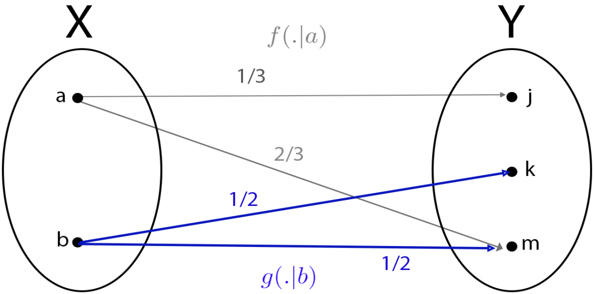
Markov Category : A 10 Minute Introduction

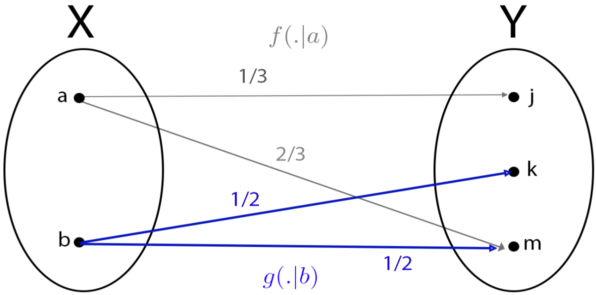
Mahdi Zamani

Nov 2023

Random Morphisms



Random Morphisms

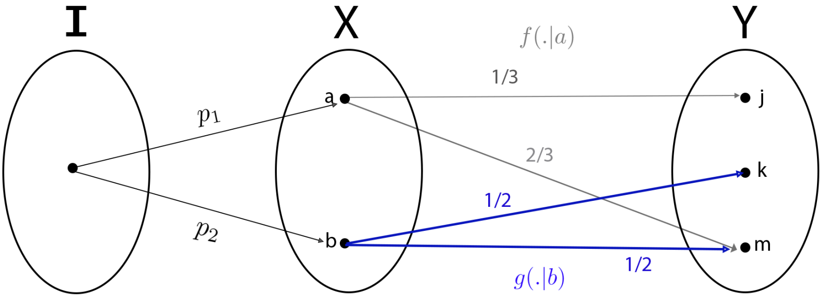


Example

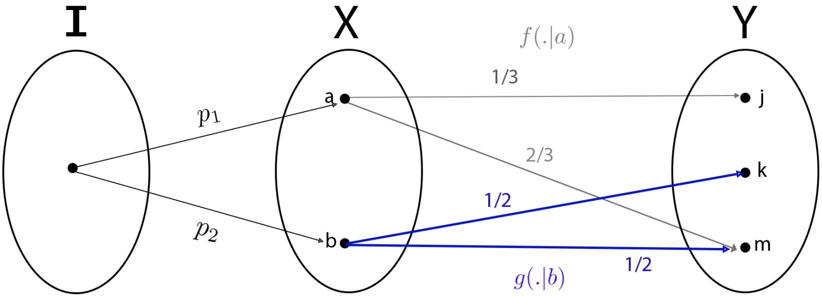
FinStoch Category, with measurable sets as objects, and random morphisms (Markov kernels for Stoch).

$$\begin{array}{cc} \swarrow & \begin{array}{cc} a & b \end{array} \\ \begin{array}{c} j \\ k \\ m \end{array} & \begin{pmatrix} 1/3 & 0 \\ 0 & 1/2 \\ 2/3 & 1/2 \end{pmatrix} \end{array} \qquad f(j|a) = 1/3$$

Probability Distribution



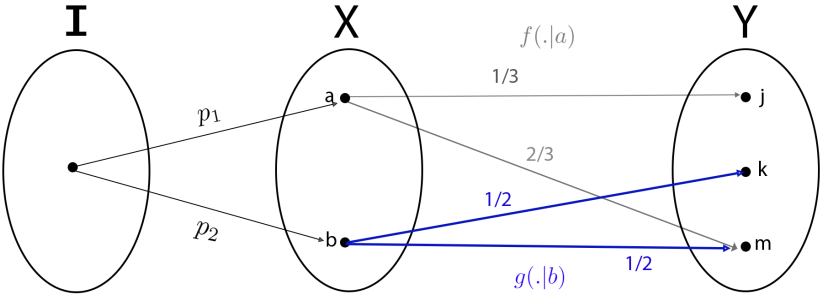
Probability Distribution



Notation

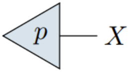
$$I \xrightarrow{P} X$$

Probability Distribution



Notation

$$I \xrightarrow{P} X$$

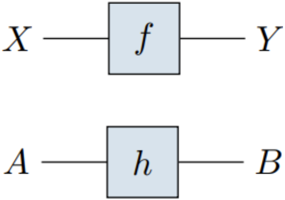


Products

$$X \; A \rightsquigarrow X \otimes A$$

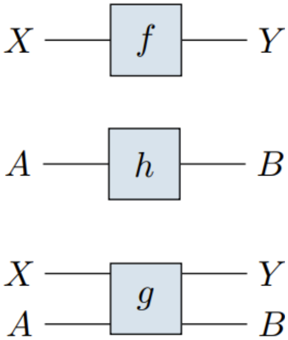
Products

$$X\ A \rightsquigarrow X \otimes A$$



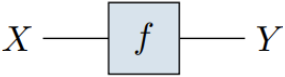
Products

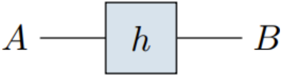
$$X\ A \rightsquigarrow X \otimes A$$

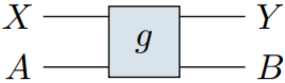


Products

$$X\ A \rightsquigarrow X \otimes A$$







$$g : X \otimes A \to Y \otimes B$$

Markov Category [2]

Markov Category [2]

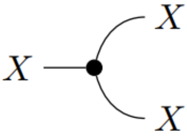
Definition

A Markov category C is a symmetric monoidal category supplied with copying and deleting morphisms on every object,

Markov Category [2]

Definition

A Markov category C is a symmetric monoidal category supplied with copying and deleting morphisms on every object,
copy : $X \to X \otimes X$

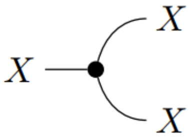


Markov Category [2]

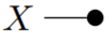
Definition

A Markov category C is a symmetric monoidal category supplied with copying and deleting morphisms on every object,

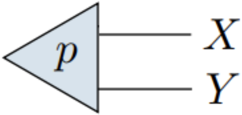
copy : $X \to X \otimes X$



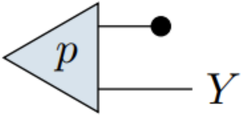
del : $X \to I$



Del

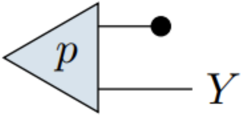


Del



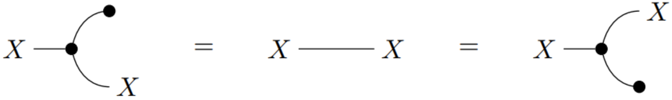
Del

$$\mathbb{P}_Y(y) = \sum_{x \in X} \mathbb{P}_{X,Y}(x, y)$$

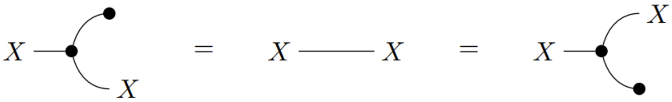


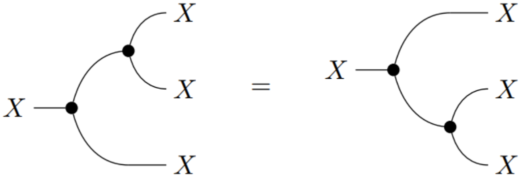
Commutative comonoid axioms

Commutative comonoid axioms

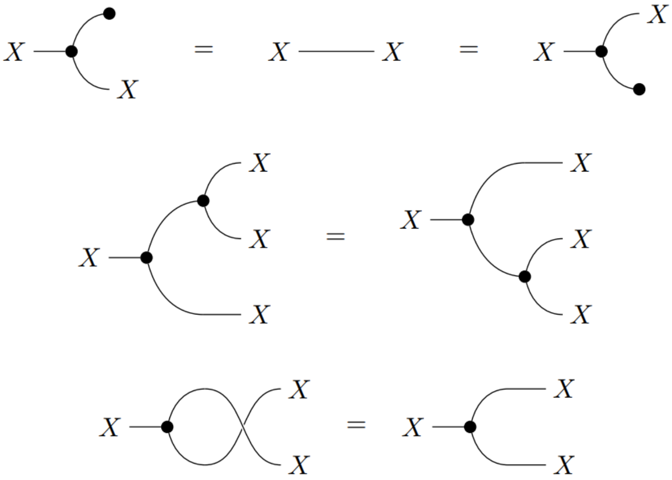


Commutative comonoid axioms





Commutative comonoid axioms



Naturality axiom

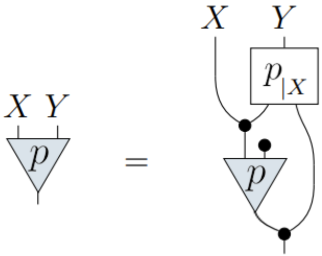
Conditional Distribution

Conditional Distribution

$p_{|X} : X \to Y$ is a conditional of p wrt to X

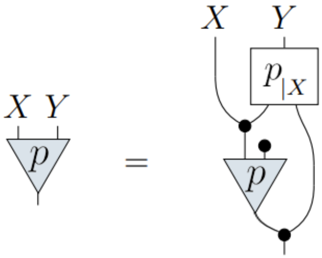
Conditional Distribution

$p_{|X} : X \to Y$ is a conditional of p wrt to X



Conditional Distribution

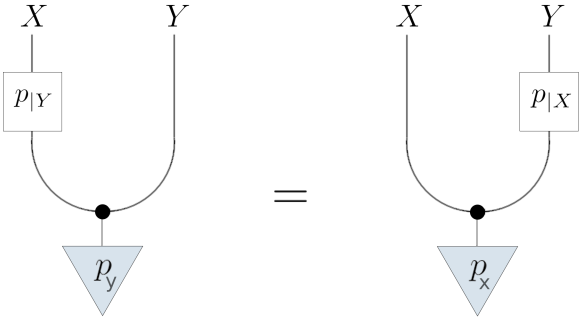
$p_{|X} : X \to Y$ is a conditional of p wrt to X



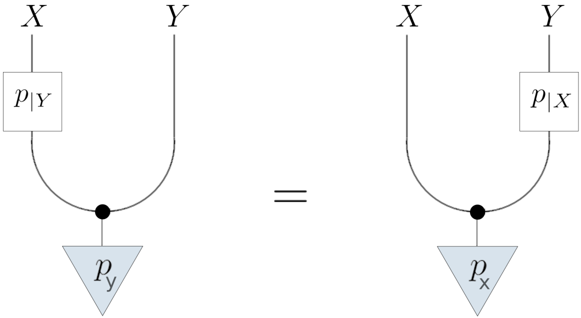
$$\mathbb{P}_{X,Y}(x,y) = \mathbb{P}_X(x)\mathbb{P}_{Y|X}(y|x)$$

Bayes

Bayes



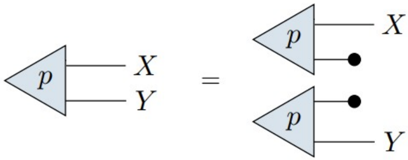
Bayes



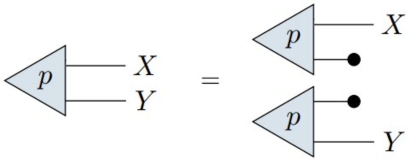
$$\mathbb{P}_Y(y)\mathbb{P}_{X|Y}(x|y) = \mathbb{P}_X(x)\mathbb{P}_{Y|X}(y|x)$$

(Conditional) Independence

(Conditional) Independence

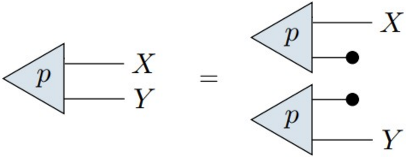


(Conditional) Independence

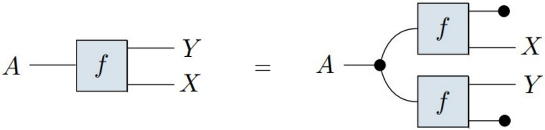


$$\mathbb{P}_{X,Y}(x, y) = \mathbb{P}_X(x)\mathbb{P}_Y(y)$$

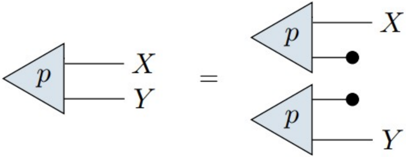
(Conditional) Independence



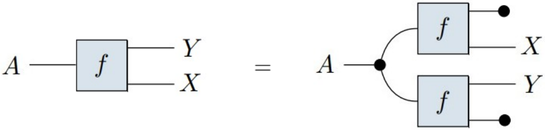
$$\mathbb{P}_{X,Y}(x, y) = \mathbb{P}_X(x)\mathbb{P}_Y(y)$$



(Conditional) Independence



$$\mathbb{P}_{X,Y}(x, y) = \mathbb{P}_X(x)\mathbb{P}_Y(y)$$



$$\mathbb{P}_{X,Y|\,A}(x, y|a) = \mathbb{P}_X(x|a)\mathbb{P}_Y(y|a)$$

Applications : Sufficient statistics [2]

Definition

Applications : Sufficient statistics [2]

Definition

A ”(parametric) statistical model” is a map $p : \Theta \to X$

Applications : Sufficient statistics [2]

Definition

A ”(parametric) statistical model” is a map $p : \Theta \to X$

A ”statistic” is a deterministic map $s : X \to V$

Applications : Sufficient statistics [2]

Definition

A ”(parametric) statistical model” is a map $p : \Theta \to X$

A ”statistic” is a deterministic map $s : X \to V$

A statistic is *sufficient* if $X \perp \Theta \mid V$

Applications : Sufficient statistics [2]

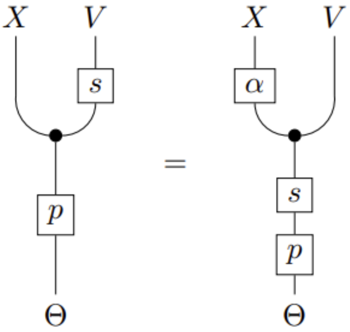
Definition

A ”(parametric) statistical model” is a map $p : \Theta \to X$

A ”statistic” is a deterministic map $s : X \to V$

A statistic is *sufficient* if $X \perp \Theta \mid V$

Which means that we have , $\alpha : V \to X$, s.t



Where to go

Where to go

- Categorical Bayesian network [1]

Where to go

- Categorical Bayesian network [1]
- De Finetti's Theorem [3]

Where to go

- Categorical Bayesian network [1]
- De Finetti's Theorem [3]
- Kolmogorov 0/1 law [4]

Where to go

- Categorical Bayesian network [1]
- De Finetti's Theorem [3]
- Kolmogorov 0/1 law [4]
- Markov categories and entropy (generalized divergence) [5]

Where to go

- Categorical Bayesian network [1]
- De Finetti's Theorem [3]
- Kolmogorov 0/1 law [4]
- Markov categories and entropy (generalized divergence) [5]
- Generalized Optimization [6]

Where to go

- Categorical Bayesian network [1]
- De Finetti's Theorem [3]
- Kolmogorov 0/1 law [4]
- Markov categories and entropy (generalized divergence) [5]
- Generalized Optimization [6]
- and much more! [7]

References

[1] Brendan Fong. *Causal Theories: A Categorical Perspective on Bayesian Networks*. 2013. eprint: arXiv:1301.6201.

[2] Tobias Fritz. “A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics”. In: (2019). DOI: 10.1016/j.aim.2020.107239. eprint: arXiv:1908.07021.

[3] Tobias Fritz, Tomáš Gonda, and Paolo Perrone. “De Finetti’s Theorem in Categorical Probability”. In: (2021). DOI: 10.31390/josa.2.4.06. eprint: arXiv:2105.02639.

[4] Tobias Fritz and Eigil Fjeldgren Rischel. “Infinite products and zero-one laws in categorical probability”. In: (2019). DOI: 10.32408/compositionality-2-3. eprint: arXiv:1912.02769.

[5] Paolo Perrone. *Markov Categories and Entropy*. 2022. eprint: arXiv:2212.11719.

[6] Dan Shiebler. *Generalized Optimization: A First Step Towards Category Theoretic Learning Theory*. 2021. eprint: arXiv:2109.10262.

[7] Dan Shiebler, Bruno Gavranović, and Paul Wilson. *Category Theory in Machine Learning*. 2021. eprint: arXiv:2106.07032.

Thank You

Questions