Discrete-Continuous Mixtures in Probabilistic Programming: Generalized Semantics and Inference Algorithms Supplementary Materials

A. Background on Measure-theoretical Probability Theory

We assume familiarity with measure-theoretic approaches to probability theory, but provide the fundamental definitions. The standard Borel σ-algebra is assumed in all the discussion. See (Durrett, 2013) and (Kallenberg, 2002) for introduction and further details.

A **measurable space** $(X, \mathcal{X})$ (space, for short) is an underlying set X paired with a σ-algebra $\mathcal{X} \subseteq 2^X$ of measurable subsets of X, i.e., a family of subsets containing the underlying set X which is closed under complements and countable unions. We'll denote the measurable space simply by $\mathcal{X}$ where no ambiguity results. A function $f\colon \mathcal{X} \to \mathcal{Y}$ between measurable spaces is measurable if measurable sets pullback to measurable sets: $f^{-1}(B) \in \mathcal{X}$ for all $B \in \mathcal{Y}$. A **measure** μ on a measurable space $\mathcal{X}$ is a function $\mu\colon \mathcal{X} \to [0, \infty]$ which satisfies countable additivity: for any countable sequence $A_1, A_2, \dots \in \mathcal{X}$ of disjoint measurable sets $\mu(\cup_i A_i) = \sum_i \mu(A_i)$. $\mathbb{P}_\mu[S]$ denotes the probability of a statement S under the base measure μ, and similarly for conditional probabilities. A probability kernel is the measure-theoretic generalization of a conditional distribution. It is commonly used to construct measures over a product space, analogously to how conditional distributions are used to define joint distributions in the chain rule.

Definition A.1. *A **probability kernel** K from one measurable space $\mathcal{X}$ to another $\mathcal{Y}$ is a function $K\colon X \times \mathcal{Y} \to [0, 1]$ such that (a) for every $x \in X$, $K(x, \cdot)$ is a probability measure over $\mathcal{Y}$, and (b) for every $B \in \mathcal{Y}$, $K(\cdot, B)$ is a measurable function from $\mathcal{X}$ to* $[0, 1]$.

Given an arbitrary index set T and spaces $\mathcal{X}_t$ for each index $t \in T$, the **product space** $\mathcal{X} = \prod_{t\in T} \mathcal{X}_t$ is the space with underlying set $X = \prod_{t\in T} X_t$ the Cartesian product of the underlying sets, adorned with the smallest σ-algebra such that the projection functions $\pi_t\colon \mathcal{X} \to \mathcal{X}_t$ are measurable.

B. MTBNs Represent Unique Measures

We prove here Theorem 3.6. Its proof requires a series of intermediate results. We first define a projective family of measures. This gives a way to recursively construct our measure μ. We define a notion of consistency such that every consistent projective family constructs a measure that M represents. We end by giving an explicit characterization of the unique consistent projective family, and thus of the unique measure M represents. The appendix contains additional technical material required in the proofs.

Intuitively, the main objective of this section is to show that an MTBN defines a unique measure that "factorizes" according to the network, as an extension to the corresponding result for Bayes Nets.

B.1. Consistent projective family of measures

Let K be a kernel from $\mathcal{X} \to \mathcal{Y}$ and L a kernel from $\mathcal{Y} \to \mathcal{Z}$. Their composition $K \circ L$ (note the ordering!) is a kernel from $\mathcal{X}$ to $\mathcal{Z}$ defined for $x \in X$, $C \in \mathcal{Z}$ by:

$$(K \circ L)(x, C) = \int K(x, dy) \int L(y, dz)\, 1_C(z). \tag{5}$$

To allow uniform notation, we will treat measurable functions and measures as special cases of kernels. A measurable function $f\colon X \to Y$ corresponds to the kernel K_f from $\mathcal{X}$ to $\mathcal{Y}$ given by $K_f(x, B) = 1(f(x) \in B)$ for $x \in X$ and $B \in \mathcal{Y}$. A measure μ on a space $\mathcal{X}$ is a kernel K_μ from 1, the one element measure space, to $\mathcal{X}$ given by $K_\mu(\cdot, A) = \mu(A)$ for $A \in \mathcal{X}$. Where this yields no confusion, we use f and μ in place of K_f and K_μ. (5) simplifies if the kernels are measures or functions. Let μ be a measure on $\mathcal{Y}_1$, K be a kernel from $\mathcal{X}_1$ to $\mathcal{Y}_1$, f be a measurable function from $\mathcal{X}_2$ to $\mathcal{X}_1$, and g be a measurable function from $\mathcal{Y}_1$ to $\mathcal{Y}_2$. Then $\mu \circ g$ is a measure on $\mathcal{Y}_2$ and $f \circ K \circ g$ is a kernel from $\mathcal{X}_2$ to $\mathcal{Y}_2$ with: $(\mu \circ g)(B) = \mu(g^{-1}(B))$, and $(f \circ K \circ g)(x, B) = K(f(x), g^{-1}(B))$.

Let Λ denote the class of upwardly closed sets: subsets of V containing all their elements' parents.

Definition B.1. *A **projective family** of measures is a family $\{\mu_U : U \in \Lambda\}$ consisting of a measure μ_U on $\mathcal{X}_U$ for every $U \in \Lambda$ such that whenever $W \subseteq U$ we have $\mu_W = \mu_U \circ \pi^U_W$, i.e., for all $A \in \mathcal{X}_W$, $\mu_W(A) = \mu_U((\pi^U_W)^{-1}(A))$.*

Def. B.1 captures the measure-theoretic version of the probability of a subset of variables being equal to the marginals obtained while "summing out" the probabilities of the other variables in a joint distribution.

Definition B.2. *Let μ be a measure on a measure space $\mathcal{X}$, and K a kernel from $\mathcal{X}$ to a measure space $\mathcal{Y}$. Then $\mu \otimes K$ is the measure on $\mathcal{X} \times \mathcal{Y}$ defined for $B \in \mathcal{X} \otimes \mathcal{Y}$ by: $(\mu \otimes K)(B) = \int \mu(dx) \int K(x, dy)\, 1_B(x, y)$.*

Def. B.2 defines the operation of composing a conditional probability with a prior on a parent, to obtain the corresponding joint distribution.

Definition B.3. *Let K_w for $w \in W$ be kernels from $\mathcal{X}_U$ to $\mathcal{X}_{\{w\}}$. Denote by $\prod_{w\in W} K_w$ the kernel from $\mathcal{X}_U$ to $\mathcal{X}_W$ defined for each $x_U \in \mathcal{X}_U$ by the infinite product of measures: $\left(\prod_{w\in W} K_w\right)(x_U, \cdot) = \otimes_{w\in W} K_w(x_U, \cdot)$.*

See (Kallenberg, 2002) 1.27 and 6.18 for definition and existence of infinite products of measures. Def. B.3 captures the kernel representation for taking the equivalent of products of conditional distributions of a set of variables with a common set U of parents.

Definition B.4. *A projective family $\{\mu_U : U \in \Lambda\}$ is **consistent with** M if for any $W, U \in \Lambda$ such that $W \subset U$ and $\mathrm{pa}(U) \subseteq W$, then: $\mu_U = \mu_W \otimes \prod_{u \in U \setminus W} (\pi^W_{\mathrm{pa}(u)} \circ K_u)$.*

Consistency in Def. B.4 captures the global condition that we would like to see in a generalization of a Bayes network. Namely, the distribution of any set of parent-closed random variables should "factorize" according to the network

A projective family $\{\mu_U : U \in \Lambda\}$ is consistent with M exactly when M represents μ_V:

Lemma B.5. *Let μ be a measure on $\mathcal{X}_V$, and define the projective family $\{\mu_U : U \in \Lambda\}$ by $\mu_U = \mu \circ \pi^V_U$. This projective family is consistent with M iff M represents μ.*

Proof. First we'll relate consistency (Def. 8) with conditional expectation and distribution properties of random variables. Take any $W, U \in \Lambda$ such that $W \subset U$ and $\mathrm{pa}(U) \subseteq W$ and observe that the following are equivalent:

- $\mu_U = \mu_W \otimes \prod_{u \in U \setminus W} (\pi^W_{\mathrm{pa}(u)} \circ K_u)$
- $\prod_{u \in U \setminus W} (\pi^W_{\mathrm{pa}(u)} \circ K_u)$ is a version of the conditional distribution of $X_{U \setminus W}$ given X_W,
- K_u is a version of the conditional distribution of X_u given $X_{\mathrm{pa}(u)}$ for all $u \in U \setminus W$, and $\{X_W, X_u : u \in U \setminus W\}$ are mutually independent conditional on $X_{\mathrm{pa}(U)}$.

The forward direction is straightforward. For the converse we use the fact that conditional independence of families of random variables holds if it holds for all finite subsets, establishing that by chaining conditional independence (see (Kallenberg, 2002) p109 and 6.8). □

Lemma B.5 shows that Def. B.4 follows iff an MTBN represents the joint distribution – in other words, it follows iff the local Markov property holds.

B.2. There exists a unique consistent family

Each vertex $v \in V$ is assigned the unique minimal ordinal $d(v)$ such that $d(u) < d(v)$ whenever $(u, v) \in E$ (see (Jech, 2003) for an introduction to ordinals). For any $U \in \Lambda$ denote by $U^\alpha = \{u \in U : v(u) < \alpha\}$ the restriction of U to vertices of depth less than α. Defining $D = \sup_{v \in V}(d(v) + 1)$, the least strict upper bound on depth, we have that $U^D = U$ for all $U \in \Lambda$. In the following, fix a limit ordinal λ.

Definition B.6. *$\{\nu_\alpha : \alpha < \lambda\}$ is a **projective sequence of measures** on $\mathcal{X}_{U_\alpha}$ if whenever $\alpha < \beta < \lambda$ we have $\nu_\alpha = \nu_\beta \circ \pi^{U_\beta}_{U_\alpha}$.*

Def. B.6 generalizes the notion of subset relationships and the marginalization operations that hold between supersets and subsets to the case of infinite dependency chains

Definition B.7. *The limit $\lim_{\alpha<\lambda} \nu_\alpha$ of a projective sequence $\{\nu_\alpha : \alpha < \lambda\}$ of measures is the unique measure on $\mathcal{X}_U$ such that $\nu_\alpha = (\lim_{\alpha<\lambda} \nu_\alpha) \circ \pi^U_{U_\alpha}$ for all $\alpha < \beta$.*

Definition B.8. *Given any $U \in \Lambda$, inductively define a measure μ^α_U on $\mathcal{X}_{U^\alpha}$ by*

$$\mu^0_U = 1,$$
$$\mu^{\alpha+1}_U = \mu^\alpha_U \otimes \prod_{v \in U : d(v) = \alpha} (\pi^{U^\alpha}_{\mathrm{pa}(v)} \circ K_v),$$
$$\mu^\lambda_U = \lim_{\alpha<\lambda} \mu^\alpha_U \qquad \text{if } \lambda \text{ is a limit ordinal.}$$

μ^α_U stabilizes for $\alpha \geq D$ to define a measure on $\mathcal{X}_U$.

The above definition is coherent as μ^α_U can be inductively shown to be a projective sequence. Lemma B.9 and B.10 allow us to show in Theorem B.11 that $\{\mu^D_U : U \in \Lambda\}$ is the unique consistent projective family of measures.

Lemma B.9. *If $W \subseteq U$ for $W, U \in \Lambda$, then for all α: $\mu^\alpha_W = \mu^\alpha_U \circ \pi^{U^\alpha}_{W^\alpha}$.*

Proof is in Appx. C.

Lemma B.10. *If $W \subset U$ where $W, U \in \Lambda$, and if $\mathrm{pa}(U) \subseteq W$, then $W^\alpha \subset U^\alpha$, $\mathrm{pa}(U^\alpha) \subseteq W^\alpha$, and $\mu^\alpha_U = \mu^\alpha_W \otimes \prod_{u \in U^\alpha \setminus W^\alpha} (\pi^{W^\alpha}_{\mathrm{pa}(u)} \circ K_u)$.*

Proof is in Appx. D.

Using the above, the following shows MTBNs satisfy the properties (1-3) mention in the beginning of Sec. 1.1:

Theorem B.11. *$\{\mu^D_U : U \in \Lambda\}$ is the unique projective family of measures consistent with M.*

Proof is in Appx. E.

Intuitively, by Lemma B.9 and Lemma B.10, we assert that consistency holds for any ordinal-bounded (prefix in terms of parent ordering) sub-network. Then the main result, Thm. B.11, follows by setting this bound appropriately. Finally Lemma B.5 and Theorem B.11 lead to Theorem 3.6.

Note that combining Thm. 3.6 and Thm. 5.6 lead to all the 4 desired properties mentioned in Sec. 1.1.

C. Proof for Lemma B.9

Proof. Proof by induction. Trivially true for $\alpha = 0$, so suppose this holds for α, and consider $\alpha + 1$. Then:

$$\begin{aligned}
\mu^{\alpha+1}_W &= \mu^\alpha_W \otimes \prod_{v \in W : d(v) = \alpha} (\pi^{W^\alpha}_{\mathrm{pa}(v)} \circ K_v) \\
&= \left(\mu^\alpha_U \circ \pi^{U^\alpha}_{W^\alpha}\right) \\
&\quad \otimes \left(\left(\prod_{v \in U : d(v) = \alpha} (\pi^{W^\alpha}_{\mathrm{pa}(v)} \circ K_v) \right) \circ \pi^{U^{\alpha+1} \setminus U^\alpha}_{W^{\alpha+1} \setminus W^\alpha} \right) \\
&= \left(\mu^\alpha_U \otimes \left(\pi^{U^\alpha}_{W^\alpha} \circ \prod_{v \in U : d(v) = \alpha} (\pi^{W^\alpha}_{\mathrm{pa}(v)} \circ K_v) \right) \right) \\
&\quad \circ (\pi^{U^\alpha}_{W^\alpha} \times \pi^{U^{\alpha+1} \setminus U^\alpha}_{W^{\alpha+1} \setminus W^\alpha}) \\
&= \left(\mu^\alpha_U \otimes \prod_{v \in U : d(v) = \alpha} (\pi^{U^\alpha}_{\mathrm{pa}(v)} \circ K_v) \right) \circ \pi^{U^{\alpha+1}}_{W^{\alpha+1}} \\
&= \mu^{\alpha+1}_U \circ \pi^{U^{\alpha+1}}_{W^{\alpha+1}}
\end{aligned}$$

The first step by Def. 12, the second by inductive hypothesis and Lemma G.11 as $\{v \in W : d(v) = \alpha\} = W^{\alpha+1} \setminus W^{\alpha}$ and $\{v \in U : d(v) = \alpha\} = U^{\alpha+1} \setminus U^{\alpha}$, the third by Lemma G.6, the fourth by Lemma G.10 since $\pi^{U^\alpha}_{\mathrm{pa}(v)} = \pi^{W^\alpha}_{\mathrm{pa}(v)} \circ \pi^{U^\alpha}_{W^\alpha}$ and by elementary properties of projections, and the fifth by Definition B.8.

Finally, suppose λ is a limit ordinal. We need to show:

$$\lim_{\alpha<\lambda} \left(\mu^\alpha_U \circ \pi^{U^\alpha}_{W^\alpha}\right) = \left(\lim_{\alpha<\lambda} \mu^\alpha_U\right) \circ \pi^{U^\lambda}_{W^\lambda}.$$

This follows from Lemma G.2 because for all $\alpha < \lambda$ we have:

$$\left(\left(\lim_{\alpha<\lambda} \mu^\alpha_U\right) \circ \pi^{U^\lambda}_{W^\lambda}\right) \circ \pi^{W^\lambda}_{W^\alpha} = \left(\left(\lim_{\alpha<\lambda} \mu^\alpha_U\right) \circ \pi^{U^\lambda}_{U^\alpha}\right) \circ \pi^{U^\alpha}_{W^\alpha} = \mu^\alpha_U \circ \pi^{U^\alpha}_{W^\alpha}$$

The first by properties of projections, the second by Lemma G.2 characterizing limits. □

D. Proof for Lemma B.10

Proof. Trivial for $\alpha = 0$, so suppose this holds for α, and consider $\alpha + 1$. Then:

$$\begin{aligned}
&\mu^{\alpha+1}_U \\
&= \mu^\alpha_U \otimes \prod_{v \in U : d(v)=\alpha} (\pi^{U^\alpha}_{\mathrm{pa}(v)} \circ K_v) \\
&= \mu^\alpha_W \otimes \prod_{u \in U^\alpha \setminus W^\alpha} (\pi^{W^\alpha}_{\mathrm{pa}(u)} \circ K_u) \otimes \prod_{v \in U : d(v)=\alpha} (\pi^{U^\alpha}_{\mathrm{pa}(v)} \circ K_v) \\
&= \mu^\alpha_W \otimes \prod_{u \in U^{\alpha+1} \setminus W^\alpha} (\pi^{W^\alpha}_{\mathrm{pa}(u)} \circ K_u) \\
&= \mu^\alpha_W \otimes \prod_{v \in W : d(v)=\alpha} (\pi^{W^\alpha}_{\mathrm{pa}(v)} \circ K_v) \\
&\qquad \otimes \prod_{u \in U^{\alpha+1} \setminus W^{\alpha+1}} (\pi^{W^{\alpha+1}}_{\mathrm{pa}(u)} \circ K_u) \\
&= \mu^{\alpha+1}_W \otimes \prod_{u \in U^{\alpha+1} \setminus W^{\alpha+1}} (\pi^{W^{\alpha+1}}_{\mathrm{pa}(u)} \circ K_u),
\end{aligned}$$

The first step by Definition B.8, the second by inductive hypothesis. The third by Lemmas G.8 and G.9 since $U^{\alpha+1} \setminus W^\alpha = U^\alpha \setminus W^\alpha \cup \{v \in U : d(v) = \alpha\}$ where the union is disjoint, and as $\mathrm{pa}(v) \subseteq W^\alpha$ when $v \in U$ and $d(v) = \alpha$ implies that $\pi^{U^\alpha}_{\mathrm{pa}(v)} = \pi^{U^\alpha}_{W^\alpha} \circ \pi^{W^\alpha}_{\mathrm{pa}(v)}$. The fourth by Lemmas G.8 and G.9 since $U^{\alpha+1} \setminus W^\alpha = U^{\alpha+1} \setminus W^{\alpha+1} \cup \{v \in W : d(v) = \alpha\}$ where the union is disjoint, and as $\mathrm{pa}(u) \subseteq W^\alpha$ when $u \in U^{\alpha+1} \setminus W^{\alpha+1}$ implies that $\pi^{W^{\alpha+1}}_{\mathrm{pa}(v)} = \pi^{W^{\alpha+1}}_{W^\alpha} \circ \pi^{W^\alpha}_{\mathrm{pa}(v)}$. Finally, the fifth by Definition B.8.

Finally, suppose λ is a limit ordinal. The result will follow from the inductive hypothesis, Definition B.8, and as limits preserve products Lemma G.7 if we can show that

$$\lim_{\alpha<\lambda} \prod_{u \in U^\alpha \setminus W^\alpha} (\pi^{W^\alpha}_{\mathrm{pa}(u)} \circ K_u) = \prod_{u \in U^\lambda \setminus W^\lambda} (\pi^{W^\lambda}_{\mathrm{pa}(u)} \circ K_u).$$

First we must show the limit on the left is well-defined. Note that the kernel inside the limit maps from $\mathcal{X}_{W^\alpha}$ to $\mathcal{X}_{U^\alpha \setminus W^\alpha}$. As W^α and $U^\alpha \setminus W^\alpha$ are both increasing sets, we verify projective sequence property by taking any $\beta > \alpha$ and observing that

$$\begin{aligned}
&\pi^{W^\beta}_{W^\alpha} \circ \prod_{u \in U^\alpha \setminus W^\alpha} (\pi^{W^\alpha}_{\mathrm{pa}(u)} \circ K_u) \\
&= \prod_{u \in U^\alpha \setminus W^\alpha} (\pi^{W^\beta}_{\mathrm{pa}(u)} \circ K_u) \\
&= \left(\prod_{u \in U^\beta \setminus W^\beta} (\pi^{W^\beta}_{\mathrm{pa}(u)} \circ K_u)\right) \circ \pi^{U^\beta \setminus W^\beta}_{U^\alpha \setminus W^\alpha}
\end{aligned}$$

the first step from Lemma G.10 and properties of projections, and the second from Lemma G.11.

Finally, we must show the expression on the right satisfies the properties characterizing the limit. However, observe this follows from our demonstration of the projective sequence property above by simply replacing β with λ. □

E. Proof for Theorem B.11

Proof. That this is a consistent projective family follows from Lemmas B.9 and B.10 since $U^D = U$ for all $U \in \Lambda$.

For uniqueness, let $\{\hat{\mu}_U : U \in \Lambda\}$ be a consistent projective family of measures, any fix any $U \in \Lambda$. We'll show inductively that $\hat{\mu}_{U^\alpha} = \mu^\alpha_U$, and thus with $\alpha = D$ that $\hat{\mu}_U = \mu_U$, giving our result. This is trivial for $\alpha = 0$, so inductively suppose it holds for α. But then:

$$\begin{aligned}
\hat{\mu}_{U^{\alpha+1}} &= \hat{\mu}_{U^\alpha} \otimes \prod_{u \in U^{\alpha+1} \setminus U^\alpha} (\pi^{U^\alpha}_{\mathrm{pa}(u)} \circ K_u) \\
&= \mu^\alpha_U \otimes \prod_{u \in U^{\alpha+1} \setminus U^\alpha} (\pi^{U^\alpha}_{\mathrm{pa}(u)} \circ K_u).
\end{aligned}$$

The first step by consistency of $\{\hat{\mu}_U\}$ (Definition B.4) since $U^\alpha \subseteq U^{\alpha+1}$ and $\mathrm{pa}(U^{\alpha+1}) \subseteq U^\alpha$, the second by inductive hypothesis, and the third by Definition B.8.

Let α be a limit ordinal. Since $\{\hat{\mu}_{U^\alpha}\}$ is a projective family and $U^\alpha = \bigcup_{\beta<\alpha} U^\beta$, by Lemma G.2 $\hat{\mu}_{U^\alpha} = \lim_{\beta<\alpha} \hat{\mu}_{U^\beta}$. By definition $\mu^\alpha_U = \lim_{\beta<\alpha} \mu^\beta_U$. Then since $\mu^\beta_U = \hat{\mu}_{U^\beta}$ for $\beta < \alpha$ inductively, $\mu^\alpha_U = \hat{\mu}_{U^\alpha}$ as the limit of this sequence is unique. □

F. Proof of Lemma 5.4

Proof. The possible world $\langle U^\sigma, I^\sigma \rangle$ is defined as follows. $U^\sigma = \langle U^\sigma_1, \ldots, U^\sigma_k \rangle$, where $U^\sigma_j = \{c_j : c_j$ is a distinct constant of type τ_j in $\mathcal{M}\} \cup \{u_{\nu,\bar{u},l} \in \mathcal{U}_\mathcal{M} : \nu$ is a number statement of type τ_j and $\sigma(V_\nu[\bar{u}]) \geq l\}$.

I^σ is defined as follows. For each function symbol $f(\bar{x})$ in $\mathcal{M}$, for each tuple $\bar{u}$ of the type of $\bar{x}$ constructed using elements of U^σ, $[f]^\sigma(\bar{u}) = \sigma(V_f[\bar{u}])$. The element $\sigma(V_f[\bar{u}])$ is a member of U^σ because of the last clause in the definition of consistent assignments (Def. 5.3) and the construction of U^σ. □

G. Additional Technical Details

For reasons of space, we present the following without their (straightforward) proofs.

Lemma G.1. *If μ is a measure on $\mathcal{X}$, and is K a kernel from $\mathcal{X}$ to $\mathcal{Y}$, then $(\mu \otimes K) \circ \pi_{\mathcal{X}}^{\mathcal{X}\times\mathcal{Y}} = \mu$.*

Lemma G.2. *A projective sequence of measures has a unique limit.*

Fix an ordinal λ, and suppose $\{U_\alpha \subseteq V : \alpha < \lambda\}$ is an increasing sequence of subsets of V, i.e., such that if $\alpha < \beta < \lambda$ then $U_\alpha \subseteq U_\beta$. Define $U = \bigcup_{\alpha<\lambda} U_\alpha$. Let $\{W_\alpha \subseteq V : \alpha < \lambda\}$ and W be another such sequence, supposing U and W are disjoint.

Definition G.3. *$\{K_\alpha : \alpha < \lambda\}$ is a **projective sequence of kernels** from $\mathcal{X}_{U_\alpha}$ to $\mathcal{X}_{W_\alpha}$ if whenever $\alpha < \beta < \lambda$ we have $\pi_{U_\alpha}^{U_\beta} \circ K_\alpha = K_\beta \circ \pi_{W_\alpha}^{W_\beta}$.*

Definition G.4. *The limit $\lim_{\alpha<\beta} K_\alpha$ of a projective sequence $\{K_\alpha : \alpha < \lambda\}$ of kernels is the unique kernel from $\mathcal{X}_U$ to $\mathcal{X}_W$ such that for all $\alpha < \lambda$ $\pi_{U_\alpha}^{U} \circ K_\alpha = (\lim_{\alpha<\beta} K_\alpha) \circ \pi_{W_\alpha}^{W}$.*

Lemma G.5. *A projective sequence of kernels has a unique limit.*

Lemma G.6. *Let $\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2, \mathcal{Y}_2$ be measurable spaces, μ be a measure on $\mathcal{X}_1$, K a kernel from $\mathcal{X}_2$ to $\mathcal{Y}_1$, $f \colon \mathcal{X}_1 \to \mathcal{X}_2$ a measurable function, and $g \colon \mathcal{Y}_1 \to \mathcal{Y}_2$ a measurable function. Then: $(\mu \otimes (f \circ K)) \circ (f \times g) = (\mu \circ f) \otimes (K \circ g)$ where $f \times g$ is the measurable function mapping (x, y) to $((f(x), g(y))$.*

Lemma G.7. *Let ν_α and K_α be as in Lemmas G.2 and G.5. Then $\lim_{\alpha<\lambda}(\nu_\alpha \otimes K_\alpha) = (\lim_{\alpha<\lambda} \nu_\alpha) \otimes (\lim_{\alpha<\lambda} K_\alpha)$.*

Lemma G.8. *μ measure on $\mathcal{X}$, K_1 a kernel from $\mathcal{X}$ to $\mathcal{Y}_1$, K_2 a kernel from $\mathcal{X}$ to $\mathcal{Y}_2$, $\mu \otimes K_1 \otimes (\pi_{\mathcal{X}}^{\mathcal{X}\times\mathcal{Y}_1} \circ K_2) = \mu \otimes \prod_{i=1,2} K_i$. where by abuse of notation $\pi_{\mathcal{X}}^{\mathcal{X}\times\mathcal{Y}_1}$ denotes the projection from $\mathcal{X} \times \mathcal{Y}_1$ to $\mathcal{X}$.*

Lemma G.9. *If $K_{i,j}$ are kernels from $\mathcal{X}$ to $\mathcal{Y}_{i,j}$ then $\prod_i \prod_j K_{i,j} = \prod_{i,j} K_{i,j}$.*

Lemma G.10. *If $f \colon \mathcal{X}' \to \mathcal{X}$ and K_i are kernels from $\mathcal{X}$ to $\mathcal{Y}_i$ then $f \circ \prod_i K_i = \prod_i f \circ K_i$.*

Lemma G.11. *If K_v for $v \in U$ are kernels from $\mathcal{X}$ to $\mathcal{X}_v$, and $W \subseteq U$ then $\left(\prod_{v\in U} K_v\right) \circ \pi_W^U = \prod_{v\in W} K_v$.*

Lemma G.12. *Let $(X, \mathcal{X})$ be a measurable space, $X, X_1, X_2, \ldots$ an iid random sequence on $\mathcal{X}$, and $w(x)$ be non-negative real-valued function of $(X, \mathcal{X})$. Then $\frac{\sum_{i=1}^n w(X_i) f(X_i)}{\sum_{i=1}^n w(X_i)} \overset{\text{a.s.}}{\to} \frac{\mathbb{E}\, w(X) f(X)}{\mathbb{E}\, w(X)}$.*

Lemma G.13. *For any measurable set E and measurable function $f(x)$: $\frac{\mathbb{E}\, P(E|X) f(X)}{\mathbb{E}\, P(E|X)} = \mathbb{E}[f(X)|E]$.*