Joint Distributions in Probabilistic Semantics

Dexter Kozen, *Cornell University*
Alexandra Silva, *Cornell University*
Erik Voogd, *University of Oslo*

MFPS 2023

Bloomington, 23 June 2023

Probabilistic Programming

- Many applications: machine learning and other statistical analyses
- Examples: Church, Anglican, Stan, Venture, ...

Probabilistic Programming

- Many applications: machine learning and other statistical analyses
- Examples: Church, Anglican, Stan, Venture, ...

- *Sampling* random values
- *Probabilistic conditioning* (*Bayesian inference*)

Probabilistic Programming

- Many applications: machine learning and other statistical analyses
- Examples: Church, Anglican, Stan, Venture, ...

- *Sampling* random values
- *Probabilistic conditioning* (*Bayesian inference*)

```
x:=normal(0,1);
y:=normal(x,1);
observe(y=1);
return(x>1);
```

Denotational Semantics

Measure theory!

- quasi-Borel spaces
- Markov kernels
- Joint measures

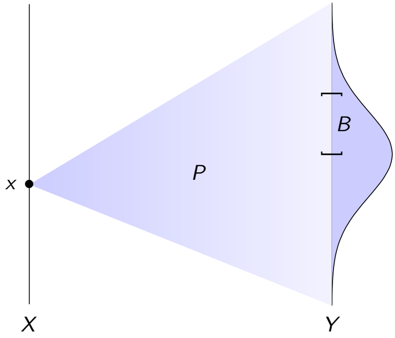
Measure theory!

- quasi-Borel spaces
- Markov kernels
- **Joint measures**

Markov Kernels

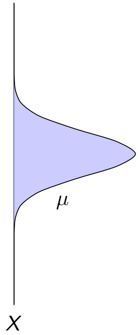
Markov kernel $(X, \Sigma_X) \multimap (Y, \Sigma_Y)$ is a map $X \times \Sigma_Y \to [0, \infty)$ s.t.

- For fixed $x \in X$, $P(x, -)$ is a finite measure on Y
- For fixed $B \in \Sigma_Y$, $P(\cdot\,, B)$ is a measurable function on X



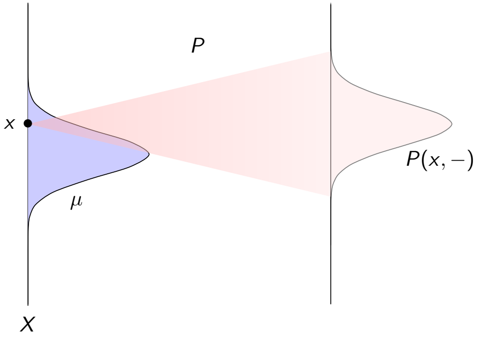
Probabilistic Programs as Markov Kernels

- `x:=normal(0,1)` builds a measure space $(\mathbb{R}, \mu)$ (Gaussian)



Probabilistic Programs as Markov Kernels

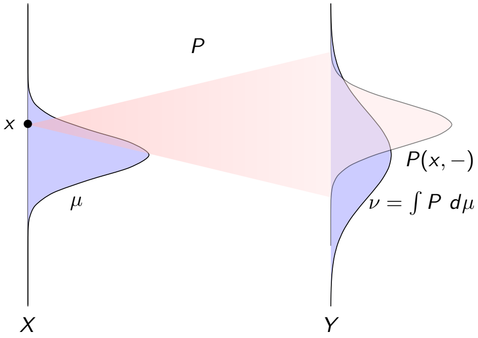
- `x:=normal(0,1)` builds a measure space $(\mathbb{R}, \mu)$ (Gaussian)
- `y:=normal(x,1)` defines kernel $P : \mathbb{R} \rightharpoonup \mathbb{R}, (x, B) \mapsto \gamma_{x,1}(B)$



Probabilistic Programs as Markov Kernels

- `x:=normal(0,1)` builds a measure space $(\mathbb{R}, \mu)$ (Gaussian)
- `y:=normal(x,1)` defines kernel $P : \mathbb{R} \to \mathbb{R}, (x, B) \mapsto \gamma_{x,1}(B)$ and builds

$$\nu(B) = \int_{x \in X} P(x, B)\ d\mu$$

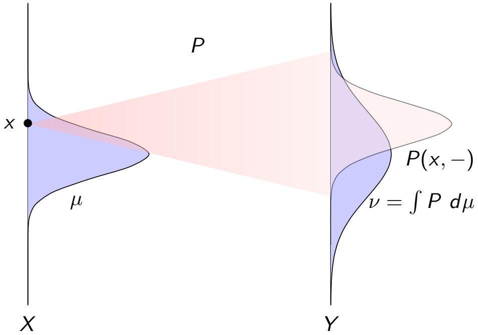


The Category of Kernels

The category **Krn**

- Objects are measure spaces (X, Σ_X, μ), (Y, Σ_Y, ν), ...
- Arrows from (X, Σ_X, μ) to (Y, Σ_Y, ν) are Markov kernels P such that

$$\nu(B) = \int_{x \in X} P(x, B)\ d\mu$$



The Category of Kernels

The category **Krn**

- Objects are measure spaces (X, Σ_X, μ), (Y, Σ_Y, ν), ...
- Arrows from (X, Σ_X, μ) to (Y, Σ_Y, ν) are Markov kernels P such that

$$\nu(B) = \int_{x \in X} P(x, B)\ d\mu$$

For example:

- Let $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ such that $\sum_i \boldsymbol{x}(i) = 1$ and $\sum_i \boldsymbol{y}(i) = 1$
- $(n, \mathcal{P}(n), \boldsymbol{x})$ and $(n, \mathcal{P}(n), \boldsymbol{y})$ are objects (probability spaces)

The Category of Kernels

The category **Krn**

- Objects are measure spaces (X, Σ_X, μ), (Y, Σ_Y, ν), ...
- Arrows from (X, Σ_X, μ) to (Y, Σ_Y, ν) are Markov kernels P such that

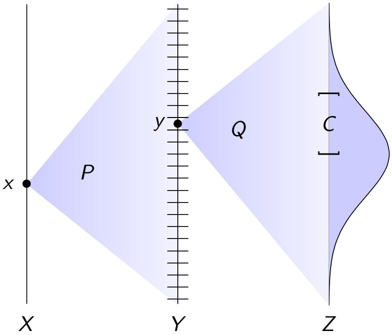
$$\nu(B) = \int_{x \in X} P(x, B) \; d\mu$$

For example:

- Let $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ such that $\sum_i \boldsymbol{x}(i) = 1$ and $\sum_i \boldsymbol{y}(i) = 1$
- $(n, \mathcal{P}(n), \boldsymbol{x})$ and $(n, \mathcal{P}(n), \boldsymbol{y})$ are objects (probability spaces)
- Arrows from $\boldsymbol{x}$ to $\boldsymbol{y}$ are matrices $A \in [0,1]^{n \times n}$ such that $\boldsymbol{y} = A^T \boldsymbol{x}$

$$\boldsymbol{y}(j) = \sum_{i \in n} A(i,j) \cdot \boldsymbol{x}(i)$$

Composition in **Krn**



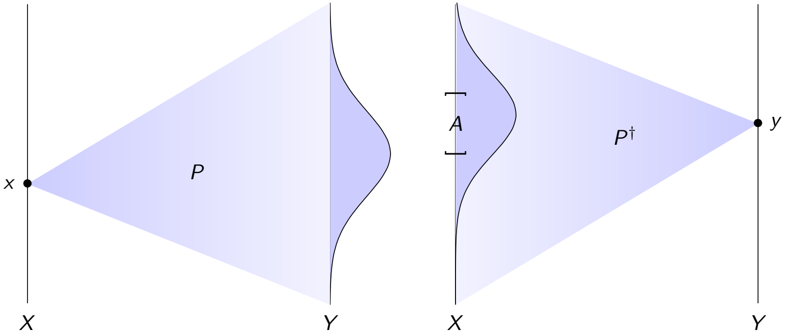
$$(P \mathbin{;} Q)(x, C) = \int_{y \in Y} Q(y, C)\ dP(x, -)$$

Bayesian Inference in **Krn**

- **Krn** is symmetric monoidal closed

- **Krn** is symmetric monoidal closed
- **Krn** is a dagger category, where $(-)^\dagger$ is Bayesian inversion

$$(Q \circ P)^\dagger = P^\dagger \circ Q^\dagger \qquad (P^\dagger)^\dagger = P$$

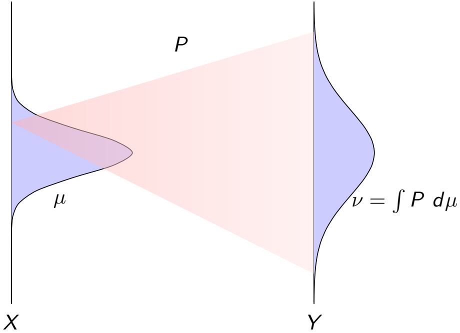


Bayesian Inference in **Krn**

- **Krn** is symmetric monoidal closed
- **Krn** is a dagger category, where $(-)^\dagger$ is Bayesian inversion

$$(Q \circ P)^\dagger = P^\dagger \circ Q^\dagger \qquad (P^\dagger)^\dagger = P$$

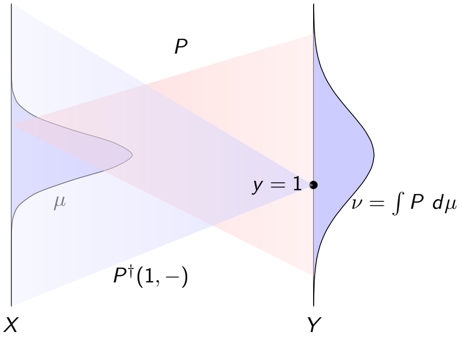
```
x:=normal(0,1);
y:=normal(x,1);
observe(y=1);
return(x>1);
```



- **Krn** is symmetric monoidal closed
- **Krn** is a dagger category, where $(-)^\dagger$ is Bayesian inversion

$$(Q \circ P)^\dagger = P^\dagger \circ Q^\dagger \qquad (P^\dagger)^\dagger = P$$

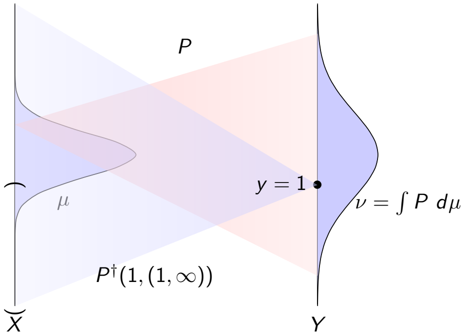
```
x:=normal(0,1);
y:=normal(x,1);
observe(y=1);
return(x>1);
```



- **Krn** is symmetric monoidal closed
- **Krn** is a dagger category, where $(-)^\dagger$ is Bayesian inversion

$$(Q \circ P)^\dagger = P^\dagger \circ Q^\dagger \qquad (P^\dagger)^\dagger = P$$

```
x:=normal(0,1);
y:=normal(x,1);
observe(y=1);
return(x>1);
```



- **Krn** is symmetric monoidal closed
- **Krn** is a dagger category, where $(-)^\dagger$ is Bayesian inversion

$$(Q \circ P)^\dagger = P^\dagger \circ Q^\dagger \qquad (P^\dagger)^\dagger = P$$

For $P : X \to Y$ a kernel, let

1. $\gamma(A \times B) := \int_{x \in X} \mathbb{1}_A(x) \cdot P(x, B)\, d\mu$ be a joint measure on $X \times Y$
    - γ has marginals μ and ν
2. Let Q be the *disintegration* of γ along π_Y
3. Define $P^\dagger := \pi_X \circ Q$

A Category of Joint Distributions

The category **PRel**

- Objects are measure spaces (X, Σ_X, μ), (Y, Σ_Y, ν), ...
- Arrows from (X, Σ_X, μ) to (Y, Σ_Y, ν) are *joint measures*

$$\gamma : \Sigma_X \otimes \Sigma_Y \to \mathbb{R}_+$$

such that $\gamma(- \times Y) \ll \mu$ and $\gamma(X \times -) \ll \nu$

A Category of Joint Distributions

The category **PRel**

- Objects are measure spaces (X, Σ_X, μ), (Y, Σ_Y, ν), ...
- Arrows from (X, Σ_X, μ) to (Y, Σ_Y, ν) are *joint measures*

$$\gamma : \Sigma_X \otimes \Sigma_Y \to \mathbb{R}_+$$

such that $\gamma(- \times Y) \ll \mu$ and $\gamma(X \times -) \ll \nu$

(X, Σ_X, μ) and (X, Σ_X, μ') are *isomorphic* iff $\mu \ll \mu'$ and $\mu' \ll \mu$

A Category of Joint Distributions

The category **PRel**

- Objects are measure spaces (X, Σ_X, μ), (Y, Σ_Y, ν), ...
- Arrows from (X, Σ_X, μ) to (Y, Σ_Y, ν) are *joint measures*

$$\gamma : \Sigma_X \otimes \Sigma_Y \to \mathbb{R}_+$$

such that $\gamma(- \times Y) = \mu$ and $\gamma(X \times -) = \nu$

(X, Σ_X, μ) and (X, Σ_X, μ') are *isomorphic* iff $\mu \ll \mu'$ and $\mu' \ll \mu$

A Category of Joint Distributions

The category **PRel**

- Objects are measure spaces (X, Σ_X, μ), (Y, Σ_Y, ν), ...
- Arrows from (X, Σ_X, μ) to (Y, Σ_Y, ν) are *joint measures*

$$\gamma : \Sigma_X \otimes \Sigma_Y \to \mathbb{R}_+$$

such that $\gamma(- \times Y) = \mu$ and $\gamma(X \times -) = \nu$

Example:

- $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ discrete probability spaces
- Joint measure $\alpha \in [0,1]^{n \times n}$ with $\alpha 1 = \boldsymbol{x}$ and $\alpha^T 1 = \boldsymbol{y}$

A Category of Joint Distributions

The category **PRel**

- Objects are measure spaces (X, Σ_X, μ), (Y, Σ_Y, ν), ...
- Arrows from (X, Σ_X, μ) to (Y, Σ_Y, ν) are *joint measures*

$$\gamma : \Sigma_X \otimes \Sigma_Y \to \mathbb{R}_+$$

such that $\gamma(- \times Y) = \mu$ and $\gamma(X \times -) = \nu$

Example:

- $\boldsymbol{x}, \boldsymbol{y} \in [0, 1]^n$ discrete probability spaces
- Joint measure $\alpha \in [0, 1]^{n \times n}$ with $\alpha 1 = \boldsymbol{x}$ and $\alpha^T 1 = \boldsymbol{y}$

$$\forall i.\ \ \boldsymbol{x}(i) = \sum_{j \in n} \alpha(i, j) \qquad \forall j.\ \ \boldsymbol{y}(j) = \sum_{i \in n} \alpha(i, j)$$

Krn and **PRel** are Isomorphic

$$\mathbf{Krn} \underset{G}{\overset{F}{\rightleftarrows}} \mathbf{PRel}$$

- For a kernel $P : (X, \mu) \to (Y, \nu)$,

$$FP(A \times B) = \int_{x \in X} \mathbb{1}_A(x) \cdot P(x, B)\ d\mu$$

Krn and **PRel** are Isomorphic

$$\mathbf{Krn} \underset{G}{\overset{F}{\rightleftarrows}} \mathbf{PRel}$$

- For a kernel $P : (X, \mu) \to (Y, \nu)$,

$$FP(A \times B) = \int_{x \in X} \mathbb{1}_A(x) \cdot P(x, B)\ d\mu$$

- Conversely, $G\gamma$ is a *disintegration* of γ
  - The resulting kernel is unique up to a μ-nullset

Krn and **PRel** are Isomorphic

$$\mathbf{Krn} \underset{G}{\overset{F}{\rightleftarrows}} \mathbf{PRel}$$

- For a kernel $P : (X, \mu) \to (Y, \nu)$,

$$FP(A \times B) = \int_{x \in X} \mathbb{1}_A(x) \cdot P(x, B)\ d\mu$$

- Conversely, $G\gamma$ is a *disintegration* of γ
  - The resulting kernel is unique up to a μ-nullset
  - Disintegration only possible if the underlying space is Polish

Krn and **PRel** are Isomorphic

$$\mathbf{Krn} \underset{G}{\overset{F}{\rightleftarrows}} \mathbf{PRel}$$

$$FP(A \times B) = \int_{x \in X} \mathbb{1}_A(x) \cdot P(x, B)\ d\mu$$

Example:

- Let $A \in [0,1]^{n \times n}$ be a kernel such that $A^T \boldsymbol{x} = \boldsymbol{y}$

$$FA = \mathrm{diag}(x) \cdot A = \begin{pmatrix} a_{11}x_1 & a_{12}x_1 & \dots & a_{1n}x_1 \\ a_{21}x_2 & a_{22}x_2 & \dots & a_{2n}x_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}x_n & a_{n2}x_n & \dots & a_{nn}x_n \end{pmatrix}$$

- FA has marginals $(FA)1 = \boldsymbol{x}$ and $(FA)^T 1 = \boldsymbol{y}$

Composition in **PRel**

PRel $\quad (X,\mu) \xrightarrow{\gamma} (Y,\nu) \xrightarrow{\theta} (Z,\rho)$

Composition in **PRel**

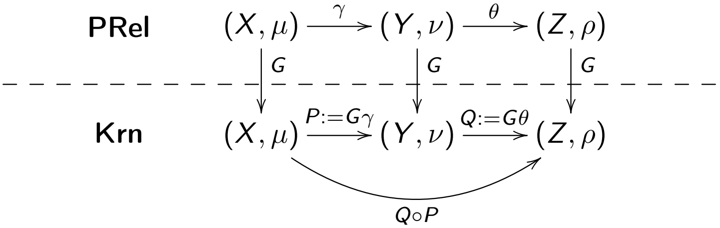
PRel	(X, μ)	$\xrightarrow{\gamma}$	(Y, ν)	$\xrightarrow{\theta}$	(Z, ρ)
Krn	(X, μ)		(Y, ν)		(Z, ρ)

Composition in **PRel**

$$
\begin{array}{llll}
\textbf{PRel} & (X,\mu) \xrightarrow{\gamma} & (Y,\nu) \xrightarrow{\theta} & (Z,\rho) \\
 & \downarrow G & \downarrow G & \downarrow G \\
\textbf{Krn} & (X,\mu) \xrightarrow{P:=G\gamma} & (Y,\nu) \xrightarrow{Q:=G\theta} & (Z,\rho)
\end{array}
$$

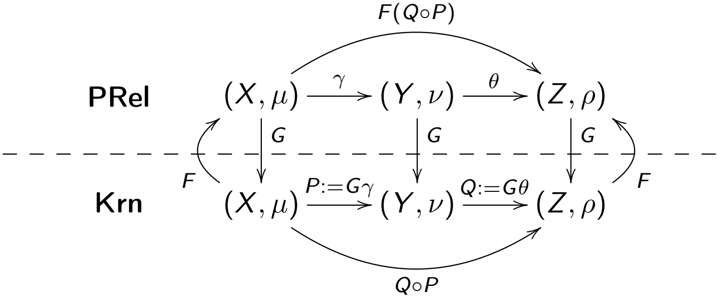
1. Disintegrate γ to P, and θ to Q

Composition in **PRel**



1. Disintegrate γ to P, and θ to Q
2. Compose P and Q in **Krn**

Composition in **PRel**



1. Disintegrate γ to P, and θ to Q
2. Compose P and Q in **Krn**
3. Reintegrate $Q \circ P$ to a joint measure

A New Category of Joint Distributions

The category **JDist**:

- Objects are *any* measure space (X, Σ_X, μ)
- Arrows $\theta : (X, \mu) \to (Y, \nu)$ are joint measures with marginals μ and ν

How to compose?

Composition in **JDist** by Example

- Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ be probability measures on $n = 3$
- Let $\boldsymbol{x} \xrightarrow{A} \boldsymbol{y}$ and $\boldsymbol{y} \xrightarrow{B} \boldsymbol{z}$ be kernels, so $A^T \boldsymbol{x} = \boldsymbol{y}$ and $B^T \boldsymbol{y} = \boldsymbol{z}$

Composition in **JDist** by Example

- Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ be probability measures on $n = 3$
- Let $\boldsymbol{x} \xrightarrow{A} \boldsymbol{y}$ and $\boldsymbol{y} \xrightarrow{B} \boldsymbol{z}$ be kernels, so $A^T \boldsymbol{x} = \boldsymbol{y}$ and $B^T \boldsymbol{y} = \boldsymbol{z}$
- The corresponding joint measures are

$$\widehat{A} = \begin{pmatrix} A_{1,1}\boldsymbol{x}_1 & A_{1,2}\boldsymbol{x}_1 & A_{1,3}\boldsymbol{x}_1 \\ A_{2,1}\boldsymbol{x}_2 & A_{2,2}\boldsymbol{x}_2 & A_{2,3}\boldsymbol{x}_2 \\ A_{3,1}\boldsymbol{x}_3 & A_{3,2}\boldsymbol{x}_3 & A_{3,3}\boldsymbol{x}_3 \end{pmatrix} \; \widehat{B} = \begin{pmatrix} B_{1,1}\boldsymbol{y}_1 & B_{1,2}\boldsymbol{y}_1 & B_{1,3}\boldsymbol{y}_1 \\ B_{2,1}\boldsymbol{y}_2 & B_{2,2}\boldsymbol{y}_2 & B_{2,3}\boldsymbol{y}_2 \\ B_{3,1}\boldsymbol{y}_3 & B_{3,2}\boldsymbol{y}_3 & B_{3,3}\boldsymbol{y}_3 \end{pmatrix}$$

Composition in **JDist** by Example

- Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ be probability measures on $n = 3$
- Let $\boldsymbol{x} \xrightarrow{A} \boldsymbol{y}$ and $\boldsymbol{y} \xrightarrow{B} \boldsymbol{z}$ be kernels, so $A^T\boldsymbol{x} = \boldsymbol{y}$ and $B^T\boldsymbol{y} = \boldsymbol{z}$
- The corresponding joint measures are

$$\widehat{A} = \begin{pmatrix} A_{1,1}\boldsymbol{x}_1 & A_{1,2}\boldsymbol{x}_1 & A_{1,3}\boldsymbol{x}_1 \\ A_{2,1}\boldsymbol{x}_2 & A_{2,2}\boldsymbol{x}_2 & A_{2,3}\boldsymbol{x}_2 \\ A_{3,1}\boldsymbol{x}_3 & A_{3,2}\boldsymbol{x}_3 & A_{3,3}\boldsymbol{x}_3 \end{pmatrix} \quad \widehat{B} = \begin{pmatrix} B_{1,1}\boldsymbol{y}_1 & B_{1,2}\boldsymbol{y}_1 & B_{1,3}\boldsymbol{y}_1 \\ B_{2,1}\boldsymbol{y}_2 & B_{2,2}\boldsymbol{y}_2 & B_{2,3}\boldsymbol{y}_2 \\ B_{3,1}\boldsymbol{y}_3 & B_{3,2}\boldsymbol{y}_3 & B_{3,3}\boldsymbol{y}_3 \end{pmatrix}$$

- For the kernel composition $C = AB$ we have $C_{i,j} = \sum_{k=1}^{3} A_{i,k}B_{k,j}$
- Its corresponding joint measure is $\widehat{C}$ with $\widehat{C}_{ij} = \sum_{k=1}^{3} A_{i,k}B_{k,j}\boldsymbol{x}_i$

Composition in **JDist** by Example

- Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ be probability measures on $n = 3$
- Let $\boldsymbol{x} \xrightarrow{A} \boldsymbol{y}$ and $\boldsymbol{y} \xrightarrow{B} \boldsymbol{z}$ be kernels, so $A^T\boldsymbol{x} = \boldsymbol{y}$ and $B^T\boldsymbol{y} = \boldsymbol{z}$
- The corresponding joint measures are

$$\widehat{A} = \begin{pmatrix} A_{1,1}\boldsymbol{x}_1 & A_{1,2}\boldsymbol{x}_1 & A_{1,3}\boldsymbol{x}_1 \\ A_{2,1}\boldsymbol{x}_2 & A_{2,2}\boldsymbol{x}_2 & A_{2,3}\boldsymbol{x}_2 \\ A_{3,1}\boldsymbol{x}_3 & A_{3,2}\boldsymbol{x}_3 & A_{3,3}\boldsymbol{x}_3 \end{pmatrix} \quad \widehat{B} = \begin{pmatrix} B_{1,1}\boldsymbol{y}_1 & B_{1,2}\boldsymbol{y}_1 & B_{1,3}\boldsymbol{y}_1 \\ B_{2,1}\boldsymbol{y}_2 & B_{2,2}\boldsymbol{y}_2 & B_{2,3}\boldsymbol{y}_2 \\ B_{3,1}\boldsymbol{y}_3 & B_{3,2}\boldsymbol{y}_3 & B_{3,3}\boldsymbol{y}_3 \end{pmatrix}$$

- For the kernel composition $C = AB$ we have $C_{i,j} = \sum_{k=1}^{3} A_{i,k}B_{k,j}$
- Its corresponding joint measure is $\widehat{C}$ with $\widehat{C}_{ij} = \sum_{k=1}^{3} A_{i,k}B_{k,j}\boldsymbol{x}_i$
- Want: $D = \widehat{B} \star \widehat{A}$ such that $D = \widehat{C}$
- Matrix multiplication gives

$$D_{i,j} = (\widehat{A}\widehat{B})_{i,j} = \left(\sum_k \widehat{A}_{i,k}\widehat{B}_{k,j}\right)_{i,j} = \left(\sum_k (A_{i,k}\boldsymbol{x}_i)(B_{k,j}\boldsymbol{y}_k)\right)_{i,j}$$

Composition in **JDist** by Example

- Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ be probability measures on $n = 3$
- Let $\boldsymbol{x} \xrightarrow{A} \boldsymbol{y}$ and $\boldsymbol{y} \xrightarrow{B} \boldsymbol{z}$ be kernels, so $A^T \boldsymbol{x} = \boldsymbol{y}$ and $B^T \boldsymbol{y} = \boldsymbol{z}$
- The corresponding joint measures are

$$\widehat{A} = \begin{pmatrix} A_{1,1}\boldsymbol{x}_1 & A_{1,2}\boldsymbol{x}_1 & A_{1,3}\boldsymbol{x}_1 \\ A_{2,1}\boldsymbol{x}_2 & A_{2,2}\boldsymbol{x}_2 & A_{2,3}\boldsymbol{x}_2 \\ A_{3,1}\boldsymbol{x}_3 & A_{3,2}\boldsymbol{x}_3 & A_{3,3}\boldsymbol{x}_3 \end{pmatrix} \quad \widehat{B} = \begin{pmatrix} B_{1,1}\boldsymbol{y}_1 & B_{1,2}\boldsymbol{y}_1 & B_{1,3}\boldsymbol{y}_1 \\ B_{2,1}\boldsymbol{y}_2 & B_{2,2}\boldsymbol{y}_2 & B_{2,3}\boldsymbol{y}_2 \\ B_{3,1}\boldsymbol{y}_3 & B_{3,2}\boldsymbol{y}_3 & B_{3,3}\boldsymbol{y}_3 \end{pmatrix}$$

- For the kernel composition $C = AB$ we have $C_{i,j} = \sum_{k=1}^{3} A_{i,k} B_{k,j}$
- Its corresponding joint measure is $\widehat{C}$ with $\widehat{C}_{ij} = \sum_{k=1}^{3} A_{i,k} B_{k,j} \boldsymbol{x}_i$
- Want: $D = \widehat{B} \star \widehat{A}$ such that $D = \widehat{C}$
- Define

$$D_{i,j} := \left(\sum_{k, \boldsymbol{y}_k > 0} \frac{\widehat{A}_{i,k} \widehat{B}_{k,j}}{\boldsymbol{y}_k} \right)_{i,j} = \widehat{C}_{i,j}$$

Composition in **JDist**

$$D(i,j) = \sum_{k, \boldsymbol{y}(k)>0} \frac{\widehat{A}(i,k) \cdot \widehat{B}(k,j)}{\boldsymbol{y}(k)}$$

- Let (X, μ), (Y, ν), and (Z, ρ) be measure spaces
- Let $\theta : (X, \mu) \to (Y, \nu)$ and $\eta : (Y, \nu) \to (Z, \rho)$ be arrows

Composition in **JDist**

$$D(i,j) = \sum_{k,\boldsymbol{y}(k)>0} \frac{\widehat{A}(i,k) \cdot \widehat{B}(k,j)}{\boldsymbol{y}(k)}$$

- Let (X,μ), (Y,ν), and (Z,ρ) be measure spaces
- Let $\theta : (X,\mu) \to (Y,\nu)$ and $\eta : (Y,\nu) \to (Z,\rho)$ be arrows
- Define

$$(\eta \circ \theta)(A \times C) = \lim_{\mathcal{D}} \sum_{B \in \mathcal{D}, \nu(B)>0} \frac{\theta(A \times B) \cdot \eta(B \times C)}{\nu(B)}$$

Composition in **JDist**

$$D(i,j) = \sum_{k,\mathbf{y}(k)>0} \frac{\widehat{A}(i,k) \cdot \widehat{B}(k,j)}{\mathbf{y}(k)}$$

- Let (X,μ), (Y,ν), and (Z,ρ) be measure spaces
- Let $\theta : (X,\mu) \to (Y,\nu)$ and $\eta : (Y,\nu) \to (Z,\rho)$ be arrows
- Define

$$(\eta \circ \theta)(A \times C) = \lim_{\mathcal{D}} \sum_{B \in \mathcal{D}, \nu(B)>0} \frac{\theta(A \times B) \cdot \eta(B \times C)}{\nu(B)}$$

- Theorem: this limit is well-defined

Composition in **JDist**

$$D(i,j) = \sum_{k,\boldsymbol{y}(k)>0} \frac{\widehat{A}(i,k)\cdot\widehat{B}(k,j)}{\boldsymbol{y}(k)}$$

- Let (X,μ), (Y,ν), and (Z,ρ) be measure spaces
- Let $\theta : (X,\mu) \to (Y,\nu)$ and $\eta : (Y,\nu) \to (Z,\rho)$ be arrows
- Define

$$(\eta \circ \theta)(A \times C) = \lim_{\mathcal{D}} \sum_{B\in\mathcal{D},\nu(B)>0} \frac{\theta(A \times B)\cdot\eta(B \times C)}{\nu(B)}$$

- Theorem: this limit is well-defined
- Theorem: this extends to a measure on $X \times Z$

Composition in **JDist**

$$D(i,j) = \sum_{k,\boldsymbol{y}(k)>0} \frac{\widehat{A}(i,k)\cdot\widehat{B}(k,j)}{\boldsymbol{y}(k)}$$

- Let (X,μ), (Y,ν), and (Z,ρ) be measure spaces
- Let $\theta : (X,\mu) \to (Y,\nu)$ and $\eta : (Y,\nu) \to (Z,\rho)$ be arrows
- Define

$$(\eta\circ\theta)(A\times C) = \lim_{\mathcal{D}} \sum_{B\in\mathcal{D},\nu(B)>0} \frac{\theta(A\times B)\cdot\eta(B\times C)}{\nu(B)}$$

- Theorem: this limit is well-defined
- Theorem: this extends to a measure on $X\times Z$
- Theorem: **Krn** embeds faithfully into **JDist**

Discussion

- **JDist** has a dagger: $\theta^\dagger(B \times A) = \theta(A \times B)$
- Composition directly defined; no disintegration
- Measure spaces need not be standard Borel

Joint Distributions in Probabilistic Semantics

Dexter Kozen, *Cornell University*
Alexandra Silva, *Cornell University*
Erik Voogd, *University of Oslo*

MFPS 2023

Bloomington, 23 June 2023