|Z) : \mathcal{Z} \overset{\mathbf{K}(X,Y|Z)}{\dashrightarrow} \mathcal{X} \times \mathcal{Y} \overset{\mathrm{id}_\mathcal{X} \times \psi}{\longrightarrow} \mathcal{X} \times \mathcal{Y}'.$$

Since $\mathcal{X}$ and $\mathcal{Y}'$ are $[0, 1]$ we can apply Thm. C.7. By Thm. C.7 we then get the regular conditional Markov kernel $\mathbf{K}(X|Y', Z)$:

$$\mathbf{K}(X|Y', Z) : \mathcal{Y}' \times \mathcal{Z} \dashrightarrow \mathcal{X}.$$

Composing $\mathbf{K}(X|Y', Z)$ with $\mathcal{Y} \times \mathcal{Z} \overset{\psi \times \mathrm{id}_\mathcal{Z}}{\longrightarrow} \mathcal{Y}' \times \mathcal{Z}$ we get the Markov kernel:

$$\mathbf{K}(X|Y, Z) : \mathcal{Y} \times \mathcal{Z} \overset{\psi \times \mathrm{id}_\mathcal{Z}}{\longrightarrow} \mathcal{Y}' \times \mathcal{Z} \overset{\mathbf{K}(X|Y',Z)}{\dashrightarrow} \mathcal{X}.$$

For any $z \in \mathcal{Z}$, $A \in \mathcal{B}_\mathcal{X}$, $B \in \mathcal{B}_\mathcal{Y}$ and any $B' \in \mathcal{B}_{\mathcal{Y}'}$ with $B = \psi^{-1}(B')$ we then get the equations:

$$\begin{aligned}
& \mathbf{K}(X \in A, Y \in B|Z = z) \\
= \quad & \mathbf{K}(X \in A, Y \in \psi^{-1}(B')|Z = z) \\
= \quad & \mathbf{K}(X \in A, \psi(Y) \in B'|Z = z) \\
= \quad & \mathbf{K}(X \in A, Y' \in B'|Z = z) \\
\overset{\text{def. } \mathbf{K}(X|Y',Z)}{=} \quad & \int \mathbb{1}_{B'}(y') \cdot \mathbf{K}(X \in A|Y' = y', Z = z)\, \mathbf{K}(Y' \in dy'|Z = z) \\
\overset{y'=\psi(y)}{=} \quad & \int \mathbb{1}_{B'}(\psi(y)) \cdot \mathbf{K}(X \in A|Y' = \psi(y), Z = z)\, \mathbf{K}(\psi(Y) \in dy'|Z = z) \\
\overset{\text{def. } \mathbf{K}(X|Y,Z)}{=} \quad & \int \mathbb{1}_{B}(y) \cdot \mathbf{K}(X \in A|Y = y, Z = z)\, \mathbf{K}(Y \in dy|Z = z).
\end{aligned}$$

So $\mathbf{K}(X|Y, Z)$ is a regular conditional Markov kernel for $\mathbf{K}(X, Y|Z)$. ☐

Theorem C.9 (Existence of regular conditional Markov kernels)**.** *Let $\mathcal{X}$ be a standard measurable space, an analytic measurable space, a countably perfect measurable space, resp., and $\mathcal{Y}$ be a countably generated measurable space and $\mathcal{Z}$ be any measurable space. Let*

$$\mathbf{K}(X, Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{X} \times \mathcal{Y},$$

be a Markov kernel in two variables. Then there exists a regular, a $\sigma\mathcal{S}$-regular, $\bullet$-regular, resp., conditional Markov kernel $\mathbf{K}(X|Y, Z)$ conditioned on Y given Z.

Proof. Since $\mathcal{X}$ is standard, analytic, countably perfect, resp., we find an injective measurable map:

$$\varphi : (\mathcal{X}, \mathcal{B}_\mathcal{X}) \hookrightarrow ([0, 1], \mathcal{B}_{[0,1]}) =: (\mathcal{X}', \mathcal{B}_{\mathcal{X}'})$$

that induces a measurable isomorphism $(\mathcal{X}, \mathcal{B}_\mathcal{X}) \cong (\varphi(\mathcal{X}), \mathcal{B}_{\mathcal{X}'|\varphi(\mathcal{X})})$ with $\varphi(\mathcal{X}) \in \mathcal{B}_{\mathcal{X}'}$ ($\varphi(\mathcal{X}) \in \mathcal{S}(\mathcal{B}_{\mathcal{X}'})$, $\varphi(\mathcal{X}) \in (\mathcal{B}_{\mathcal{X}'})_\bullet$, resp.). So we can consider the push-forward Markov kernel $\mathbf{K}(X', Y|Z) := \mathbf{K}(\varphi(X), Y|Z)$:

$$\mathbf{K}(X', Y|Z) : (\mathcal{Z}, \mathcal{B}_\mathcal{Z}) \overset{\mathbf{K}(X,Y|Z)}{\dashrightarrow} (\mathcal{X} \times \mathcal{Y}, \mathcal{B}_\mathcal{X} \otimes \mathcal{B}_\mathcal{Y}) \overset{\varphi \times \mathrm{id}_\mathcal{Y}}{\longrightarrow} (\mathcal{X}' \times \mathcal{Y}, \mathcal{B}_{\mathcal{X}'} \otimes \mathcal{B}_\mathcal{Y}).$$

Since $\mathcal{X}' = [0,1]$ and $\mathcal{B}_\mathcal{Y}$ is countably generated we can apply Prp. C.8 and we then get the regular conditional Markov kernel $\mathbf{K}(X'|Y,Z)$:

$$\mathbf{K}(X'|Y,Z):\ (\mathcal{Y}\times\mathcal{Z},\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z}) \dashrightarrow (\mathcal{X}',\mathcal{B}_{\mathcal{X}'}).$$

If we put $A' := \mathcal{X}'\backslash\varphi(\mathcal{X})$ we have $A' \in \mathcal{B}_{\mathcal{X}'}$ ($A' \in \sigma\mathcal{S}(\mathcal{B}_{\mathcal{X}'})$, $A' \in (\mathcal{B}_{\mathcal{X}'})_\bullet$, resp.). So $\mathbf{K}(X' \in A'|Y=y,Z=z)$ is well-defined for every $(y,z)\in\mathcal{Y}\times\mathcal{Z}$. Consider the set:

$$D := \{(y,z)\in\mathcal{Y}\times\mathcal{Z}\,|\,\mathbf{K}(X'\in A'|Y=y,Z=z)>0\}.$$

We first show that $D\in\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z}$ ($D\in\sigma\mathcal{S}(\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})$, $D\in(\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})_\bullet$, resp.). We now consider the Markov kernel $\mathbf{K}(X'|Y,Z)$ as the measurable map:

$$\mathbf{K}(X'|Y,Z):\ (\mathcal{Y}\times\mathcal{Z},\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z}) \longrightarrow (\mathcal{P}(\mathcal{B}_{\mathcal{X}'}),\mathcal{B}_{\mathcal{P}(\mathcal{B}_{\mathcal{X}'})}).$$

Now consider the map:

$$j_{A'}:\ \mathcal{P}(\mathcal{B}_{\mathcal{X}'})\to[0,1],\quad \mu\mapsto\mu(A').$$

By Theorems B.38 and B.37 the map $j_{A'}$ is measurable w.r.t. $\mathcal{B}_{\mathcal{P}(\mathcal{B}_{\mathcal{X}'})}$ (w.r.t. $\sigma\mathcal{S}(\mathcal{B}_{\mathcal{P}(\mathcal{B}_{\mathcal{X}'})})$, w.r.t. $(\mathcal{B}_{\mathcal{P}(\mathcal{B}_{\mathcal{X}'})})_\bullet$, resp.). Then the composition:

$$H:\ (\mathcal{Y}\times\mathcal{Z},\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z}) \overset{\mathbf{K}(X'|Y,Z)}{\longrightarrow} (\mathcal{P}(\mathcal{B}_{\mathcal{X}'}),\mathcal{B}_{\mathcal{P}(\mathcal{B}_{\mathcal{X}'})}) \overset{j_{A'}}{\longrightarrow} ([0,1],\mathcal{B}_{[0,1]})$$

is measurable w.r.t. $\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z}$ (w.r.t. $\sigma\mathcal{S}(\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})$, w.r.t. $(\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})_\bullet$, resp.). It follows that $D=H^{-1}((0,1])\in\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z}$ ($D\in\sigma\mathcal{S}(\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})$, $D\in(\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})_\bullet$, resp.). Since $y\mapsto(y,z)$ is $\mathcal{B}_\mathcal{Y}$-$(\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})$-measurable it follows that $D_z\in(\mathcal{B}_\mathcal{Y})_\bullet$. So we can evaluate $\mathbf{K}(Y\in D_z|Z=z)=\mathbf{K}((Y,Z)\in D|Z=z)$ for every $z\in\mathcal{Z}$ and the map:

$$\mathcal{Z}\overset{\mathbf{K}(Y,Z|Z)}{\longrightarrow}\mathcal{P}(\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})\overset{j_D}{\longrightarrow}[0,1],\quad z\mapsto\mathbf{K}(Y\in D_z|Z=z),$$

is measurable w.r.t. $\mathcal{B}_\mathcal{Z}$ (w.r.t. $\sigma\mathcal{S}(\mathcal{B}_\mathcal{Z})$, w.r.t. $(\mathcal{B}_\mathcal{Z})_\bullet$, resp.) again by Theorems B.38 and B.37. So we can integrate:

$$\begin{aligned} 0 &= \mathbf{K}(\varphi(X)\in A', Y\in D_z|Z=z)\\ &= \int \mathbb{1}_{D_z}(y)\cdot\mathbf{K}(X'\in A'|Y=y,Z=z)\,\mathbf{K}(dy|Z=z). \end{aligned}$$

It follows that for all $z\in\mathcal{Z}$ we have: $\mathbf{K}(Y\in D_z|Z=z)=0$.
For $A\in\mathcal{B}_\mathcal{X}$ let $\tilde{A}\in\mathcal{B}_{\mathcal{X}'}$ such that $A=\varphi^{-1}(\tilde{A})$. Since φ is injective we have $\varphi(A)=\tilde{A}\cap f(\mathcal{X})\in\mathcal{B}_{\mathcal{X}'}$ (in $\sigma\mathcal{S}(\mathcal{B}_{\mathcal{X}'})$, in $(\mathcal{B}_{\mathcal{X}'})_\bullet$, resp.). So we can define:

$$\begin{aligned} &\mathbf{K}(X\in A|Y=y,Z=z)\\ :=\ &\mathbf{K}(X'\in\varphi(A)|Y=y,Z=z)\cdot\mathbb{1}_{D^c}(y,z)+\mathbf{K}_0(X\in A)\cdot\mathbb{1}_D(y,z), \end{aligned}$$

with any probability distribution $\mathbf{K}_0$. So we get the Markov kernel:

$$\mathbf{K}(X|Y,Z):\ (\mathcal{Y}\times\mathcal{Z},\mathcal{B}_\mathcal{Y}\otimes\mathcal{B}_\mathcal{Z})\dashrightarrow(\mathcal{X},\mathcal{B}_\mathcal{X}),$$

(with $\sigma\mathcal{S}(\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z})$-measurability, $(\mathcal{B}_\mathcal{Y} \otimes \mathcal{B}_\mathcal{Z})_\bullet$-measurability, resp., again by Theorems B.38 and B.37. Furthermore, we have that for all $A \in \mathcal{B}_\mathcal{X}$, $B \in \mathcal{B}_\mathcal{Y}$ and $z \in \mathcal{Z}$:

$$\mathbf{K}(X \in A|Y = y, Z = z) = \int \mathbb{1}_B(y) \cdot \mathbf{K}(X \in A|Y = y, Z = z)\, \mathbf{K}(dy|Z = z).$$

This shows the claim. ☐

Theorem C.10 (Existence of $\bullet$-regular conditional Markov kernels)**.** *Let $\mathcal{X}$ be a universal measurable space, $\mathcal{Y}$ be a universally (countably) generated measurable space and $\mathcal{Z}$ be any measurable space. Let*

$$\mathbf{K}(X, Y|Z) : \mathcal{Z} \dashrightarrow \mathcal{X} \times \mathcal{Y},$$

be a Markov kernel in two variables. Then there exists a $\bullet$-regular conditional Markov kernel $\mathbf{K}(X|Y, Z)$ conditioned on Y given Z.

Proof. Since $\mathcal{Y}$ is universally (countably) generated by Corollary B.14, by Definition B.15, resp., there exists a universally measurable map $\psi : \mathcal{Y} \to [0, 1] =: \mathcal{Y}'$ such that $(\mathcal{B}_\mathcal{Y})_\bullet = (\psi^*(\mathcal{B}_{\mathcal{Y}'})_\bullet)_\bullet$.
Since $\mathcal{X}$ is a universal measurable space there exists a universally measurable injective map: $\varphi : \mathcal{X} \to [0, 1] =: \mathcal{X}'$ such that $\varphi(\mathcal{X}) \in (\mathcal{B}_{\mathcal{X}'})_\bullet$ and $(\mathcal{B}_\mathcal{X})_\bullet = (\varphi^*\mathcal{B}_{\mathcal{X}'})_\bullet$. By push-forward and Theorem B.38 we get a Markov kernel:

$$\mathbf{K}(X', Y'|Z) : \mathcal{Z}_\bullet \dashrightarrow \mathcal{X}' \times \mathcal{Y}'.$$

By Proposition C.7 we get a regular conditional Markov kernel:

$$\mathbf{K}(X'|Y', Z) : \mathcal{Y}' \times \mathcal{Z}_\bullet \dashrightarrow \mathcal{X}',$$

which we can compose with universally measurable $\psi \times \mathrm{id}$ to get:

$$\mathbf{K}(X'|Y, Z) : \mathcal{Y}_\bullet \times \mathcal{Z}_\bullet \to \mathcal{Y}'_\bullet \times \mathcal{Z}_\bullet \dashrightarrow \mathcal{X}'.$$

To check that this is a $\bullet$-regular conditional Markov kernel, one needs to evaluate on the product on $z \in \mathcal{Z}$, $A \in \mathcal{B}_{\mathcal{X}'}$ and $B \in (\mathcal{B}_\mathcal{Y})_\bullet$. Since $(\mathcal{B}_\mathcal{Y})_\bullet = (\psi^*(\mathcal{B}_{\mathcal{Y}'})_\bullet)_\bullet$ we only need to evaluate on $B \in \psi^*(\mathcal{B}_{\mathcal{Y}'})_\bullet$, i.e. $B = \psi^{-1}(B')$ with $B' \in (\mathcal{B}_{\mathcal{Y}'})_\bullet$. The same computation as in Proposition C.8 shows that this is the case.
Now one needs to check that $\mathbf{K}(X'|Y, Z)$ restricts to $\mathcal{X}$ by intersection with $\varphi(\mathcal{X}) \in (\mathcal{B}_{\mathcal{X}'})_\bullet$. This then gives a Markov kernel by Theorem B.38:

$$\mathbf{K}(X|Y, Z) : (\mathcal{Y} \times \mathcal{Z})_\bullet \dashrightarrow \mathcal{X}.$$

The same computations as in Theorem C.9 shows that this is a $\bullet$-regular conditional Markov kernel of $\mathbf{K}(X, Y|Z)$. ☐

The above results inspire for a new definition:

Definition C.11 (Universal disintegration space)**.** *A measurable space $\mathcal{X}$ is called* universal disintegration space *if for every universally generated measurable space $\mathcal{Y}$, every measurable space $\mathcal{Z}$ and every Markov kernel*

$$\mathbf{K}(X,Y|Z):\ \mathcal{Z} \dashrightarrow \mathcal{X}\times\mathcal{Y},$$

there exists a $\bullet$-regular conditional Markov kernel conditioned on Y given Z:

$$\mathbf{K}(X|Y,Z):\ (\mathcal{Y}\times\mathcal{Z})_\bullet \dashrightarrow \mathcal{X}.$$

Remark C.12. *All the results above show that every standard, analytic, countably perfect and every universal measurable space is a universal disintegration space.*

Lemma C.13. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ be measurable spaces and:*

$$\mathbf{K}(X,Y|Z):\ \mathcal{Z} \dashrightarrow \mathcal{X}\times\mathcal{Y},$$

a Markov kernel in two variables. Assume one of the following:

1. *$X \precsim_{\mathbf{K}} (Y,Z)$.*
2. *$Y \precsim_{\mathbf{K}} Z$.*

then there exists a regular conditional Markov kernel conditioned on Y given Z.

Proof. 1.) If $X \precsim_{\mathbf{K}} (Y,Z)$ then there exists a measurable map $\varphi:\ \mathcal{Y}\times\mathcal{Z}\to\mathcal{X}$ such that:

$$\mathbf{K}(X,Y|Z) = \boldsymbol{\delta}_\varphi(X|Y,Z)\otimes\mathbf{K}(Y|Z).$$

So $\mathbf{K}(X|Y,Z) := \boldsymbol{\delta}_\varphi(X|Y,Z)$ is a regular conditional Markov kernel.
2.) If $Y \precsim_{\mathbf{K}} Z$ then there exists a measurable map $\varphi:\ \mathcal{Z}\to\mathcal{Y}$ such that:

$$\begin{aligned}\mathbf{K}(X,Y|Z) &= \boldsymbol{\delta}_\varphi(Y|Z)\otimes\mathbf{K}(X|Z)\\ &= \mathbf{K}(X|Z)\otimes\boldsymbol{\delta}_\varphi(Y|Z)\\ &= \mathbf{K}(X|Z)\otimes\mathbf{K}(Y|Z).\end{aligned}$$

So $\mathbf{K}(X|Y,Z) := \mathbf{K}(X|Z)$ is a regular conditional Markov kernel. □

Remark C.14. *To get further existence results for $\bullet$-regular conditional Markov kernels one could use the following slightly weaker conditions:*

1. *$X \precsim_{\mathbf{K},\bullet} (Y,Z)$, i.e. if there exists a universally measurable map $\varphi:\ \mathcal{Y}\times\mathcal{Z}\to\mathcal{X}$ such that:*

$$\mathbf{K}(X,Y|Z) = \boldsymbol{\delta}_\varphi(X|Y,Z)\otimes\mathbf{K}(Y|Z).$$

2. *$Y \precsim_{\mathbf{K},\bullet} Z$, i.e. if there exists a universally measurable map $\varphi:\ \mathcal{Z}\to\mathcal{Y}$ such that:*

$$\mathbf{K}(X,Y|Z) = \boldsymbol{\delta}_\varphi(Y|Z)\otimes\mathbf{K}(X|Z).$$

The proof would follow the same lines as in Lemma C.13.

D. Proofs - Join-Semi-Lattice Rules for Transitional Random Variables

In this section we will collect properties of the relation $\precsim_{\mathbf{K}}$ introduced in the main paper in Notations 2.19. For this let $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ be a transition probability space and $\mathbf{X} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$ and $\mathbf{Y} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Y}$ and $\mathbf{Z} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Z}$ and $\mathbf{U} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{U}$ be transitional random variables, i.e. Markov kernels. We put:

$$\mathbf{K}(X, Y, Z, U|T) := (\mathbf{X}(X|W, T) \otimes \mathbf{Y}(Y|W, T) \otimes \mathbf{Z}(Z|W, T) \otimes \mathbf{U}(U|W, T)) \circ \mathbf{K}(W|T).$$

The relation $\precsim_{\mathbf{K}}$ will be a main ingredient to show that *transitional conditional independence*, see Definition 3.1, forms a *T-$*$-separoid*, see Definition A.3 and Appendix E. According to Remark A.1 in Appendix A we also need to check the compatibility of $\precsim_{\mathbf{K}}$ with the equivalence relation, $\cong$, of isomorphisms of measurable spaces. This will be done in Appendix E in Lemma E.2.

Remark D.1. *1. In general we do* not *have:* $\mathbf{X} \precsim_{\mathbf{K}} \mathbf{X}$ *for arbitrary Markov kernels.*

2. In general we do not *have anti-symmetry, i.e. that:*

$$\mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y} \precsim_{\mathbf{K}} \mathbf{X} \implies \mathbf{X} = \mathbf{Y}.$$

Notation D.2. *Recall that we write:*

1. $\mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y}$ *if there exists a measurable map* $\varphi : \mathcal{Y} \to \mathcal{X}$ *such that:*

$$\mathbf{K}(X, Y|T) = \boldsymbol{\delta}_{\varphi}(X|Y) \otimes \mathbf{K}(Y|T).$$

We further define:

2. $\mathbf{X} \approx_{\mathbf{K}} \mathbf{Y} \; :\iff \; \mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y} \precsim_{\mathbf{K}} \mathbf{X}$.

The next Lemma D.3 is crucial for most of following results where $\precsim_{\mathbf{K}}$ is involved. Note that a similar result for $\precsim_{\mathbf{K}}^*$ would not hold, i.e. where $\boldsymbol{\delta}_{\varphi}(Y|Z, T)$ would be replaced by an arbitrary Markov kernel $\mathbf{Y}(Y|Z, T)$.

Lemma D.3. *Consider a Markov kernel* $\mathbf{K}(X, Y, Z|T)$ *such that the marginal factorizes:*

$$\mathbf{K}(Y, Z|T) = \boldsymbol{\delta}_{\varphi}(Y|Z, T) \otimes \mathbf{K}(Z|T),$$

with a measurable map $\varphi : \mathcal{Z} \times \mathcal{T} \to \mathcal{Y}$. *Then the joint Markov kernel also factorizes:*

$$\mathbf{K}(X, Y, Z|T) = \boldsymbol{\delta}_{\varphi}(Y|Z, T) \otimes \mathbf{K}(X, Z|T).$$

Proof. By assumption the set:

$$M := \{(y, z, t) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{T} \,|\, y \neq \varphi(z, t)\}$$

is a $\mathbf{K}(Y, Z|T)$-null set. Indeed:

$$\begin{aligned}
&= \mathbf{K}((Y, Z) \in M_t | T = t) \\
&= (\boldsymbol{\delta}_\varphi(Y|Z, T = t) \otimes \mathbf{K}(Z|T = t))\,(M_t) \\
&= \int \boldsymbol{\delta}_\varphi(Y \in M_{z,t}|Z = z, T = t)\,\mathbf{K}(Z \in dz|T = t) \\
&= \int \mathbb{1}_{M_{z,t}}(\varphi(z, t))\,\mathbf{K}(Z \in dz|T = t) \\
&= \int \mathbb{1}_{\{y|y \neq \varphi(z,t)\}}(\varphi(z, t))\,\mathbf{K}(Z \in dz|T = t) \\
&= \int \mathbb{1}_{\varphi(z,t) \neq \varphi(z,t)}\,\mathbf{K}(Z \in dz|T = t) \\
&= 0.
\end{aligned}$$

Then $N := \mathcal{X} \times \mathcal{Y} \times M$ is a $\mathbf{K}(X, \tilde{Y}, Y, Z|T)$-null set, where $\tilde{Y} = \varphi(Z, T)$. This implies:

$$\begin{aligned}
&(\boldsymbol{\delta}_\varphi(Y|Z, T) \otimes \mathbf{K}(X, Z|T))\,(A \times B \times C, t) \\
&= \mathbf{K}(X \in A, \varphi(Z, T) \in B, Z \in C|T = t) \\
&= \mathbf{K}(X \in A, \varphi(Z, T) \in B, Y \in \mathcal{Y}, Z \in C|T = t) \\
&\overset{N}{=} \mathbf{K}(X \in A, \varphi(Z, T) \in \mathcal{Y}, Y \in B, Z \in C|T = t) \\
&= \mathbf{K}(X \in A, Y \in B, Z \in C|T = t).
\end{aligned}$$

This shows the claim. □

Lemma D.4 (Product extension)**.** *We have the implication:*

$$\mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y} \implies \mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y} \otimes \mathbf{Z}.$$

Proof. By assumption we have:

$$\begin{aligned}
&\mathbf{K}(X, Y|T) = \boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y|T) \\
\xRightarrow{\text{Lemma } D.3}\quad &\mathbf{K}(X, Y, Z|T) = \boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y, Z|T).
\end{aligned}$$

□

Lemma D.5 (Bottom element)**.** *We always have:*

$$\boldsymbol{\delta}_* \precsim_{\mathbf{K}} \mathbf{X}.$$

Proof. We have:

$$\mathbf{K}(*, X|T) = \boldsymbol{\delta}_* \otimes \mathbf{K}(X|T).$$

□

Lemma D.6 (Transitivity)**.** *We have the implication:*

$$\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \implies \mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z}.$$

Proof. We have:

$$\begin{aligned}
& & \mathbf{K}(Y, Z|T) &= \boldsymbol{\delta}_\psi(Y|Z) \otimes \mathbf{K}(Z|T), \\
& & \mathbf{K}(X, Y|T) &= \boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y|T) \\
&\xRightarrow{\text{Lemma } D.3} & \mathbf{K}(X, Y, Z|T) &= \boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y, Z|T) \\
& & &= \boldsymbol{\delta}_\varphi(X|Y) \otimes \boldsymbol{\delta}_\psi(Y|Z) \otimes \mathbf{K}(Z|T) \\
&\implies & \mathbf{K}(X, Z|T) &= (\boldsymbol{\delta}_\varphi(X|Y) \circ \boldsymbol{\delta}_\psi(Y|Z)) \otimes \mathbf{K}(Z|T) \\
& & &= \boldsymbol{\delta}_{\varphi\circ\psi}(X|Z) \otimes \mathbf{K}(Z|T)
\end{aligned}$$

□

Lemma D.7 (Restricted reflexivity)**.** *If $X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}$ is a measurable map and $\mathbf{X} = \boldsymbol{\delta}(X|W, T)$. Then we have:*

$$\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{X}.$$

Proof. Let X_1, X_2 be copies of X. Then using $\varphi := \mathrm{id} : \mathcal{X} \to \mathcal{X}$ gives:

$$\boldsymbol{\delta}_{\mathrm{id}}(X_1|X_2) \otimes \mathbf{K}(X_2|T).$$

□

Lemma D.8 (Product stays bounded)**.** *We have the implication:*

$$\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z} \quad \wedge \quad \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \quad \implies \quad \mathbf{X} \otimes \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z}.$$

Proof. By the assumptions we have:

$$\begin{aligned}
& & \mathbf{K}(Y, Z|T) &= \boldsymbol{\delta}_\psi(Y|Z) \otimes \mathbf{K}(Z|T), \\
& & \mathbf{K}(X, Z|T) &= \boldsymbol{\delta}_\varphi(X|Z) \otimes \mathbf{K}(Z|T) \\
&\xRightarrow{\text{Lemma } D.3} & \mathbf{K}(X, Y, Z|T) &= \boldsymbol{\delta}_\varphi(X|Z) \otimes \mathbf{K}(Y, Z|T) \\
& & &= \boldsymbol{\delta}_\varphi(X|Z) \otimes \boldsymbol{\delta}_\psi(Y|Z) \otimes \mathbf{K}(Z|T) \\
& & &= \boldsymbol{\delta}_{\varphi\times\psi}(X, Y|Z) \otimes \mathbf{K}(Z|T).
\end{aligned}$$

□

Lemma D.9 (Product compatibility)**.** *We have the implication:*

$$\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z} \quad \wedge \quad \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{U} \quad \implies \quad \mathbf{X} \otimes \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \otimes \mathbf{U}.$$

Proof. $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z}$ implies $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z} \otimes \mathbf{U}$ by Lemma D.4. Similarly, $\mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{U}$ implies $\mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \otimes \mathbf{U}$. By Lemma D.8 we then get the claim: $\mathbf{X} \otimes \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \otimes \mathbf{U}$. □

Theorem D.10 (The join-semi-lattice of transitional random variables)**.** *Let $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ be a transition probability space and $X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}$ and $Y : \mathcal{W} \times \mathcal{T} \to \mathcal{Y}$ and $Z : \mathcal{W} \times \mathcal{T} \to \mathcal{Z}$ be transitional random variables. We put: $\mathbf{X} := \boldsymbol{\delta}(X|W,T)$ and $\mathbf{Y} := \boldsymbol{\delta}(Y|W,T)$ and $\mathbf{Z} := \boldsymbol{\delta}(Z|W,T)$ and:*

$$\mathbf{K}(X,Y,Z|T) := (\mathbf{X}(X|W,T) \otimes \mathbf{Y}(Y|W,T) \otimes \mathbf{Z}(Z|W,T)) \circ \mathbf{K}(W|T).$$

We then have:

1. *Reflexivity:* $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{X}$.
2. *Transitivity:* $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \implies \mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z}$.
3. *Almost-sure anti-symmetry (per definition):*

$$\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{X} \implies: \mathbf{X} \approx_{\mathbf{K}} \mathbf{Y}.$$

4. *Join:* $\mathbf{X} \otimes \mathbf{Y} = \boldsymbol{\delta}(X,Y|W,T)$.
5. *Join is upper bound:* $\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{X} \otimes \mathbf{Y}$ *and* $\mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{X} \otimes \mathbf{Y}$.
6. *Join is smallest upper bound:*

$$\mathbf{X} \lesssim_{\mathbf{K}} \mathbf{Z} \quad \wedge \quad \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z} \quad \implies \quad \mathbf{X} \otimes \mathbf{Y} \lesssim_{\mathbf{K}} \mathbf{Z}.$$

7. *Bottom element:* $\boldsymbol{\delta}_* \lesssim_{\mathbf{K}} \mathbf{X}$.
8. *Bottom element is neutral:* $\mathbf{X} \approx_{\mathbf{K}} \boldsymbol{\delta}_* \otimes \mathbf{X}$.
9. *Idempotent:* $\mathbf{X} \approx_{\mathbf{K}} \mathbf{X} \otimes \mathbf{X}$.

So the class of transitional random variables of the form $\mathbf{X} = \boldsymbol{\delta}(X|W,T)$, $\mathbf{Y} = \boldsymbol{\delta}(Y|W,T)$, $\mathbf{Z} = \boldsymbol{\delta}(Z|W,T)$, etc., for some measurable maps X, Y, Z, etc., on the transition probability space $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ together with the relation $\lesssim_{\mathbf{K}}$, join $\otimes$ and bottom element $\boldsymbol{\delta}_*$ forms a join-semi-lattice *modulo almost-sure anti-symmetry (and up to the fact that such a class might not be a set).*

Proof. Reflexivity: See Lemma D.7.
Transitivity: See Lemma D.6.
Almost-sure anti-symmetry: This holds per definition. The next point is just a straight forward reformulation.
Join is upper bound: This follows from reflexivity, Lemma D.7, and Lemma D.4.
Join is smallest upper bound: See Lemma D.8.
Bottom element: See Lemma D.5.

Bottom element is neural: $\mathbf{X} \precsim_{\mathbf{K}} \boldsymbol{\delta}_* \otimes \mathbf{X}$ is clear. For the other direction, $\boldsymbol{\delta}_* \otimes \mathbf{X} \precsim_{\mathbf{K}} \mathbf{X}$, note:

$$\mathbf{K}(*, X_1, X_2|T) = \boldsymbol{\delta}_{*\times \mathrm{id}}(*, X_1|X_2) \otimes \mathbf{K}(X_2|T).$$

Idempotent: $\mathbf{X} \precsim_{\mathbf{K}} \mathbf{X} \otimes \mathbf{X}$ is clear. For the other direction, $\mathbf{X} \otimes \mathbf{X} \precsim_{\mathbf{K}} \mathbf{X}$, note:

$$\mathbf{K}(X_1, X_2, X_3|T) = \boldsymbol{\delta}_{\mathrm{id}\times \mathrm{id}}(X_1, X_2|X_3) \otimes \mathbf{K}(X_3|T).$$

□

E. Proofs - Separoid Rules for Transitional Conditional Independence

In this section we want to prove that the class of *transitional random variables* together with the equivalence relation, $\cong$, isomorphism of measurable spaces, the relation $\precsim_{\mathbf{K}}$, the product $\otimes$ and the ternary relation $\perp\!\!\!\perp_{\mathbf{K}}$ of *transitional conditional independence*, see Definition 3.1, forms an $\mathbf{T}$-$\boldsymbol{\delta}_*$-*separoid* (or in different symbols: T-$*$-*separoid*), see Definition A.3 in Appendix A, at least when restricted to codomains that are standard measurable spaces (or universal measurable spaces, when one replaces measurability with universal measurability everywhere).

For this let $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ be a transition probability space and $\mathbf{X} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$ and $\mathbf{Y} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Y}$ and $\mathbf{Z} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Z}$ and $\mathbf{U} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{U}$ be Markov kernels. We denote by $T : \mathcal{W} \times \mathcal{T} \to \mathcal{T}$ the canonical projection map and $\mathbf{T}(T|W,T) := \boldsymbol{\delta}(T|W,T)$. We also consider the constant map $* : \mathcal{W} \times \mathcal{T} \to \circledast := \{*\}$ and $\boldsymbol{\delta}_* = \boldsymbol{\delta}(*|W,T)$ the corresponding Markov kernel. We put:

$$\mathbf{K}(X, Y, Z, U|T) := (\mathbf{X}(X|W,T) \otimes \mathbf{Y}(Y|W,T) \otimes \mathbf{Z}(Z|W,T) \otimes \mathbf{U}(U|W,T)) \circ \mathbf{K}(W|T),$$

or similarly if more or other Markov kernels are involved.

Recall that we say that $\mathbf{X}$ is *transitionally independent of* $\mathbf{Y}$ *conditioned on* $\mathbf{Z}$ *w.r.t.* $\mathbf{K} = \mathbf{K}(W|T)$, in symbols:

$$\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z},$$

if there exists a Markov kernel $\mathbf{Q}(X|Z)$ such that:

$$\mathbf{K}(X, Y, Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{K}(Y, Z|T),$$

where $\mathbf{K}(Y, Z|T)$ is the marginal of $\mathbf{K}(X, Y, Z|T)$.

Notation E.1. *Recall that we write:*

1. $\mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y}$ *if there exists a measurable map* $\varphi : \mathcal{Y} \to \mathcal{X}$ *such that:*

$$\mathbf{K}(X, Y|T) = \boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y|T).$$

2. $\mathbf{X} \approx_{\mathbf{K}} \mathbf{Y} \;:\iff\; \mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y} \precsim_{\mathbf{K}} \mathbf{X}.$

We further define:

3. $\mathbf{X} \cong \mathbf{Y}$ *if there exists a measurable isomorphism* $\varphi : \mathcal{Y} \to \mathcal{X}$*, i.e. a bijective measurable map with a measurable inverse, such that:* $\varphi_* \mathbf{Y} = \mathbf{X}$.

According to Remark A.1 we first need to check that $\precsim_{\mathbf{K}}$, $\otimes$, $\cong$, $\boldsymbol{\delta}_*$, $\perp\!\!\!\perp_{\mathbf{K}}$ are all sufficiently compatible with each other. This will be done in the next Lemma.

Lemma E.2 (Compatibility of $\precsim_{\mathbf{K}}$, $\otimes$, $\cong$, $\boldsymbol{\delta}_*$, $\perp\!\!\!\perp_{\mathbf{K}}$, see A.1)**.** *We have the following:*

1. $(\mathbf{X} \cong \mathbf{X}') \wedge (\mathbf{Y} \cong \mathbf{Y}') \implies (\mathbf{X} \otimes \mathbf{Y}) \cong (\mathbf{X}' \otimes \mathbf{Y}')$.

   Proof. *With isomorphisms* $\varphi : \mathcal{X} \cong \mathcal{X}'$ *and* $\psi : \mathcal{Y} \cong \mathcal{Y}'$ *with* $\varphi_* \mathbf{X} = \mathbf{X}'$ *and* $\psi_* \mathbf{Y} = \mathbf{Y}'$ *we get:* $(\varphi \times \psi)_* (\mathbf{X} \otimes \mathbf{Y}) = (\varphi_* \mathbf{X}) \otimes (\psi_* \mathbf{Y}) = \mathbf{X}' \otimes \mathbf{Y}'$.

2. $(\mathbf{X} \otimes \mathbf{Y}) \cong (\mathbf{Y} \otimes \mathbf{X})$.

   Proof. *Use the isomorphism:* $\mathcal{X} \times \mathcal{Y} \cong \mathcal{Y} \times \mathcal{X}$ *with* $(x, y) \mapsto (y, x)$.

3. $(\mathbf{X} \otimes \mathbf{Y}) \otimes \mathbf{Z} \cong \mathbf{X} \otimes (\mathbf{Y} \otimes \mathbf{Z})$.

   Proof. *Use the isomorphism:* $\mathrm{id} : (\mathcal{X} \times \mathcal{Y}) \times \mathcal{Z} \cong \mathcal{X} \times (\mathcal{Y} \times \mathcal{Z})$.

4. $(\mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y}) \wedge (\mathbf{X} \cong \mathbf{X}') \wedge (\mathbf{Y} \cong \mathbf{Y}') \implies (\mathbf{X}' \precsim_{\mathbf{K}} \mathbf{Y}')$.

   Proof. *Consider* $\mathbf{X}' = \xi_* \mathbf{X}$ *and* $\mathbf{Y}' = \zeta_* \mathbf{Y}$ *with isomorphisms* ξ, ζ.

   Let $\varphi : \mathcal{Y} \to \mathcal{X}$ *such that:* $\mathbf{K}(X, Y|T) = \boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y|T)$. *Then:*

$$\begin{aligned}
\mathbf{K}(X', Y'|T) &= (\boldsymbol{\delta}_\xi(X'|X) \otimes \boldsymbol{\delta}_\zeta(Y'|Y)) \circ \mathbf{K}(X, Y|T) \\
&= (\boldsymbol{\delta}_\xi(X'|X) \otimes \boldsymbol{\delta}_\zeta(Y'|Y)) \circ (\boldsymbol{\delta}_\varphi(X|Y) \otimes \mathbf{K}(Y|T)) \\
&= (\boldsymbol{\delta}_{\xi \circ \varphi}(X'|Y) \otimes \boldsymbol{\delta}_\zeta(Y'|Y)) \circ \mathbf{K}(Y|T) \\
&= \boldsymbol{\delta}_{\xi \circ \varphi \circ \zeta^{-1}}(X'|Y') \otimes \mathbf{K}(Y'|T).
\end{aligned}$$

5. $(\mathbf{X} \precsim_{\mathbf{K}} \mathbf{Y}) \implies (\mathbf{X} \precsim_{\mathbf{K}} (\mathbf{Y} \otimes \mathbf{Z}))$.

   Proof. *This is proven in Lemmata D.3 and D.4.*

6. $(\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{Z}) \wedge (\mathbf{X} \cong \mathbf{X}') \wedge (\mathbf{Y} \cong \mathbf{Y}') \wedge (\mathbf{Z} \cong \mathbf{Z}') \implies (\mathbf{X}' \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y}' \,|\, \mathbf{Z}')$.

   Proof. *If* $\mathbf{X}' = \varphi_* \mathbf{X}$ *and* $\mathbf{Y}' = \psi_* \mathbf{Y}$ *and* $\mathbf{Z}' = \xi_* \mathbf{Z}$ *and:*

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

   Then we get:

$$\mathbf{K}(X', Y', Z'|T) = \mathbf{Q}(\varphi(X)|Z = \xi^{-1}(Z')) \otimes \mathbf{K}(Y', Z'|T).$$

7. $\boldsymbol{\delta}_* \precsim_{\mathbf{K}} \mathbf{X}$.

   Proof. $\mathbf{K}(*, X|T) = \boldsymbol{\delta}_* \otimes \mathbf{K}(X|T)$.

8. $\boldsymbol{\delta}_* \otimes \mathbf{X} \cong \mathbf{X}$.

   Proof. *Use isomorphism:* $\{*\} \times \mathcal{X} \cong \mathcal{X}$.

E.1. Core Separoid Rules for Transitional Conditional Independence

Lemma E.3 (Left Redundancy). *We have for any* $\mathbf{Y}$ *the implication:*

$$\mathbf{U} \precsim_{\mathbf{K}} \mathbf{Z} \implies \mathbf{U} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}.$$

Proof. The assumption implies the existence of a factorization:

$$\mathbf{K}(U, Z|T) = \boldsymbol{\delta}_\varphi(U|Z) \otimes \mathbf{K}(Z|T).$$

Lemma D.3 then shows that this extends to:

$$\mathbf{K}(U, Y, Z|T) = \boldsymbol{\delta}_\varphi(U|Z) \otimes \mathbf{K}(Y, Z|T),$$

which shows the claim. □

Lemma E.4 ($\mathbf{T}$-Restricted Right Redundancy). *Let* $\mathcal{X}$ *be standard and* $\mathcal{Z}$ *be countably generated. Then:*

$$\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \boldsymbol{\delta}_* \,|\, \mathbf{Z} \otimes \mathbf{T}.$$

Proof. Since $\mathcal{X}$ is standard and $\{*\} \times \mathcal{Z}$ is countably generated by Theorem C.9 we get the factorization:

$$\mathbf{K}(X, *, Z|T) = \mathbf{K}(X|*, Z, T) \otimes \mathbf{K}(*, Z|T).$$

Multiplying both sides with $\mathbf{T} = \boldsymbol{\delta}(T|T)$ gives:

$$\mathbf{K}(X, *, T, Z|T) = \mathbf{K}(X|*, Z, T) \otimes \mathbf{K}(*, T, Z|T).$$

This shows the claim. □

Lemma E.5 (Left Decomposition).

$$\mathbf{X} \otimes \mathbf{U} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \implies \mathbf{U} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}.$$

Proof. By assumption we have the factorization:

$$\mathbf{K}(X, U, Y, Z|T) = \mathbf{Q}(X, U|Z) \otimes \mathbf{K}(Y, Z|T).$$

Marginalizing out X gives:

$$\mathbf{K}(U, Y, Z|T) = \mathbf{Q}(U|Z) \otimes \mathbf{K}(Y, Z|T).$$

This shows the claim. □

Lemma E.6 (Right Decomposition).

$$\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z} \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{U} \,|\, \mathbf{Z}.$$

Proof. By assumption we have the factorization:

$$\mathbf{K}(X,U,Y,Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{K}(Y,U,Z|T).$$

Marginalizing out Y gives:

$$\mathbf{K}(X,U,Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{K}(U,Z|T).$$

This shows the claim. □

Lemma E.7 (**T**-Inverted Right Decomposition)**.**

$$\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{T} \otimes \mathbf{Y} \,|\, \mathbf{Z}.$$

Proof. By the assumption $\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{Z}$ we have a factorization:

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

Multiplying both sides with $\mathbf{T} = \boldsymbol{\delta}(T|T)$ gives:

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

This shows the claim. □

Lemma E.8 (Left Weak Union)**.** *Let $\mathcal{X}$ be standard and $\mathcal{U}$ be countably generated. Then:*

$$\mathbf{X} \otimes \mathbf{U} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}.$$

Proof. By assumption we have:

$$\mathbf{K}(X,U,Y,Z|T) = \mathbf{Q}(X,U|Z) \otimes \mathbf{K}(Y,Z|T),$$

for some Markov kernel $\mathbf{Q}(X,U|Z)$. If we marginalize out X we get:

$$\mathbf{K}(U,Y,Z|T) = \mathbf{Q}(U|Z) \otimes \mathbf{K}(Y,Z|T).$$

Because $\mathcal{X}$ is standard and $\mathcal{U}$ countably generated we have a factorization:

$$\mathbf{Q}(X,U|Z) = \mathbf{Q}(X|U,Z) \otimes \mathbf{Q}(U|Z),$$

with the conditional Markov kernel $\mathbf{Q}(X|U,Z)$ (via Theorem C.9).
Putting these equations together we get:

$$\begin{aligned}\mathbf{K}(X,U,Y,Z|T) &= \mathbf{Q}(X,U|Z) \otimes \mathbf{K}(Y,Z|T)\\ &= \mathbf{Q}(X|U,Z) \otimes \mathbf{Q}(U|Z) \otimes \mathbf{K}(Y,Z|T)\\ &= \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(U,Y,Z|T).\end{aligned}$$

This shows the claim. □

Remark E.9. *Left Weak Union E.8 was formulated to rely on the assumption that $\mathcal{X}$ is a standard and $\mathcal{U}$ a countably generated measurable space. The reason was that we relied on the existence of a regular conditional Markov kernel, Theorem C.9. If we instead use the Theorem C.10 with the weaker assumption of universal measurable $\mathcal{X}$ and universally (countably) generated $\mathcal{U}$ we would get a (universally measurable) $\bullet$-regular conditional Markov kernel instead. So if we replaced measurability for all maps and Markov kernels everywhere with the weaker notion of universal measurability we could conclude similarly, but with weaker assumptions and for a bigger class of measurable spaces, so would be much more general. The price one has to pay is rather small, besides dealing with slightly more technical definitions, since measurablity always implies universal measurability, and also because every probability measure can uniquely be extended to measure all universally measurable subsets.*

If one does not want to make any assumptions about the underlying measurable spaces one could resort to the following:

Lemma E.10 (Restricted Left Weak Union)**.**

$$\left(\mathbf{X} \otimes \mathbf{U} \mathrel{\underset{\mathbf{K}}{\perp\!\!\!\perp}} \mathbf{Y} \,|\, \mathbf{Z}\right) \wedge \left(\mathbf{X} \mathrel{\underset{\mathbf{K}}{\perp\!\!\!\perp}} \boldsymbol{\delta}_* \,|\, \mathbf{U} \otimes \mathbf{Z}\right) \implies \mathbf{X} \mathrel{\underset{\mathbf{K}}{\perp\!\!\!\perp}} \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}.$$

Proof. By assumption we have:

$$\mathbf{K}(X, U, Y, Z|T) = \mathbf{Q}(X, U|Z) \otimes \mathbf{K}(Y, Z|T),$$
$$\mathbf{K}(X, U, Z|T) = \mathbf{P}(X|U, Z) \otimes \mathbf{K}(U, Z|T),$$

for some Markov kernels $\mathbf{Q}(X, U|Z)$, $\mathbf{P}(X|U, Z)$. If we marginalize out Y and then X in the first equation we get:

$$\mathbf{K}(X, U, Z|T) = \mathbf{Q}(X, U|Z) \otimes \mathbf{K}(Z|T),$$
$$\mathbf{K}(U, Z|T) = \mathbf{Q}(U|Z) \otimes \mathbf{K}(Z|T).$$

This together with the second equation gives:

$$\mathbf{K}(X, U, Z|T) = \mathbf{P}(X|U, Z) \otimes \mathbf{Q}(U|Z) \otimes \mathbf{K}(Z|T).$$

Comparing this to the above equation we get:

$$\mathbf{Q}(X, U|Z) \otimes \mathbf{K}(Z|T) = \mathbf{P}(X|U, Z) \otimes \mathbf{Q}(U|Z) \otimes \mathbf{K}(Z|T).$$

By the essential uniqueness (see Lemma C.2) of such factorization we get that for every $A \in \mathcal{B}_{\mathcal{X}}$ and $D \in \mathcal{B}_{\mathcal{U}}$:

$$\mathbf{Q}(X \in A, U \in D|Z) = \int_B \mathbf{P}(X \in A|U = u, Z)\, \mathbf{Q}(U \in du|Z) \quad \mathbf{K}(Z|T)\text{-a.s.}$$

Then this holds also $\mathbf{K}(Y, Z|T)$-a.s. Plugging this back into the first equation we get:

$$\mathbf{K}(X, U, Y, Z|T) = \mathbf{P}(X|U, Z) \otimes \mathbf{Q}(U|Z) \otimes \mathbf{K}(Y, Z|T).$$

Marginalizing X out gives:

$$\mathbf{K}(U, Y, Z|T) = \mathbf{Q}(U|Z) \otimes \mathbf{K}(Y, Z|T).$$

Plugging that back in finally gives:

$$\begin{aligned}\mathbf{K}(X, U, Y, Z|T) &= \mathbf{P}(X|U, Z) \otimes \mathbf{Q}(U|Z) \otimes \mathbf{K}(Y, Z|T)\\ &= \mathbf{P}(X|U, Z) \otimes \mathbf{K}(U, Y, Z|T).\end{aligned}$$

This shows the claim. □

Lemma E.11 (Right Weak Union)**.**

$$\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z} \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}.$$

Proof. We have the factorization:

$$\mathbf{K}(X, Y, U, Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{K}(Y, U, Z|T),$$

with some Markov kernel $\mathbf{Q}(X|Z)$. If we view $\mathbf{Q}(X|Z)$ as a function in (u, z) via:

$$(u, z) \mapsto \mathbf{Q}(X|Z = z),$$

by just ignoring the argument u then the claim follows from the same factorization above. □

Lemma E.12 (Left Contraction)**.**

$$(\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \wedge (\mathbf{U} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}) \implies \mathbf{X} \otimes \mathbf{U} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}.$$

Proof. By assumption we have the two factorizations:

$$\begin{aligned}\mathbf{K}(X, Y, U, Z|T) &= \mathbf{Q}(X|U, Z) \otimes \mathbf{K}(Y, U, Z|T),\\ \mathbf{K}(Y, U, Z|T) &= \mathbf{P}(U|Z) \otimes \mathbf{K}(Y, Z|T),\end{aligned}$$

with some Markov kernels $\mathbf{Q}(X|U, Z)$, $\mathbf{P}(U|Z)$. Putting these equations together using $\mathbf{Q}(X|U, Z) \otimes \mathbf{P}(U|Z)$ we get:

$$\mathbf{K}(X, Y, U, Z|T) = (\mathbf{Q}(X|U, Z) \otimes \mathbf{P}(U|Z)) \otimes \mathbf{K}(Y, Z|T).$$

This shows the claim. □

Lemma E.13 (Right Contraction)**.**

$$(\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \wedge (\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{U} \,|\, \mathbf{Z}) \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z}.$$

Proof. By assumption we have the two factorizations:

$$\begin{aligned}\mathbf{K}(X,Y,U,Z|T) &= \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(Y,U,Z|T),\\ \mathbf{K}(X,U,Z|T) &= \mathbf{P}(X|Z) \otimes \mathbf{K}(U,Z|T),\end{aligned}$$

with some Markov kernels $\mathbf{Q}(X|U,Z)$, $\mathbf{P}(X|Z)$.
Marginalizing out Y we get the equalities:

$$\begin{aligned}\mathbf{K}(X,U,Z|T) &= \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(U,Z|T),\\ \mathbf{K}(X,U,Z|T) &= \mathbf{P}(X|Z) \otimes \mathbf{K}(U,Z|T).\end{aligned}$$

By the essential uniqueness (see Lemma C.2) of such factorization we get that for every $A \in \mathcal{B}_{\mathcal{X}}$:

$$\mathbf{Q}(X \in A|U,Z) = \mathbf{P}(X \in A|Z) \qquad \mathbf{K}(U,Z|T)\text{-a.s.}$$

The same equation then holds also $\mathbf{K}(Y,U,Z|T)$-a.s. (by ignoring argument y). Plugging that back into the first equation gives:

$$\mathbf{K}(X,Y,U,Z|T) = \mathbf{P}(X|Z) \otimes \mathbf{K}(Y,U,Z|T).$$

This shows the claim. ☐

Lemma E.14 (Right Cross Contraction)**.**

$$(\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \wedge (\mathbf{U} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{X} \,|\, \mathbf{Z}) \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z}.$$

Proof. By assumption we have the two factorizations:

$$\mathbf{K}(X,Y,U,Z|T) = \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(Y,U,Z|T), \tag{4}$$
$$\mathbf{K}(X,U,Z|T) = \mathbf{P}(U|Z) \otimes \mathbf{K}(X,Z|T), \tag{5}$$

with some Markov kernels $\mathbf{Q}(X|U,Z)$, $\mathbf{P}(U|Z)$.
We then define the Markov kernel:

$$\mathbf{R}(X,U|Z) := \mathbf{Q}(X|U,Z) \otimes \mathbf{P}(U|Z). \tag{6}$$

We will now show that its marginal:

$$\mathbf{R}(X|Z) = \mathbf{Q}(X|U,Z) \circ \mathbf{P}(U|Z). \tag{7}$$

will satisfy the claim.
If we marginalize out Y from equation 4 we get:

$$\mathbf{K}(X,U,Z|T) = \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(U,Z|T). \tag{8}$$

Equating equations 5 and 8 gives:

$$\mathbf{P}(U|Z) \otimes \mathbf{K}(X,Z|T) = \mathbf{K}(X,U,Z|T) = \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(U,Z|T). \tag{9}$$

Marginalizing out X in equation 9 on both sides gives:

$$\mathbf{K}(U, Z|T) = \mathbf{P}(U|Z) \otimes \mathbf{K}(Z|T). \tag{10}$$

If we now plug equation 10 into 8 then we get:

$$\mathbf{K}(X, U, Z|T) = \mathbf{Q}(X|U, Z) \otimes \mathbf{P}(U|Z) \otimes \mathbf{K}(Z|T) \tag{11}$$

$$\overset{6}{=} \mathbf{R}(X, U|Z) \otimes \mathbf{K}(Z|T). \tag{12}$$

If we marginalize out U in equation 12 and use definition 7 we arrive at:

$$\mathbf{K}(X, Z|T) = \mathbf{R}(X|Z) \otimes \mathbf{K}(Z|T). \tag{13}$$

We now get:

$$\mathbf{Q}(X|U, Z) \otimes \mathbf{K}(U, Z|T) \overset{8}{=} \mathbf{K}(X, U, Z|T) \tag{14}$$

$$\overset{5}{=} \mathbf{P}(U|Z) \otimes \mathbf{K}(X, Z|T) \tag{15}$$

$$\overset{13}{=} \mathbf{P}(U|Z) \otimes \mathbf{R}(X|Z) \otimes \mathbf{K}(Z|T) \tag{16}$$

$$= \mathbf{R}(X|Z) \otimes \mathbf{P}(U|Z) \otimes \mathbf{K}(Z|T) \tag{17}$$

$$\overset{10}{=} \mathbf{R}(X|Z) \otimes \mathbf{K}(U, Z|T). \tag{18}$$

By the essential uniqueness (see Lemma C.2) of such a factorization we get that for every $A \in \mathcal{B}_{\mathcal{X}}$:

$$\mathbf{Q}(X \in A|U, Z) = \mathbf{R}(X \in A|Z) \qquad \mathbf{K}(U, Z|T)\text{-a.s.} \tag{19}$$

The same equation then holds also $\mathbf{K}(Y, U, Z|T)$-a.s. (by ignoring the non-occuring argument y). Plugging 19 back into the equation 4 we get:

$$\mathbf{K}(X, Y, U, Z|T) = \mathbf{Q}(X|U, Z) \otimes \mathbf{K}(Y, U, Z|T), \tag{20}$$

$$= \mathbf{R}(X|Z) \otimes \mathbf{K}(Y, U, Z|T). \tag{21}$$

This shows the claim. □

Lemma E.15 (Flipped Left Cross Contraction)**.**

$$(\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{U} \otimes \mathbf{Z}) \wedge (\mathbf{Y} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{U} \,|\, \mathbf{Z}) \implies \mathbf{Y} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{X} \otimes \mathbf{U} \,|\, \mathbf{Z}.$$

Proof. By assumption we have the two factorizations:

$$\mathbf{K}(X, Y, U, Z|T) = \mathbf{Q}(X|U, Z) \otimes \mathbf{K}(Y, U, Z|T),$$
$$\mathbf{K}(Y, U, Z|T) = \mathbf{P}(Y|Z) \otimes \mathbf{K}(U, Z|T),$$

with some Markov kernels $\mathbf{Q}(X|U,Z)$, $\mathbf{P}(Y|Z)$.
Marginalizing out Y in the first equation we get the equality:

$$\mathbf{K}(X,U,Z|T) = \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(U,Z|T).$$

Plugging all three equations into each other we get:

$$\begin{aligned}\mathbf{K}(X,Y,U,Z|T) &= \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(Y,U,Z|T)\\ &= \mathbf{Q}(X|U,Z) \otimes \mathbf{P}(Y|Z) \otimes \mathbf{K}(U,Z|T)\\ &= \mathbf{P}(Y|Z) \otimes \mathbf{Q}(X|U,Z) \otimes \mathbf{K}(U,Z|T)\\ &= \mathbf{P}(Y|Z) \otimes \mathbf{K}(X,U,Z|T).\end{aligned}$$

This shows the claim. □

Corollary E.16. *The class of all transitional random variables with standard measurable spaces as codomain on transition probability space $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ together with the product of Markov kernels $\otimes$, the equivalence of isomorphisms of measurable spaces $\cong$, the relation $\precsim_{\mathbf{K}}$, the one-point Markov kernel $\boldsymbol{\delta}_*$ and transitional conditional independence $\perp\!\!\!\perp_{\mathbf{K}}$ forms a $\mathbf{T}$-$\boldsymbol{\delta}_*$-separoid (or in simpler symbols, a T-$*$-separoid).*
Furthermore, the class of transitional random variables of the form $\boldsymbol{\delta}(X|W,T)$ with standard measurable spaces as codomains on $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ form a join-semi-lattice (with $\otimes$ and $\precsim_{\mathbf{K}}$) up to almost-sure anti-symmetry and a T-$*$-separoid.

Remark E.17. *Similarly, if one replaces measurability with universal measurability everywhere and standard measurable spaces with universal measurable spaces, then also the class of transitional random variables with universal measurable spaces as codomains form a $\mathbf{T}$-$\boldsymbol{\delta}_*$-separoid. Furthermore, the ones of the form $\boldsymbol{\delta}(X|W,T)$ with universal measurable spaces as codomains form a join-semi-lattice (with $\otimes$ and $\precsim_{\mathbf{K},\bullet}$) up to almost-sure anti-symmetry and a T-$*$-separoid.*

E.2. Derived Separoid Rules for Transitional Conditional Independence

Most the following rules follow directly from the general T-$*$-separoid rules proven in the last subsection and the general theory developed in Appendix A. Nonetheless, since we have to track which of the spaces are standard or countably generated we will go through those proofs carefully again.

Lemma E.18 (Extended $\mathbf{T}$-Restricted Right Redundancy)**.** *Let $\mathcal{X}$ be standard and $\mathcal{Z}$ be countably generated. Then:*

$$\mathbf{T} \precsim_{\mathbf{K}} \mathbf{Z} \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \boldsymbol{\delta}_* \,|\, \mathbf{Z}.$$

Proof. Extended $\mathbf{T}$-Restricted Right Redundancy E.18 can be proven using $\mathbf{T}$-Restricted Right Redundancy E.4 ($\implies \mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \boldsymbol{\delta}_* \,|\, \mathbf{Z} \otimes \mathbf{T}$) together with Left Redundancy E.3 ($\implies \mathbf{T} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{X} \,|\, \mathbf{Z}$) and Right Cross Contraction E.14 ($\implies \mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{T} \otimes \boldsymbol{\delta}_* \,|\, \mathbf{Z}$) and then Right Decomposition E.6 ($\implies \mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \boldsymbol{\delta}_* \,|\, \mathbf{Z}$). □

Lemma E.19 (Restricted Symmetry)**.**

$$(\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}) \wedge (\mathbf{Y} \underset{\mathbf{K}}{\perp\!\!\!\perp} \boldsymbol{\delta}_* \,|\, \mathbf{Z}) \implies \mathbf{Y} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{X} \,|\, \mathbf{Z}.$$

Proof. This follows from Flipped Left Cross Contraction E.15 with $\mathbf{U} = \boldsymbol{\delta}_*$. □

Lemma E.20 ($\mathbf{T}$-Restricted Symmetry)**.** *Let $\mathcal{Y}$ be standard and $\mathcal{Z}$ be countably generated. Then:*

$$\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \otimes \mathbf{T} \implies \mathbf{Y} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{X} \,|\, \mathbf{Z} \otimes \mathbf{T}.$$

Proof. Since $\mathcal{Y}$ is standard and $\mathcal{Z}$ countably generated we get by $\mathbf{T}$-Restricted Right Redundancy E.4:

$$\mathbf{Y} \underset{\mathbf{K}}{\perp\!\!\!\perp} \boldsymbol{\delta}_* \,|\, \mathbf{Z} \otimes \mathbf{T}.$$

Together with $\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{Z} \otimes \mathbf{T}$ and Restricted Symmetry E.19 we get:

$$\mathbf{Y} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{X} \,|\, \mathbf{Z} \otimes \mathbf{T}.$$

□

Lemma E.21 (Symmetry)**.** *Let $\mathcal{Y}$ be standard and $\mathcal{Z}$ be countably generated and $\mathcal{T} = \circledast = \{*\}$ the one-point space. Then:*

$$\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \implies \mathbf{Y} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{X} \,|\, \mathbf{Z}.$$

Proof. This follows similarly to $\mathbf{T}$-Restricted Symmetry E.20 with $T = *$. □

Lemma E.22 (Inverted Left Decomposition)**.**

$$\left(\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}\right) \wedge (\mathbf{U} \precsim_{\mathbf{K}} \mathbf{X} \otimes \mathbf{Z}) \implies \mathbf{X} \otimes \mathbf{U} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}.$$

Proof. Inverted Left Decomposition E.22 can be proven using Left Redundancy E.3 ($\implies \mathbf{U} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{X} \otimes \mathbf{Z}$) together with the assumption ($\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{Z}$) and Left Contraction E.12 ($\implies \mathbf{X} \otimes \mathbf{U} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{Y} \,|\, \mathbf{Z}$). □

Lemma E.23 ($\mathbf{T}$-Extended Inverted Right Decomposition)**.**

$$\left(\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}\right) \wedge (\mathbf{U} \precsim_{\mathbf{K}} \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{Z}) \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z}.$$

Proof. $\mathbf{T}$-Extended Inverted Right Decomposition E.23 can be proven using $\mathbf{T}$-Inverted Right Decomposition E.7 ($\implies \mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{T} \otimes \mathbf{Y} \,|\, \mathbf{Z}$) in combination with Left Redundancy E.3 ($\implies \mathbf{U} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{X} \,|\, \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{Z}$) and Flipped Left Cross Contraction E.15 ($\implies \mathbf{X} \perp\!\!\!\perp_{\mathbf{K}} \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{U} \,|\, \mathbf{Z}$). □

Lemma E.24 (Equivalent Exchange)**.**

$$\left(\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}\right) \wedge (\mathbf{Z} \approx_{\mathbf{K}} \mathbf{Z}') \implies \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}'.$$

Proof. We get:

$$
\begin{array}{lcll}
\mathbf{Z}' \precsim_{\mathbf{K}} \mathbf{Z} & \xRightarrow{\text{Lemma D.4}} & \mathbf{Z}' \precsim_{\mathbf{K}} \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{Z}, & \\
\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} & \xRightarrow{\mathbf{T}\text{-Ext. Inv. Right Decomposition E.23}} & \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{Z}' \,|\, \mathbf{Z} & \\
 & \xRightarrow{\text{Right Decomposition E.6}} & \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \otimes \mathbf{Z}' \,|\, \mathbf{Z} & \\
 & \xRightarrow{\text{Right Weak Union E.11}} & \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}' \otimes \mathbf{Z}, & (a) \\
\mathbf{Z} \precsim_{\mathbf{K}} \mathbf{Z}' & \xRightarrow{\text{Left Redundancy E.3}} & \mathbf{Z} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}', & (b) \\
(a) \wedge (b) & \xRightarrow{\text{Left Contraction E.12}} & \mathbf{X} \otimes \mathbf{Z} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}' & \\
 & \xRightarrow{\text{Left Decomposition E.5}} & \mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}'. &
\end{array}
$$

□

Lemma E.25 (Full Equivalent Exchange)**.** *If* $\mathbf{X}' \approx_{\mathbf{K}} \mathbf{X}$ *and* $\mathbf{Y}' \approx_{\mathbf{K}} \mathbf{Y}$ *and* $\mathbf{Z}' \approx_{\mathbf{K}} \mathbf{Z}$ *then we have the equivalence:*

$$\mathbf{X} \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \qquad \iff \qquad \mathbf{X}' \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y}' \,|\, \mathbf{Z}'.$$

Proof.

$$
\begin{array}{lcl}
\mathbf{X}' \precsim_{\mathbf{K}} \mathbf{X} & \xRightarrow{\text{Lemma D.4}} & \mathbf{X}' \precsim_{\mathbf{K}} \mathbf{X} \otimes \mathbf{Z} \\
 & \xRightarrow{\text{Inverted Left Decomposition E.22}} & \mathbf{X} \otimes \mathbf{X}' \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \\
 & \xRightarrow{\text{Left Decomposition E.5}} & \mathbf{X}' \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}, \\
\mathbf{Y}' \precsim_{\mathbf{K}} \mathbf{Y} & \xRightarrow{\text{Lemma D.4}} & \mathbf{Y}' \precsim_{\mathbf{K}} \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{Z} \\
 & \xRightarrow{\mathbf{T}\text{-Ext. Inv. Right Decomposition E.23}} & \mathbf{X}' \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{T} \otimes \mathbf{Y} \otimes \mathbf{Y}' \,|\, \mathbf{Z} \\
 & \xRightarrow{\text{Right Decomposition E.6}} & \mathbf{X}' \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y}' \,|\, \mathbf{Z}, \\
\mathbf{Z}' \approx_{\mathbf{K}} \mathbf{Z} & \xRightarrow{\text{Equivalent Exchange E.24}} & \mathbf{X}' \underset{\mathbf{K}}{\perp\!\!\!\perp} \mathbf{Y}' \,|\, \mathbf{Z}'.
\end{array}
$$

The other direction works similarly. □

F. Proofs - Applications to Statistical Theory

Next we will give a proof that the classical Fisher-Neyman factorization criterion (see [Fis22, Ney36, HS49, Bur61]) implies sufficiency reformulated as transitional conditional independence.

Theorem F.1 (Fisher-Neyman). *Consider a statistical model, witten as a Markov kernel, $\mathbf{P}(X|\Theta)$. Assume that we can write it in the following form:*

$$\mathbf{P}(X \in A|\Theta = \theta) = \int_A h(x) \cdot g(S(x); \theta)\, d\boldsymbol{\mu}(x),$$

where X takes values in a standard measurable space $\mathcal{X}$, $S : \mathcal{X} \to \mathcal{S}$ is measurable, $\mathcal{S}$ countably generated, $\boldsymbol{\mu}$ is a measure on $\mathcal{X}$, $g, h \geqslant 0$ are measurable and h is $\boldsymbol{\mu}$-integrable (e.g. exponential families are of such form).
Then S is a sufficient statistic of X for $\mathbf{P}(X|\Theta)$: $\quad X \underset{\mathbf{P}(X|\Theta)}{\perp\!\!\!\perp} \Theta \,|\, S.$

Proof. Consider the function: $f := \left(\int h\, d\boldsymbol{\mu}\right) \cdot g$ and the probability measure $\mathbf{Q}(X)$ given by:

$$\mathbf{Q}(X \in A) := \frac{1}{\int h\, d\boldsymbol{\mu}} \int_A h(x)\, \boldsymbol{\mu}(dx), \qquad \text{i.e.} \quad \frac{d\mathbf{Q}}{d\boldsymbol{\mu}}(x) = \frac{h(x)}{\int h\, d\boldsymbol{\mu}}.$$

Then we have the joint distribution $\mathbf{Q}(X, S)$ and a regular conditional probability distribution $\mathbf{Q}(X|S)$, by Theorem C.9. With this we get:

$$\begin{aligned}
\mathbf{P}(S \in B|\Theta = \theta) &= \mathbf{P}(X \in S^{-1}(B)|\Theta = \theta) \\
&= \int \mathbb{1}_B(S(x)) \cdot g(S(x); \theta) \cdot h(x)\, \boldsymbol{\mu}(dx) \\
&= \int \mathbb{1}_B(S(x)) \cdot f(S(x); \theta)\, \mathbf{Q}(X \in dx) \\
&= \int \mathbb{1}_B(s) \cdot f(s; \theta)\, \mathbf{Q}(S \in ds).
\end{aligned}$$

This means:

$$\frac{\mathbf{P}(S \in ds|\Theta = \theta)}{\mathbf{Q}(S \in ds)}(s) = f(s; \theta).$$

With this we get:

$$\begin{aligned}
\mathbf{P}(X \in A, S \in B|\Theta = \theta) &= \mathbf{P}(X \in A \cap S^{-1}(B)|\Theta = \theta) \\
&= \int \mathbb{1}_A(x) \cdot \mathbb{1}_B(S(x)) \cdot g(S(x); \theta) \cdot h(x)\, \boldsymbol{\mu}(dx) \\
&= \int \mathbb{1}_A(x) \cdot \mathbb{1}_B(S(x)) \cdot f(S(x); \theta)\, \mathbf{Q}(X \in dx) \\
&= \int \int \mathbb{1}_A(x) \cdot \mathbb{1}_B(s) \cdot f(s; \theta)\, \mathbf{Q}(X \in dx, S \in ds) \\
&= \int \int \mathbb{1}_A(x) \cdot \mathbb{1}_B(s) \cdot f(s; \theta)\, \mathbf{Q}(X \in dx|S = s)\, \mathbf{Q}(S \in ds) \\
&= \int \mathbf{Q}(X \in A|S = s) \cdot \mathbb{1}_B(s) \cdot f(s; \theta)\, \mathbf{Q}(S \in ds) \\
&= \int \mathbf{Q}(X \in A|S = s) \cdot \mathbb{1}_B(s)\, \mathbf{P}(S \in ds|\Theta = \theta) \\
&= \left(\mathbf{Q}(X|S) \otimes \mathbf{P}(S|\Theta)\right)(A \times B, \theta).
\end{aligned}$$

This implies:

$$\mathbf{P}(X, S|\Theta) = \mathbf{Q}(X|S) \otimes \mathbf{P}(S|\Theta),$$

and thus:

$$X \underset{\mathbf{P}(X|\Theta)}{\perp\!\!\!\perp} \Theta \,|\, S.$$

□

This is one direction of the Fisher-Neyman factorization theorem for sufficiency (see [Fis22, Ney36, HS49]). Our definition of conditional independence generalizes the factorization theorem to Markov kernels (per definition) without the necessity of densities and/or reference measures.

The next theorem contains the proof for the propensity score of Section 4.4 and the weak likelihood principle of Section 4.5.

Theorem F.2. *Let* $\mathbf{P}(Y|X)$ *be a Markov kernel. For* $x \in \mathcal{X}$ *we put:*

$$E(x) := \mathbf{P}(Y|X = x) \;\in\; \mathcal{P}(\mathcal{Y}).$$

Then $E : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *is measurable,* $E \precsim X$ *and we have:*

$$Y \underset{\mathbf{P}(Y|X)}{\perp\!\!\!\perp} X \,|\, E.$$

Furthermore, if $S : \mathcal{X} \to \mathcal{S}$ *is another measurable map* ($S \precsim X$) *with:*

$$Y \underset{\mathbf{P}(Y|X)}{\perp\!\!\!\perp} X \,|\, S,$$

then:

$$E \precsim S \precsim X.$$

Proof. For the first claim we define $\mathcal{E} := \mathcal{P}(\mathcal{Y})$ and the Markov kernel:

$$\mathbf{Q}(Y|E) : \mathcal{E} \dashrightarrow \mathcal{Y}, \quad \mathbf{Q}(Y \in B|E = e) := e(B),$$

for $e \in \mathcal{E} = \mathcal{P}(\mathcal{Y})$ and $B \in \mathcal{B}_{\mathcal{Y}}$. Evaluating this gives:

$$\begin{aligned}
&(\mathbf{Q}(Y|E) \otimes \mathbf{P}(E|X))\,(B \times C, x) \\
&= \int_C \mathbf{Q}(Y \in B|E = e)\, \boldsymbol{\delta}(E \in de|X = x) \\
&= \mathbf{Q}(Y \in B|E = E(x)) \cdot \mathbb{1}_C(E(x)) \\
&= E(x)(B) \cdot \boldsymbol{\delta}(E \in C|X = x) \\
&= \mathbf{P}(Y \in B|X = x) \cdot \boldsymbol{\delta}(E \in C|X = x) \\
&= \mathbf{P}(Y \in B, E \in C|X = x) \\
&= \mathbf{P}(Y, E|X)(B \times C, x).
\end{aligned}$$

Multiplying both sides with $\boldsymbol{\delta}(X|X)$ implies:

$$\mathbf{Q}(Y|E) \otimes \mathbf{P}(X, E|X) = \mathbf{P}(Y, X, E|X),$$

and thus:

$$Y \underset{\mathbf{P}(Y|X)}{\perp\!\!\!\perp} X \,|\, E.$$

This shows the first claim.

For the second claim, the conditional independence:

$$Y \underset{\mathbf{P}(Y|X)}{\perp\!\!\!\perp} X \,|\, S.$$

implies that there exists a measurable function:

$$\mathbf{P}(Y|S, \not{X}) : \mathcal{S} \to \mathcal{P}(\mathcal{Y}),$$

such that:

$$\mathbf{P}(Y, S|X) = \mathbf{P}(Y|S, \not{X}) \otimes \mathbf{P}(S|X).$$

Noting that $\mathbf{P}(S|X) = \boldsymbol{\delta}(S|X)$ and marginalizing S out we get:

$$E(x) = \mathbf{P}(Y|X = x) = \mathbf{P}(Y|S = S(x), \not{X}),$$

for every $x \in \mathcal{X}$. This shows the second claim. □

We now turn to Bayesian statistics.

Theorem F.3 (Bayesian statistics)**.** *Let* $\mathbf{P}(X|\Theta)$ *be a Markov kernel between standard measurable spaces and* $\mathbf{P}(\Theta|\Pi)$ *be another Markov kernel. Then put:*

$$\mathbf{P}(X, \Theta|\Pi) := \mathbf{P}(X|\Theta) \otimes \mathbf{P}(\Theta|\Pi).$$

Then by Theorem C.9 we have a (regular) conditional Markov kernel:

$$\mathbf{P}(\Theta|X, \Pi),$$

which is unique up to $\mathbf{P}(X|\Pi)$*-null set. We now define the transitional random variable:*

$$Z(x, \pi) := \mathbf{P}(\Theta|X = x, \Pi = \pi),$$

which gives us a joint (transition) probability distribution: $\mathbf{P}(X, \Theta, Z|\Pi)$*.*
With the above notations we have the conditional independence:

$$\Theta \underset{\mathbf{P}(X,\Theta|\Pi)}{\perp\!\!\!\perp} X \,|\, Z.$$

Furthermore, if S *is another deterministic measurable function in* (X, Π) *such that:*

$$\Theta \underset{\mathbf{P}(X,\Theta|\Pi)}{\perp\!\!\!\perp} X \,|\, S,$$

then:

$$Z \precsim S \quad \mathbf{P}(X|\Pi)\text{-}a.s.$$

Proof. For the first statement consider the Markov kernel given by:

$$\mathbf{K}(\Theta \in D|Z = z) := z(D).$$

Then we get:

$$\begin{aligned}
&(\mathbf{K}(\Theta|Z) \otimes \mathbf{P}(Z, X|\Pi))\,(B \times C \times A, \pi) \\
&= \int_{C \times A} \mathbf{K}(\Theta \in B|Z = z)\,\mathbf{P}(Z \in dz, X \in dx|\Pi = \pi) \\
&= \int_A \int_C \mathbf{K}(\Theta \in B|Z = z)\,\boldsymbol{\delta}(Z \in dz|X = x, \Pi = \pi)\,\mathbf{P}(X \in dx|\Pi = \pi) \\
&= \int_A \mathbb{1}_C(Z(x, \pi)) \cdot \mathbf{P}(\Theta \in B|X = x, \Pi = \pi)\,\mathbf{P}(X \in dx|\Pi = \pi) \\
&= \int_A \int_C \mathbf{P}(\Theta \in B|X = x, \Pi = \pi)\,\boldsymbol{\delta}(Z \in dz|X = x, \Pi = \pi)\,\mathbf{P}(X \in dx|\Pi = \pi) \\
&= \mathbf{P}(\Theta \in B, Z \in C, X \in A|\Pi = \pi).
\end{aligned}$$

This shows the first claim.
The conditional independence in the second claim gives us the factorization:

$$\begin{aligned}
\mathbf{P}(\Theta, S, X|\Pi) &= \mathbf{P}(\Theta|S, \cancel{X, \Pi}) \otimes \mathbf{P}(S, X|\Pi) \\
&= \mathbf{P}(\Theta|S, \cancel{X, \Pi}) \otimes \boldsymbol{\delta}(S|X, \Pi) \otimes \mathbf{P}(X|\Pi).
\end{aligned}$$

On the other hand we have:

$$\mathbf{P}(\Theta, S, X|\Pi) = \mathbf{P}(\Theta|X, \Pi) \otimes \boldsymbol{\delta}(S|X, \Pi) \otimes \mathbf{P}(X|\Pi).$$

Marginalizing S out in those equations gives:

$$\mathbf{P}(\Theta|S = S(X, \Pi), \cancel{X, \Pi}) \otimes \mathbf{P}(X|\Pi) = \mathbf{P}(\Theta|X, \Pi) \otimes \mathbf{P}(X|\Pi).$$

Because conditional Markov kernels are essentially unique by C.2 we get:

$$\mathbf{P}(\Theta|S = S(X, \Pi), \cancel{X, \Pi})) = \mathbf{P}(\Theta|X, \Pi) = Z(X, \Pi) \qquad \mathbf{P}(X|\Pi)\text{-a.s..}$$

This shows:

$$Z \precsim S \quad \mathbf{P}(X|\Pi)\text{-a.s.}$$

□

Theorem F.4 (A weak likelihood principle for Bayesian statistics)**.** *We also have the transitional conditional independence with $L(\theta) := \mathbf{P}(X|\Theta = \theta)$:*

$$X \underset{\mathbf{P}(X,\Theta|\Pi)}{\perp\!\!\!\perp} \Theta, \Pi \,|\, L.$$

Furthermore, if $\mathcal{X}$ is countably generated, then any other measurable map S in Θ with:

$$X \underset{\mathbf{P}(X,\Theta|\Pi)}{\perp\!\!\!\perp} \Theta, \Pi \,|\, S,$$

satisfies:

$$L \precsim S \precsim \Theta \quad \mathbf{P}(\Theta|\Pi)\text{-}a.s.$$

Proof. We have:

$$\mathbf{P}(X, L, \Theta, \Pi|\Pi) = \mathbf{Q}(X|L) \otimes \mathbf{P}(L, \Theta, \Pi|\Pi),$$

with $\mathbf{Q}(X \in A|L = \ell) := \ell(A)$ for $\ell \in \mathcal{P}(\mathcal{X})$.
For the reverse direction we get a factorization:

$$\begin{aligned}\mathbf{P}(S, X, \Theta|\Pi) &= \mathbf{K}(X|S) \otimes \mathbf{P}(S, \Theta|\Pi) \\ &= \mathbf{K}(X|S) \otimes \boldsymbol{\delta}(S|\Theta) \otimes \mathbf{P}(\Theta|\Pi).\end{aligned}$$

Marginalizing S out on both sides gives:

$$\mathbf{P}(X|\Theta) \otimes \mathbf{P}(\Theta|\Pi) = \mathbf{K}(X|S = S(\Theta)) \otimes \mathbf{P}(\Theta|\Pi).$$

Since such factorizations are essentially unique by Lemma C.2 and $\mathcal{X}$ is countably generated, we have that for $\mathbf{P}(\Theta|\Pi)$-almost-all (θ, π) we get:

$$L(\theta) = \mathbf{P}(X|\Theta = \theta) = \mathbf{K}(X|S = S(\theta)).$$

This shows:

$$L \precsim S \precsim \Theta \quad \mathbf{P}(\Theta|\Pi)\text{-a.s.}$$

□

G. Proofs - Reparameterization of Transitional Random Variables

In this section we generalize a few folklore results via now standard techniques that were introduced after [Dar53].

Lemma G.1. *Let $\bar{\mathbb{R}} := [-\infty, \infty]$ be endowed with the usual ordering and Borel σ-algebra. Let $\mathbf{P}$ be a probability measure on $\bar{\mathbb{R}}$ and $F(x) := \mathbf{P}([-\infty, x])$. Then $F : \bar{\mathbb{R}} \to [0, 1]$ is non-decreasing, right-continous with at most countably many discontinuities and $F(\infty) = 1$. So $R(t) := \inf F^{-1}([t, 1])$ is a well-defined map $R : [0, 1] \to \bar{\mathbb{R}}$, non-decreasing, left-continous with at most countably many discontinuities and $R(0) = -\infty$. Furthermore, for $x \in \bar{\mathbb{R}}$ and $t \in [0, 1]$ we have:*

$$t \leqslant F(x) \iff R(t) \leqslant x.$$

In particular, we have $F(R(t)) \geqslant t$, thus $R(t) \in F^{-1}([t, 1])$ the minimal element. We also have $R(F(x)) \leqslant x$, with equality if and only if $x \in R([0, 1])$. Furthermore, F and R are measurable and $R_*\boldsymbol{\lambda} = \mathbf{P}$. We also have that R is a reflexive generalized inverse of F, i.e.:

$$F \circ R \circ F = F, \qquad R \circ F \circ R = R.$$

Proof. From the properties of $\mathbf{P}$ it is clear that F is non-decreasing, right-continuous and $F(\infty) = 1$.
Let $D_F \subseteq \bar{\mathbb{R}}$ be the set of discontinuities of F and $x \in D_F$. Then there exists a $q(x) \in \mathbb{Q}$ such that $F_-(x) < q(x) < F_+(x)$. If now $x_1 < x_2$ are two such points we get: $q(x_1) < F_+(x_1) \leqslant F_-(x_2) < q(x_2)$. So the map $q : D_F \to \mathbb{Q}$ is injective. Thus D_F is countable.
Next, we show that $R(t) \in F^{-1}([t, 1])$, thus $R(t) = \min F^{-1}([t, 1])$. For this let $(x_n)_{n\in\mathbb{N}} \subseteq F^{-1}([t, 1])$ be a non-increasing sequence converging to $R(t)$. Then by the right-continuity $F(x_n)$ converges to $F(R(t))$ from above. So we have:

$$F(R(t)) = \inf_{n\in\mathbb{N}} F(x_n) \geqslant t.$$

It follows that $F(R(t)) \geqslant t$ and thus $R(t) \in F^{-1}([t, 1])$. This shows the claim.
R is clearly non-decreasing, thus has only a countable set of discontiuities $D_R \subseteq [0, 1]$ by the same arguments before, and $R(0) = -\infty$. To see that $R(t)$ is left-continuous let $t \in [0, 1]$ and $(t_n)_{n\in\mathbb{N}}$ a non-decreasing sequence converging to t from below. Then by the monotonicity of R we have $\sup_{n\in\mathbb{N}} R(t_n) \leqslant R(t)$. On the other hand we have:

$$t = \sup_{n\in\mathbb{N}} t_n \leqslant \sup_{n\in\mathbb{N}} F(R(t_n)) \leqslant F(\sup_{n\in\mathbb{N}} R(t_n)),$$

implying: $\sup_{n\in\mathbb{N}} R(t_n) \in F^{-1}([t, 1])$ and thus $\sup_{n\in\mathbb{N}} R(t_n) \geqslant R(t)$, leading to equality, which shows the claim.
For any $x \in \bar{\mathbb{R}}$ we have the implication:

$$x \geqslant R(t) \implies F(x) \geqslant F(R(t)) \geqslant t.$$

For any $x \in \bar{\mathbb{R}}$ and any $t \in [0, 1]$ we have the implications:

$$\begin{aligned} t \leqslant F(x) &\iff F(x) \in [t, 1] \\ &\iff x \in F^{-1}([t, 1]) \\ &\implies x \geqslant \inf F^{-1}([t, 1]) = R(t). \end{aligned}$$

Together this shows for any $x \in \bar{\mathbb{R}}$ and $t \in [0, 1]$ the equivalence:

$$t \leqslant F(x) \iff R(t) \leqslant x.$$

Since $F(x) \leqslant F(x)$ we get $R(F(x)) \leqslant x$ for all $x \in \bar{\mathbb{R}}$. If equality holds then $x \in R([0, 1])$. And, if $x = R(t)$ for some $t \in [0, 1]$ then we use the inequalities $x \geqslant R(F(x))$ and $F(R(t)) \geqslant t$ to conclude:

$$x \geqslant R(F(x)) = R(F(R(t))) \geqslant R(t) = x,$$

showing equality, and that:

$$R \circ F \circ R = R.$$

Similarly for $t = F(x)$ we get:

$$t \leqslant F(R(t)) = F(R(F(x))) \leqslant F(x) = t,$$

showing

$$F \circ R \circ F = F.$$

Now consider the uniform distribution $\boldsymbol{\lambda}$ on $[0,1]$ and any $x \in \bar{\mathbb{R}}$. Then we have:

$$\begin{aligned}
(R_*\boldsymbol{\lambda})([-\infty, x]) &= \boldsymbol{\lambda}(R^{-1}([-\infty, x])) \\
&= \boldsymbol{\lambda}(t \in [0,1] \,|\, R(t) \leqslant x) \\
&= \boldsymbol{\lambda}(t \in [0,1] \,|\, t \leqslant F(x)) \\
&= \boldsymbol{\lambda}([0, F(x)]) \\
&= F(x) \\
&= \mathbf{P}([-\infty, x]).
\end{aligned}$$

It follows that: $R_*\boldsymbol{\lambda} = \mathbf{P}$. □

Lemma G.2. *Let the notation be like in G.1. For $u \in [0,1]$ and $x \in \bar{\mathbb{R}}$ define:*

$$F_u(x) := E(x; u) := \mathbf{P}([-\infty, x)) + u \cdot \mathbf{P}(\{x\}).$$

Then $E : \bar{\mathbb{R}} \times [0,1] \to [0,1]$ is measurable, non-decreasing in both arguments with $F_0(-\infty) = 0$, $F_1(\infty) = 1$, F_0 is left-continuous and

$$F_u(\tilde{x}) \leqslant F_1(\tilde{x}) \leqslant F_0(x) \leqslant F_u(x)$$

for any $\tilde{x} < x$, $u \in [0,1]$. We further have for every $u \in (0,1]$:

$$R \circ F_u \circ R = R,$$

and $R \circ F_u = \mathrm{id}_{\bar{\mathbb{R}}}$ $\mathbf{P}$-almost-surely for any $u \in (0,1]$.

Proof. Most of the properties are clear from its definition. Let $x < \tilde{x}$ then $[-\infty, x] \subseteq [-\infty, \tilde{x})$ and thus $F_1(x) \leqslant F_0(\tilde{x})$.
To show $R \circ F_u \circ R = R$ fix a $t \in [0,1]$, $u \in (0,1]$ and let $x := R(t)$. If F_1 is continuous in x then $F_u = F_1$ and the claim $R \circ F_1 \circ R = R$ was already shown using the inequalities:

$$x \geqslant R(F_1(x)) = R(F_1(R(t))) \geqslant R(t) = x.$$

So let us assume that F_1 is discontinuous in $x = R(t)$. Then $F_u(x) \in (F_0(x), F_1(x)]$. We have:

$$R(F_u(x)) = \min\{\tilde{x} \in \bar{\mathbb{R}} \,|\, F_1(\tilde{x}) \geqslant F_u(x)\}.$$

If $F_1(\tilde{x}) \geqslant F_u(x) > F_0(x)$ then $\tilde{x} \geqslant x$, otherwise $\tilde{x} < x$ leads to the contradiction $F_1(\tilde{x}) \leqslant F_0(x)$. Since clearly $F_1(x) \geqslant F_u(x)$ we must have:

$$R(F_u(x)) = x,$$

with $x = R(t)$, which proves the claim: $R \circ F_u \circ R = R$ for $u \in (0,1]$.
We now want to show that $R \circ F_u = \mathrm{id}_{\bar{\mathbb{R}}}$ $\mathbf{P}$-a.s. for $u \in (0,1]$. From $R \circ F_u \circ R = R$ we already see, that $R \circ F_u|_{R([0,1])} = \mathrm{id}_{R([0,1])}$. We will see below that $C := \bar{\mathbb{R}} \backslash R([0,1])$ is measurable and $\mathbf{P}(C) = 0$, which will prove the claim.

In the following we will only need $F = F_1$. First, by G.1 we know that for any $x \in \bar{\mathbb{R}}$ we have $R(F(x)) \leqslant x$ with equality if and only if $x \in R([0,1])$. So this gives us the equivalence:

$$x \in C \iff x > R(F(x)).$$

We now claim that $(R(F(x)), x] \subseteq C$ for every $x \in C$: Indeed, If $\tilde{x} \in (R(F(x)), x]$ then:

$$F(x) = F(R(F(x))) \leqslant F(\tilde{x}) \leqslant F(x)$$

and thus $F(\tilde{x}) = F(x)$, from which follows that $R(F(\tilde{x})) = R(F(x)) < \tilde{x}$ and ergo $\tilde{x} \in C$. It follows that C is the union of such intervals $(R(F(x)), x]$ with $x \in C$. Furthermore, $F(C)$ is contained in the set of discontinuities D_R of R: otherwise there would be an $x \in C$ and a $t \geqslant F(x)$ such that $R(t) \in (R(F(x)), x] \subseteq C$, which is a contradiction. Since D_R is countable it must follow that $F(C)$ and thus also $R(F(C))$ is at most countable. Write $R(F(C)) = \{x_n \,|\, n \in \mathbb{N}\}$, which is the set of the possible left end-points of the above intervals. For each fixed $n \in \mathbb{N}$ let

$$C_n := \{x \in C \,|\, R(F(x)) = x_n\},$$

which is, as a union of intervals $(x_n, x]$, $x \in C_n$, either of the form $(x_n, \bar{x}_n]$ or $(x_n, \bar{x}_n)$ with $\bar{x}_n := \sup C_n$. In both cases we can cover C_n by $C_{n,m} := (x_n, x_{n,m}]$ with $x_{n,m} \in C_n$ either equal to $\bar{x}_n$ or converging to it from below for running m. So we can write C as the countable union:

$$C = \bigcup_{n,m \in \mathbb{N}} C_{n,m}.$$

We now have for each $x = x_{n,m}$:

$$\mathbf{P}(C_{n,m}) = \mathbf{P}((x_n, x]) = \mathbf{P}((R(F(x)), x]) = F(x) - F(R(F(x))) = F(x) - F(x) = 0.$$

This implies:

$$\mathbf{P}(C) = \mathbf{P}\left(\bigcup_{n,m \in \mathbb{N}} C_{n,m}\right) \leqslant \sum_{n,m \in \mathbb{N}} \mathbf{P}(C_{n,m}) = 0,$$

showing that $\mathbf{P}(C) = 0$ and thus:

$$R \circ F_u = \mathrm{id}_{\bar{\mathbb{R}}} \qquad \mathbf{P}\text{-a.s.}$$

for $u \in (0,1]$. □

Lemma G.3. *Let the notations be like in G.1 and G.2. Let $\boldsymbol{\lambda}$ be the uniform distribution on $[0,1]$ and $\bar{\mathbf{P}} := \mathbf{P} \otimes \boldsymbol{\lambda}$ the product distribution on $\bar{\mathbb{R}} \times [0,1]$. For every $e \in [0,1]$ define the event:*

$$\{E \leqslant e\} := \{(x,u) \in \bar{\mathbb{R}} \times [0,1] \mid E(x;u) \leqslant e\}.$$

Then $\bar{\mathbf{P}}(E \leqslant e) = e$. In other words, the random variable:

$$\begin{array}{rccl} E & : & \bar{\mathbb{R}} \times [0,1] & \to & [0,1], \\ & & (x,u) & \mapsto & \mathbf{P}([-\infty, x)) + u \cdot \mathbf{P}(\{x\}), \end{array}$$

is uniformly distributed under $\bar{\mathbf{P}} = \mathbf{P} \otimes \boldsymbol{\lambda}$.
Furthermore, $R(E) = X$ $\bar{\mathbf{P}}$*-a.s., where* $X : \bar{\mathbb{R}} \times [0,1] \to \bar{\mathbb{R}}$ *is the canonical projection onto the first factor:* $X(x,u) := x$*, and which has distribution* $\mathbf{P}$*.*

Proof. First, since $\boldsymbol{\lambda}(\{0\}) = 0$ we can w.l.o.g. exclude $u = 0$ and restrict $\bar{\mathbf{P}}$ to $\bar{\mathbb{R}} \times (0,1]$. We have seen in G.2 that $R \circ F_u \circ R = R$ for $u \in (0,1]$, which translates to:

$$R \circ E|_{R([0,1]) \times (0,1]} = X|_{R([0,1]) \times (0,1]}.$$

Also with $C := \bar{\mathbb{R}} \backslash R([0,1])$ we get:

$$\bar{\mathbf{P}}(C \times (0,1]) = \mathbf{P}(C) \cdot \boldsymbol{\lambda}((0,1]) = 0 \cdot 1 = 0.$$

So we get the second claim that:

$$R \circ E = X \qquad \bar{\mathbf{P}}\text{-a.s.}$$

Now we turn to $\{E \leqslant e\}$ for $e \in [0,1]$. We abbreviate $U : \bar{\mathbb{R}} \times [0,1] \to [0,1]$ to be the projection onto the second factor: $U(x,u) := u$, which is uniformly distributed under $\bar{\mathbf{P}}$, and also $p(x) := \mathbf{P}(\{x\}) = F_1(x) - F_0(x)$. With these notations: $E = F_0(X) + U \cdot p(X)$. First, we show that $\bar{\mathbf{P}}(E = e) = 0$ for all $e \in [0,1]$. For this let $x := R(e)$. Then by the above ($R(E) = X$ $\bar{\mathbf{P}}$-a.s.) we have:

$$\bar{\mathbf{P}}(E = e) = \bar{\mathbf{P}}(E = e,\ X = x).$$

We have to distinguish between two cases: $p(x) = 0$ and $p(x) > 0$.
Case $p(x) = 0$: We have:

$$\begin{aligned} \bar{\mathbf{P}}(E = e) &= \bar{\mathbf{P}}(E = e, X = x) \\ &\leqslant \bar{\mathbf{P}}(X = x) \\ &= p(x) \\ &= 0. \end{aligned}$$

Case $p(x) > 0$: We get:

$$\begin{aligned} \bar{\mathbf{P}}(E = e) &= \bar{\mathbf{P}}(E = e, X = x) \\ &= \bar{\mathbf{P}}(F_0(X) + U \cdot p(X) = e,\ X = x) \\ &= \bar{\mathbf{P}}\left(U = \frac{e - F_0(x)}{p(x)},\ X = x\right) \\ &= \boldsymbol{\lambda}\left(\left\{\frac{e - F_0(x)}{p(x)}\right\}\right) \cdot p(x) \\ &= 0. \end{aligned}$$

To prove $\bar{\mathbf{P}}(E \leqslant e) = e$ for $e \in [0,1]$ we have several cases:
Case $e \in F_1(\bar{\mathbb{R}})$: Let $\tilde{x}$ be any element in $\bar{\mathbb{R}}$ with $e = F_1(\tilde{x})$ (e.g. $\tilde{x} = R(e)$). Then we

get:

$$\begin{aligned}\bar{\mathbf{P}}(E \leqslant e) &= \bar{\mathbf{P}}(E \leqslant F_1(\tilde{x})) \\ &= \bar{\mathbf{P}}(R(E) \leqslant \tilde{x}) \\ &\overset{R\circ E=X}{=} \bar{\mathbf{P}}(X \leqslant \tilde{x}) \\ &= \mathbf{P}([-\infty, \tilde{x}]) \cdot \boldsymbol{\lambda}((0,1]) \\ &= F_1(\tilde{x}) \cdot 1 \\ &= e.\end{aligned}$$

For the cases $e \notin F_1(\bar{\mathbb{R}})$ we put $x := R(e)$ and $\tilde{e} := F_0(x)$.
Then by definition, x is minimal with $F_1(x) \geqslant e$. We also have $\tilde{e} = F_0(x) \leqslant e$. Otherwise: $e < F_0(x) = \sup_{\tilde{x}<x} F_1(\tilde{x})$ implied that there existed $\tilde{x} < x$ with $e < F_1(\tilde{x}) \leqslant F_0(x)$, which is a contradiction to the minimality of $x = R(e)$. Since $\tilde{e} \leqslant e$ we can decompose:

$$\bar{\mathbf{P}}(E \leqslant e) = \bar{\mathbf{P}}(E < \tilde{e}) + \bar{\mathbf{P}}(E = \tilde{e}) + \bar{\mathbf{P}}(\tilde{e} < E \leqslant e).$$

We have already seen that the second term $\bar{\mathbf{P}}(E = \tilde{e}) = 0$ vanishes.
For the first term we have:

$$\begin{aligned}\bar{\mathbf{P}}(E < \tilde{e}) &= \bar{\mathbf{P}}(E < F_0(x)) \\ &= \bar{\mathbf{P}}(E < F_0(x)) \\ &= \bar{\mathbf{P}}(E < \sup_{\tilde{x}<x} F_1(\tilde{x})) \\ &= \sup_{\tilde{x}<x} \bar{\mathbf{P}}(E \leqslant F_1(\tilde{x})) \\ &\overset{(*)}{=} \sup_{\tilde{x}<x} F_1(\tilde{x}) \\ &= F_0(x) \\ &= \tilde{e}.\end{aligned}$$

Equation (*) comes from the previous case for $F_1(\tilde{x}) \in F_1(\bar{\mathbb{R}})$.
For the third term $\bar{\mathbf{P}}(\tilde{e} < E \leqslant e)$ first note that $E \in (\tilde{e}, e]$ implies that $X = x$ $\bar{\mathbf{P}}$-a.s. by applying R: Indeed, every element $t \in (\tilde{e}, e] \subseteq (F_0(x), F_1(x)]$ can be written as $t = F_{\tilde{u}}(x)$ for an $\tilde{u} \in (0, 1]$ and we can use:

$$R(t) = R(F_{\tilde{u}}(R(e)) = R(e) = x.$$

For $p(x) > 0$ and the above we get:

$$\begin{aligned}
\bar{\mathbf{P}}(\tilde{e} < E \leqslant e) &= \bar{\mathbf{P}}(\tilde{e} < E \leqslant e,\, X = x) \\
&= \bar{\mathbf{P}}(0 < F_0(X) + U \cdot p(X) - F_0(x) \leqslant e - \tilde{e}, X = x) \\
&= \bar{\mathbf{P}}(0 < U \leqslant \frac{e - \tilde{e}}{p(x)}, X = x) \\
&= \boldsymbol{\lambda}\left(\left(0, \frac{e - \tilde{e}}{p(x)}\right]\right) \cdot \mathbf{P}(\{x\}) \\
&= \frac{e - \tilde{e}}{p(x)} \cdot p(x) \\
&= e - \tilde{e}.
\end{aligned}$$

For the case $p(x) = 0$, the first row can be upper bounded by $\bar{\mathbf{P}}(X = x) = p(x) = 0$ as before, but we also have $\tilde{e} - e = 0$ in this case, and the equality stays trivially true as well.
Putting all together we get:

$$\begin{aligned}
\bar{\mathbf{P}}(E \leqslant e) &= \bar{\mathbf{P}}(E < \tilde{e}) + \bar{\mathbf{P}}(E = \tilde{e}) + \bar{\mathbf{P}}(\tilde{e} < E \leqslant e) \\
&= \tilde{e} + 0 + e - \tilde{e} \\
&= e.
\end{aligned}$$

This shows the claim. ☐

Theorem G.4. *Let $\mathcal{Z}$ be any measurable space and $\mathcal{X}$ be a standard measurable space with a fixed embedding $\iota : \mathcal{X} \hookrightarrow \bar{\mathbb{R}} = [-\infty, \infty]$ onto a Borel subset (which always exists, so w.l.o.g. $\mathcal{X} = \bar{\mathbb{R}}$ endowed with the Borel σ-algebra). Let $\mathbf{K}(X|Z) : \mathcal{Z} \dashrightarrow \mathcal{X}$ be a Markov kernel. Furthermore, let $\mathcal{U} := [0, 1]$ and $\mathbf{K}(U)$ be the uniform distribution/Markov kernel on $\mathcal{U}$. We write:*

$$\mathbf{K}(U, X|Z) := \mathbf{K}(U) \otimes \mathbf{K}(X|Z).$$

Also put:

$$\begin{aligned}
F(x; u|z) &:= \mathbf{K}(X < x|Z = z) + u \cdot \mathbf{K}(X = x|Z = z) \\
R(e|z) &:= \inf \left\{\tilde{x} \in \mathcal{X} \,|F(\tilde{x}; 1|z) \geqslant e\right\}.
\end{aligned}$$

Let $E := F(X; U|Z)$. We consider X, U, Z, E as the measurable maps:

$$\begin{aligned}
X : \quad \mathcal{X} \times \mathcal{U} \times \mathcal{Z} &\to \mathcal{X}, \\
(x, u, z) &\mapsto x, \\
U : \quad \mathcal{X} \times \mathcal{U} \times \mathcal{Z} &\to \mathcal{U}, \\
(x, u, z) &\mapsto u, \\
Z : \quad \mathcal{X} \times \mathcal{U} \times \mathcal{Z} &\to \mathcal{Z}, \\
(x, u, z) &\mapsto z, \\
E : \quad \mathcal{X} \times \mathcal{U} \times \mathcal{Z} &\to \mathcal{E} := [0, 1], \\
(x, u, z) &\mapsto F(x; u|z).
\end{aligned}$$

Then for all $e \in \mathcal{E}$ and $z \in \mathcal{Z}$ we have:

$$\mathbf{K}(E \leqslant e|Z = z) = e,$$

implying $E \perp\!\!\!\perp_{\mathbf{K}(U,X|Z)} Z$. Furthermore, we have:

$$X = R(E|Z) \quad \mathbf{K}(U, X|Z)\text{-a.s.}.$$

Proof. After the measurabilities are checked the statement directly follows from G.3 by applying it for every z separately. ☐

Corollary G.5. *Let X and Z be random variables with values in any standard measurable spaces $\mathcal{X}$ and $\mathcal{Z}$, resp., and with a joint distribution $\mathbf{P}(X, Z)$. Then there exists a uniformly distributed random variable E on $[0,1]$ that is $\mathbf{P}$-independent of Z and a measurable function g such that $X = g(E, Z)$ $\mathbf{P}$-almost-surely. Furthermore, E can be constructed via a deterministic measurable function in X and Z and (uniformly distributed) independent noise U (on $[0,1]$).*

Proof. The regular conditional probability distribution $\mathbf{P}(X|Z)$ exists for standard measurable spaces (and is unique up to a $\mathbf{P}(Z)$-zero-set), and is a Markov kernel. Then apply the result from above for $\mathbf{K}(X|Z) := \mathbf{P}(X|Z)$ to get $g(e, z) := R(e|z)$ and E. ☐

Remark G.6. *Any Polish space, i.e. any completely metrizable topological space with a countable dense subset (separable), is a standard measurable space in its Borel σ-algebra. These are fundamental theorems in classical descriptive set theory, see [Bog07, Fre15, Kec95]. Examples of Polish and thus standard measurable spaces are $[0,1]$, $\mathbb{R}$, $\mathbb{R}^d$, $\mathbb{N}$, any (discrete) finite or countable set, any topological or smooth manifold $\mathcal{X}$, any finite (or even countable) CW-complex $\mathcal{Y}$, etc., (in its usual Borel σ-algebra). So these are all measurably isomorphic to a Borel subset of $[0,1]$ (or $\bar{\mathbb{R}}$), and measurably isomorphic to $[0,1]$ (or $\bar{\mathbb{R}}$) itself if non-countable (excluding finite and countable sets).*

H. Operations on Graphs

In this section we present how one can create new CDMGs from old ones through operations like hard interventions, extensions by soft intervention nodes, marginalization and acyclification.

Definition H.1 (Hard interventional CDMG)**.** *Let $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J)) = (J, V, E, L)$ be CDMG and $W \subseteq J \cup V$. Then the* hard interventional CDMG intervened on W *is $\mathbf{G}(V \setminus W|\,\mathrm{do}(J \cup W)) := (J \cup W, V \setminus W, E_{\mathrm{do}(W)}, L_{\mathrm{do}(W)})$, where:*

1. $E_{\mathrm{do}(W)} := \{v_1 \rightarrowtail v_2 \in E \,|\, v_1 \in J \cup V, v_2 \in V \setminus W\}$,
2. $L_{\mathrm{do}(W)} := \{v_1 \leftrightarrow v_2 \in L \,|\, v_1, v_2 \in V \setminus W\}$.

Definition H.2 (Soft interventional CDMG)**.** *Let $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J)) = (J, V, E, L)$ be CDMG and $W \subseteq J \cup V$. Then the* soft interventional CDMG extended on W *is $\mathbf{G}(V|\,\mathrm{do}(J, I_W)) := (J \,\dot\cup\, I_W, V, E_{\mathrm{do}(I_W)}, L)$, where:*

1. $I_W := \{I_w \,|\, w \in W\}$ *is a set of new input nodes, one* I_w *for each* $w \in W$.

2. $E_{\mathrm{do}(I_W)} := E \,\dot\cup\, \{I_w \rightarrowtail w \,|\, w \in W \backslash J\}$.

Definition H.3 (Marginalized CDMG, latent projection)**.** *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J)) = (J, V, E, L)$ *be CDMG and* $W \subseteq V$*. Then the* latent projection of $\mathbf{G}$ onto $V \backslash W$ *or the* marginal CDMG*, where* W *is marginalized out, is the CDMG* $\mathbf{G}(V \backslash W|\,\mathrm{do}(J)) := (J, V \backslash W, E^{\backslash W}, L^{\backslash W})$*, where:*

1. $\overline{v} \rightarrowtail \underline{v} \in E^{\backslash W}$ *iff there exists* $n \geqslant 1$, $w_1, \dots, w_{n-1} \in W$ *and a directed walk in* $\mathbf{G}$*:*

$$\overline{v} = w_0 \rightarrowtail \cdots \rightarrowtail w_n = \underline{v}.$$

2. $\overline{v} \leftrightarrow \underline{v} \in L^{\backslash W}$ *iff there exists* $n \geqslant 1$, $w_1, \dots, w_{n-1} \in W$ *and a walk in* $\mathbf{G}$ *of the form (of a so called trek or arc):*

$$\overline{v} = w_0 \leftarrowtail \cdots \leftarrowtail w_{k-1} \ast\!\!\rightarrow w_k \rightarrowtail \cdots \rightarrowtail w_n = \underline{v},$$

where there are arrow heads pointing to $\overline{v}$ *and* $\underline{v}$ *and every other node has at most one arrow head pointing to it. Note that these include:* $\overline{v} \leftrightarrow \underline{v}$, $\overline{v} \leftarrowtail w_1 \rightarrowtail \underline{v}$, $\overline{v} \leftrightarrow w_1 \rightarrowtail \underline{v}$, $\overline{v} \leftarrowtail w_1 \leftrightarrow \underline{v}$*, etc., but not e.g.:* $\overline{v} \leftrightarrow w_1 \leftrightarrow \underline{v}$.

Remark H.4 (Marginalization preserves ancestral relations and acyclicity)**.** *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG and* $\mathbf{G}' = \mathbf{G}(V \backslash W|\,\mathrm{do}(J))$ *its marginalization.*

1. *Then for* $v_1, v_2 \in \mathbf{G}$ *with* $v_1, v_2 \notin W$ *we have the equivalence:*

$$v_1 \in \mathrm{Anc}^{\mathbf{G}}(v_2) \qquad \Longleftrightarrow \qquad v_1 \in \mathrm{Anc}^{\mathbf{G}'}(v_2).$$

2. *If the CDMG* $\mathbf{G}(V|\,\mathrm{do}(J))$ *is acyclic then also* $\mathbf{G}(V \backslash W|\,\mathrm{do}(J))$ *and a topological order* $<$ *of* $\mathbf{G}(V|\,\mathrm{do}(J))$ *induces a topological order on* $\mathbf{G}(V \backslash W|\,\mathrm{do}(J))$ *(by just ignoring the nodes from* W*).*

Lemma H.5 (Marginalization preserves σ-separation, see [FM17, FM18, FM20])**.** *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG. Let* $A, B, C \subseteq J \,\dot\cup\, V$ *and* $W \subseteq V$ *such that* $(A \cup B \cup C) \cap W = \emptyset$*. Then we have the equivalence:*

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \qquad \Longleftrightarrow \qquad A \underset{\mathbf{G}(V \backslash W|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C.$$

Notation H.6. *Let* $\mathbf{G}(V|\,\mathrm{do}(J)) = (J, V, E, L)$ *be a CDMG. Let us write:* $\mathbf{G}(V \,\dot\cup\, J|\,\mathrm{do}(\emptyset))$ *for the CDMG:* $(\emptyset, J \,\dot\cup\, V, E, L)$*, where we interprete all nodes from* J *the same as the nodes from* V*. Then we have by definition:*

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \qquad \Longleftrightarrow \qquad A \underset{\mathbf{G}(V \,\dot\cup\, J|\,\mathrm{do}(\emptyset))}{\overset{\sigma}{\perp}} J \cup B \,|\, C.$$

Definition H.7 (Acyclification of CDMGs, see [FM17,FM18,FM20]). *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG. We define the* acyclification $\mathbf{G}^{\mathrm{acy}} = \mathbf{G}^{\mathrm{acy}}(V|\,\mathrm{do}(J)) = (J, V, E^{\mathrm{acy}}, L^{\mathrm{acy}})$ *of* $\mathbf{G}(V|\,\mathrm{do}(J))$ *as follows:*

1. $v \rightarrowtail w \in \mathbf{G}^{\mathrm{acy}}$ *iff* $v \notin \mathrm{Sc}^{\mathbf{G}}(w)$ *and there exists* $\tilde{w} \in \mathrm{Sc}^{\mathbf{G}}(w)$ *such that* $v \rightarrowtail \tilde{w} \in \mathbf{G}$*:*

$$E^{\mathrm{acy}} := \left\{ v \rightarrowtail w \,|\, v \notin \mathrm{Sc}^{\mathbf{G}}(w) \,\wedge\, \exists \tilde{w} \in \mathrm{Sc}^{\mathbf{G}}(w) :\; v \rightarrowtail \tilde{w} \in \mathbf{G} \right\}.$$

2. $v \leftrightarrow w \in \mathbf{G}^{\mathrm{acy}}$ *iff either* $v \in \mathrm{Sc}^{\mathbf{G}}(w)$ *or if there exists* $\tilde{v} \in \mathrm{Sc}^{\mathbf{G}}(v)$ *and* $\tilde{w} \in \mathrm{Sc}^{\mathbf{G}}(w)$ *such that* $\tilde{v} \leftrightarrow \tilde{w} \in \mathbf{G}$*:*

$$L^{\mathrm{acy}} := \left\{ v \leftrightarrow w \,|\, \left(v \in \mathrm{Sc}^{\mathbf{G}}(w)\right) \,\vee\, \left(\exists \tilde{v} \in \mathrm{Sc}^{\mathbf{G}}(v) \,\wedge\, \exists \tilde{w} \in \mathrm{Sc}^{\mathbf{G}}(w) :\; \tilde{v} \leftrightarrow \tilde{w} \in \mathbf{G}\right) \right\}.$$

Theorem H.8 (See [FM17,FM18,FM20]). *Let* $\mathbf{G} = \mathbf{G}(V|\,\mathrm{do}(J))$ *be a CDMG, then its acyclification* $\mathbf{G}^{\mathrm{acy}}(V|\,\mathrm{do}(J))$ *is a conditional* acyclic *directed mixed graph (CDAMG) and for all subsets* $A, B, C \subseteq J \cup V$ *we have the equivalence:*

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \quad\Longleftrightarrow\quad A \underset{\mathbf{G}^{\mathrm{acy}}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \quad\Longleftrightarrow\quad A \underset{\mathbf{G}^{\mathrm{acy}}(V \,\dot{\cup}\, J|\,\mathrm{do}(\emptyset))}{\overset{\sigma}{\perp}} J \cup B \,|\, C,$$

where the both right hand relations can, due their acyclicity, ignore the the extra conditions $v_k \notin \mathrm{Sc}^{\mathbf{G}}(v_{\pm k})$ *in the Definition 5.8 of* σ*-separation.*

I. Proofs - Separoid Rules for Sigma-Separation

In the following let $\mathbf{G}(V|\,\mathrm{do}(J))$ be a CDMG and $A, B, C, D \subseteq J \cup V$ (not necessarily disjoint) subsets of nodes.

Since by Theorem H.8 we have the equivalence:

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \quad\Longleftrightarrow\quad A \underset{\mathbf{G}^{\mathrm{acy}}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \quad\Longleftrightarrow\quad A \underset{\mathbf{G}^{\mathrm{acy}}(V \,\dot{\cup}\, J|\,\mathrm{do}(\emptyset))}{\overset{\sigma}{\perp}} J \cup B \,|\, C,$$

we can restrict ourselves to *acyclic* graphs. It is known that for acylic graphs without input nodes like $\mathbf{G}^{\mathrm{acy}}(V \,\dot{\cup}\, J|\,\mathrm{do}(\emptyset))$ the relation $\perp^{\sigma}_{\mathbf{G}^{\mathrm{acy}}(V \,\dot{\cup}\, J|\,\mathrm{do}(\emptyset))}$ is a (symmetric) separoid, see e.g. [Ric03,FM17]. By the general theory in Appendix A Remark A.4 we then know that $\perp^{\sigma}_{\mathbf{G}^{\mathrm{acy}}(V \,\dot{\cup}\, J|\,\mathrm{do}(\emptyset))}$ is a $\emptyset$-$\emptyset$-separoid, see Definition A.3. By Theorem A.11 we then know that the relation $\perp^{\sigma}_{\mathbf{G}^{\mathrm{acy}}(V|\,\mathrm{do}(J))}$ is a J-$\emptyset$-separoid and thus satisfies all the separoid rules from Theorem 5.11. This already completes the proof.

Since we are interested in proving the global Markov property for (*acyclic*) causal Bayesian networks, we will, for completeness sake, in the following give detailed proofs for all separoid rules again, by assuming, w.l.o.g. that $\mathbf{G} = \mathbf{G}^{\mathrm{acy}}(V|\,\mathrm{do}(J))$ is acyclic. For that recall that for a CDMG $\mathbf{G}$ we say that A *is* σ*-separated from* B *given* C in $\mathbf{G}$, in symbols:

$$A \underset{\mathbf{G}}{\overset{\sigma}{\perp}} B \,|\, C,$$

if every walk from a node in A to a node in $J \cup B$ (sic!) is σ-blocked by C.
For *acyclic* CDMGs, a walk π is *σ-blocked by C* if it either contains a non-collider (i.e. either an end node, fork, left/right chain) in C or a collider not in C.

We abbreviate the ternary relations in the following: $\perp := \overset{\sigma}{\underset{\mathbf{G}}{\perp}} = \overset{\sigma}{\underset{\mathbf{G}^{\mathrm{acy}}(V \mid \mathrm{do}(J))}{\perp}}$.

I.1. Core Separoid Rules for Sigma-Separation

Lemma I.1 (Left Redundancy).

$$A \subseteq C \implies A \perp B \mid C.$$

Proof. If π is a walk from a node v in A to a node w in $J \cup B$ then its first end node is in A, so π is σ-blocked by A. ☐

Lemma I.2 (J-Restricted Right Redundancy).

$$A \perp \emptyset \mid C \cup J \qquad \textit{always holds.}$$

Proof. If π is a walk from a node v in A to a node w in J then its last end node is in $C \cup J$, so π is σ-blocked by A. ☐

Lemma I.3 (Left Decomposition).

$$A \cup D \perp B \mid C \implies D \perp B \mid C.$$

Proof. If π is a walk from a node v in D to a node w in $J \cup B$, then π is a walk from $A \cup D$ to $J \cup B$, which by assumption is σ-blocked by C. ☐

Lemma I.4 (Right Decomposition).

$$A \perp B \cup D \mid C \implies A \perp D \mid C.$$

Proof. If π is a walk from a node v in A to a node w in $J \cup D$, then π is a walk from A to $J \cup B \cup D$, which by assumption is σ-blocked by C. ☐

Lemma I.5 (J-Inverted Right Decomposition).

$$A \perp B \mid C \implies A \perp J \cup B \mid C.$$

Proof. If π is a walk from a node v in A to a node w in $J \cup J \cup B$ then $w \in J \cup B$. If $w \in J \cup B$ then by assumption π is σ-blocked by C. ☐

Lemma I.6 (Left Weak Union).

$$A \cup D \perp B \mid C \implies A \perp B \mid D \cup C.$$

Proof. Lets assume the contrary: $A \not\perp B \,|\, D \cup C$. Then there exists a shortest $(D \cup C)$-σ-open walk π from a node v in A to a node w in $J \cup B$ in $\mathbf{G}$. Then every collider of π is in $D \cup C$ and every non-collider of π is not in $D \cup C$.
If now π does not contain any node from $D \setminus C$ then every collider of π lies in C. This implies that π is C-σ-open, which contradicts the assumption: $A \cup D \perp B \,|\, C$.
So we can assume now that π contains a node in $D \setminus C$. Then consider the shortest sub-walk $\tilde{\pi}$ in π from $w \in J \cup B$ to a node $u \in D \setminus C$. This means that $\tilde{\pi}$ does not contain any collider in $D \setminus C$, so they are all in C. So $\tilde{\pi}$ is C-σ-open walk from $A \cup D$ to $J \cup B$. This contradicts the assumption: $A \cup D \perp B \,|\, C$. □

Lemma I.7 (Right Weak Union)**.**

$$A \perp B \cup D \,|\, C \implies A \perp B \,|\, D \cup C.$$

Proof. Follow the same steps as in Left Weak Union I.6, but this time get a contradiction with: $A \perp B \cup D \,|\, C$. Then again we can assume that π contains a node in $D \setminus C$. Then consider the shortest sub-walk $\tilde{\pi}$ in π from $v \in A$ to a node $u \in D \setminus C$. This means that $\tilde{\pi}$ does not contain any collider in $D \setminus C$, so they are all in C. So $\tilde{\pi}$ is C-σ-open walk from A to $J \cup B \cup D$. This contradicts the assumption: $A \perp B \cup D \,|\, C$. □

Lemma I.8 (Left Contraction)**.**

$$(A \perp B \,|\, D \cup C) \wedge (D \perp B \,|\, C) \implies A \cup D \perp B \,|\, C.$$

Proof. Lets assume the contrary: $A \cup D \not\perp B \,|\, C$. Then there exists a shortest C-σ-open walk π from a node v in $A \cup D$ to a node w in $J \cup B$ in $\mathbf{G}$. So every collider of π lies in C and every non-collider lies not in C and v is the only node of π that lies in $(A \cup D) \setminus C$ (otherwise π could be shortend).
Also v cannot lie in $D \setminus C$ as it would contradict the assumption: $D \perp B \,|\, C$. Thus $v \in A \setminus C$ and π is a walk from A to $J \cup B$ whose colliders all lie in $C \subseteq D \cup C$ and all non-collider outside of $D \cup C$. But this contradicts the other assumption: $A \perp B \,|\, D \cup C$. □

Lemma I.9 (Right Contraction)**.**

$$(A \perp B \,|\, D \cup C) \wedge (A \perp D \,|\, C) \implies A \perp B \cup D \,|\, C.$$

Proof. Lets assume the contrary: $A \not\perp B \cup D \,|\, C$. Then there exists a shortest C-σ-open walk π from a node v in A to a node w in $J \cup B \cup D$ in $\mathbf{G}$. So every collider of π lies in C and every non-collider outside C and w is the only node of π that lies in $(J \cup B \cup D) \setminus C$ (otherwise π could be shortend).
Also w cannot lie in $D \setminus C$ as it would contradict the assumption: $A \perp D \,|\, C$. Thus $w \in (J \cup B) \setminus C$ and π is a walk from A to $J \cup B$ whose colliders all lie in $C \subseteq D \cup C$ and all non-colliders outside of $D \cup C$. But this contradicts the other assumption: $A \perp B \,|\, D \cup C$. □

Lemma I.10 (Right Cross Contraction)**.**

$$(A \perp B \,|\, D \cup C) \wedge (D \perp A \,|\, C) \implies A \perp B \cup D \,|\, C.$$

Proof. Verbatim the same as Right Contraction I.9, only the first contradiction is with: $D \perp A \,|\, C$. ☐

Lemma I.11 (Flipped Left Cross Contraction)**.**

$$(A \perp B \,|\, D \cup C) \wedge (B \perp D \,|\, C) \implies B \perp A \cup D \,|\, C.$$

Proof. Lets assume the contrary: $B \not\perp A \cup D \,|\, C$. Then there exists a shortest C-σ-open walk π from a node v in B to a node w in $J \cup A \cup D$ in $\mathbf{G}$. So every collider of π lies in C and every non-collider outside C and w is the only node of π that lies in $(J \cup A \cup D)\backslash C$ (otherwise π could be shortend).
Also w cannot lie in $(J \cup D)\backslash C$ as it would contradict the assumption: $B \perp D \,|\, C$. Thus $w \in A\backslash C$ and the walk π (in reverse direction) is a walk from A to B whose colliders all lie in $C \subseteq D \cup C$ and all non-colliders outside of $D \cup C$. But this contradicts the other assumption: $A \perp B \,|\, D \cup C$. ☐

I.2. Further Separoid Rules for Sigma-Separation

Lemma I.12 (Left Composition)**.**

$$(A \perp B \,|\, C) \wedge (D \perp B \,|\, C) \implies A \cup D \perp B \,|\, C.$$

Proof. Let π be a walk from a node v in $A \cup D$ to a node w in $J \cup B$. If $v \in A$ then π is σ-blocked by C by assumption: $A \perp B \,|\, C$. If $v \in D$ then π is σ-blocked by C by assumption: $D \perp B \,|\, C$. ☐

Lemma I.13 (Right Composition)**.**

$$(A \perp B \,|\, C) \wedge (A \perp D \,|\, C) \implies A \perp B \cup D \,|\, C.$$

Proof. Let π be a walk from a node v in A to a node w in $J \cup B \cup D$. If $w \in J \cup B$ then π is σ-blocked by C by assumption: $A \perp B \,|\, C$. If $w \in J \cup D$ then π is σ-blocked by C by assumption: $A \perp D \,|\, C$. ☐

Lemma I.14 (Left Intersection)**.** *Assume that $A \cap D = \emptyset$, then:*

$$(A \perp B \,|\, D \cup C) \wedge (D \perp B \,|\, A \cup C) \implies A \cup D \perp B \,|\, C.$$

Proof. Lets assume the contrary: $A \cup D \not\perp B \,|\, C$. Then there exists a shortest C-σ-open walk π from a node v in $A \cup D$ to a node w in $J \cup B$ in $\mathbf{G}$. So every collider of π lies in C and every non-collider outside C and v is the only node of π that lies in $(A \cup D)\backslash C$ (otherwise π could be shortend).
If $v \in A$ then by the disjointness of A and D we have that $v \notin D$. Then π is a walk from A to $J \cup B$ whose colliders lie in $C \subseteq D \cup C$ and all non-colliders outside of $(D\backslash C) \cup C = D \cup C$. This contradicts the assumption: $A \perp B \,|\, D \cup C$.
If $v \in D$ then similarly we get a contradiction: $D \perp B \,|\, A \cup C$ ☐

Lemma I.15 (Right Intersection)**.** *Assume that $B \cap D = \emptyset$, then:*

$$(A \perp B \,|\, D \cup C) \wedge (A \perp D \,|\, B \cup C) \implies A \perp B \cup D \,|\, C.$$

Proof. Lets assume the contrary: $A \not\perp B \cup D \,|\, C$. Then there exists a shortest C-σ-open walk π from a node v in A to a node w in $J \cup B \cup D$ in $\mathbf{G}$. So every collider of π lies in C and every non-collider outside C and w is the only node of π that lies in $(J \cup B \cup D) \backslash C$ (otherwise π could be shortend).
If $w \notin B$ then $w \in J \cup D$. In this case π is a walk from A to $J \cup D$ where every collider lies in $C \subseteq B \cup C$ and all non-colliders are outside of $(B \backslash C) \cup C = B \cup C$. So π is a $(B \cup C)$-σ-open walk from A to $J \cup D$. This contradicts the assumption: $A \perp D \,|\, B \cup C$.
If $w \notin D$ then $w \in J \cup B$. In this case π is a walk from A to $J \cup B$ where every collider lies in $C \subseteq D \cup C$ and all non-colliders are outside of $D \cup C$. So π is a $(D \cup C)$-σ-open walk from A to $J \cup D$. This contradicts the assumption: $A \perp B \,|\, D \cup C$.
Since $B \cap D = \emptyset$ there are no other cases ($B^{\mathsf{c}} \cup D^{\mathsf{c}} = J \cup V$) and we are done. ☐

I.3. Derived Separoid Rules for Sigma-Separation

Lemma I.16 (Restricted Symmetry)**.**

$$(A \perp B \,|\, C) \wedge (B \perp \emptyset \,|\, C) \implies B \perp A \,|\, C.$$

Proof. Follows from Flipped Left Cross Contraction I.11 with $D = \emptyset$. ☐

Lemma I.17 (J-Restricted Symmetry)**.**

$$A \perp B \,|\, C \cup J \implies B \perp A \,|\, C \cup J.$$

Proof. Follows from Restricted Symmetry I.16 and J-Restricted Right Redundancy I.2. ☐

Lemma I.18 (Symmetry)**.** *If $J = \emptyset$ then we have:*

$$A \perp B \,|\, C \implies B \perp A \,|\, C.$$

Proof. Follows directly from J-Restricted Symmetry I.17. ☐

Lemma I.19 (More Redundancies)**.**

$$A \perp B \,|\, C \iff (A \backslash C) \perp (B \backslash C) \,|\, C \iff A \cup C \perp J \cup B \cup C \,|\, C.$$

J. Proofs - Global Markov Property

The proof of the global Markov property follows similar arguments as used in [LDLL90, Ver93, Ric03, FM17, FM18, RERS17], namely chaining the separoid rules together in an inductive way. The main difference here is that we never rely on the Symmetry property but instead use the left and right versions of the separoid rules separately.

Theorem J.1 (Global Markov property for causal Bayesian networks)**.** *Consider a causal Bayesian network* $\mathbf{M}(V|\,\mathrm{do}(J))$ *with CDAG* $\mathbf{G}(V|\,\mathrm{do}(J))$ *and joint Markov kernel* $\mathbf{P}(X_V|\,\mathrm{do}(X_J))$*. Then for all* $A, B, C \subseteq J \cup V$ *(not-necessarily disjoint) we have the implication:*

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B\,|\,C \qquad \Longrightarrow \qquad X_A \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B\,|\,X_C.$$

If one wants to make the implicit dependence on J *more explicit one can equivalently also write:*

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B\,|\,C \qquad \Longrightarrow \qquad X_A \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_J, X_B\,|\,X_C.$$

Proof. We do induction by $\#V$.

0.) Induction start: $V = \emptyset$. This means that $A, B, C \subseteq J$. The assumption:

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B\,|\,C,$$

implies that we must have that $A \subseteq C$. Otherwise a trivial walk from $A \subseteq J$ to $J \cup B$ would be C-open. Since $A, B, C \subseteq J$ we have the factorization:

$$\mathbf{P}(X_A, X_B, X_C|\,\mathrm{do}(X_J)) = \underbrace{\bigotimes_{w\in A} \boldsymbol{\delta}(X_w|X_w)}_{=:\mathbf{Q}(X_A|X_C)} \otimes \underbrace{\bigotimes_{w\in B} \boldsymbol{\delta}(X_w|X_w) \otimes \bigotimes_{w\in C} \boldsymbol{\delta}(X_w|X_w)}_{=\mathbf{P}(X_B,X_C|\,\mathrm{do}(X_J))}.$$

Because $A \subseteq C$ the Markov kernel $\mathbf{Q}(X_A|X_C) := \bigotimes_{w\in A} \boldsymbol{\delta}(X_w|X_w)$ really is a Markov kernel from $\mathcal{X}_C \dashrightarrow \mathcal{X}_A$. This already shows:

$$X_A \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B|X_C.$$

(IND): Induction assumption: The global Markov property holds for all causal Bayesian networks (with input variables) with $\#V < n$ (and arbitrary J).

1.) Now assume: $\#V = n > 0$ and $A \perp^{\sigma}_{\mathbf{G}(V|\,\mathrm{do}(J))} J \cup B\,|\,C$.

Since G is acyclic we can find a topological order $<$ for G where the elements of J are ordered first. Let $v \in V$ be its last element, which is thus childless.
Note that, since $\mathrm{Ch}^{\mathbf{G}}(v) = \emptyset$, the marginalization $\mathbf{G}(V\setminus\{v\}|\,\mathrm{do}(J))$ has no bi-directed edges and thus induces again a causal Bayesian network without latent variables with $\#V^{\setminus\{v\}} = n - 1 < n$.
Furthermore, we have the factorization:

$$\mathbf{P}(X_V|\,\mathrm{do}(X_J)) = \mathbf{P}_v(X_v|\,\mathrm{do}(X_{\mathrm{Pa}^{\mathbf{G}}(v)})) \otimes \underbrace{\bigotimes_{w\in \mathrm{Pred}^{\mathbf{G}}_{<}(v)\setminus J} \mathbf{P}_w(X_w|\,\mathrm{do}(X_{\mathrm{Pa}^{\mathbf{G}}(w)}))}_{\mathbf{P}(X_{\mathrm{Pred}^{\mathbf{G}}_{<}(v)\setminus J}|\,\mathrm{do}(X_J))},$$

where $\mathrm{Pred}^{\mathbf{G}}_{<}(v) = J \dot{\cup} V \setminus \{v\}$ is the set of predesessors of $\{v\}$. This factorization implies that we already have the conditional independence:

$$X_v \underset{\mathbf{P}(X_V \mid \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{\mathrm{Pred}^{\mathbf{G}}_{<}(v)} \mid X_D,$$

where we put $D := \mathrm{Pa}^{\mathbf{G}}(v)$.

In the following we will distinguish between 4 cases:

A.) $v \in A \setminus C$,

B.) $v \in B \setminus C$,

C.) $v \in C$,

D.) $v \notin A \cup J \cup B \cup C$,

Note that $v \in V$, thus $v \notin J$, which shows that the above cover all possible cases. Further note that:

$$A \underset{\mathbf{G}(V \mid \mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \mid C,$$

implies that:

$$A \cap (J \cup B) \subseteq C.$$

Otherwise a trivial walk from A to $J \cup B$ would be C-open. This shows that $A \setminus C$, $(J \cup B) \setminus C$ and C are pairwise disjoint.

Case D.): $v \notin A \cup J \cup B \cup C$. Then we can marginalize out v and use the equivalence:

$$A \underset{\mathbf{G}(V \mid \mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \mid C \quad \Longleftrightarrow \quad A \underset{\mathbf{G}(V \setminus \{v\} \mid \mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \mid C.$$

With $\# V^{\setminus \{v\}} < n$ and induction (IND) we then get:

$$X_A \underset{\mathbf{P}(X_V \mid \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \mid X_C.$$

This shows the claim in case D.

Case A.): $v \in A \setminus C$. Then we can write:

$$\begin{aligned} A &= A' \dot{\cup} (A \cap C) \dot{\cup} \{v\}, \\ B &= B' \dot{\cup} (B \cap C), \end{aligned}$$

with some disjoint $A' \subseteq A \backslash C$ and $B' \subseteq B \backslash C$. We then have the implications:

$$A \underset{\mathbf{G}(V|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B \,|\, C \xRightarrow{\text{Right Decomposition I.4}} A \underset{\mathbf{G}(V|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C$$

$$\xRightarrow{\text{Left Decomposition I.3}} A' \underset{\mathbf{G}(V|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C$$

$$\xRightarrow{\text{marginalization, } v \notin A' \cup J \cup B' \cup C} A' \underset{\mathbf{G}(V \backslash \{v\}|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C$$

$$\xRightarrow{\text{induction (IND)}} X_{A'} \underset{\mathbf{P}(X_V|\operatorname{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_C. \qquad (\#1)$$

On the other hand we have with $D = \mathrm{Pa}^{\mathbf{G}}(v)$:

$$A \underset{\mathbf{G}(V|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B \,|\, C \xRightarrow{\text{Right Decomposition I.4, } B' \subseteq B} A \underset{\mathbf{G}(V|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C$$

$$\xRightarrow{\text{Left Weak Union I.6, } A = A' \,\dot\cup\, (A \cap C) \,\dot\cup\, \{v\}} \{v\} \underset{\mathbf{G}(V|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, A' \,\dot\cup\, C.$$

$$\xRightarrow{(*)\text{, see below}} D \underset{\mathbf{G}(V|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, A' \,\dot\cup\, C$$

$$\xRightarrow{\text{marginalization, } v \notin D \cup J \cup B' \cup A' \cup C} D \underset{\mathbf{G}(V \backslash \{v\}|\operatorname{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, A' \,\dot\cup\, C$$

$$\xRightarrow{\text{induction (IND)}} X_D \underset{\mathbf{P}(X_V|\operatorname{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_{A' \,\dot\cup\, C}$$

$$\xRightarrow{A' \,\dot\cup\, C} X_D \underset{\mathbf{P}(X_V|\operatorname{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_{A'}, X_C. \qquad (\#2)$$

(*) holds since every $(A' \,\dot\cup\, C)$-open walk $w \ast\!\!\rightarrow\!\!\ast \cdots$ from a $w \in D = \mathrm{Pa}^{\mathbf{G}}(v)$ to $J \cup B'$ extends to an $(A' \,\dot\cup\, C)$-open walk from v to $J \cup B'$ via $v \leftarrow w \ast\!\!\rightarrow\!\!\ast \cdots$, as w stays a non-collider in the extended walk (not in $A' \,\dot\cup\, C$) and $v \notin A' \,\dot\cup\, C$.

As discussed above we also already have the conditional independence:

$$X_v \underset{\mathbf{P}(X_V|\operatorname{do}(X_J))}{\perp\!\!\!\perp} X_{\mathrm{Pred}^{\mathbf{G}}_{<}(v)} \,|\, X_D.$$

With this and $A' \dot\cup B' \dot\cup C \subseteq \mathrm{Pred}^{\mathbf{G}}_{<}(v)$ we get the implications:

$$
\begin{aligned}
& & X_v \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{\mathrm{Pred}^{\mathbf{G}}_{<}(v)} \,|\, X_D \\
&\xRightarrow{\text{Right Decomposition E.6}} & X_v \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{A'}, X_{B'}, X_C \,|\, X_D \\
&\xRightarrow{\text{Right Weak Union E.11}} & X_v \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_{A'}, X_C, X_D \\
&\xRightarrow{\text{Left Contraction E.12, (\#2)}} & X_v, X_D \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_{A'}, X_C \\
&\xRightarrow{\text{Left Decomposition E.5}} & X_v \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_{A'}, X_C \\
&\xRightarrow{\text{Left Contraction E.12, (\#1)}} & X_{A'}, X_v \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_C \\
&\xRightarrow{X_J\text{-Inverted Right Decomposition E.7}} & X_{A'}, X_v \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_J, X_{B'}, X_C \,|\, X_C \\
&\xRightarrow{\text{Right Decompositon E.6, } B \subseteq B' \dot\cup C} & X_{A'}, X_v \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C. \qquad (\#3)
\end{aligned}
$$

By Left Redundancy E.3 we have:

$$X_{A'}, X_v, X_C \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_{A'}, X_v, X_C.$$

With this we get the implications:

$$
\begin{aligned}
& & X_{A'}, X_v, X_C \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_{A'}, X_v, X_C \\
&\xRightarrow{\text{Left Contraction E.12, (\#3)}} & X_{A'}, X_v, X_{A'}, X_v, X_C \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C \\
&\xRightarrow{\text{Left Decomposition E.5, } A \subseteq A' \dot\cup \{v\} \dot\cup C} & X_A \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C.
\end{aligned}
$$

This shows the claim in case A.

Case B.): $v \in B \setminus C$. Then we can write:

$$
\begin{aligned}
A &= A' \dot\cup (A \cap C), \\
B &= B' \dot\cup (B \cap C) \dot\cup \{v\},
\end{aligned}
$$

with some disjoint $A' \subseteq A \setminus C$ and $B' \subseteq B \setminus C$.

We then have the implications:

$$
\begin{aligned}
A \underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \,|\, C &\xrightarrow{\text{Left Decomposition I.3}} A' \underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \,|\, C \\
&\xrightarrow{\text{Right Decomposition I.4}} A' \underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C \\
&\xrightarrow{\text{marginalization, } v \notin A' \cup J \cup B' \cup C} A' \underset{\mathbf{G}(V \setminus \{v\}\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C \\
&\xrightarrow{\text{induction (IND)}} X_{A'} \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_C. \qquad (\#1')
\end{aligned}
$$

Again with $D = \mathrm{Pa}^{\mathbf{G}}(v)$ we get:

$$
\begin{aligned}
A \underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \,|\, C &\xrightarrow{\text{Left Decomposition I.3}} A' \underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \,|\, C \\
&\xrightarrow{\text{Right Decomposition I.4}} A' \underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \cup \{v\} \,|\, C \\
&\xrightarrow{\text{Right Weak Union I.7}} A' \underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup \{v\} \,|\, B' \dot\cup C \\
&\xrightarrow{(\bullet)\text{, see below}} A' \underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup D \,|\, B' \dot\cup C \\
&\xrightarrow{\text{induction (IND)}} X_{A'} \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_D \,|\, X_{B' \dot\cup C} \\
&\xrightarrow{B' \dot\cup C} X_{A'} \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_D \,|\, X_{B'}, X_C. \qquad (\#2')
\end{aligned}
$$

$(\bullet)$ holds since every $(B' \dot\cup C)$-open walk $\cdots \ast\!\!-\!\!\ast\, w$ from A' to a $w \in J \cup D$ extends to a $(B' \dot\cup C)$-open walk from A' to $J \cup \{v\}$, either because $w \in J$ or via $\cdots \ast\!\!-\!\!\ast\, w \rightarrowtail v$ if $w \in D = \mathrm{Pa}^{\mathbf{G}}(v)$. Note again that w stays a non-collider in the extended walk (outside of $B' \dot\cup C$) and $v \notin B' \dot\cup C$.

As before we will use the following conditional independence:

$$X_v \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{\mathrm{Pred}^{\mathbf{G}}_{\le}(v)} \,|\, X_D.$$

With this and $A' \cup J \cup B' \cup C \subseteq \mathrm{Pred}^{\mathbf{G}}_{<}(v)$ we get the implications:

$$
\begin{aligned}
& X_v \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{\mathrm{Pred}^{\mathbf{G}}_{<}(v)} \,|\, X_D \\
\xRightarrow{\text{Right Decomposition E.6}} \quad & X_v \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{A'}, X_{B'}, X_C \,|\, X_D \\
\xRightarrow{\text{Right Weak Union E.11}} \quad & X_v \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{A'} \,|\, X_{B'}, X_C, X_D \\
\xRightarrow{\text{Flipped Left Cross Contraction E.15, (\#2')}} \quad & X_{A'} \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_D, X_v \,|\, X_{B'}, X_C \\
\xRightarrow{\text{Right Decomposition E.6}} \quad & X_{A'} \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_v \,|\, X_{B'}, X_C \\
\xRightarrow{\text{Right Contraction E.13, (\#1')}} \quad & X_{A'} \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'}, X_v \,|\, X_C \\
\xRightarrow{X_J\text{-Inverted Right Decomposition E.7}} \quad & X_{A'} \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_J, X_{B'}, X_v, X_C \,|\, X_C \\
\xRightarrow{\text{Right Decomposition E.6, } B \subseteq B' \,\dot\cup\, \{v\} \,\dot\cup\, C} \quad & X_{A'} \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C. \qquad (\#3')
\end{aligned}
$$

By Left Redundancy E.3 we have:

$$X_{A'}, X_C \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_{A'}, X_C.$$

With this we get the implications:

$$
\begin{aligned}
& X_{A'}, X_C \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_{A'}, X_C \\
\xRightarrow{\text{Left Contraction E.12, (\#3')}} \quad & X_{A'}, X_{A'}, X_C \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C. \\
\xRightarrow{\text{Left Decomposition E.5, } A \subseteq A' \,\dot\cup\, C} \quad & X_A \underset{\mathbf{P}(X_V \,|\, \mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C.
\end{aligned}
$$

This shows the claim in case B.

Case C.): $v \in C$. Then we can write:

$$
\begin{aligned}
A &= A' \,\dot\cup\, (A \cap C), \\
B &= B' \,\dot\cup\, (B \cap C), \\
C &= C' \,\dot\cup\, \{v\},
\end{aligned}
$$

with some pairwise disjoint $A' \subseteq A \backslash C$, $B' \subseteq B \backslash C$ and $C' \subseteq C$.

We then get the implications.

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \,|\, C \xRightarrow{\text{Left Decomposition I.3}} A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B \,|\, C$$

$$\xRightarrow{\text{Right Decomposition I.4}} A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C$$

$$\xRightarrow{C = C' \,\dot\cup\, \{v\}} A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C' \,\dot\cup\, \{v\}$$

We now claim that:

$$A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C' \,\dot\cup\, \{v\}$$

implies that one of the following statements holds:

$$A' \,\dot\cup\, \{v\} \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C' \qquad \vee \qquad A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup (B' \,\dot\cup\, \{v\}) \,|\, C'.$$

Assume the contrary:

$$A' \,\dot\cup\, \{v\} \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\not\perp}} J \cup B' \,|\, C' \qquad \wedge \qquad A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\not\perp}} J \cup (B' \,\dot\cup\, \{v\}) \,|\, C'.$$

So there exist shortest C'-open walks π_1 and π_2 in $\mathbf{G}(V|\,\mathrm{do}(J))$:

$$\pi_1 : \quad A' \cup \{v\} \ni u_0 \ast\!\!-\!\!\ast \cdots \ast\!\!-\!\!\ast u_k \in J \cup B',$$

and:

$$\pi_2 : \quad A' \ni w_0 \ast\!\!-\!\!\ast \cdots \ast\!\!-\!\!\ast w_m \in J \cup (B' \,\dot\cup\, \{v\}).$$

So all colliders of π_1 and π_2 are in C' and all non-colliders outside of C'. Since we consider shortest walks and $v \notin C'$ at most an end node of π_1 and π_2 could be equal to v. Otherwise one could shorten the walk.
Then note that $v \notin A'$ and $v \notin J \cup B'$, thus: $u_k \neq v$ and $w_0 \neq v$.
If now π_i does not contain v as an (end) node, then π_i would be $(C' \,\dot\cup\, \{v\})$-open, which is a contradiction to the assumption:

$$A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C' \,\dot\cup\, \{v\}.$$

So we can assume that the other end nodes equal v, i.e.: $u_0 = v$ and $w_m = v$.
Furthermore, both π_1 and π_2 are non-trivial walks, since $u_0 \neq u_k$ and $w_0 \neq w_m$. Since v is childless and $k, m \geqslant 1$ we have that the π_i are of the forms:

$$\pi_1 : \quad v \leftarrow\!\!\prec u_1 \ast\!\!-\!\!\ast \cdots \ast\!\!-\!\!\ast u_k,$$

and:

$$\pi_2 : \quad w_0 \ast\!\!-\!\!\ast \cdots \ast\!\!-\!\!\ast w_{m-1} \succ\!\!\rightarrow v,$$

with $u_1, w_{m-1} \in D = \mathrm{Pa}^{\mathbf{G}}(v)$. Then the following walk:

$$A' \ni w_0 \ast\!\!-\!\!\ast \cdots \ast\!\!-\!\!\ast w_{m-1} \rightarrowtail v \leftarrowtail u_1 \ast\!\!-\!\!\ast \cdots \ast\!\!-\!\!\ast u_k \in J \cup B',$$

is a $(C' \dot\cup \{v\})$-open walk from A' to $J \cup B'$, in contradiction to:

$$A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C' \dot\cup \{v\}.$$

So the claim:

$$A' \dot\cup \{v\} \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup B' \,|\, C' \qquad \vee \qquad A' \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} J \cup (B' \dot\cup \{v\}) \,|\, C',$$

must be true. So we reduced case C. to case A. or case B., which then imply:

$$X_{A'}, X_v \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_{C'} \qquad \vee \qquad X_{A'} \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'}, X_v \,|\, X_{C'}.$$

If we apply Left Weak Union E.8 to the left and Right Weak Union E.11 to the right we get:

$$X_{A'} \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_{B'} \,|\, X_{C'}, X_v,$$

which - as before - implies:

$$X_A \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C.$$

This shows the claim in case C. □

Remark J.2. 1. *Note that we only needed to use Left Weak Union E.8 for nodes $\{v\} \cup A'$, which were in V. So no assumptions about standard measurable spaces for $\mathcal{X}_j$, $j \in J$, were needed.*

2. *All results would also hold under the weaker assumptions where all measurable maps are replaced by universally measurable maps, all countably generated measurable spaces by universally (countably) generated ones and all standard measurable spaces by universal measurable spaces. Analytic versions would also be possible.*

K. Causal Models

K.1. Causal Bayesian Networks - More General

We deviate here from the Section 6.1 of the main paper a bit and give here a slightly more general definition of causal Bayesian networks that allow, besides input variables, also for latent variables. The reason is that we want to demonstrate how the correspondence of operations on graphs and operations on causal Bayesian networks, like interventions, marginalizations, etc., immediately implies the global Markov property, Theorem 6.3 and Appendix J, for such more general cases.

Definition K.1 (Causal Bayesian network)**.** *A* causal Bayesian network (CBN) $\mathbf{M}$ *consists of:*

1. *finite pairwise disjoint index sets: J, V, U,*

   corresponding to (non-stochastic) input variables X_j, $j \in J$, (stochastic) observed output variables X_v, $v \in V$, and stochastic unobserved/latent output variables X_u, $u \in U$.

2. *a conditional directed acyclic graph (CDAG):* $\mathbf{G} = \mathbf{G}(U \dot\cup V | \operatorname{do}(J)) = (J, U \dot\cup V, E, \emptyset)$,

   i.e. an acyclic CDMG without bi-directed edges: $L = \emptyset$,

3. *a measurable space $\mathcal{X}_v$ for every $v \in J \dot\cup U \dot\cup V$, where $\mathcal{X}_v$ is standard*[7] *if $v \in U \dot\cup V$,*

4. *a Markov kernel, suggestively written as:* $\mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right)$*:*

$$\begin{array}{rcl} \mathcal{X}_{\mathrm{Pa}^{\mathbf{G}}(v)} & \dashrightarrow & \mathcal{X}_v, \\ (A, x_{\mathrm{Pa}^{\mathbf{G}}(v)}) & \mapsto & \mathbf{P}_v\left(X_v \in A \middle| \operatorname{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)} = x_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right), \end{array}$$

   for every $v \in U \dot\cup V$, where we write for $D \subseteq J \dot\cup U \dot\cup V$:

$$\begin{aligned} \mathcal{X}_D &:= \prod_{v \in D} \mathcal{X}_v, & \mathcal{X}_\emptyset &:= \ast = \{\ast\}, \\ X_D &:= (X_v)_{v \in D}, & X_\emptyset &:= \ast, \\ x_D &:= (x_v)_{v \in D}, & x_\emptyset &:= \ast. \end{aligned}$$

By abuse of notation, we denote the causal Bayesian network as:

$$\mathbf{M}(U, V | \operatorname{do}(J)) = \left(\mathbf{G}(U \dot\cup V | \operatorname{do}(J)), \left(\mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right)\right)_{v \in U \dot\cup V}\right)$$

Definition K.2. *Let* $\mathbf{M} = \mathbf{M}(U, V | \operatorname{do}(J))$ *be a CBN.*

1. *We write the* marginal CDMG *of* $\mathbf{G} = \mathbf{G}(U \dot\cup V | \operatorname{do}(J))$*, where the latent nodes U are marginalized out, as* $\mathbf{G}(V | \operatorname{do}(J)) = (J, V, E, L)$*, which now might have bi-directed edges.*

2. *The CBN* $\mathbf{M}$ *comes with the* joint Markov kernel*:*

$$\mathbf{P}(X_U, X_V | \operatorname{do}(X_J)) : \mathcal{X}_J \dashrightarrow \mathcal{X}_{U \dot\cup V},$$

   given by:

$$\mathbf{P}(X_U, X_V | \operatorname{do}(X_J)) := \overset{>}{\bigotimes_{v \in U \dot\cup V}} \mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right),$$

[7] We could also work under the weaker assumption of universal measurable spaces if we would replace all mentionings of measurability with the weaker notion of universal measurability.

where the product $\otimes$ is taken in reverse order of a fixed topological order $<$, i.e. parents proceed children variables. Note that by remark 2.8 about associativity and (restricted) commutativity of the product the joint Markov kernels does actually not depend on the topological order.

3. *The marginal Markov kernel* $\mathbf{P}\left(X_V \middle| \operatorname{do}\left(X_J\right)\right)$ *is called the* observational Markov kernel *of* $\mathbf{M}$.

K.2. Operations on Causal Bayesian Networks

Definition K.3 (Hard interventions on CBNs)**.** *Let* $\mathbf{M} = \mathbf{M}(U, V | \operatorname{do}(J))$ *be a CBN and* $W \subseteq J \cup V$. *Then the* hard interventional causal Bayesian network intervened on *W is given by:* $\mathbf{M}(U, V \setminus W | \operatorname{do}(J \cup W)) :=$

$$\left(\mathbf{G}(U \,\dot\cup\, V \setminus W | \operatorname{do}(J \cup W)), \left(\mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right)\right)_{v \in U \,\dot\cup\, V \setminus W}\right).$$

It then comes with joint Markov kernel:

$$\mathbf{P}\left(X_U, X_{V\setminus W} \middle| \operatorname{do}\left(X_{J \cup W}\right)\right) = \bigotimes_{v \in U \,\dot\cup\, V \setminus W}^{>} \mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right).$$

Definition K.4 (Marginalizing CBNs)**.** *Let* $\mathbf{M} = \mathbf{M}(U, V | \operatorname{do}(J))$ *be a CBN and* $W \subseteq V$. *Then the* marginal causal Bayesian network*, where W is marginalized out, is given by:* $\mathbf{M}(U \,\dot\cup\, W, V \setminus W | \operatorname{do}(J)) :=$

$$\left(\mathbf{G}(U \,\dot\cup\, V | \operatorname{do}(J)), \left(\mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right)\right)_{v \in U \,\dot\cup\, V}\right),$$

with the only difference that nodes from W are moved from V to U, i.e. are declared latent. It then comes with the (same) joint Markov kernel:

$$\mathbf{P}\left(X_{U \,\dot\cup\, W}, X_{V\setminus W} \middle| \operatorname{do}\left(X_J\right)\right) = \bigotimes_{v \in U \,\dot\cup\, V}^{>} \mathbf{P}_v\left(X_v \middle| \operatorname{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right),$$

but with the marginalized CDMG as marginal CDMG: $\mathbf{G}(V \setminus W | \operatorname{do}(J))$*, and the marginalized Markov kernel as marginal Markov kernel:* $\mathbf{P}(X_{V\setminus W} | \operatorname{do}(X_J))$.

Remark K.5 (Modelling soft interventions)**.** *If we wanted to model* soft interventions *onto some variables given by index set W in a CBN* $\mathbf{M}(U, V | \operatorname{do}(J))$ *with CDMG* $\mathbf{G} = \mathbf{G}(U \,\dot\cup\, V | \operatorname{do}(J))$ *we would introduce soft intervention variables X_{I_w} for each $w \in W$ and then model the effect of the variable X_{I_w} onto X_w together with its "natural" causes $X_{\mathrm{Pa}^{\mathbf{G}}(w)}$. This would result in specifying a Markov kernel* $\mathbf{P}(X_w | \operatorname{do}(X_{\mathrm{Pa}^{\mathbf{G}}(w)}, X_{I_w}))$ *for each $w \in W$ and leave the Markov kernels* $\mathbf{P}(X_v | \operatorname{do}(X_{\mathrm{Pa}^{\mathbf{G}}(v)}))$ *for all other variables the same. This then would constitute a new CBN* $\mathbf{M}(U, V | \operatorname{do}(J, I_W))$ *with CDMG* $\mathbf{G}(U \,\dot\cup\, V | \operatorname{do}(J, I_W))$.

Example K.6 (Modelling hard interventions as soft interventions)**.** *If we now wanted to model hard interventions in a CBN* $\mathbf{M}(U, V|\,\mathrm{do}(J))$ *on variables corresponding to* $W \in V$ *as soft interventions we would define for* $w \in W$*:*

1. $\mathcal{X}_{I_w} := \mathcal{X}_w \,\dot{\cup}\, \{\star\}$ *with a new symbol* $\star$ *that represents "no intervention",*
2. *Markov kernels:* $\mathbf{P}(X_w \in A|\,\mathrm{do}(X_{\mathrm{Pa}^{\mathbf{G}}(w)} = x_{\mathrm{Pa}^{\mathbf{G}}(w)}, X_{I_w} = x_{I_w})) :=$

$$\begin{cases} \mathbf{P}(X_w \in A|\,\mathrm{do}(X_{\mathrm{Pa}^{\mathbf{G}}(w)} = x_{\mathrm{Pa}^{\mathbf{G}}(w)})) & \text{if} \quad x_{I_w} = \star, \\ \boldsymbol{\delta}(X_w \in A|X_w = x_{I_w}) = \mathbb{1}_A(x_{I_w}) & \text{if} \quad x_{I_w} \in \mathcal{X}_w. \end{cases}$$

This then defines a CBN $\mathbf{M}(U, V|\,\mathrm{do}(J, I_W))$ *with CDMG* $\mathbf{G}(U \,\dot{\cup}\, V|\,\mathrm{do}(J, I_W))$*.*

K.3. Global Markov Property for More General Causal Models

Here we now extend the global Markov property from causal Baysian networks with input nodes but without latent variables, see Theorem 6.3 and Appendix J, to causal Bayesian networks with both, input variables and latent variables. We also shortly discuss and indicate how one could even go beyond and use the results and techniques from this paper to prove a global Markov property based on transitional conditional independence for, say, structural causal models with input variables, latent variables, variables introducing selection bias and that also allows for (negative) feedback cycles.

Theorem K.7 (Global Markov property for causal Bayesian networks)**.** *Consider a causal Bayesian network* $\mathbf{M}(U, V|\,\mathrm{do}(J))$ *with marginal CADMG:* $\mathbf{G}(V|\,\mathrm{do}(J))$*, and observational Markov kernel:* $\mathbf{P}(X_V|\,\mathrm{do}(X_J))$*. Then for all* $A, B, C \subseteq J \,\dot{\cup}\, V$ *(not-necessarily disjoint) we have the implication:*

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B\,|\,C \qquad \Longrightarrow \qquad X_A \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B\,|\,X_C.$$

Recall that we have - per definition - an implicit dependence on J*,* X_J*, resp., in the second argument on each side.*

Proof. Because σ-separation is preserved under marginalization we have the equivalence:

$$A \underset{\mathbf{G}(V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B\,|\,C \quad \Longleftrightarrow \quad A \underset{\mathbf{G}(U \,\dot{\cup}\, V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B\,|\,C.$$

Note that the CBN $\mathbf{M}(U, V|\,\mathrm{do}(J))$ can mathematically also be considered as the CBN $\mathbf{M}(\emptyset, U \,\dot{\cup}\, V|\,\mathrm{do}(J))$, where all latent variables are treated like observed ones. Then the global Markov property for $\mathbf{M}(\emptyset, U \,\dot{\cup}\, V|\,\mathrm{do}(J))$, proven in Appendix J, implies the mentioned transitional conditional independence:

$$A \underset{\mathbf{G}(U \,\dot{\cup}\, V|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B\,|\,C \qquad \Longrightarrow \qquad X_A \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B\,|\,X_C.$$

For this note that the CBN $\mathbf{M}(\emptyset, U \,\dot{\cup}\, V|\,\mathrm{do}(J))$ has the same graph $\mathbf{G}(U \,\dot{\cup}\, V|\,\mathrm{do}(J))$ as the original CBN $\mathbf{M}(U, V|\,\mathrm{do}(J))$. □

Remark K.8. *The above proof technique would work similarly if we would include another set of nodes S and variables X_S, representing* selection bias. *Then we would get a graph that we could denote with $\mathbf{G}(V|S,\mathrm{do}(J))$, either representing S directly or allowing for undirected edges. An extension of σ-separation would then be adjusted, see e.g. [FM18, FM20, BFPM21], such that:*

$$A \underset{\mathbf{G}(V|S,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \quad \iff \quad A \underset{\mathbf{G}(U \,\dot\cup\, V \,\dot\cup\, S|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, S \cup C.$$

Then we could just directly apply the proven global Markov property to the right hand side and we would be done:

$$A \underset{\mathbf{G}(U \,\dot\cup\, V \,\dot\cup\, S|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, S \cup C \quad \Longrightarrow \quad X_A \underset{\mathbf{P}(X_V, X_S|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_S, X_C.$$

One could even go further and allow for input, latent, selection variables and cycles in a notion of structural causal models *(SCM), e.g. see again [FM18, FM20, BFPM21], since the compatibility conditions for such cyclic SCM lead for every interventional setting to an acyclic resolution and we would get:*

$$A \underset{\mathbf{G}(V|S,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, C \quad \iff \quad A \underset{\mathbf{G}^{\mathrm{acy}}(U \,\dot\cup\, V \,\dot\cup\, S|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, S \cup C.$$

Again the proven global Markov property would apply to the right hand side:

$$A \underset{\mathbf{G}^{\mathrm{acy}}(U \,\dot\cup\, V \,\dot\cup\, S|\,\mathrm{do}(J))}{\overset{\sigma}{\perp}} B \,|\, S \cup C \quad \Longrightarrow \quad X_A \underset{\mathbf{P}(X_V, X_S|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_J, X_B \,|\, X_S, X_C,$$

showing the global Markov property based on transitional conditional independence even for such general cases.

K.4. Causal/Do-Calculus

In this section we will prove the 3 main rules of *do-calculus*, see [Pea09, Pea93a, Pea93b, FM20]. for causal Bayesian networks (CBN) that allow for input variables and latent variables. We will demonstrate how the use of transitional conditional independence circumvents the usual measure-theoretic problems of matching functions/densities on null-sets, etc. Such questions were investigated e.g. also in [GR01]. First recall the following.

Remark K.9 (Marginal conditional interventional Markov kernels)**.** *Consider a causal Bayesian network:*

$$\mathbf{M}(U, V|\,\mathrm{do}(J)) = \left(\mathbf{G} = \mathbf{G}(U \,\dot\cup\, V|\,\mathrm{do}(J)), \left(\mathbf{P}_v\left(X_v \middle|\, \mathrm{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right)\right)_{v \in U \,\dot\cup\, V}\right),$$

$<$ a fixed topological order for $\mathbf{G}$. For any disjoint subsets $D \subseteq J \cup V$, $W \subseteq V$ we then have the hard and soft interventional Markov kernel $\mathbf{P}\left(X_U, X_{V \setminus D} \middle|\, \mathrm{do}\left(X_{D \cup J}, X_{I_W}\right)\right) =$

$$\bigotimes^{>}_{v \in U \,\dot\cup\, V \setminus (D \cup W)} \mathbf{P}_v\left(X_v \middle|\, \mathrm{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right) \otimes^{>} \bigotimes^{>}_{v \in W} \mathbf{P}_v\left(X_v \middle|\, \mathrm{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}, X_{I_v}\right)\right),$$

where we reorder all factors in reverse order of $<$, see Section K.2. If, furthermore, $A, B \subseteq J \cup V$ then we can extend, marginalize and condition to get the marginal conditional interventional Markov kernels, see Sections 2.2, 2.7:

$$\mathbf{P}\left(X_A \middle| X_B, \mathrm{do}\left(X_{D \cup J}, X_{I_W}\right)\right),$$

which are unique up to $\mathbf{P}\left(X_B \middle| \mathrm{do}\left(X_{D \cup J}, X_{I_W}\right)\right)$*-null set. These are related to each other by the graphical structure of the CDMG* $\mathbf{G}(V|\,\mathrm{do}(J, I_W))$ *as we will see next.*

We now prove the do-calculus rules, see [Pea09], for causal Bayesian networks, which were most of the time proposed and proven only for discrete spaces, but postulated for arbitrary ones. Even when densities were used the measure theoretical difficulties coming from dealing with the null-sets were typically ignored. Here we want to resolve these problems by making use of the our strong version of the global Markov property for causal Bayesian networks, Thereom K.7, in combination with transitional conditional independence, Definition 3.1. It turns out the occuring Markov kernels will automatically aline all the null sets.

Corollary K.10 (Do-calculus)**.** *Consider a causal Bayesian network:*

$$\mathbf{M}(U, V|\,\mathrm{do}(J)) = \left(\mathbf{G} = \mathbf{G}(U \,\dot{\cup}\, V|\,\mathrm{do}(J)), \left(\mathbf{P}_v\left(X_v \middle| \mathrm{do}\left(X_{\mathrm{Pa}^{\mathbf{G}}(v)}\right)\right)\right)_{v \in U \,\dot{\cup}\, V}\right),$$

Let $A, B, C \subseteq V$ and $D \subseteq J \cup V$. For simplicity, we assume that A, B, C, D are pairwise disjoint. Then we have the following rules relating marginal conditional interventional Markov kernels to each other. Note that for a more suggestive notation we reorder the conditioning and do*-parts in the Markov kernels arbitrarily.*

1. *Insertion/deletion of observation: If we have:*

$$A \underset{\mathbf{G}(V \setminus D|\,\mathrm{do}(D \cup J))}{\overset{\sigma}{\perp}} B \,|\, C \,\dot{\cup}\, D,$$

then there exists a Markov kernel:

$$\mathbf{P}\left(X_A | \cancel{X_B}, X_C, \mathrm{do}(X_D, \cancel{X_{J \setminus D}})\right) : \mathcal{X}_C \times \mathcal{X}_D \dashrightarrow \mathcal{X}_A$$

that is a version of:

$$\mathbf{P}\left(X_A | X_{\tilde{B}}, X_C, \mathrm{do}(X_{D \cup J})\right),$$

for every subset $\tilde{B} \subseteq B$ simultaneously. Such a Markov kernel is unique up to $\mathbf{P}\left(X_C |\,\mathrm{do}(X_{D \cup J})\right)$*-null set. In particular and in short we could write:*

$$\mathbf{P}\left(X_A | X_B, X_C, \mathrm{do}(X_D)\right) = \mathbf{P}\left(X_A | X_C, \mathrm{do}(X_D)\right).$$

2. *Action/observation exchange: If we have:*

$$A \underset{\mathbf{G}(V\setminus D|\,\mathrm{do}(D\cup J, I_B))}{\overset{\sigma}{\perp}} I_B \,|\, B \dot\cup C \dot\cup D,$$

then there exists a Markov kernel:

$$\mathbf{P}\left(X_A|\cancel{\mathrm{do}}(X_B), X_C, \mathrm{do}(X_D, \cancel{X_{J\setminus D}})\right) : \mathcal{X}_B \times \mathcal{X}_C \times \mathcal{X}_D \dashrightarrow \mathcal{X}_A$$

that is a version of:

$$\mathbf{P}\left(X_A|X_{B_1}, \mathrm{do}(X_{B_2}), X_C, \mathrm{do}(X_{D\cup J})\right),$$

for every disjoint decomposition: $B = B_1 \dot\cup B_2$, simultaneously. Such a Markov kernel is unique up to $\mathbf{P}\left(X_{B\dot\cup C}|\,\mathrm{do}(X_{D\cup J}, X_{I_B})\right)$-null set. In particular and in short we could write:

$$\mathbf{P}\left(X_A|\,\mathrm{do}(X_B), X_C, \mathrm{do}(X_D)\right) = \mathbf{P}\left(X_A|X_B, X_C, \mathrm{do}(X_D)\right).$$

3. *Insertion/deletion of action: If we have:*

$$A \underset{\mathbf{G}(V\setminus D|\,\mathrm{do}(D\cup J, I_B))}{\overset{\sigma}{\perp}} I_B \,|\, C \dot\cup D,$$

then there exists a Markov kernel:

$$\mathbf{P}\left(X_A|\cancel{\mathrm{do}(X_B)}, X_C, \mathrm{do}(X_D, \cancel{X_{J\setminus D}})\right) : \mathcal{X}_C \times \mathcal{X}_D \dashrightarrow \mathcal{X}_A$$

that is a version of:

$$\mathbf{P}\left(X_A|\,\mathrm{do}(X_{\tilde{B}}), X_C, \mathrm{do}(X_{D\cup J})\right)$$

for every subset $\tilde{B} \subseteq B$ simultaneously. Such a Markov kernel is unique up to $\mathbf{P}\left(X_C|\,\mathrm{do}(X_{D\cup J}, X_{I_B})\right)$-null set. In particular and in short we could write:

$$\mathbf{P}\left(X_A|\,\mathrm{do}(X_B), X_C, \mathrm{do}(X_D)\right) = \mathbf{P}\left(X_A|X_C, \mathrm{do}(X_D)\right).$$

Proof. We make use of the global Markov property (GMP), theorem K.7.

Point 1.) The assumption:

$$A \underset{\mathbf{G}(V\setminus D|\,\mathrm{do}(D\cup J))}{\overset{\sigma}{\perp}} B \,|\, C \dot\cup D,$$

implies the following transitional conditional independence by the global Markov property K.7:

$$X_A \underset{\mathbf{P}(X_V|\,\mathrm{do}(X_{D\cup J}))}{\perp\!\!\!\perp} X_B \,|\, X_C, X_D.$$

So we get the following factorization:

$$\mathbf{P}\left(X_A, X_B, X_C, X_D \middle| \operatorname{do}(X_{D\cup J})\right) = \mathbf{Q}(X_A|X_C, X_D) \otimes \mathbf{P}\left(X_B, X_C, X_D \middle| \operatorname{do}(X_{D\cup J})\right),$$

for some Markov kernel $\mathbf{Q}(X_A|X_C, X_D)$. If we marginalize out the deterministic variables X_D we get:

$$\mathbf{P}\left(X_A, X_B, X_C \middle| \operatorname{do}(X_{D\cup J})\right) = \mathbf{Q}(X_A|X_C, X_D) \otimes \mathbf{P}\left(X_B, X_C \middle| \operatorname{do}(X_{D\cup J})\right).$$

Clearly $\mathbf{Q}(X_A|X_C, X_D)$ serves here as a regular version of the conditional Markov kernel:

$$\mathbf{P}\left(X_A | X_B, X_C, \operatorname{do}(X_{D\cup J})\right).$$

If we marginalize out X_{B_1} for any decomposition $B = B_1 \dot\cup B_2$ in the above factorization we also get:

$$\mathbf{P}\left(X_A, X_C, X_{B_2} \middle| \operatorname{do}(X_{D\cup J})\right) = \mathbf{Q}(X_A|X_C, X_D) \otimes \mathbf{P}\left(X_C, X_{B_2} \middle| \operatorname{do}(X_{D\cup J})\right),$$

showing that $\mathbf{Q}(X_A|X_C, X_D)$ is also a version of:

$$\mathbf{P}\left(X_A | X_{B_2}, X_C, \operatorname{do}(X_{D\cup J})\right).$$

In particular, this holds for $B_2 = \varnothing$. This shows point 1.

Point 2.) The assumption:

$$A \underset{\mathbf{G}(V\setminus D|\operatorname{do}(D\cup J, I_B))}{\overset{\sigma}{\perp}} I_B \mid B \dot\cup C \dot\cup D,$$

implies the following transitional conditional independence by the global Markov property K.7:

$$X_A \underset{\mathbf{P}(X_V|\operatorname{do}(X_{D\cup J}, X_{I_B}))}{\perp\!\!\!\perp} X_{I_B} \mid X_B, X_C, X_D.$$

As before, we get the following factorization:

$$\mathbf{P}\left(X_A, X_B, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{I_B})\right) = \mathbf{Q}(X_A|X_B, X_C, X_D) \otimes \mathbf{P}\left(X_B, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{I_B})\right),$$

for some Markov kernel $\mathbf{Q}(X_A|X_B, X_C, X_D)$. This then serves as a version of a regular conditional Markov kernel for:

$$\mathbf{P}\left(X_A | X_B, X_C, \operatorname{do}(X_{D\cup J}, X_{I_B})\right),$$

and which is independent of X_{I_B}.
We can now look at the different input values for any decomposition: $B = B_1 \dot\cup B_2$. For this we put: $X_{I_{B_1}} = x_{B_1} \in \mathcal{X}_{B_1}$ and $X_{I_{B_2}} = \star = (\star)_{v\in B_2}$. This implies:

$$\begin{aligned}
&\mathbf{P}\left(X_A, X_B, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1} = x_{B_1})\right) \\
&= \mathbf{P}\left(X_A, X_B, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{I_{B_1}} = x_{B_1}, X_{I_{B_2}} = \star)\right) \\
&= \mathbf{Q}(X_A|X_B, X_C, X_D) \otimes \mathbf{P}\left(X_B, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{I_{B_1}} = x_{B_1}, X_{I_{B_2}} = \star)\right), \\
&= \mathbf{Q}(X_A|X_B, X_C, X_D) \otimes \mathbf{P}\left(X_B, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1} = x_{B_1})\right).
\end{aligned}$$

So we get:

$$\mathbf{P}\left(X_A, X_B, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1})\right) = \mathbf{Q}(X_A|X_B, X_C, X_D) \otimes \mathbf{P}\left(X_B, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1})\right),$$

where we can further marginalize out the deterministic X_{B_1} to get:

$$\mathbf{P}\left(X_A, X_{B_2}, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1})\right) = \mathbf{Q}(X_A|X_B, X_C, X_D) \otimes \mathbf{P}\left(X_{B_2}, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1})\right).$$

This implies that $\mathbf{Q}(X_A|X_B, X_C, X_D)$ is a regular version of the conditional Markov kernel:

$$\mathbf{P}\left(X_A|X_{B_2}, X_C, \operatorname{do}(X_{D\cup J}, X_{B_1})\right),$$

for every decomposition: $B = B_1 \dot\cup B_2$, simultaneously, in particular for the two cases $B_1 = \emptyset$ and $B_1 = B$. This shows point 2.

Point 3.) The assumption:

$$A \underset{\mathbf{G}(V\setminus D|\operatorname{do}(D\cup J, I_B))}{\overset{\sigma}{\perp}} I_B \,|\, C \dot\cup D,$$

implies the transitional conditional independence by the global Markov property K.7:

$$X_A \underset{\mathbf{P}(X_V|\operatorname{do}(X_{D\cup J}, X_{I_B}))}{\perp\!\!\!\perp} X_{I_B} \,|\, X_C, X_D.$$

So we have the following factorization:

$$\mathbf{P}\left(X_A, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{I_B})\right) = \mathbf{Q}(X_A|X_C, X_D) \otimes \mathbf{P}\left(X_C \middle| \operatorname{do}(X_{D\cup J}, X_{I_B})\right),$$

for some Markov kernel $\mathbf{Q}(X_A|X_C, X_D)$, which serves as a regular version of the conditional Markov kernel:

$$\mathbf{P}\left(X_A|X_C, \operatorname{do}(X_{D\cup J}, X_{I_B})\right),$$

and which is independent of X_{I_B}.
We can now look at the different input values for any decomposition: $B = B_1 \dot\cup B_2$. For this we put: $X_{I_{B_1}} = x_{B_1} \in \mathcal{X}_{B_1}$ and $X_{I_{B_2}} = \star = (\star)_{v\in B_2}$. This implies:

$$\begin{aligned}
&\mathbf{P}\left(X_A, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1} = x_{B_1})\right) \\
&= \mathbf{P}\left(X_A, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{I_{B_1}} = x_{B_1}, X_{I_{B_2}} = \star)\right) \\
&= \mathbf{Q}(X_A|X_C, X_D) \otimes \mathbf{P}\left(X_C \middle| \operatorname{do}(X_{D\cup J}, X_{I_{B_1}} = x_{B_1}, X_{I_{B_2}} = \star)\right), \\
&= \mathbf{Q}(X_A|X_C, X_D) \otimes \mathbf{P}\left(X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1} = x_{B_1})\right).
\end{aligned}$$

So we get:

$$\mathbf{P}\left(X_A, X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1})\right) = \mathbf{Q}(X_A|X_C, X_D) \otimes \mathbf{P}\left(X_C \middle| \operatorname{do}(X_{D\cup J}, X_{B_1})\right),$$

which shows that $\mathbf{Q}(X_A|X_C, X_D)$ is a regular version of the conditional Markov kernel:

$$\mathbf{P}\left(X_A|X_C, \operatorname{do}(X_{D\cup J}, X_{B_1})\right)$$

for every decomposition: $B = B_1 \dot\cup B_2$, simultaneously, in particular for the two corner cases: $B_1 = \emptyset$ and $B_1 = B$. This shows point 3. □

Remark K.11. *1. Please note in Corollary K.10 how the null sets are handled. By having one Markov kernel that serves as versions for all mentioned conditional interventional Markov kernels at once circumvents the problem of having different versions and then trying to simultaneously match them by changing them on families of compatible null sets.*

2. *The 3 do-calculus rules, Corollary K.10, follow directly from the global Markov property, Theorem K.7.*

3. *Slightly more general rules can be derived from the global Markov property Theorem K.7 without the assumption of disjointness of A, B, C, D and by combining such graphical separation criteria. This will become less suggestive and readable though.*

4. *In the 3 rules of do-calculus in Corollary K.10 we were not restricted to assume that D contains J or is disjoint from it. This is thanks to the Definition 3.1 of transitional conditional independence, which will automatically provide the correct Markov kernel. This is in contrast to e.g. [FM20], where a weaker version of extended conditional independence was used. The strategy there was to first construct a suitable Markov kernel, case by case, before checking the conditional independence. For this they needed to make the restrictive assumption that $J \subseteq D$.*

K.5. Backdoor Covariate Adjustment Formula

Here we will restate the well-known backdoor covariate adjustment rules for causal Bayesian networks, see [Pea93a,Pea93b,Pea09,FM20], which were usually only stated for discrete spaces and probability distributions. Here we want to show how this rule works out in the general measure theoretic setting and highlight how the difficulties of matching occurring Markov kernels are addressed. We are not aiming at finding new adjustment rules, for extensions we refer to: [BTP14,PP14,SVR10,PTKM15,CB17,FM20]. For the identification of causal effects we refer to: [Pea09,GP95,Tia02,TP02,Tia04,SP06,HV06, HV08,RERS17,FM20].

Corollary K.12 (Conditional backdoor covariate adjustment formula)**.** *Consider a causal Bayesian network:*

$$\mathbf{M}(U, V|\,\mathrm{do}(J)) = \left(\mathbf{G}(U \,\dot{\cup}\, V|\,\mathrm{do}(J)), \quad \left(\mathbf{P}_v\left(X_v\middle|\,\mathrm{do}\left(X_{\mathrm{Pa}^{G^+}(v)}\right)\right)\right)_{v \in U \,\dot{\cup}\, V}\right).$$

Consider $A, B, C, F \subseteq V$ and $D \subseteq J \cup V$ with $J \subseteq D$ that are pairwise disjoint. Assume that the conditional backdoor criterion *in the CDGM $\mathbf{G}(V \setminus D|\,\mathrm{do}(D, I_B))$ holds:*

1. $F \underset{\mathbf{G}(V\setminus D|\,\mathrm{do}(D,I_B))}{\overset{\sigma}{\perp}} I_B\,|(C \cup D)$, *and:*

2. $A \underset{\mathbf{G}(V\setminus D|\,\mathrm{do}(D,I_B))}{\overset{\sigma}{\perp}} I_B\,|(B \cup F \cup C \cup D)$.

Then there exists a Markov kernel:

$$\mathbf{P}(X_A|X_F, X_C, \cancel{\mathrm{do}}(X_B), \mathrm{do}(X_D))$$

that simultaneously is a regular version of:

$$\mathbf{P}(X_A|X_F, X_C, \mathrm{do}(X_B, X_D)) \qquad \textit{and} \qquad \mathbf{P}(X_A|X_F, X_C, X_B, \mathrm{do}(X_D));$$

and there exists a Markov kernel:

$$\mathbf{P}(X_F|X_C, \cancel{\mathrm{do}(X_B)}, \mathrm{do}(X_D))$$

that simultaneously is a regular version of:

$$\mathbf{P}(X_F|X_C, \mathrm{do}(X_B, X_D)) \qquad \textit{and} \qquad \mathbf{P}(X_F|X_C, \mathrm{do}(X_D)).$$

Furthermore, their composition:

$$\mathbf{P}(X_A|X_F, X_C, \cancel{\mathrm{do}}(X_B), \mathrm{do}(X_D)) \circ \mathbf{P}(X_F|X_C, \cancel{\mathrm{do}(X_B)}, \mathrm{do}(X_D)),$$

is then a version of: $\mathbf{P}(X_A|X_C, \mathrm{do}(X_B, X_D))$.

Remark K.13. 1. *The proof of Corollary K.12 goes along the same lines as Rule 2 and Rule 3 in Corollary K.10.*

2. *The meaning of the backdoor covariate adjustment criterion in Corollary K.12 is that when one wants to estimate the (conditional) causal effect* $\mathbf{P}(X_A|\,\mathrm{do}(X_B))$*, for simplicity we use* $C = D = \emptyset$*, and one has a set of measured covariates/features* X_F *of the data that shields off* X_A *from "backdoor" common confounders of* X_A *and* X_B *then one can estimate the causal effect* $\mathbf{P}(X_A|\,\mathrm{do}(X_B))$ *by only using* observational *data from* $\mathbf{P}(X_F)$ *and* $\mathbf{P}(X_A|X_F, X_B)$*.*

L. Comparison to Other Notions of Conditional Independence

In this section we want to look at other notions of conditional independence and compare them to transitional conditional independence.

Recall that for transition probability space $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ and transitional random variables $\mathbf{X} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$ and $\mathbf{Y} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Y}$ and $\mathbf{Z} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Z}$ with joint Markov kernel:

$$\mathbf{K}(X, Y, Z|T) := (\mathbf{X}(X|W, T) \otimes \mathbf{Y}(Y|W, T) \otimes \mathbf{Z}(Z|W, T)) \circ \mathbf{K}(W|T),$$

we define the *transitional conditional independence* of $\mathbf{X}$ from $\mathbf{Y}$ given $\mathbf{Z}$:

$$\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \quad :\iff \quad \exists \mathbf{Q}(X|Z) : \ \mathbf{K}(X, Y, Z|T) = \mathbf{Q}(X|Z) \otimes \mathbf{K}(Y, Z|T).$$

L.1. Variation Conditional Independence

We follow [CD17a, Daw01b] and their supplementary material [CD17b] to review variation conditional independence and then comment on some possible generalizations.

For this let $\mathcal{W}$ be a set and $X : \mathcal{W} \to \mathcal{X}$, $Y : \mathcal{W} \to \mathcal{Y}$, $Z : \mathcal{W} \to \mathcal{Z}$, $U : \mathcal{W} \to \mathcal{U}$ be maps.

Notation L.1 (See [CD17a] §2.2)**.** *We define:*

$$\mathcal{R}(X) := X(\mathcal{W}) = \{X(w) \,|\, w \in \mathcal{W}\} \in 2^{\mathcal{X}},$$

and for $z \in \mathcal{Z}$:

$$\mathcal{R}(X|Z = z) := X\left(Z^{-1}(z)\right) = \{X(w) \,|\, w \in \mathcal{W}, Z(w) = z\} \in 2^{\mathcal{X}}.$$

We then define the map:

$$\mathcal{R}(X|Z) : \mathcal{Z} \to 2^{\mathcal{X}}, \quad z \mapsto \mathcal{R}(X|Z = z).$$

In this sense we then can also make sense of:

$$\mathcal{R}(X, Y|Z, U) : \mathcal{Z} \times \mathcal{U} \to 2^{\mathcal{X}\times\mathcal{Y}},$$

$$(z, u) \mapsto \mathcal{R}(X, Y|Z = z, U = u) := \{(X(w), Y(w)) \,|\, w \in \mathcal{W}, Z(w) = z, U(w) = u\}.$$

Definition L.2 (Variation conditional independence)**.** *We will say that* X is variation conditionally independent of Y given Z *if:*

$$\forall (y, z) \in \mathcal{R}(Y, Z) : \mathcal{R}(X|Y = y, Z = z) = \mathcal{R}(X|Z = z).$$

In symbols we will write then:

$$X \underset{v}{\perp\!\!\!\perp} Y \,|\, Z.$$

Notation L.3. *We will write:*

$$X \precsim_v Y$$

if there exists a map $\varphi : \mathcal{Y} \to \mathcal{X}$ such that $\varphi \circ Y = X$. Note that we use a slightly simpler, but equivalent relation, than [CD17a] §2.2, Prop. 2.6.

Remark L.4 (See [CD17a] Thm. 2.7., [CD17b])**.** *The ternary relation $\perp\!\!\!\perp_v$ together with $\precsim_v$ and $X \vee Y := (X, Y)$ is a (symmetric) separoid.*

Remark L.5. *The relation between variation conditional independence and stochastic conditional independence for random variables seems rather on the structural side, i.e. both follow similar functorial relations. In short, if one wants to go from variation to stochastic conditional independence one could start by replacing $2^{\mathcal{X}}$ with $\mathcal{P}(\mathcal{X})$ and maps with measurable maps, etc. Then maps $\mathcal{Z} \to 2^{\mathcal{X}}$ become measurable maps $\mathcal{Z} \to \mathcal{P}(\mathcal{X})$,* *which are nothing else but Markov kernels, reflecting the approach we went down for transitional conditional independence. So* $\mathcal{R}(X,Y,Z)$ *can be represented as the (constant) map:*

$$\mathcal{R}(X,Y,Z): \circledast \to 2^{\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}}, \quad * \mapsto \mathcal{R}(X,Y,Z).$$

where $\circledast = \{*\}$ *is the one-point space. We also have the "marginal":*

$$\mathcal{R}(Y,Z): \circledast \to 2^{\mathcal{Y}\times\mathcal{Z}}, \quad * \mapsto \mathcal{R}(Y,Z) = \mathrm{pr}_{\mathcal{Y}\times\mathcal{Z}}\left(\mathcal{R}(X,Y,Z)\right),$$

which is received by ignoring the X*-entries.*
It is then easily seen that we have: $X \perp\!\!\!\perp_v Y \,|\, Z$ *iff there exists a map:*

$$\mathcal{Q}(X|Z): \mathcal{Z} \to 2^{\mathcal{X}},$$

such that:

$$\mathcal{R}(X,Y,Z) = \mathcal{Q}(X|Z) \otimes_v \mathcal{R}(Y,Z),$$

where we put:

$$\mathcal{Q}(X|Z) \otimes_v \mathcal{R}(Y,Z) := \bigcup_{(y,z)\in\mathcal{R}(Y,Z)} \; \bigcup_{x\in\mathcal{Q}(X|Z=z)} \{(x,y,z)\}.$$

Example L.6. *We can now apply the above to* X,Y,Z *with common domain* $\mathcal{W}\times\mathcal{T}$ *and* $T: \mathcal{W}\times\mathcal{T}$ *the canonical projection map. Then we get:*

$$\begin{aligned} & X \underset{v,T}{\perp\!\!\!\perp} Y \,|\, Z \\ :\iff \quad & X \underset{v}{\perp\!\!\!\perp} T, Y \,|\, Z \\ \iff \quad & \exists \mathcal{Q}(X|Z): \mathcal{Z} \to 2^{\mathcal{X}}: \quad \mathcal{R}(X,Y,Z|T) = \mathcal{Q}(X|Z) \otimes_v \mathcal{R}(Y,Z|T), \end{aligned}$$

where we again now have:

$$\mathcal{R}(X,Y,Z|T=t) = \{(x,y,z) \,|\, \exists w \in \mathcal{W}: X(w,t) = x, Y(w,t) = y, Z(w,t) = z\}.$$

This shows the close formal relationship between variation conditional independence and transitional conditional independence. It then follows from Remark L.4 and the general theory in Appendix A with Theorem A.11 that $\perp\!\!\!\perp_{v,T}$ *forms a* T*-*$*$*-separoid.*

It seems, more generally, that one can formulate a (transitional) conditional independence relation in any monad with products and some extra structure. We leave this for future research.

L.2. Transitional Conditional Independence for Random Variables

If we wanted to re-define the notion of *independent* random variables X and Y on a probability space $(\mathcal{W}, \mathbf{P}(W))$ we would have a hard time coming up with something else than the classical definition of:

$$\mathbf{P}(X,Y) = \mathbf{P}(X) \otimes \mathbf{P}(Y), \tag{22}$$

where $\mathbf{P}(X)$ and $\mathbf{P}(Y)$ are the marginals. This is in contrast to *conditionally independent* random variables X and Y given a third Z, where many nuances can play a role. For instance, the direct analogon to relation 22 would read like:

$$\mathbf{P}(X, Y|Z) = \mathbf{P}(X|Z) \otimes \mathbf{P}(Y|Z) \qquad \mathbf{P}(Z)\text{-a.s.} \tag{23}$$

The problem with definition 23 is that the conditional probability distributions, like $\mathbf{P}(X, Y|Z)$, may not exist on general measurable spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, in contrast to the marginals $\mathbf{P}(X)$, $\mathbf{P}(Y)$ in equation 22. This then forces one to restrict oneself to only work with measurable spaces where regular conditional probability distributions exist, like standard measurable spaces. But even if the existence were guaranteed, they would only be unique up to some null sets. One then either ends up with a notion of conditional independence that would depend on the choices made or, as the better alternative, one would work with almost-sure equations like we already indicated in equation 23.
If one wanted to work with more general measurable spaces one could demand that equation 23 only holds for every $A \in \mathcal{B}_\mathcal{X}$ and $B \in \mathcal{B}_\mathcal{Y}$ individually. Furthermore, one then could replace $\mathbf{P}(X \in A, Y \in B|Z)$ with conditional expectations $\mathbb{E}[\mathbb{1}_A(X) \cdot \mathbb{1}_B(Y)|Z]$, etc., which exist on all measurable spaces. One then arrives at the most general and *weak form of conditional independence* for random variables:

$$\forall A \in \mathcal{B}_\mathcal{X}\, \forall B \in \mathcal{B}_\mathcal{Y} :\ \mathbb{E}[\mathbb{1}_A(X) \cdot \mathbb{1}_B(Y)|Z] = \mathbb{E}[\mathbb{1}_A(X)|Z] \cdot \mathbb{E}[\mathbb{1}_B(Y)|Z] \quad \mathbf{P}(W)\text{-a.s.}, \tag{24}$$

which can, equivalently, but more compactly, also be written as:

$$\forall A \in \mathcal{B}_\mathcal{X} :\ \mathbb{E}[\mathbb{1}_A(X)|Y, Z] = \mathbb{E}[\mathbb{1}_A(X)|Z] \qquad \mathbf{P}(W)\text{-a.s.} \tag{25}$$

We will use the following symbols for weak conditional independence:

$$X \underset{\mathbf{P}(W)}{\overset{\omega}{\perp\!\!\!\perp}} Y \,|\, Z.$$

Furthermore, if we used definition 24 or 25 on standard measurable spaces, where regular conditional probability distributions like $\mathbf{P}(X, Y|Z)$ exist, the equation 23 would automatically be implied. So the equations 24 or 25 seem to be the way to go, as one does not need to bother with existence questions, and when existence is secured the above versions are equivalent anyways. The only downside is that this definition does not provide one with a meaningful factorization. Furthermore, the conditional expectations, like $\mathbb{E}[\mathbb{1}_A(X)|Z]$, are only defined for each event A separately and thus might not be countably additive in A. So we are not given an object like a conditional distribution that we could use to further work with.
In contrast, our definition of *transitional conditional independence* $X \perp\!\!\!\perp_{\mathbf{P}(W)} Y \,|\, Z$, when restricted to random variables, would read like:

$$\exists \mathbf{P}(X|Z) : \quad \mathbf{P}(X, Y, Z) = \mathbf{P}(X|Z) \otimes \mathbf{P}(Y, Z), \tag{26}$$

where clearly $\mathbf{P}(X|Z)$ would be a regular conditional probability distribution of X given Z. So the existence of one of the regular conditional probability distributions $\mathbf{P}(X|Z)$ and a proper factorization of the joint distribution are directly built into the definition of transitional conditional independence. This makes this notion also meaningful for general measurable spaces, with the tendency that random variables are declared conditional dependent if such a regular conditional probability distribution does not even exist. Certainly, all three definitions 26, 25, 23 are equivalent on standard measurable spaces. Note that transitional conditional independence 26 is asymmetric in nature, which at this level might look like a flaw, but which allows one to generalize the definition of transitional conditional independence to *transitional random variables*, where dependencies are asymmetric from the start.

L.3. Transitional Conditional Independence for Deterministic Variables

Theorem L.7 (Transitional conditional independence for deterministic variables)**.** *Let $F : \mathcal{T} \to \mathcal{F}$ and $H : \mathcal{T} \to \mathcal{H}$ be measurable maps with $\mathcal{F}$ standard. We now consider them as (deterministic) transitional random variables on the transition probability space $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$. Let $\mathbf{Y} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Y}$ be another transitional random variable. Then the following statements are equivalent:*

1. $F \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, H$.
2. *There exists a measurable function $\varphi : \mathcal{H} \to \mathcal{F}$ such that $F = \varphi \circ H$.*

Proof. "$\Longleftarrow$": This direction follows from Left Redundancy E.3.
"$\Longrightarrow$": Since F and H are deterministic and only dependent on T we get that:

$$\mathbf{K}(F, Y, H|T) = \boldsymbol{\delta}(F|T) \otimes \boldsymbol{\delta}(H|T) \otimes \mathbf{K}(Y|T).$$

By the conditional independence we now have a Markov kernel $\mathbf{Q}(F|H)$ such that we have the factorization:

$$\mathbf{K}(F, Y, H|T) = \mathbf{Q}(F|H) \otimes \mathbf{K}(Y, H|T) = \mathbf{Q}(F|H) \otimes \boldsymbol{\delta}(H|T) \otimes \mathbf{K}(Y|T).$$

Marginalizing out Y, H and taking $T = t$ we get from these equations:

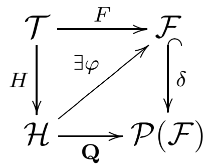
$$\boldsymbol{\delta}_{F(t)} = \boldsymbol{\delta}(F|T = t) = \mathbf{Q}(F|H(t)),$$

which is a Dirac measure centered at $F(t)$. We can now define the mapping:

$$\varphi : H(\mathcal{T}) \to \mathcal{F}, \quad H(t) \mapsto F(t),$$

which is well-defined, because $h := H(t_1) = H(t_2)$ implies that $\mathbf{Q}(F|H = h)$ is a Dirac measure centered at $F(t_1)$ and $F(t_2)$. Since $\mathcal{B}_{\mathcal{F}}$ separates points ($\mathcal{F}$ is standard) we get: $F(t_1) = F(t_2)$. φ is measurable. Indeed, its composition with $\delta : \mathcal{F} \to \mathcal{P}(\mathcal{F})$, $z \mapsto \boldsymbol{\delta}_z$

equals $\mathbf{Q}(F|H)$, which is measurable. Since $\mathcal{B}_{\mathcal{F}} = \delta^*\mathcal{B}_{\mathcal{P}(\mathcal{F})}$, see Lemma B.35, also φ is measurable. Since $\mathcal{F}$ is a standard measurable space, φ extends to a measurable mapping $\varphi : \mathcal{H} \to \mathcal{F}$ by Kuratowski's extension theorem for standard measurable spaces (see [Kec95] 12.2 and Theorem B.50). Finally, note that we have $F(t) = \varphi(H(t))$ for all $(w,t) \in \mathcal{W} \times \mathcal{T}$, which shows the claim.



□

Remark L.8. *Note that the proof of Theorem L.7 also works for universal measurable spaces $\mathcal{F}$ if one replaces measurability with universal measurability everywhere. One then needs to use Kuratowski's extension theorem for univeral measurable spaces, Theorem B.52.*

L.4. Equivalent Formulations of Transitional Conditional Independence

Our groundwork of developing the framework of transition probability spaces, transitional random variables and transitional conditional independence now allows us to rigorously compare different notions of extended conditional indpendence in the literature. We will compare to three of them, namely the one from [CD17a, RERS17, FM20].

To relate transitional conditional independence to other notion of conditional independence it is useful to reformulate transitional conditional independence in other terms. The main result for this will be the next theorem.

Theorem L.9. *Let $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ be a transition probability space and $\mathbf{X} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{X}$ and $\mathbf{Y} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Y}$ and $\mathbf{Z} : \mathcal{W} \times \mathcal{T} \dashrightarrow \mathcal{Z}$ transitional random variables. We put:*

$$\mathbf{K}(X,Y,Z|T) := (\mathbf{X}(X|W,T) \otimes \mathbf{Y}(Y|W,T) \otimes \mathbf{Z}(Z|W,T)) \circ \mathbf{K}(W|T),$$

We will write $\mathbf{K}(X|Z,\cancel{T})$ for a fixed version of the Markov kernel appearing in the conditional independence $\mathbf{X} \perp\!\!\!\perp_{\mathbf{K}(W|T)} \boldsymbol{\delta}_* \,|\, \mathbf{Z}$ (only in case it holds).
With these notations, the following are equivalent:

1. $\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z}$,

2. $\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{T} \otimes \mathbf{Y} \,|\, \mathbf{Z}$,

3. $\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \boldsymbol{\delta}_* \,|\, \mathbf{Z}$ *and* $\mathbf{K}(X,Y,Z|T) = \mathbf{K}(X|Z,\cancel{T}) \otimes \mathbf{K}(Y,Z|T)$.

4. *$\mathbf{X} \precsim^*_{\mathbf{K}} \mathbf{Z}$ and for every $t \in \mathcal{T}$ we have: $X_t \overset{\omega}{\underset{\mathbf{K}(X,Y,Z|T=t)}{\perp\!\!\!\perp}} Y_t \,|\, Z_t$ (in the weak sense).*

Furthermore, any of those points implies the following:

5. *For every probability distribution $\mathbf{Q}(T) \in \mathcal{P}(\mathcal{T})$ we have the conditional independence:*

$$\mathbf{X} \underset{\mathbf{K}(W|T)\otimes\mathbf{Q}(T)}{\perp\!\!\!\perp} \mathbf{T} \otimes \mathbf{Y} \,|\, \mathbf{Z}.$$

Proof. 3. $\implies$ 1. is clear by definition.
1. $\implies$ 2.: by $\mathbf{T}$-Inverted Right Decomposition E.7.
2. $\implies$ 4.,5.: By assumption we have the factorization:

$$\mathbf{K}(X, Y, Z, T|T) = \mathbf{K}(X|Z) \otimes \mathbf{K}(Y, Z, T|T),$$

for some Markov kernel $\mathbf{K}(X|Z)$. Via marginalization and multiplication this implies the two equations:

$$\begin{aligned} \mathbf{K}(X, Z, T|T) &= \mathbf{K}(X|Z) \otimes \mathbf{K}(Z, T|T), \\ \underbrace{\mathbf{K}(X, Y, Z|T) \otimes \mathbf{Q}(T)}_{=:\mathbf{Q}(X,Y,Z,T)} &= \mathbf{K}(X|Z) \otimes \underbrace{\mathbf{K}(Y, Z|T) \otimes \mathbf{Q}(T)}_{=\mathbf{Q}(Y,Z,T)}, \end{aligned}$$

for every $\mathbf{Q}(T) \in \mathcal{P}(\mathcal{T})$. The last equation shows 5.
If we take $\mathbf{Q}(T) = \boldsymbol{\delta}_t$ we get:

$$\mathbf{K}(X, Y, Z|T = t) = \mathbf{K}(X|Z) \otimes \mathbf{K}(Y, Z|T = t).$$

Together with the first of the above equations this shows 4.
4. $\implies$ 3.: By $\mathbf{X} \precsim^*_{\mathbf{K}} \mathbf{Z}$ we have a factorization:

$$\mathbf{K}(X, Z|T) = \mathbf{P}(X|Z) \otimes \mathbf{K}(Z|T).$$

This means that for every $t \in \mathcal{T}$ and every measurable $A \subseteq \mathcal{X}$, $C \subseteq \mathcal{Z}$ we have:

$$\mathbb{E}_t \left[\mathbb{1}_A(X_t) \cdot \mathbb{1}_C(Z_t)\right] = \mathbb{E}_t \left[\mathbf{P}(X \in A|Z_t) \cdot \mathbb{1}_C(Z_t)\right],$$

where the expectation $\mathbb{E}_t$ is w.r.t. $\mathbf{K}(X, Y, Z|T = t)$. This shows that $\mathbf{P}(X \in A|Z_t)$ is a version of $\mathbb{E}_t[\mathbb{1}_A(X_t)|Z_t]$ for every $t \in \mathcal{T}$, by the defining properties of conditional expectation.
By the assumption $X_t \perp\!\!\!\perp^{\omega}_{\mathbf{K}(X,Y,Z|T=t)} Y_t \,|\, Z_t$ we then have for every fixed $t \in \mathcal{T}$ and measurable $A \subseteq \mathcal{X}$:

$$\mathbb{E}_t[\mathbb{1}_A(X_t)|Y_t, Z_t] = \mathbb{E}_t[\mathbb{1}_A(X_t)|Z_t] = \mathbf{P}(X \in A|Z_t) \qquad \mathbf{K}(X, Y, Z|T = t)\text{-a.s.}$$

By the defining properties of conditional expectation for $\mathbb{E}_t[\mathbb{1}_A(X_t)|Y_t, Z_t]$ we then get that for every measurable $A \subseteq \mathcal{X}$, $B \subseteq \mathcal{Y}$, $C \subseteq \mathcal{Z}$:

$$\mathbb{E}_t \left[\mathbb{1}_A(X_t) \cdot \mathbb{1}_B(Y_t) \cdot \mathbb{1}_C(Z_t)\right] = \mathbb{E}_t \left[\mathbf{P}(X \in A|Z_t) \cdot \mathbb{1}_B(Y_t) \cdot \mathbb{1}_C(Z_t)\right].$$

Since this holds for every $t \in \mathcal{T}$ we get:

$$\mathbf{K}(X, Y, Z|T) = \mathbf{P}(X|Z) \otimes \mathbf{K}(Y, Z|T),$$

which shows the claim. □

Corollary L.10. *If $\mathcal{X}$ is standard and $\mathcal{Z}$ countably generated (e.g. also standard) then we have the equivalence:*

$$\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \otimes \mathbf{T} \qquad \Longleftrightarrow \qquad \forall t \in \mathcal{T}: \quad X_t \underset{\mathbf{K}(X,Y,Z|T=t)}{\perp\!\!\!\perp} Y_t \,|\, Z_t.$$

Proof. This directly follows from Theorem L.9 4. with (Z, T) in the role of Z and Remark 3.6 to get the first part of 4. □

L.5. The Extended Conditional Independence

We shortly review the definition of *extended conditional independence* introduced in [CD17a].

Definition L.11 (Extendend conditional independence, see [CD17a] Def. 3.2)**.** *Let $\mathcal{W}$ and $\mathcal{T}$ be measurable spaces and $\mathcal{E} = (\mathbf{P}_t(W))_{t \in \mathcal{T}}$ be a family of probability measures on $\mathcal{W}$. Let X, Y, Z be measurable maps on $\mathcal{W}$ and Φ, Θ measurable maps on $\mathcal{T}$ such that the joint map (Φ, Θ) is injective. For these cases* extended conditional independence *was defined as:*

$$X \underset{\mathcal{E}}{\perp\!\!\!\perp} (Y, \Theta) \,|\, (Z, \Phi)$$

if for all $\phi \in \Phi(\mathcal{T})$ and all real bounded measurable h there exists a function $g_{h,\phi}$ such that for all $t \in \Phi^{-1}(\phi)$ we have that:

$$\mathbb{E}_t \left[h(X) | Y, Z \right] = g_{h,\phi}(Z) \quad \mathbf{P}_t(W)\text{-a.s.},$$

where the conditional expectation $\mathbb{E}_t$ is w.r.t. $\mathbf{P}_t(W)$.

We now show that when (Y, Θ) and (Z, Φ) are considered as transitional random variables on transition probability space $(\mathcal{W} \times \mathcal{T}, \mathbf{P}(W|T))$, where we put $\mathbf{P}(W|T = t) := \mathbf{P}_t(W)$, then transitional conditional independence implies extended conditional independence.

Lemma L.12. *We have the implication:*

$$X \underset{\mathbf{P}(W|T)}{\perp\!\!\!\perp} (Y, \Theta) \,|\, (Z, \Phi) \qquad \Longrightarrow \qquad X \underset{\mathcal{E}}{\perp\!\!\!\perp} (Y, \Theta) \,|\, (Z, \Phi).$$

Proof. Indeed, by the assumption we get a Markov kernel $\mathbf{Q}(X|Z, \Phi)$ such that:

$$\begin{aligned} \mathbf{P}(X, Y, \Theta, Z, \Phi|T) &= \mathbf{Q}(X|Z, \Phi) \otimes \mathbf{P}(Y, \Theta, Z, \Phi|T) \\ &= \mathbf{Q}(X|Z, \Phi) \otimes \boldsymbol{\delta}(\Theta|T) \otimes \boldsymbol{\delta}(\Phi|T) \otimes \mathbf{P}(Y, Z|T). \end{aligned}$$

Marginalizing out Θ and Φ gives:

$$\mathbf{P}(X, Y, Z|T = t) = \mathbf{Q}(X|Z, \Phi = \Phi(t)) \otimes \mathbf{P}(Y, Z|T = t).$$

For any ϕ and function h we then define:

$$g_{h,\phi}(z) := \int_{\mathcal{X}} h(x)\, \mathbf{Q}(X \in dx|Z = z, \Phi = \phi).$$

Then for each $t \in \Phi^{-1}(\phi)$ and $B \in \mathcal{B}_{\mathcal{Y}}$ and $C \in \mathcal{B}_{\mathcal{Z}}$ we get:

$$\begin{aligned}
&\int_C \int_B \int_{\mathcal{X}} h(x)\, \mathbf{P}(X \in dx, Y \in dy, Z \in dz|T = t) \\
&= \int_C \int_B \int_{\mathcal{X}} h(x)\, \mathbf{Q}(X \in dx|Z = z, \Phi = \Phi(t))\, \mathbf{P}(Y \in dy, Z \in dz|T = t) \\
&= \int_C \int_B g_{h,\phi}(z)\, \mathbf{P}(Y \in dy, Z \in dz|T = t).
\end{aligned}$$

Since this is the defining equation for the conditional expectation we get the claim:

$$\mathbb{E}_t\left[h(X)|Y, Z\right] = g_{h,\phi}(Z) \quad \mathbf{P}_t(W)\text{-a.s..}$$

□

Remark L.13. *So transitional conditional independence is the stronger notion and implies extended conditional independence, but it works for all transitional random variables, not just of the restricted type in Definition L.11. Furthermore and in contrast to extended conditional independence, transitional conditional independence satisfies all the (asymmetric) separoid rules, Theorem 3.11, for all measurable spaces, except Left Weak Union and T-Restricted Right Redundancy, which hold when one can ensure the existence of regular conditional Markov kernels, e.g. on standard measurable spaces, see Theorem 2.24. This makes transitional conditional independence a better fit for the use in graphical models, see, for instance, the global Markov property, Theorem 6.3. Note that both notions only have a restricted direct relation to variation conditional independence, see Theorem 3.8 and [CD17a, CD17b]. A formal analogy between variation conditional independence and transitional conditional independence was discussed in Remark L.5 and Example L.6.*

L.6. Symmetric Extended Conditional Independence

If we wanted to arrive at a symmetric version of extended conditional independence that satisfies all (symmetric) separoid rules (at least when restricted so standard or universal measurable spaces) we could just use *symmetrized transitional conditional independence*:

$$\mathbf{X} \underset{\mathbf{K}(W|T)}{\overset{\vee}{\perp\!\!\!\perp}} \mathbf{Y} \,|\, \mathbf{Z} \qquad :\Longleftrightarrow \qquad \mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \quad \vee \quad \mathbf{Y} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{X} \,|\, \mathbf{Z}.$$

Since $\perp\!\!\!\perp_{\mathbf{K}(W|T)}$ forms a T-$*$-separoid it is immediate that $\perp\!\!\!\perp^{\vee}_{\mathbf{K}(W|T)}$ is a symmetric separoid by the general theory of τ-κ-separoids, see Theorem A.12. We clearly have the implication:

$$\mathbf{X} \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} \mathbf{Y} \,|\, \mathbf{Z} \qquad \Longrightarrow \qquad \mathbf{X} \underset{\mathbf{K}(W|T)}{\overset{\vee}{\perp\!\!\!\perp}} \mathbf{Y} \,|\, \mathbf{Z},$$

showing that the asymmetric version is stronger than the symmetrized version, where the latter might have lost some information about the interplay between $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{T}$. Furthermore, because of its symmetry it can not equivalently express classical statistical concepts like ancillarity, sufficiency, adequacy, etc., see Section 4. A symmetric notion for discrete Markov kernels was intoduced in [RERS17] without the measure theoretic overhead.

L.7. Extended Conditional Independence for Families of Probability Distributions

In this subsection we will introduce a strikingly simple and powerful form of extended conditional independence that works for all measurable spaces and satisfies all the separoid rules. For this consider a transition probability space $(\mathcal{W} \times \mathcal{T}, \mathbf{K}(W|T))$ and transitional random variables $X : \mathcal{W} \times \mathcal{T} \to \mathcal{X}$, $Y : \mathcal{W} \times \mathcal{T} \to \mathcal{Y}$, $Z : \mathcal{W} \times \mathcal{T} \to \mathcal{Z}$. Furthermore, fix a set $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{T})$ of probability measures on $\mathcal{T}$, e.g. $\mathcal{Q} = \mathcal{P}(\mathcal{T})$ or $\mathcal{Q} = \{\boldsymbol{\delta}_t \,|\, t \in \mathcal{T}\}$. Then we can define *$\mathcal{Q}$-extended conditional independence* as:

$$X \underset{\mathbf{K}(W|T)\otimes\mathcal{Q}}{\overset{\omega}{\perp\!\!\!\perp}} Y \,|\, Z \qquad :\Longleftrightarrow \qquad \forall \mathbf{Q}(T) \in \mathcal{Q} : \quad X \underset{\mathbf{K}(W|T)\otimes\mathbf{Q}(T)}{\overset{\omega}{\perp\!\!\!\perp}} Y \,|\, Z.$$

It is easily seen that the usual weak conditional independence $\perp\!\!\!\perp^{\omega}$ satisfies all separoid axioms [Daw01a, CD17a] for arbitrary measurable spaces. Furthermore, if one combines several separoids by conjunction $\wedge$ then one gets another separoid, see [Daw01a]. So $\perp\!\!\!\perp^{\omega}_{\mathbf{K}(W|T)\otimes\mathcal{Q}}$ clearly satisfies all (symmetric) separoid axioms for arbitrary measurable spaces. By Theorem L.9 we have the implications:

$$X \underset{\mathbf{K}(W|T)}{\perp\!\!\!\perp} Y \,|\, Z \quad \Longrightarrow \quad X \underset{\mathbf{K}(W|T)\otimes\mathcal{Q}}{\overset{\omega}{\perp\!\!\!\perp}} T, Y \,|\, Z \quad \Longrightarrow \quad X \underset{\mathbf{K}(W|T)\otimes\mathcal{Q}}{\overset{\omega}{\perp\!\!\!\perp}} Y \,|\, Z.$$

The middle ternary relation in X, Y, Z satisfy the asymmetric separoid rules from Theorem 3.11, but without any requirement for standard or countably generated measurable spaces, in contrast to transitional conditional indpendence on the left. The asymmetric separoid rules for the middle relation follow from the right relation and Theorem A.11.

It seems that $\mathcal{Q}$-extended conditional independence checks all boxes that one would like to have from a notion of extended conditional independence. It is certainly simpler than most other notions. It just comes with one drawback: it does not provide one with the existence or factorization of certain Markov kernels. When the reverse implication holds is stated in Theorem L.9, e.g. if $\boldsymbol{\delta}_t \in \mathcal{Q}$ for all $t \in \mathcal{T}$ and $X \precsim^*_{\mathbf{K}} Z$, where the latter encodes the existence of a certain Markov kernel, which is thus the main obstruction to arrive at transitional conditional independence.

To ellaborate further, a specialized version of this $\mathcal{Q}$-extended conditional independence was first introduced in [FM20], where it was used to derive the causal do-calculus rules, see [Pea09, FM20], for certain structural causal models. In their proofs they had to construct certain Markov kernels and then check for $\mathcal{Q}$-extended conditional independence. Since the construction of such Markov kernels became complicated many corner cases could not been proved. The main ingredient of their proof was a *global Markov property* for $\mathcal{Q}$-extended conditional independence:

$$A \overset{\sigma}{\underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\perp}} J \cup B \,|\, C \qquad \Longrightarrow \qquad X_A \overset{\omega}{\underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))\otimes\mathcal{Q}}{\perp\!\!\!\perp}} X_J, X_B \,|\, X_C.$$

Here, $\mathcal{Q}$-extended conditional independence was not strong enough to produce the needed Markov kernels. This is in contrast to transitional conditional independence, whose global Markov property, Theorem 6.3 now gives:

$$A \overset{\sigma}{\underset{\mathbf{G}(V\,|\,\mathrm{do}(J))}{\perp}} B \,|\, C \qquad \Longrightarrow \qquad X_A \underset{\mathbf{P}(X_V\,|\,\mathrm{do}(X_J))}{\perp\!\!\!\perp} X_B \,|\, X_C,$$

which is a stronger conclusion and provides us with the needed Markov kernels for free. This was one of the core motivation for developing transitional conditional independence. For a proof of the do-calculus rules for causal Bayesian networks with non-stochastic input variables and latent variables that also covers such corner cases easily see Corollary K.10, which is based on the global Markov property, Theorem 6.3.