The Mondrian Kernel

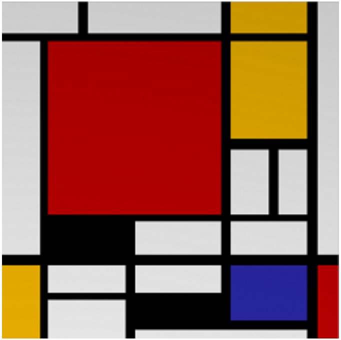
Matej Balog

Balaji Lakshminarayanan

Zoubin Ghahramani

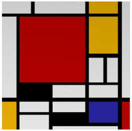
Daniel M. Roy

Yee Whye Teh



Mondrian
kernel

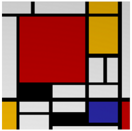
$$k(\cdot, \cdot) \ : \ \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$$



Mondrian
process

↓

Mondrian
kernel

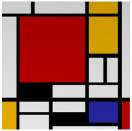


Mondrian process

↓

Mondrian kernel ⟶ Laplace kernel

$$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x} - \mathbf{x}'\|_1}$$

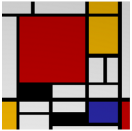


Mondrian process

Mondrian kernel

Laplace kernel

$$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x} - \mathbf{x}'\|_1}$$



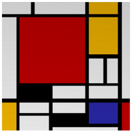
Mondrian
process

Mondrian
kernel

Laplace
kernel

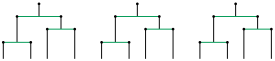
$$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$$

inverse width



Mondrian process

Mondrian forest

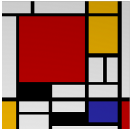


Mondrian kernel

Laplace kernel

$$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$$

inverse width (↑ λ)

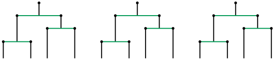


Mondrian process

Mondrian forest

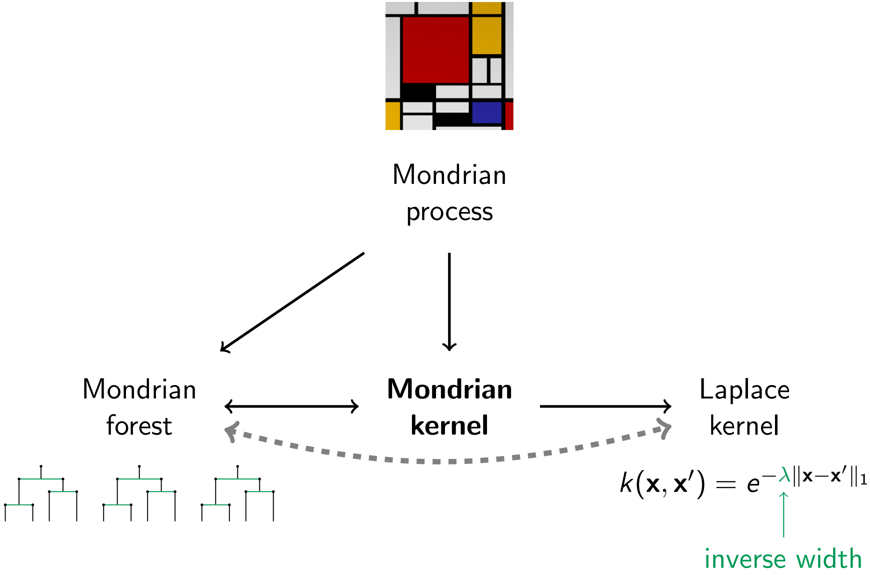
Mondrian kernel

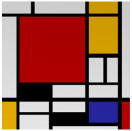
Laplace kernel



$$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$$

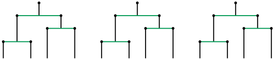
inverse width


Mondrian process
Mondrian forest
Mondrian kernel
Laplace kernel
$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$
inverse width



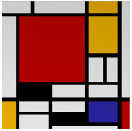
1 Mondrian process

Mondrian forest ⟷ **Mondrian kernel** ⟶ Laplace kernel



$$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$$

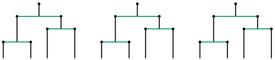
λ: inverse width



1 Mondrian process

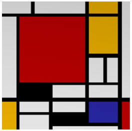
2 **Mondrian kernel**

Mondrian forest ⟷ **Mondrian kernel** ⟶ Laplace kernel



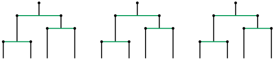
$$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$$

λ: inverse width



1 Mondrian process

Mondrian forest



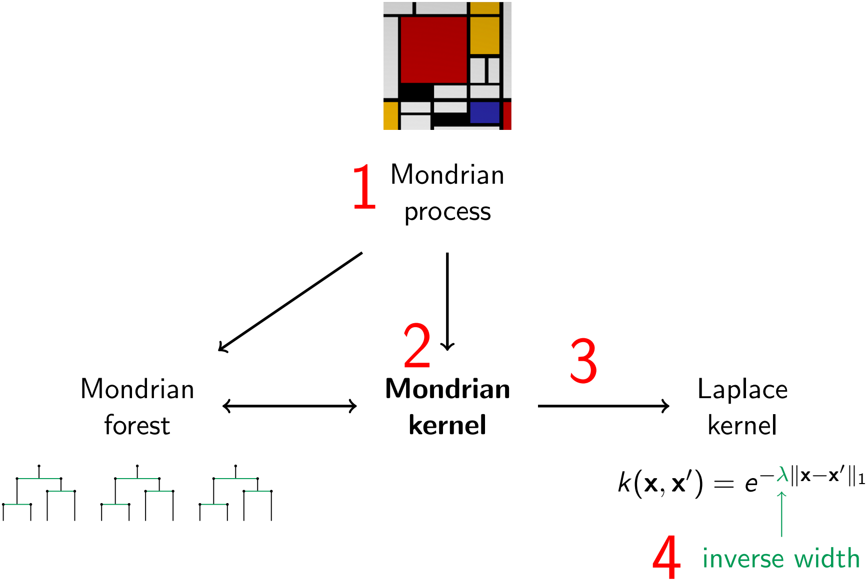
2 **Mondrian kernel**

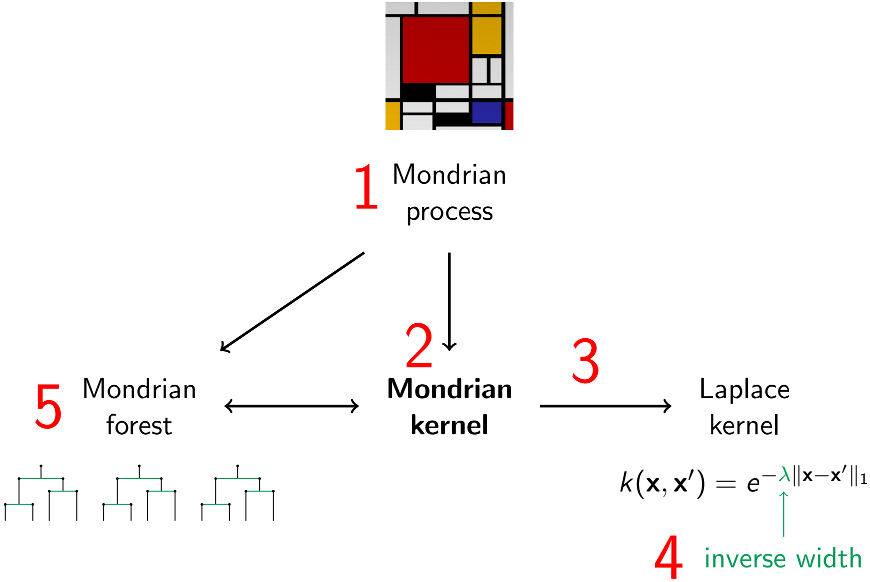
3

Laplace kernel

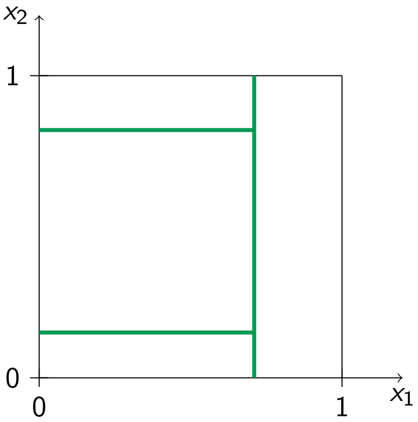
$$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x} - \mathbf{x}'\|_1}$$

λ: inverse width

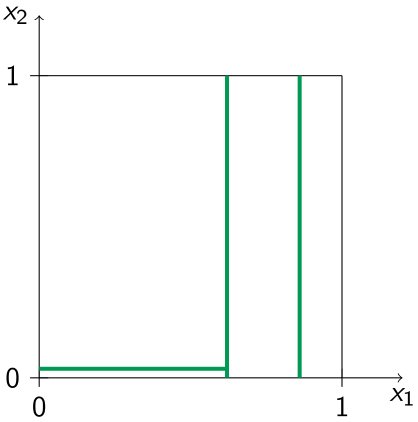

1 Mondrian process
2 Mondrian kernel
3
Mondrian forest
Laplace kernel
$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$
4 inverse width


1 Mondrian process
2 Mondrian kernel
3
Laplace kernel
$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$
4 inverse width
5 Mondrian forest

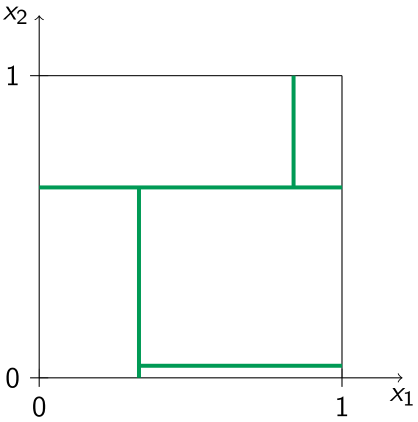
Mondrian Process



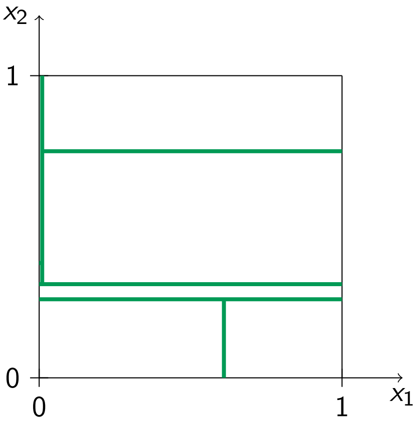
Mondrian Process



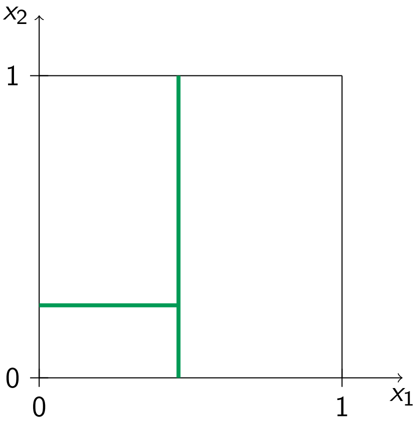
Mondrian Process



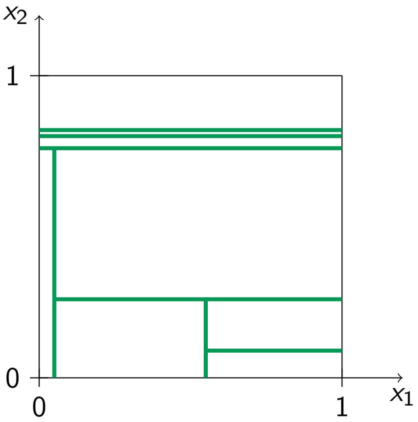
Mondrian Process



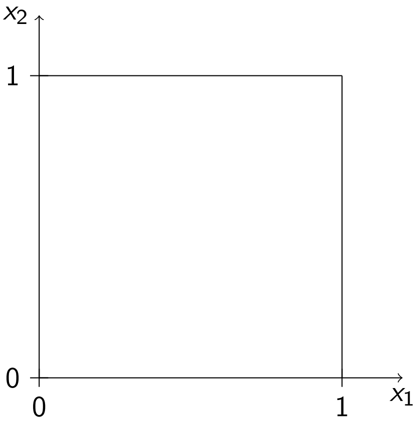
Mondrian Process



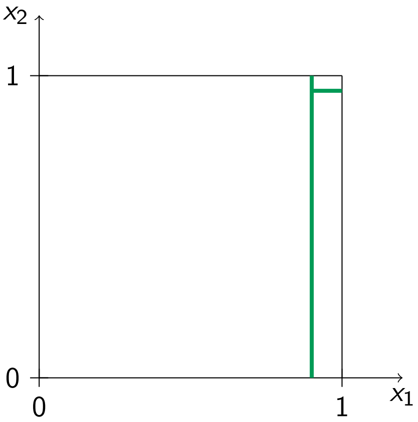
Mondrian Process



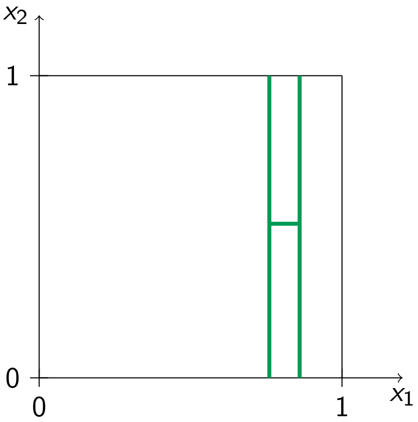
Mondrian Process



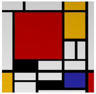
Mondrian Process

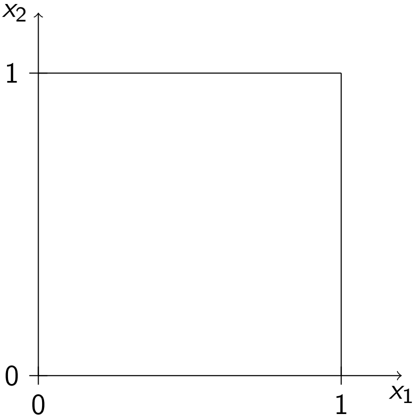


Mondrian Process

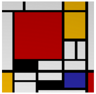


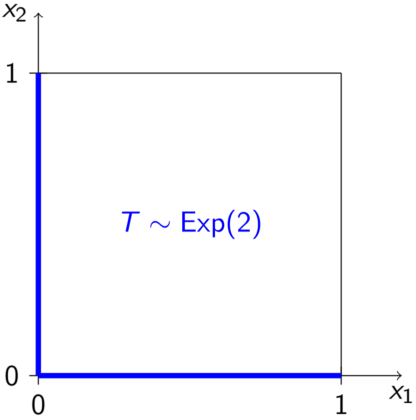
Mondrian Process



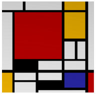


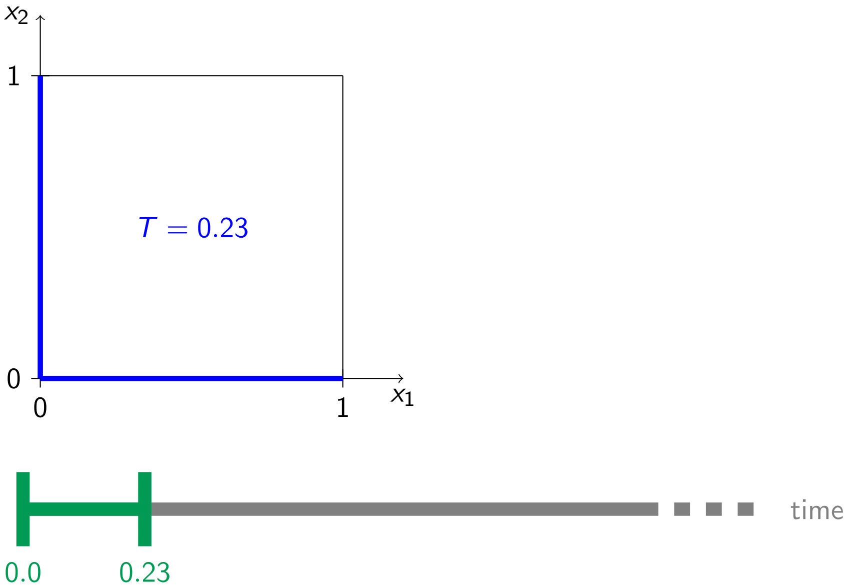
Mondrian Process



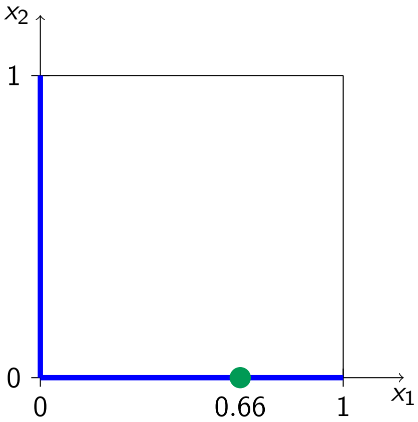


Mondrian Process

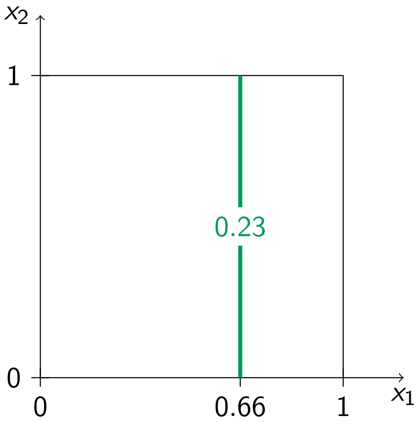




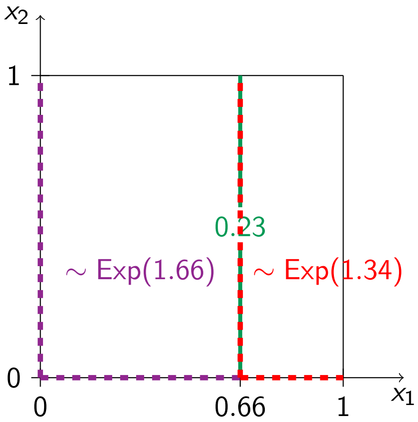
Mondrian Process



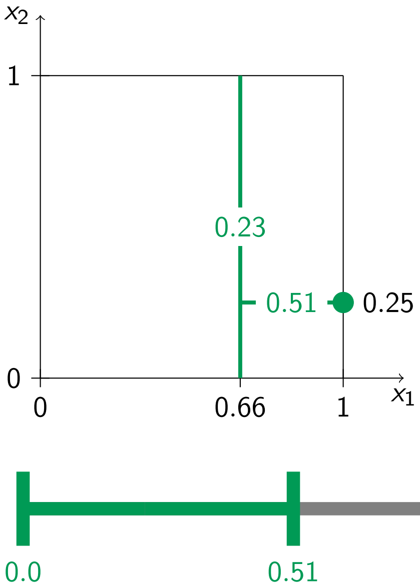
Mondrian Process



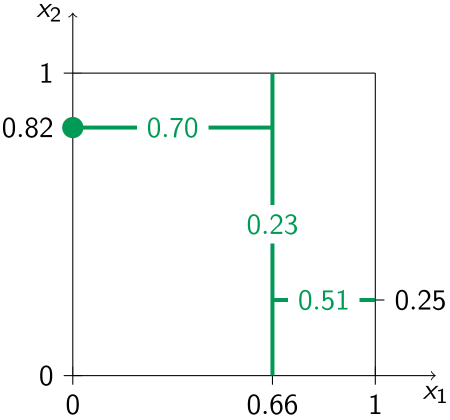
Mondrian Process



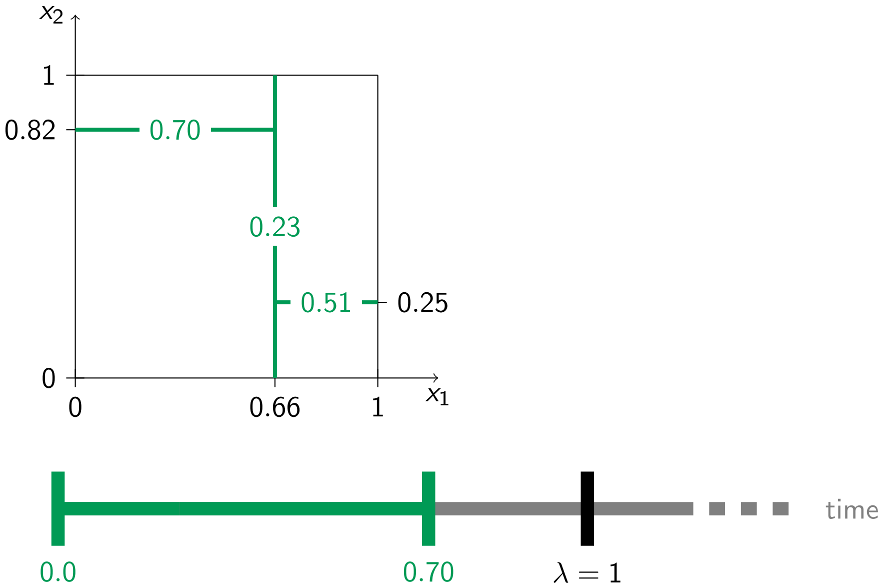
Mondrian Process



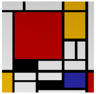
Mondrian Process

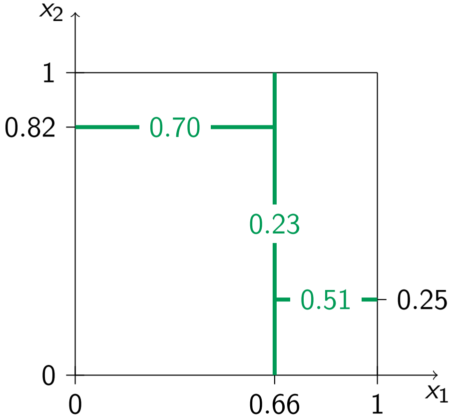


Mondrian Process

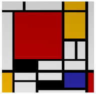


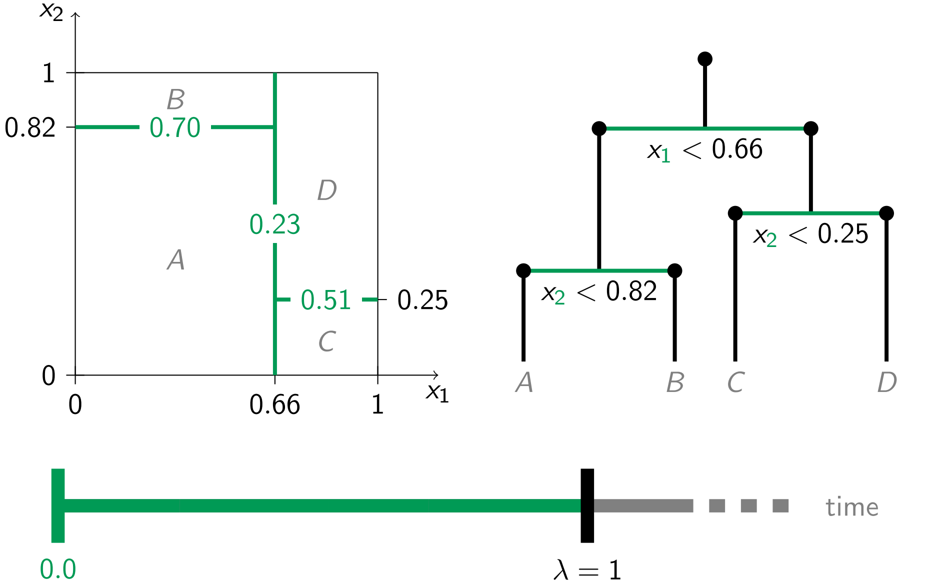
Mondrian Process



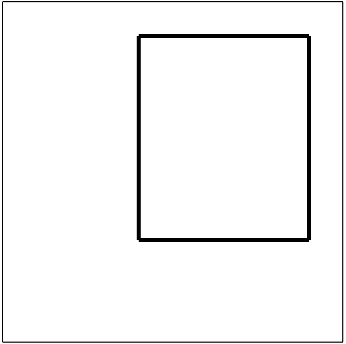


Mondrian Process

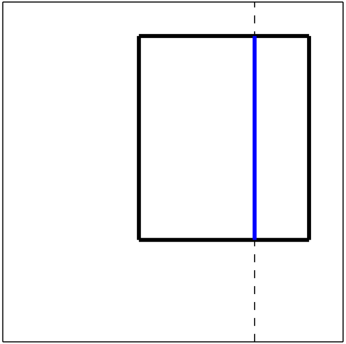




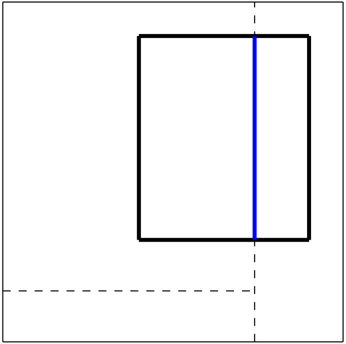
Mondrian Process – projectivity



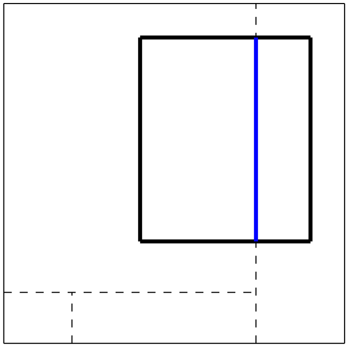
Mondrian Process – projectivity



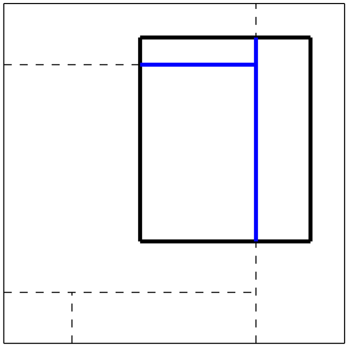
Mondrian Process – projectivity



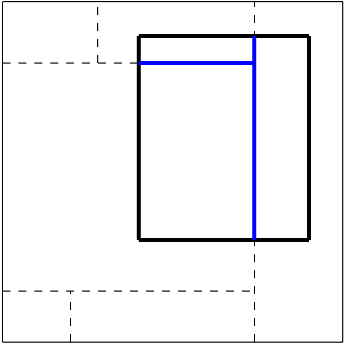
Mondrian Process – projectivity



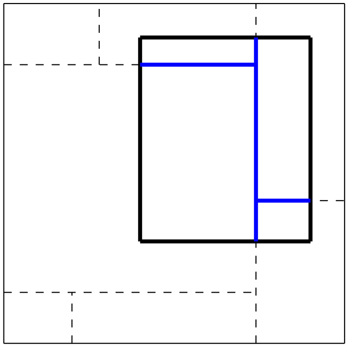
Mondrian Process – projectivity



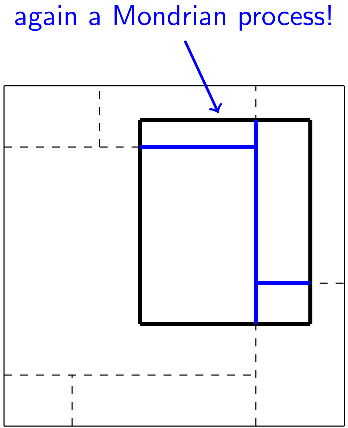
Mondrian Process – projectivity



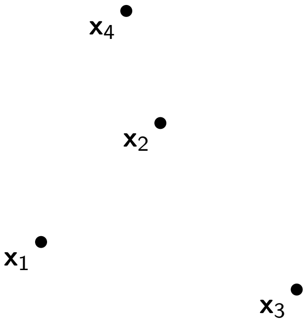
Mondrian Process – projectivity



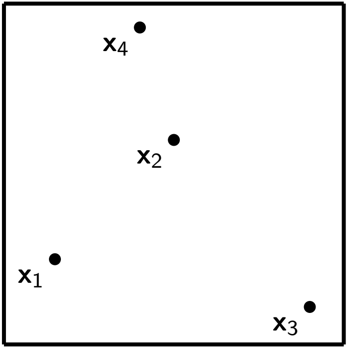
Mondrian Process – projectivity



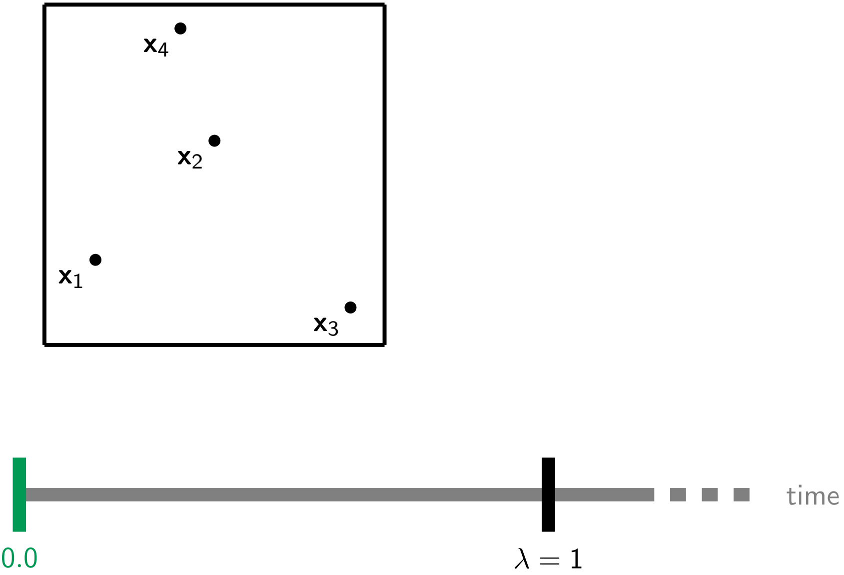
Mondrian kernel



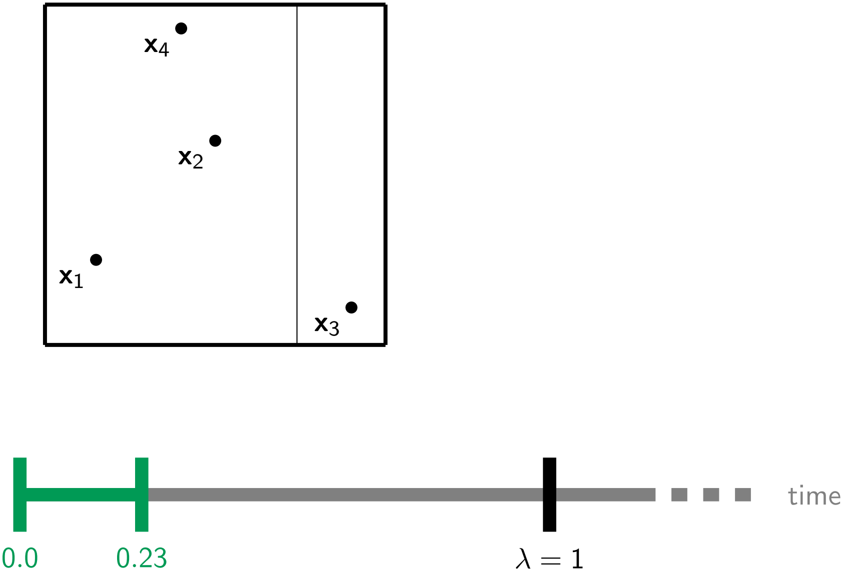
Mondrian kernel



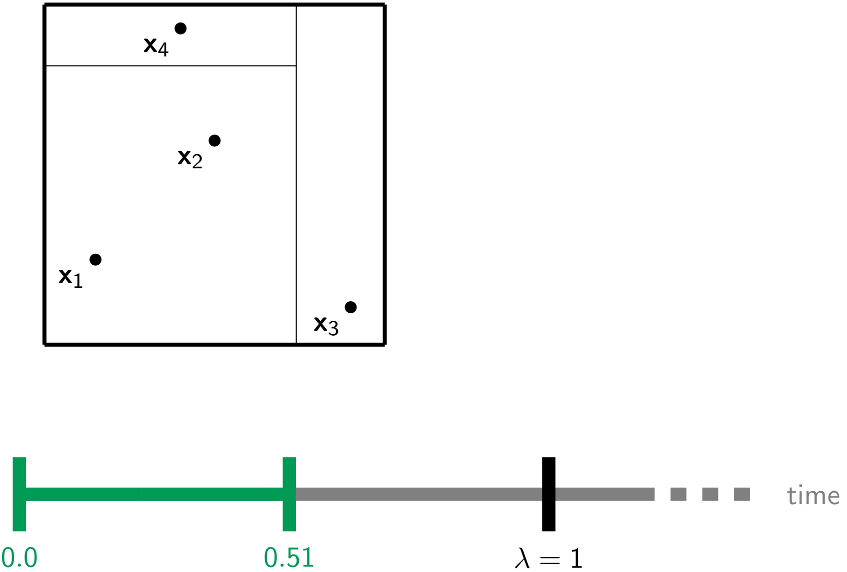
Mondrian kernel



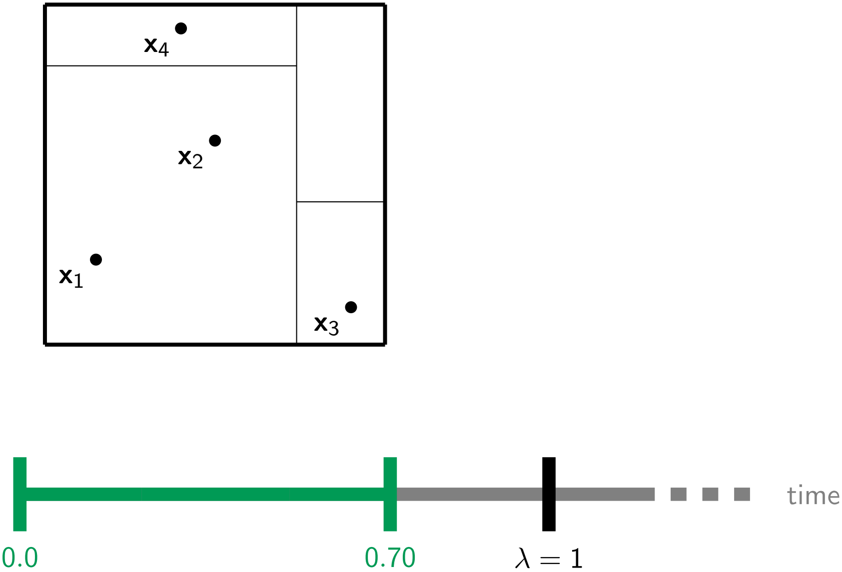
Mondrian kernel



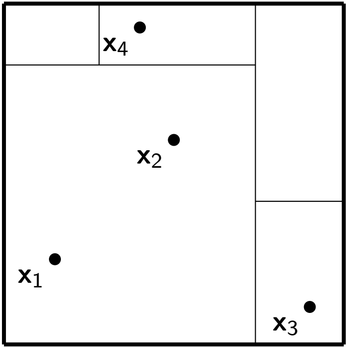
Mondrian kernel



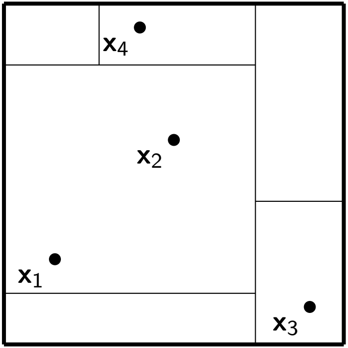
Mondrian kernel



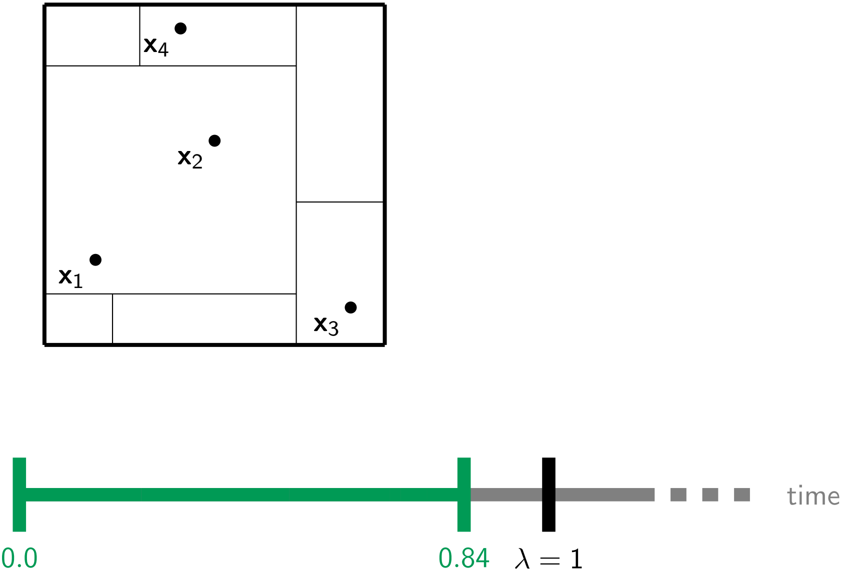
Mondrian kernel



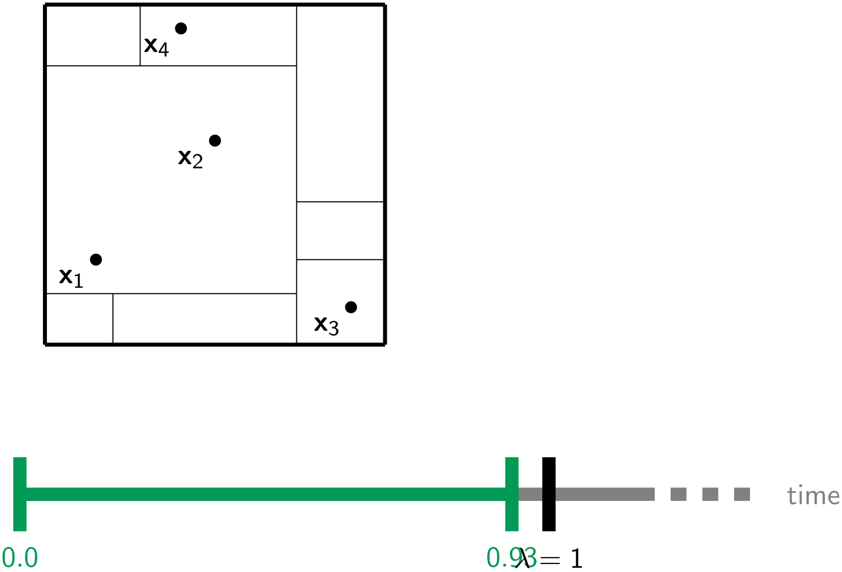
Mondrian kernel



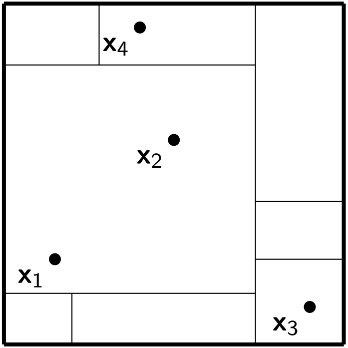
Mondrian kernel



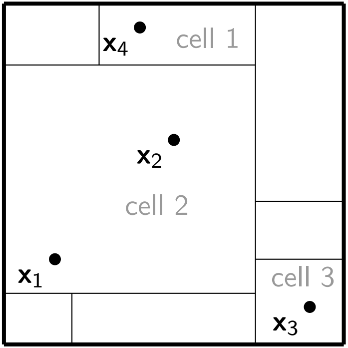
Mondrian kernel



Mondrian kernel

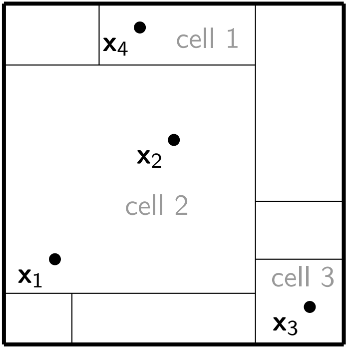


Mondrian kernel



$\mathbf{x}$	$\phi(\mathbf{x})$
$\mathbf{x}_1$	[0 1 0]
$\mathbf{x}_2$	[0 1 0]
$\mathbf{x}_3$	[0 0 1]
$\mathbf{x}_4$	[1 0 0]

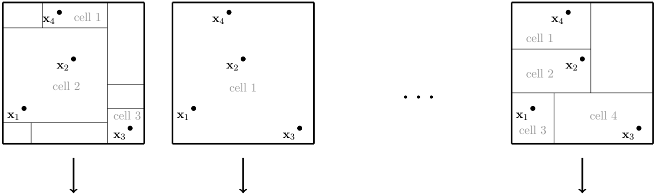
Mondrian kernel (of order 1)



$\mathbf{x}$	$\phi(\mathbf{x})$
$\mathbf{x}_1$	[0 1 0]
$\mathbf{x}_2$	[0 1 0]
$\mathbf{x}_3$	[0 0 1]
$\mathbf{x}_4$	[1 0 0]

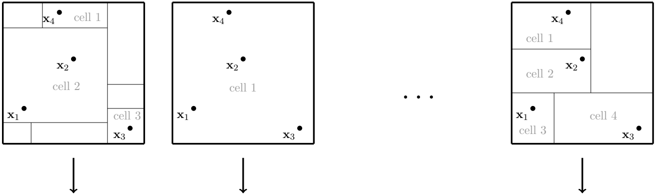
$$k_1(\mathbf{x}, \mathbf{x}') := \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle = \begin{cases} 1 & \text{if } \mathbf{x}, \mathbf{x}' \text{ in the same cell} \\ 0 & \text{otherwise} \end{cases}$$

Mondrian kernel (of order M)



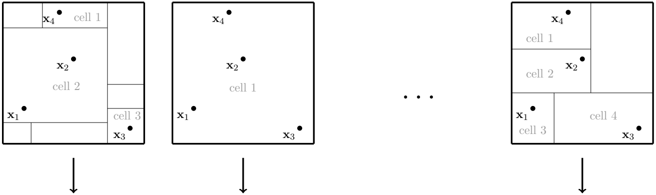
[0 1 0]	[1]	$\cdots$	[0 0 1 0]
[0 1 0]	[1]	$\cdots$	[0 1 0 0]
[0 0 1]	[1]	$\cdots$	[0 0 0 1]
[1 0 0]	[1]	$\cdots$	[1 0 0 0]

Mondrian kernel (of order M)



$$\phi(\mathbf{x}_1) = \frac{1}{\sqrt{M}} [\quad 0\ 1\ 0 \quad 1 \quad \cdots \quad 0\ 0\ 1\ 0 \quad]$$
$$\phi(\mathbf{x}_2) = \frac{1}{\sqrt{M}} [\quad 0\ 1\ 0 \quad 1 \quad \cdots \quad 0\ 1\ 0\ 0 \quad]$$
$$\phi(\mathbf{x}_3) = \frac{1}{\sqrt{M}} [\quad 0\ 0\ 1 \quad 1 \quad \cdots \quad 0\ 0\ 0\ 1 \quad]$$
$$\phi(\mathbf{x}_4) = \frac{1}{\sqrt{M}} [\quad 1\ 0\ 0 \quad 1 \quad \cdots \quad 1\ 0\ 0\ 0 \quad]$$

Mondrian kernel (of order M)



$$\phi(\mathbf{x}_1) = \frac{1}{\sqrt{M}} [\quad 0\ 1\ 0 \qquad 1 \qquad \cdots \qquad 0\ 0\ 1\ 0 \quad]$$

$$\phi(\mathbf{x}_2) = \frac{1}{\sqrt{M}} [\quad 0\ 1\ 0 \qquad 1 \qquad \cdots \qquad 0\ 1\ 0\ 0 \quad]$$

$$\phi(\mathbf{x}_3) = \frac{1}{\sqrt{M}} [\quad 0\ 0\ 1 \qquad 1 \qquad \cdots \qquad 0\ 0\ 0\ 1 \quad]$$

$$\phi(\mathbf{x}_4) = \frac{1}{\sqrt{M}} [\quad 1\ 0\ 0 \qquad 1 \qquad \cdots \qquad 1\ 0\ 0\ 0 \quad]$$

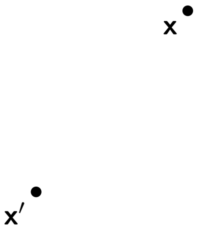
$$k_M(\mathbf{x}, \mathbf{x}') := \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x}, \mathbf{x}' \text{ in same cell of } m\text{-th partition}\}}$$

Mondrian-Laplace connection

$$k_M(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x}, \mathbf{x}' \text{ in same cell } | m\}} \xrightarrow{M \to \infty} \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \text{ in same cell }\right]$$

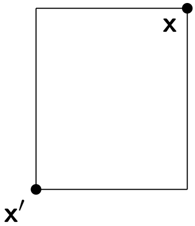
Mondrian-Laplace connection

$$k_M(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x}, \mathbf{x}' \text{ in same cell } | m\}} \xrightarrow{M \to \infty} \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \text{ in same cell }\right]$$



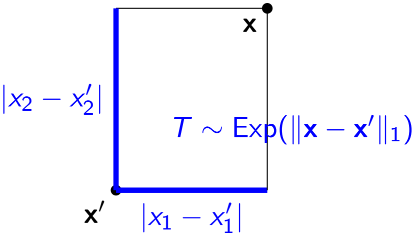
Mondrian-Laplace connection

$$k_M(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x}, \mathbf{x}' \text{ in same cell } | m\}} \xrightarrow{M \to \infty} \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \text{ in same cell }\right]$$



Mondrian-Laplace connection

$$k_M(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x}, \mathbf{x}' \text{ in same cell } | m\}} \xrightarrow{M \to \infty} \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \text{ in same cell }\right]$$

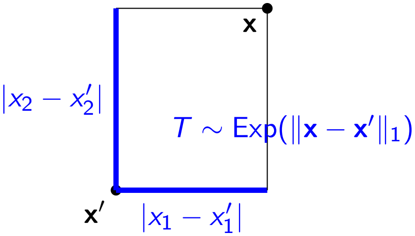


Mondrian-Laplace connection

$$k_M(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x},\mathbf{x}' \text{ in same cell } | m\}} \xrightarrow{M \to \infty} \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \text{ in same cell }\right]$$

$$\parallel$$

$$\mathbb{P}\left(T > \lambda\right) \text{ where } T \sim \operatorname{Exp}\left(\|\mathbf{x} - \mathbf{x}'\|_1\right)$$



Mondrian-Laplace connection

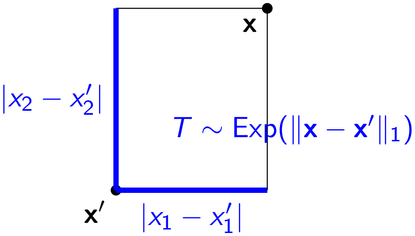
$$k_M(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x},\mathbf{x}' \text{ in same cell } | m\}} \xrightarrow{M \to \infty} \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \text{ in same cell }\right]$$

$$||$$

$$\mathbb{P}\left(T > \lambda\right) \text{ where } T \sim \text{Exp}\left(\|\mathbf{x} - \mathbf{x}'\|_1\right)$$

$$||$$

$$\exp\left(-\lambda \|\mathbf{x} - \mathbf{x}'\|_1\right)$$



Mondrian-Laplace connection

$$k_M(\mathbf{x}, \mathbf{x}') = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x}, \mathbf{x}' \text{ in same cell } | m\}} \xrightarrow{M \to \infty} \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \text{ in same cell }\right]$$

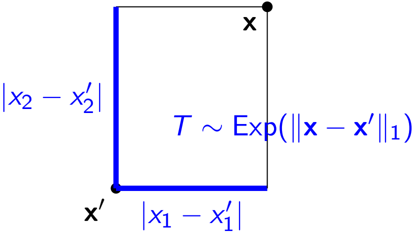
$$||$$

$$\mathbb{P}\left(T > \lambda\right) \text{ where } T \sim \text{Exp}\left(\|\mathbf{x} - \mathbf{x}'\|_1\right)$$

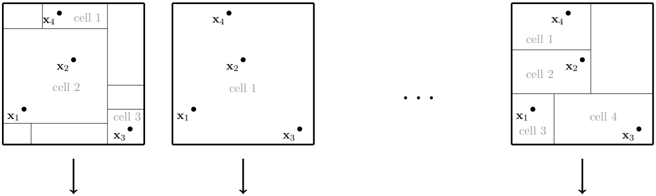
$$||$$

$$\exp\left(-\lambda \|\mathbf{x} - \mathbf{x}'\|_1\right)$$

Mondrian process lifetime



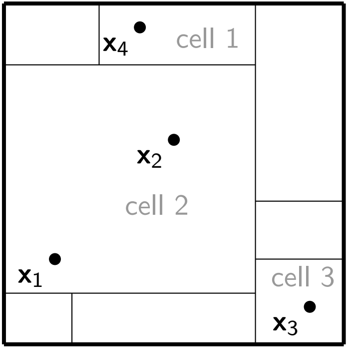
Mondrian kernel (of order M)



$$\phi(\mathbf{x}_1) = \frac{1}{\sqrt{M}} [\quad 0\ 1\ 0 \qquad 1 \qquad \cdots \qquad 0\ 0\ 1\ 0 \quad]$$
$$\phi(\mathbf{x}_2) = \frac{1}{\sqrt{M}} [\quad 0\ 1\ 0 \qquad 1 \qquad \cdots \qquad 0\ 1\ 0\ 0 \quad]$$
$$\phi(\mathbf{x}_3) = \frac{1}{\sqrt{M}} [\quad 0\ 0\ 1 \qquad 1 \qquad \cdots \qquad 0\ 0\ 0\ 1 \quad]$$
$$\phi(\mathbf{x}_4) = \frac{1}{\sqrt{M}} [\quad 1\ 0\ 0 \qquad 1 \qquad \cdots \qquad 1\ 0\ 0\ 0 \quad]$$

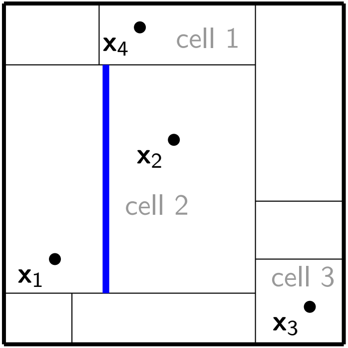
$$k_M(\mathbf{x}, \mathbf{x}') := \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\mathbf{x},\mathbf{x}' \text{ in same cell of } m\text{-th partition}\}}$$
$$\xrightarrow{M \to \infty} \exp\left(-\lambda \|\mathbf{x} - \mathbf{x}'\|_1\right)$$

Kernel width selection



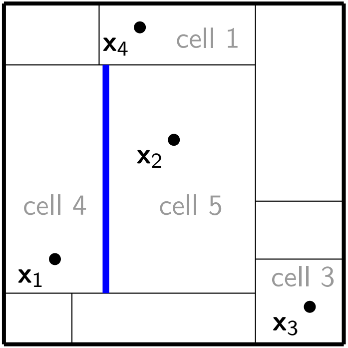
$\mathbf{x}$	$\phi(\mathbf{x})$
$\mathbf{x}_1$	[0 1 0]
$\mathbf{x}_2$	[0 1 0]
$\mathbf{x}_3$	[0 0 1]
$\mathbf{x}_4$	[1 0 0]

Kernel width selection



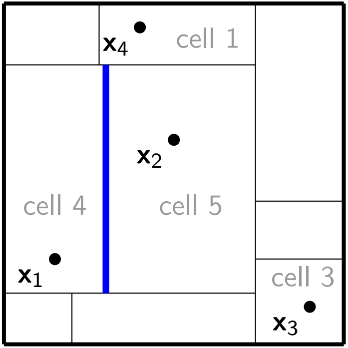
$\mathbf{x}$	$\phi(\mathbf{x})$
$\mathbf{x}_1$	[0 1 0]
$\mathbf{x}_2$	[0 1 0]
$\mathbf{x}_3$	[0 0 1]
$\mathbf{x}_4$	[1 0 0]

Kernel width selection



$\mathbf{x}$	$\phi(\mathbf{x})$
$\mathbf{x}_1$	[0 *1* 0 **1** **0**]
$\mathbf{x}_2$	[0 *1* 0 **0** **1**]
$\mathbf{x}_3$	[0 *0* 1 **0** **0**]
$\mathbf{x}_4$	[1 *0* 0 **0** **0**]

Kernel width selection



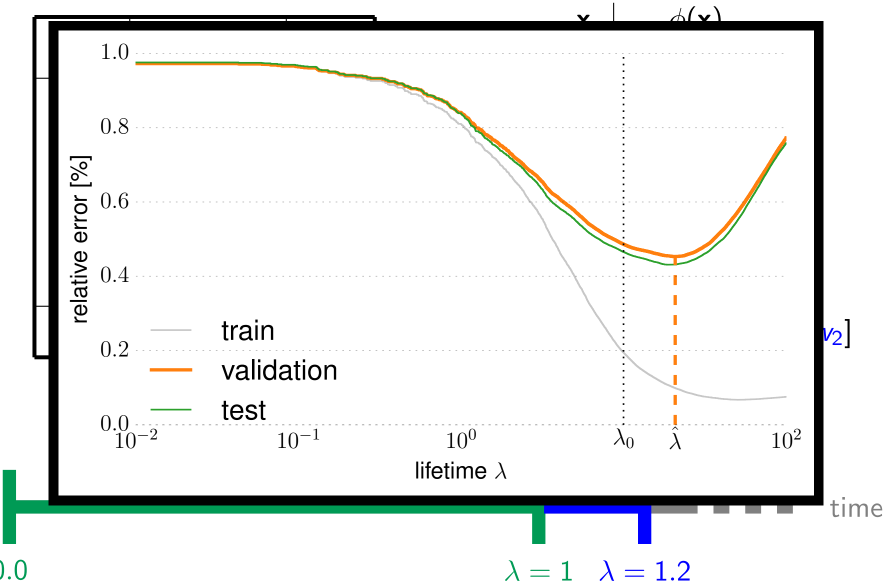
$\mathbf{x}$	$\phi(\mathbf{x})$
$\mathbf{x}_1$	[0 *1* 0 **1** **0**]
$\mathbf{x}_2$	[0 *1* 0 **0** **1**]
$\mathbf{x}_3$	[0 *0* 1 **0** **0**]
$\mathbf{x}_4$	[1 *0* 0 **0** **0**]

$$\mathbf{w}_{\text{opt}} = [w_1 \;\; w_2 \;\; w_3]$$

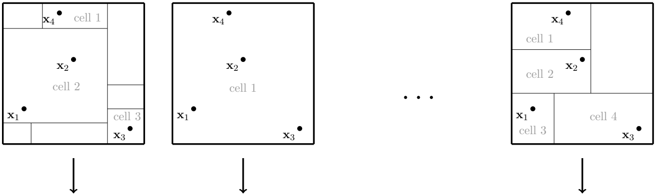
$$\downarrow$$

$$\mathbf{w}_{\text{init}} = [w_1 \;\; \cancel{w_2} \;\; w_3 \;\; w_2 \;\; w_2]$$

Kernel width selection



Mondrian kernel vs Mondrian forest

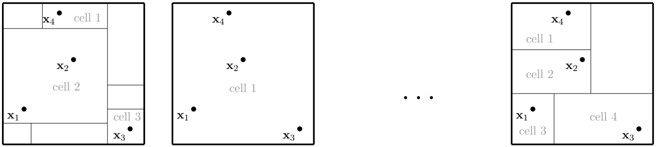


$$\phi(\mathbf{x}_1) = \frac{1}{\sqrt{M}} \big[\quad 0\,1\,0 \quad 1 \quad \cdots \quad 0\,0\,0\,1 \quad \big]$$
$$\phi(\mathbf{x}_2) = \frac{1}{\sqrt{M}} \big[\quad 0\,1\,0 \quad 1 \quad \cdots \quad 0\,1\,0\,0 \quad \big]$$
$$\phi(\mathbf{x}_3) = \frac{1}{\sqrt{M}} \big[\quad 0\,0\,1 \quad 1 \quad \cdots \quad 0\,0\,1\,0 \quad \big]$$
$$\phi(\mathbf{x}_4) = \frac{1}{\sqrt{M}} \big[\quad 1\,0\,0 \quad 1 \quad \cdots \quad 1\,0\,0\,0 \quad \big]$$

kernel

feature weights fit jointly: $\min_{\mathbf{w}} \|\mathbf{\Phi}\mathbf{w} - \mathbf{y}\|_2^2$

Mondrian kernel vs Mondrian forest



$$\phi(\mathbf{x}_1) = \frac{1}{\sqrt{M}}\left[\; 0\,1\,0 \quad 1 \quad \cdots \quad 0\,0\,0\,1 \;\right]$$

$$\phi(\mathbf{x}_2) = \frac{1}{\sqrt{M}}\left[\; 0\,1\,0 \quad 1 \quad \cdots \quad 0\,1\,0\,0 \;\right]$$

$$\phi(\mathbf{x}_3) = \frac{1}{\sqrt{M}}\left[\; 0\,0\,1 \quad 1 \quad \cdots \quad 0\,0\,1\,0 \;\right]$$

$$\phi(\mathbf{x}_4) = \frac{1}{\sqrt{M}}\left[\; 1\,0\,0 \quad 1 \quad \cdots \quad 1\,0\,0\,0 \;\right]$$

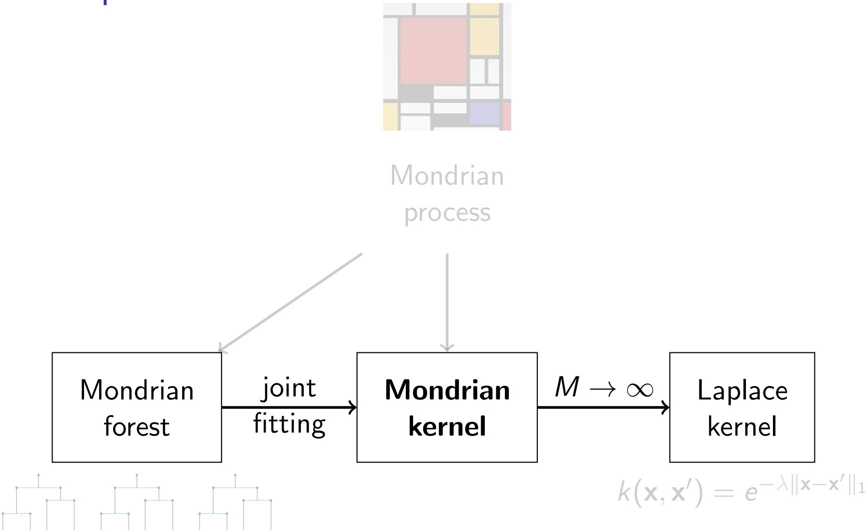
kernel

feature weights fit jointly: $\min_{\mathbf{w}} \|\mathbf{\Phi}\mathbf{w} - \mathbf{y}\|_2^2$

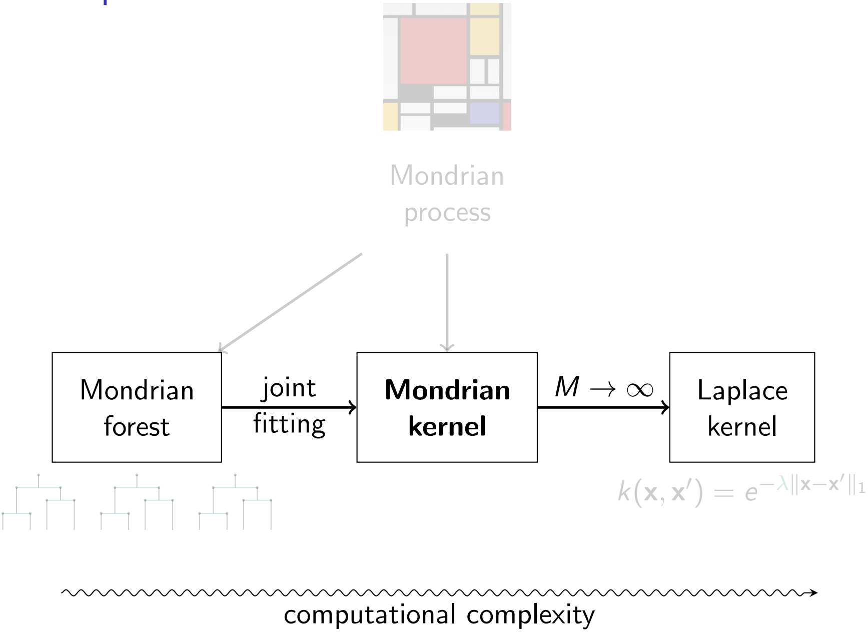
forest

fit trees independently

Landscape


Mondrian process
Mondrian forest
joint fitting
Mondrian kernel
$M \to \infty$
Laplace kernel
$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$

Landscape


Mondrian process
Mondrian forest
joint fitting
Mondrian kernel
$M \to \infty$
Laplace kernel
$k(\mathbf{x}, \mathbf{x}') = e^{-\lambda \|\mathbf{x}-\mathbf{x}'\|_1}$
computational complexity

Thank you

Matej Balog
Dept. of Engineering
University of Cambridge
& MPI Tübingen

Balaji Lakshminarayanan
Gatsby Unit
University College London

Zoubin Ghahramani
Dept. of Engineering
University of Cambridge

Daniel M. Roy
Dept. of Statistical Sciences
University of Toronto

Yee Whye Teh
Dept. of Statistics
University of Oxford

References

- Matej Balog et al. "The Mondrian Kernel". In: *32nd Conference on Uncertainty in Artificial Intelligence (UAI)*. 2016.
- Daniel M. Roy and Yee Whye Teh. "The Mondrian process". In: *Adv. Neural Information Proc. Systems (NIPS)*. 2009.
- Daniel M. Roy. "Computability, inference and modeling in probabilistic programming". PhD thesis. Massachusetts Institute of Technology, 2011.
- Ali Rahimi and Benjamin Recht. "Random features for large-scale kernel machines". In: *Adv. Neural Information Proc. Systems (NIPS)*. 2007.
- Balaji Lakshminarayanan, Daniel M. Roy, and Yee Whye Teh. "Mondrian forests: Efficient online random forests". In: *Adv. Neural Information Proc. Systems (NIPS)*. 2014.