

# Statistical Regular Pavings in Bayesian Nonparametric Density Estimation

**Raazesh Sainudiin**

joint work with: Jennifer Harlow, Dominic Lee, Carey Priebe, Gloria Teng and Warwick Tucker

Department of Mathematics and Statistics, University of Canterbury,  
Christchurch, New Zealand

**March 4, 2014,**

Scalable Bayesian Computation Workshop, Banff, Canada, March 4, 2014

## Massive Metric Data Streams

**Air Traffic Examples** (Teng, Kuhn and S., *Jnl. Aerospace Comp., Inf. & Commun.*, [acc.] 2012)

**Synthetic Examples** (Teng, Harlow, Lee and S., *ACM Trans. Mod. & Comp. Sim.*, [r. 2] 2012)

## Theory of Regular Pavings (RPs)

## Theory of Real Mapped Regular Pavings ( $\mathbb{R}$ -MRPs)

## Statistical Regular Pavings (SRPs)

## Adaptive Histograms

S.E.B. Priority Queue –  $L_1$  Consistent Initialization

## Smoothing by Averaging

## Posterior Expectation over Histograms in $\mathbb{S}_{0:\infty}$

## Examples - good, bad and ugly

## Conclusions and References

# Massive Metric Data Streams – Introduction

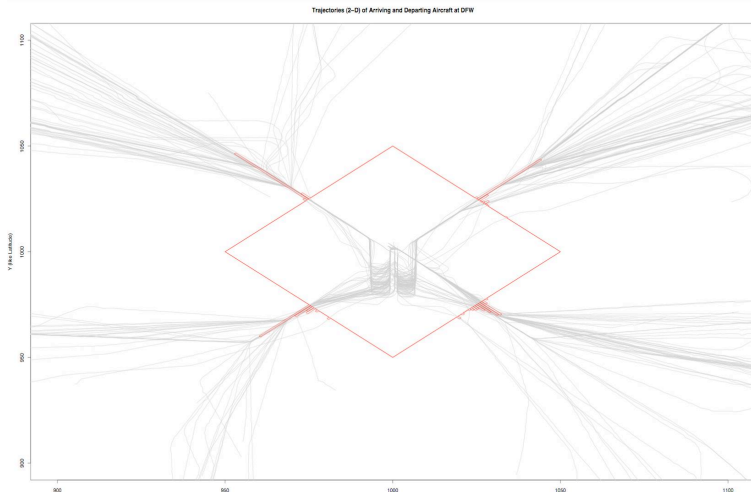
- ▶ A massive metric data stream is:

$$\dots, X_{-3}, X_{-2}, X_{-1}, X_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_n, X_{n+1}, \dots \sim F, \quad X_i \in \mathbb{R}^d.$$

- ▶ Large Dimension:  $1 \leq d \leq 1000$
- ▶ Huge Observations:  $10^6 \leq n \leq 10^{10}$
- ▶ Need an **efficient** and **sufficient** multi-dimensional metric data-structure for non-parametric inference that is capable of:
  1.  $L_1$ -consistent density estimation – adaptive histograms
  2. Extend arithmetic over a dense class of histograms with different partitions

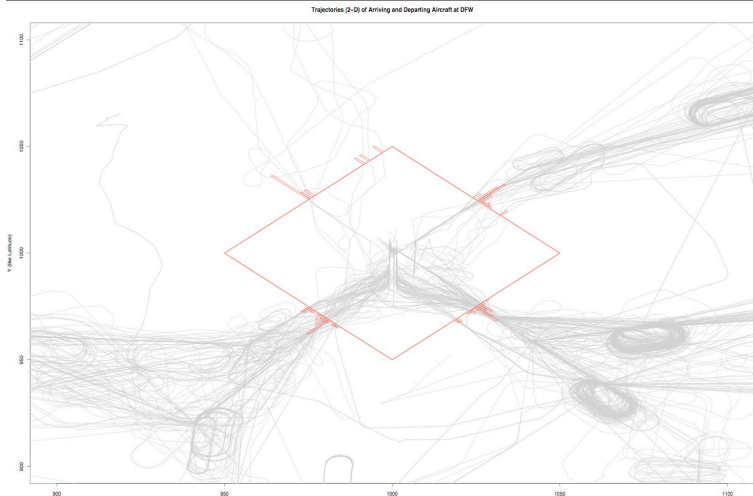
# Massive Metric Data Streams – Air Traffic Example

## On a Sunny Day



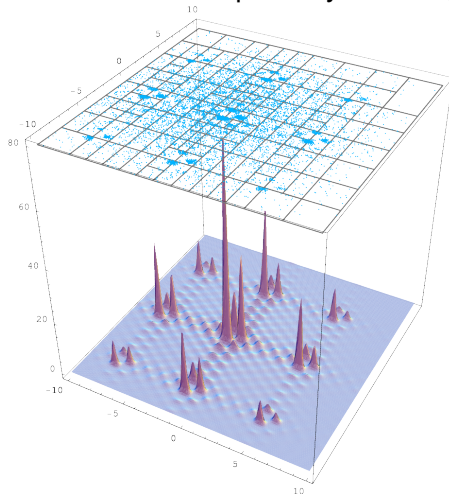
# Massive Metric Data Streams – Air Traffic Example

## On a Rainy Day



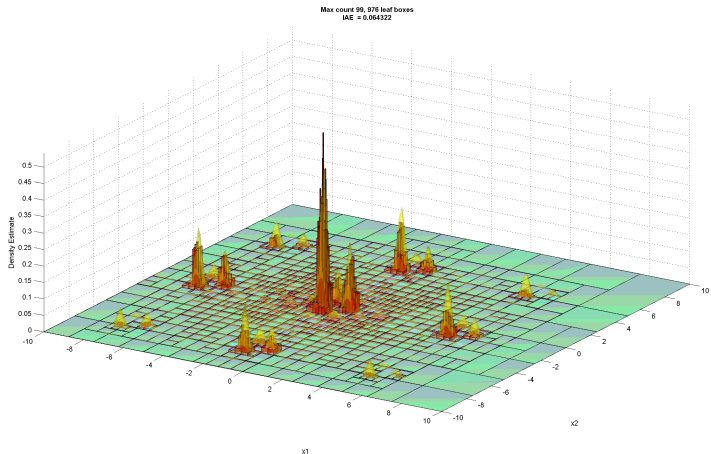
# Massive Metric Data Streams – Synthetic Examples

Take millions of **realizations** of a possibly ‘challenging’ density



# Massive Metric Data Streams – Synthetic Examples

and produce a consistent estimate of the density



# Intervals and Boxes in $\mathbb{R}^d$

*Intervals and Boxes as interval vectors:*

$$\mathbf{x} = [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2] \times \dots \times [\underline{x}_d, \bar{x}_d], \quad \underline{x}_i \leq \bar{x}_i .$$

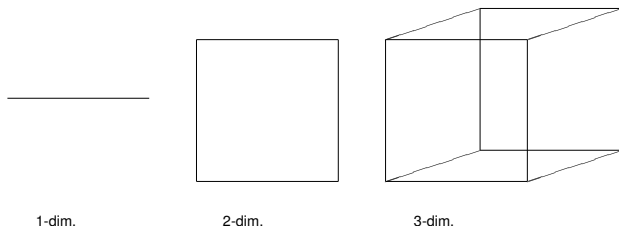


Figure : Boxes in 1D, 2D, and 3D.

## An RP tree a root interval $\mathbf{x}_\rho \in \mathbb{IR}^d$

The **regularly paved boxes** of  $\mathbf{x}_\rho$  can be represented by nodes of **finite rooted binary (frb-trees)** of **geometric group theory**

An operation of bisection on a box is equivalent to performing the operation on its corresponding node in the tree:

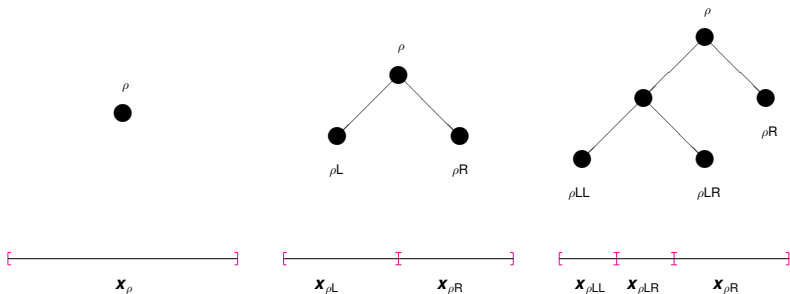
# An RP tree a root interval $\mathbf{x}_\rho \in \mathbb{IR}^d$

The **regularly paved boxes** of  $\mathbf{x}_\rho$  can be represented by nodes of **finite rooted binary (frb-trees)** of **geometric group theory**

An operation of bisection on a box is equivalent to performing the operation on its corresponding node in the tree:

Leaf boxes of RP tree partition the root interval  $\mathbf{x}_\rho \in \mathbb{IR}^1$

**Bisect at the midpoint of the chosen leaf interval**



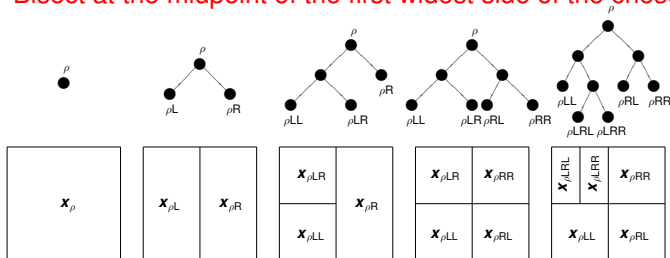
# An RP tree a root interval $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^d$

The **regularly paved boxes** of  $\mathbf{x}_\rho$  can be represented by nodes of **finite rooted binary (frb-trees)** of **geometric group theory**

An operation of bisection on a box is equivalent to performing the operation on its corresponding node in the tree:

Leaf boxes of RP tree partition the root interval  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^2$

**Bisect at the midpoint of the first widest side of the chosen leaf box**



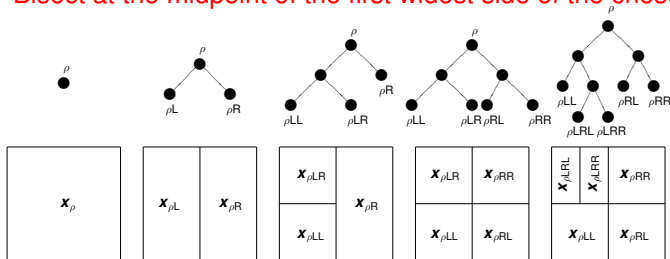
# An RP tree a root interval $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^d$

The **regularly paved boxes** of  $\mathbf{x}_\rho$  can be represented by nodes of **finite rooted binary (frb-trees)** of **geometric group theory**

An operation of bisection on a box is equivalent to performing the operation on its corresponding node in the tree:

Leaf boxes of RP tree partition the root interval  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^2$

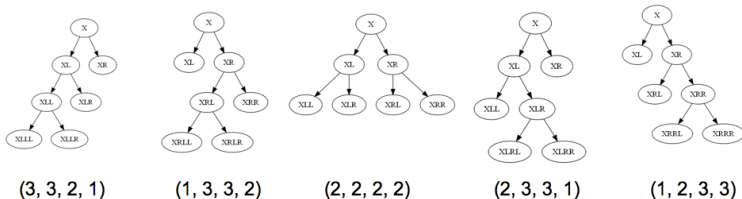
**Bisect at the midpoint of the first widest side of the chosen leaf box**



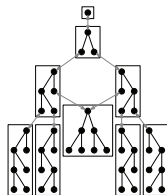
By this “RP Peano’s curve” frb-trees encode partitions of  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^d$

# Algebraic Structure and Combinatorics of RPs

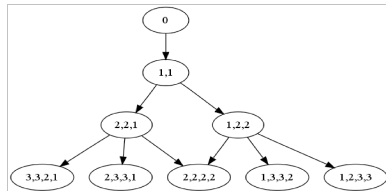
## Leaf-depth encoded RPs



There are  $C_k$  RPs with  $k$  splits

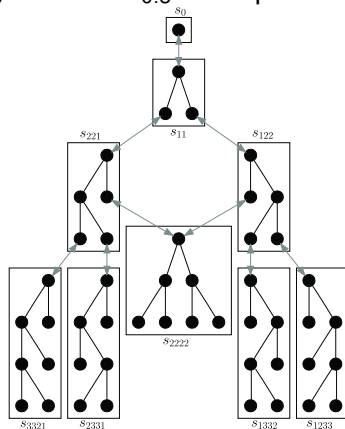


- $C_0 = 1$
- $C_1 = 1$
- $C_2 = 2$
- $C_3 = 5$
- $C_4 = 14$
- $C_5 = 42$
- $\dots = \dots$
- $C_k = \frac{(2k)!}{(k+1)!k!}$
- $\dots = \dots$
- $C_{15} = 9694845$
- $\dots = \dots$
- $C_{20} = 6564120420$



# Hasse (transition) Diagram of Regular Pavings

Transition diagram over  $\mathbb{S}_{0:3}$  with split/reunion operations

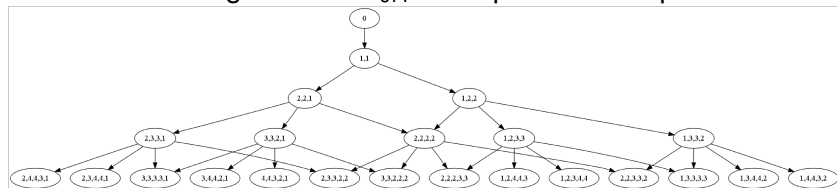


RS, W.Taylor and G.Teng, [Catalan Coefficients, Sequence A185155 in The On-Line Encyclopedia of Integer](#)

[Sequences, 2012](#), <http://oeis.org>

# Hasse (transition) Diagram of Regular Pavings

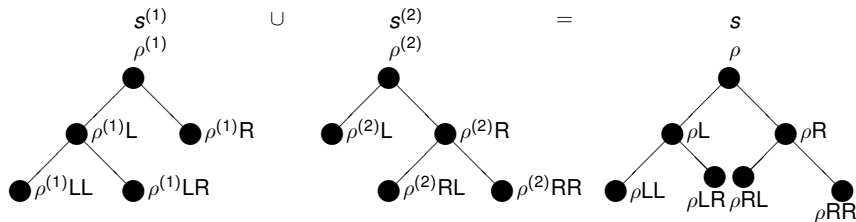
Transition diagram over  $\mathbb{S}_{0:4}$  with split/reunion operations



1. The above state space is denoted by  $\mathbb{S}_{0:4}$
2. Number of RPs with  $k$  splits is the Catalan number  $C_k$
3. There is more than one way to reach a RP by  $k$  splits
4. Randomized algorithms are Markov chains on  $\mathbb{S}_{0:\infty}$

# RPs are closed under union operations

$s^{(1)} \cup s^{(2)} = s$  is union of two RPs  $s^{(1)}$  and  $s^{(2)}$  of  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^2$ .



$\mathbf{x}_{\rho^{(1)LR}}$	$\mathbf{x}_{\rho^{(1)R}}$
$\mathbf{x}_{\rho^{(1)LL}}$	

$\mathbf{x}_{\rho^{(2)L}}$	$\mathbf{x}_{\rho^{(2)RR}}$
	$\mathbf{x}_{\rho^{(2)RL}}$

$\mathbf{x}_{\rho LR}$	$\mathbf{x}_{\rho RR}$
$\mathbf{x}_{\rho LL}$	$\mathbf{x}_{\rho RL}$

## RPs are closed under union operations

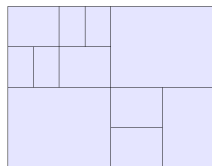
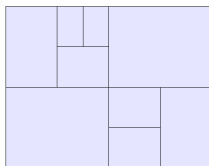
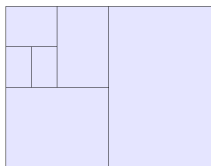
**Lemma 1:** The algebraic structure of frb-trees (underlying Thompson's group) is closed under union operations.

## RPs are closed under union operations

**Lemma 1:** The algebraic structure of frb-trees (underlying Thompson's group) is closed under union operations.

**Proof:** by a “transparency overlay process” argument (cf. Meier 2008).

$s^{(1)} \cup s^{(2)} = s$  is union of two RPs  $s^{(1)}$  and  $s^{(2)}$  of  $\mathbf{x}_\rho \in \mathbb{R}^2$ .



---

## Algorithm 1: $\text{RPUnion}(\rho^{(1)}, \rho^{(2)})$

---

**input** : Root nodes  $\rho^{(1)}$  and  $\rho^{(2)}$  of RPs  $s^{(1)}$  and  $s^{(2)}$ , respectively, with root box  $\mathbf{x}_{\rho^{(1)}} = \mathbf{x}_{\rho^{(2)}}$

**output** : Root node  $\rho$  of RP  $s = s^{(1)} \cup s^{(2)}$

**if**  $\text{IsLeaf}(\rho^{(1)}) \ \& \ \text{IsLeaf}(\rho^{(2)})$  **then**

$\rho \leftarrow \text{Copy}(\rho^{(1)})$

**return**  $\rho$

**end**

**else if**  $!\text{IsLeaf}(\rho^{(1)}) \ \& \ \text{IsLeaf}(\rho^{(2)})$  **then**

$\rho \leftarrow \text{Copy}(\rho^{(1)})$

**return**  $\rho$

**end**

**else if**  $\text{IsLeaf}(\rho^{(1)}) \ \& \ !\text{IsLeaf}(\rho^{(2)})$  **then**

$\rho \leftarrow \text{Copy}(\rho^{(2)})$

**return**  $\rho$

**end**

**else**

$!\text{IsLeaf}(\rho^{(1)}) \ \& \ !\text{IsLeaf}(\rho^{(2)})$

**end**

Make  $\rho$  as a node with  $\mathbf{x}_{\rho} \leftarrow \mathbf{x}_{\rho^{(1)}}$

Graft onto  $\rho$  as left child the node  $\text{RPUnion}(\rho^{(1)}\text{L}, \rho^{(2)}\text{L})$

Graft onto  $\rho$  as right child the node  $\text{RPUnion}(\rho^{(1)}\text{R}, \rho^{(2)}\text{R})$

**return**  $\rho$

---

Note: this is not the minimal union of the (Boolean mapped) RPs of Jaulin et. al. 2001

## Dfn: Real Mapped Regular Paving ( $\mathbb{R}$ -MRP)

- ▶ Let  $\mathbf{s} \in \mathbb{S}_{0:\infty}$  be an RP with root node  $\rho$  and root box  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^d$

## Dfn: Real Mapped Regular Paving ( $\mathbb{R}$ -MRP)

- ▶ Let  $s \in \mathbb{S}_{0:\infty}$  be an RP with root node  $\rho$  and root box  $\mathbf{x}_\rho \in \mathbb{IR}^d$
- ▶ Let  $\mathbb{V}(s)$  and  $\mathbb{L}(s)$  denote the sets all nodes and leaf nodes of  $s$ , respectively.

## Dfn: Real Mapped Regular Paving ( $\mathbb{R}$ -MRP)

- ▶ Let  $s \in \mathbb{S}_{0:\infty}$  be an RP with root node  $\rho$  and root box  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^d$
- ▶ Let  $\mathbb{V}(s)$  and  $\mathbb{L}(s)$  denote the sets all nodes and leaf nodes of  $s$ , respectively.
- ▶ Let  $f : \mathbb{V}(s) \rightarrow \mathbb{R}$  map each node of  $s$  to an element in  $\mathbb{R}$  as follows:

$$\{\rho\mathbf{v} \mapsto f_{\rho\mathbf{v}} : \rho\mathbf{v} \in \mathbb{V}(s), f_{\rho\mathbf{v}} \in \mathbb{R}\} .$$

## Dfn: Real Mapped Regular Paving ( $\mathbb{R}$ -MRP)

- ▶ Let  $s \in \mathbb{S}_{0:\infty}$  be an RP with root node  $\rho$  and root box  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^d$
- ▶ Let  $\mathbb{V}(s)$  and  $\mathbb{L}(s)$  denote the sets all nodes and leaf nodes of  $s$ , respectively.
- ▶ Let  $f : \mathbb{V}(s) \rightarrow \mathbb{R}$  map each node of  $s$  to an element in  $\mathbb{R}$  as follows:

$$\{\rho\mathbf{v} \mapsto f_{\rho\mathbf{v}} : \rho\mathbf{v} \in \mathbb{V}(s), f_{\rho\mathbf{v}} \in \mathbb{R}\} .$$

- ▶ Such a map  $f$  is called a  $\mathbb{R}$ -mapped regular paving ( $\mathbb{R}$ -MRP).

## Dfn: Real Mapped Regular Paving ( $\mathbb{R}$ -MRP)

- ▶ Let  $s \in \mathbb{S}_{0:\infty}$  be an RP with root node  $\rho$  and root box  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^d$
- ▶ Let  $\mathbb{V}(s)$  and  $\mathbb{L}(s)$  denote the sets all nodes and leaf nodes of  $s$ , respectively.
- ▶ Let  $f : \mathbb{V}(s) \rightarrow \mathbb{R}$  map each node of  $s$  to an element in  $\mathbb{R}$  as follows:

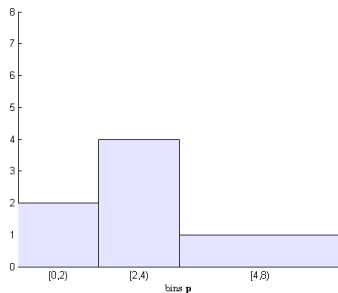
$$\{\rho\mathbf{v} \mapsto f_{\rho\mathbf{v}} : \rho\mathbf{v} \in \mathbb{V}(s), f_{\rho\mathbf{v}} \in \mathbb{R}\} .$$

- ▶ Such a map  $f$  is called a  *$\mathbb{R}$ -mapped regular paving ( $\mathbb{R}$ -MRP)*.
- ▶ Thus, a  $\mathbb{R}$ -MRP  $f$  is obtained by augmenting each node  $\rho\mathbf{v}$  of the RP tree  $s$  with an additional data member  $f_{\rho\mathbf{v}}$ .

# Example of an $\mathbb{R}$ -MRP

Simple functions over an RP tree partition

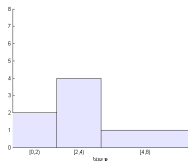
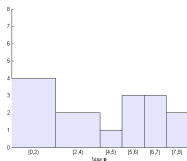
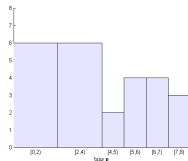
$\mathbb{R}$ -MRP over  $s_{221}$  with  $x_\rho = [0, 8]$



# $\mathbb{R}$ -MRP Arithmetic

If  $\star : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  then we can extend  $\star$  point-wise to two  $\mathbb{R}$ -MRPs  $f$  and  $g$  with root nodes  $\rho^{(1)}$  and  $\rho^{(2)}$  via  $\text{MRPOperate}(\rho^{(1)}, \rho^{(2)}, \star)$ .

This is done using  $\text{MRPOperate}(\rho^{(1)}, \rho^{(2)}, +)$

 $f$  $g$  $f + g$ 

## $\mathbb{R}$ -MRP Addition by $\text{MRPOperate}(\rho^{(1)}, \rho^{(2)}, +)$

adding two piece-wise constant functions or  $\mathbb{R}$ -MRPs

**Algorithm 2:**  $\text{MRPOperate}(\rho^{(1)}, \rho^{(2)}, \star)$ 

**input** : two root nodes  $\rho^{(1)}$  and  $\rho^{(2)}$  with same root box  $\mathbf{x}_{\rho^{(1)}} = \mathbf{x}_{\rho^{(2)}}$  and binary operation  $\star$ .

**output** : the root node  $\rho$  of  $\mathbb{R}$ -MRP  $h = f \star g$ .

Make a new node  $\rho$  with box and image

$\mathbf{x}_{\rho} \leftarrow \mathbf{x}_{\rho^{(1)}}; h_{\rho} \leftarrow f_{\rho^{(1)}} \star g_{\rho^{(2)}}$

**if**  $\text{IsLeaf}(\rho^{(1)}) \ \& \ !\text{IsLeaf}(\rho^{(2)})$  **then**

    Make temporary nodes  $L', R'$

$\mathbf{x}_{L'} \leftarrow \mathbf{x}_{\rho^{(1) L}}; \mathbf{x}_{R'} \leftarrow \mathbf{x}_{\rho^{(1) R}}$

$f_{L'} \leftarrow f_{\rho^{(1)}}, f_{R'} \leftarrow f_{\rho^{(1)}}$

    Graft onto  $\rho$  as left child the node  $\text{MRPOperate}(L', \rho^{(2) L}, \star)$

    Graft onto  $\rho$  as right child the node  $\text{MRPOperate}(R', \rho^{(2) R}, \star)$

**end**

**else if**  $\text{IsLeaf}(\rho^{(1)}) \ \& \ \text{IsLeaf}(\rho^{(2)})$  **then**

    Make temporary nodes  $L', R'$

$\mathbf{x}_{L'} \leftarrow \mathbf{x}_{\rho^{(2) L}}; \mathbf{x}_{R'} \leftarrow \mathbf{x}_{\rho^{(2) R}}$

$g_{L'} \leftarrow g_{\rho^{(2)}}, g_{R'} \leftarrow g_{\rho^{(2)}}$

    Graft onto  $\rho$  as left child the node  $\text{MRPOperate}(\rho^{(1) L}, L', \star)$

    Graft onto  $\rho$  as right child the node  $\text{MRPOperate}(\rho^{(1) R}, R', \star)$

**end**

**else if**  $\text{IsLeaf}(\rho^{(1)}) \ \& \ !\text{IsLeaf}(\rho^{(2)})$  **then**

    Graft onto  $\rho$  as left child the node  $\text{MRPOperate}(\rho^{(1) L}, \rho^{(2) L}, \star)$

    Graft onto  $\rho$  as right child the node  $\text{MRPOperate}(\rho^{(1) R}, \rho^{(2) R}, \star)$

**end**

**return**  $\rho$

## Unary transformations are easy too

Let  $\text{MRPTransform}(\rho, \tau)$  apply the unary transformation  $\tau : \mathbb{R} \rightarrow \mathbb{R}$  to a given  $\mathbb{R}$ -MRP  $f$  with root node  $\rho$  as follows:

- ▶ copy  $f$  to  $g$
- ▶ recursively set  $f_{\rho v} = \tau(f_{\rho v})$  for each node  $\rho v$  in  $g$
- ▶ return  $g$  as  $\tau(f)$

# Minimal Representation of $\mathbb{R}$ -MRP

---

## Algorithm 3: MinimiseLeaves( $\rho$ )

---

**input** :  $\rho$ , the root node of  $\mathbb{R}$ -MRP  $f$ .

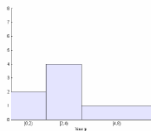
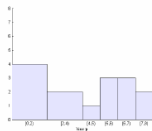
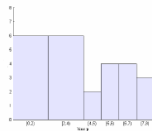
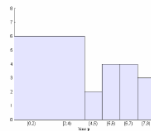
**output** : Modify  $f$  into  $\succ(f)$ , the unique  $\mathbb{R}$ -MRP with fewest leaves.

```

if !IsLeaf( $\rho$ ) then
  MinimiseLeaves( $\rho$ L)
  MinimiseLeaves( $\rho$ R)
  if IsCherry( $\rho$ ) & ( $f_{\rho L} = f_{\rho R}$ ) then
     $f_{\rho} \leftarrow f_{\rho L}$ 
    Prune( $\rho$ L)
    Prune( $\rho$ R)
  end
end

```

---

(a)  $f$ (b)  $g$ (c)  $f + g$ (d)  $\succ(f + g)$

## Arithmetic and Algebra of $\mathbb{R}$ -MRPs

Thus, we can obtain any  $\mathbb{R}$ -MRP arithmetical expression that is specified by finitely many sub-expressions involving:

1. constant  $\mathbb{R}$ -MRP,

## Arithmetic and Algebra of $\mathbb{R}$ -MRPs

Thus, we can obtain any  $\mathbb{R}$ -MRP arithmetical expression that is specified by finitely many sub-expressions involving:

1. constant  $\mathbb{R}$ -MRP,
2. binary arithmetic operation  $\star \in \{+, -, \cdot, /\}$  over two  $\mathbb{R}$ -MRPs,

## Arithmetic and Algebra of $\mathbb{R}$ -MRPs

Thus, we can obtain any  $\mathbb{R}$ -MRP arithmetical expression that is specified by finitely many sub-expressions involving:

1. constant  $\mathbb{R}$ -MRP,
2. binary arithmetic operation  $\star \in \{+, -, \cdot, /\}$  over two  $\mathbb{R}$ -MRPs,
3. standard transformations of  $\mathbb{R}$ -MRPs by elements of  $\mathfrak{G} := \{\exp, \sin, \cos, \tan, \dots\}$  and

## Arithmetic and Algebra of $\mathbb{R}$ -MRPs

Thus, we can obtain any  $\mathbb{R}$ -MRP arithmetical expression that is specified by finitely many sub-expressions involving:

1. constant  $\mathbb{R}$ -MRP,
2. binary arithmetic operation  $\star \in \{+, -, \cdot, /\}$  over two  $\mathbb{R}$ -MRPs,
3. standard transformations of  $\mathbb{R}$ -MRPs by elements of  $\mathcal{G} := \{\exp, \sin, \cos, \tan, \dots\}$  and
4. their compositions.

# Stone-Wierstrass Theorem: $\mathbb{R}$ -MRPs Dense in $C(\mathbf{x}_\rho, \mathbb{R})$

## Theorem

*Let  $\mathcal{F}$  be the class of  $\mathbb{R}$ -MRPs with the same root box  $\mathbf{x}_\rho$ . Then  $\mathcal{F}$  is dense in  $C(\mathbf{x}_\rho, \mathbb{R})$ , the algebra of real-valued continuous functions on  $\mathbf{x}_\rho$ .*

## Stone-Wierstrass Theorem: $\mathbb{R}$ -MRPs Dense in $C(\mathbf{x}_\rho, \mathbb{R})$

### Theorem

*Let  $\mathcal{F}$  be the class of  $\mathbb{R}$ -MRPs with the same root box  $\mathbf{x}_\rho$ . Then  $\mathcal{F}$  is dense in  $C(\mathbf{x}_\rho, \mathbb{R})$ , the algebra of real-valued continuous functions on  $\mathbf{x}_\rho$ .*

### Proof:

Since  $\mathbf{x}_\rho \in \mathbb{I}\mathbb{R}^d$  is a compact Hausdorff space, by the Stone-Weierstrass theorem we can establish that  $\mathcal{F}$  is dense in  $C(\mathbf{x}_\rho, \mathbb{R})$  with the topology of uniform convergence, provided that  $\mathcal{F}$  is a sub-algebra of  $C(\mathbf{x}_\rho, \mathbb{R})$  that separates points in  $\mathbf{x}_\rho$  and which contains a non-zero constant function.

## Stone-Weierstrass Theorem: ℝ-MRPs Dense in $C(\mathbf{x}_\rho, \mathbb{R})$

### Theorem

*Let  $\mathcal{F}$  be the class of ℝ-MRPs with the same root box  $\mathbf{x}_\rho$ . Then  $\mathcal{F}$  is dense in  $C(\mathbf{x}_\rho, \mathbb{R})$ , the algebra of real-valued continuous functions on  $\mathbf{x}_\rho$ .*

### Proof:

Since  $\mathbf{x}_\rho \in \mathbb{R}^d$  is a compact Hausdorff space, by the Stone-Weierstrass theorem we can establish that  $\mathcal{F}$  is dense in  $C(\mathbf{x}_\rho, \mathbb{R})$  with the topology of uniform convergence, provided that  $\mathcal{F}$  is a sub-algebra of  $C(\mathbf{x}_\rho, \mathbb{R})$  that separates points in  $\mathbf{x}_\rho$  and which contains a non-zero constant function.

We will show all these conditions are satisfied by  $\mathcal{F}$

## Stone-Wierstrass Theorem Contd.: $\mathbb{R}$ -MRPs Dense in $C(\mathbf{x}_\rho, \mathbb{R})$

- ▶  $\mathcal{F}$  is a sub-algebra of  $C(\mathbf{x}_\rho, \mathbb{R})$  since it is closed under addition and scalar multiplication.

## Stone-Wierstrass Theorem Contd.: $\mathbb{R}$ -MRPs Dense in $C(\mathbf{x}_\rho, \mathbb{R})$

- ▶  $\mathcal{F}$  is a sub-algebra of  $C(\mathbf{x}_\rho, \mathbb{R})$  since it is closed under addition and scalar multiplication.
- ▶  $\mathcal{F}$  contains non-zero constant functions

## Stone-Wierstrass Theorem Contd.: $\mathbb{R}$ -MRPs Dense in $C(\mathbf{x}_\rho, \mathbb{R})$

- ▶  $\mathcal{F}$  is a sub-algebra of  $C(\mathbf{x}_\rho, \mathbb{R})$  since it is closed under addition and scalar multiplication.
- ▶  $\mathcal{F}$  contains non-zero constant functions
- ▶ Finally, RPs can clearly separate distinct points  $x, x' \in \mathbf{x}_\rho$  into distinct leaf boxes by splitting deeply enough.

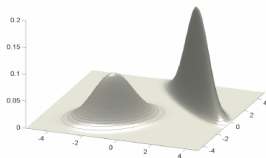
## Stone-Wierstrass Theorem Contd.: $\mathbb{R}$ -MRPs Dense in $C(\mathbf{x}_\rho, \mathbb{R})$

- ▶  $\mathcal{F}$  is a sub-algebra of  $C(\mathbf{x}_\rho, \mathbb{R})$  since it is closed under addition and scalar multiplication.
- ▶  $\mathcal{F}$  contains non-zero constant functions
- ▶ Finally, RPs can clearly separate distinct points  $x, x' \in \mathbf{x}_\rho$  into distinct leaf boxes by splitting deeply enough.
- ▶ Thus,  $\mathcal{F}$ , the class of  $\mathbb{R}$ -MRPs with the same root box  $\mathbf{x}_\rho$ , is dense in  $C(\mathbf{x}_\rho, \mathbb{R})$ , the algebra of real-valued continuous functions on  $\mathbf{x}_\rho$ .

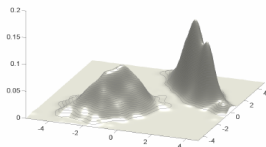
## Stone-Wierstrass Theorem Contd.: $\mathbb{R}$ -MRPs Dense in $C(\mathbf{x}_\rho, \mathbb{R})$

- ▶  $\mathcal{F}$  is a sub-algebra of  $C(\mathbf{x}_\rho, \mathbb{R})$  since it is closed under addition and scalar multiplication.
- ▶  $\mathcal{F}$  contains non-zero constant functions
- ▶ Finally, RPs can clearly separate distinct points  $x, x' \in \mathbf{x}_\rho$  into distinct leaf boxes by splitting deeply enough.
- ▶ Thus,  $\mathcal{F}$ , the class of  $\mathbb{R}$ -MRPs with the same root box  $\mathbf{x}_\rho$ , is dense in  $C(\mathbf{x}_\rho, \mathbb{R})$ , the algebra of real-valued continuous functions on  $\mathbf{x}_\rho$ .
- ▶ Q.E.D.

# Kernel Density Estimate (visualization of a procedure)

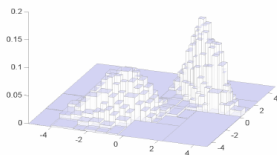


(a) True density.

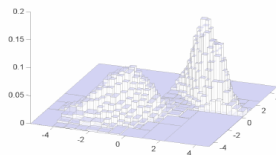


(c) MCMC bandwidth KDE.

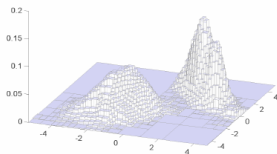
# Approximating Kernel Density Estimates by $\mathbb{R}$ -MRPs



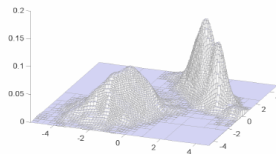
(a)  $\bar{\psi} = 0.001$  (187 leaves).



(b)  $\bar{\psi} = 0.005$  (316 leaves).



(c)  $\bar{\psi} = 0.0001$  (919 leaves).



(d)  $\bar{\psi} = 0.00001$  (4420 leaves).

Approximating Kernel Density Estimates by  $\mathbb{R}$ -MRPsTable J.4: 5- $d$  case: estimated errors for KDE and RMRP-KDE approximations.

	$\hat{d}_{KL}$	$\hat{L}_1$ error	Time (s)	Leaves
KDE ( $n_K = 2,000$ )	0.41	0.66	7,350–8,880	$n/a$
RMRP-KDE approximations				
$\overline{\psi} = 0.0001$	5.06	0.96	1.0	2,363
$\overline{\psi} = 0.00005$	4.85	0.91	2.3	4,639
$\overline{\psi} = 0.00001$	4.51	0.85	8.7	17,759
$\overline{\psi} = 0.000005$	4.49	0.84	17.2	31,335
$\overline{\psi} = 0.000001$	3.33	0.76	66.1	133,493
$\overline{\psi} = 0.0000005$	3.31	0.75	131.0	237,561
$\overline{\psi} = 0.0000001$	3.54	0.74	470.0	895,012

## Finding image of $\mathbb{R}$ -MRP is by fast look-ups

---

### Algorithm 4: PointwiseImage( $\rho, x$ )

---

**input** :  $\rho$  with box  $\mathbf{x}_\rho$ , the root node of  $\mathbb{R}$ -MRP  $f$  with RP  $s$ , and a point  $x \in \mathbf{x}_\rho$ .  
**output** : Return  $f_{\eta(x)}$  at the leaf node  $\eta(x)$  that is associated with the box  $\mathbf{x}_{\eta(x)}$  containing  $x$ .

```

if IsLeaf( $\rho$ ) then
  |   return  $f_\rho$ 
end
else
  |   if  $x \in \mathbf{x}_{\rho R}$  then
  |   |   PointwiseImage( $\rho R, x$ )
  |   end
  |   else
  |   |   PointwiseImage( $\rho L, x$ )
  |   end
end

```

---

## Finding image of $\mathbb{R}$ -MRP is by fast look-ups

---

### Algorithm 5: PointWiseImage( $\rho, x$ )

---

**input** :  $\rho$  with box  $\mathbf{x}_\rho$ , the root node of  $\mathbb{R}$ -MRP  $f$  with RP  $s$ , and a point  $x \in \mathbf{x}_\rho$ .  
**output** : Return  $f_{\eta(x)}$  at the leaf node  $\eta(x)$  that is associated with the box  $\mathbf{x}_{\eta(x)}$  containing  $x$ .

```

if IsLeaf( $\rho$ ) then
  |   return  $f_\rho$ 
end
else
  |   if  $x \in \mathbf{x}_{\rho R}$  then
  |   |   PointWiseImage( $\rho R, x$ )
  |   end
  |   else
  |   |   PointWiseImage( $\rho L, x$ )
  |   end
end

```

---

- ▶ Cost of KDE image  $\sim O(n)$  **KFLOPs** (FLOPs for kernel evaluation procedure)

## Finding image of $\mathbb{R}$ -MRP is by fast look-ups

---

### Algorithm 6: PointWiseImage( $\rho, x$ )

---

**input** :  $\rho$  with box  $\mathbf{x}_\rho$ , the root node of  $\mathbb{R}$ -MRP  $f$  with RP  $s$ , and a point  $x \in \mathbf{x}_\rho$ .  
**output** : Return  $f_{\eta(x)}$  at the leaf node  $\eta(x)$  that is associated with the box  $\mathbf{x}_{\eta(x)}$  containing  $x$ .

```

if IsLeaf( $\rho$ ) then
  |   return  $f_\rho$ 
end
else
  |   if  $x \in \mathbf{x}_{\rho R}$  then
  |   |   PointWiseImage( $\rho R, x$ )
  |   end
  |   else
  |   |   PointWiseImage( $\rho L, x$ )
  |   end
end

```

---

- ▶ Cost of KDE image  $\sim O(n)$  **KFLOPs** (FLOPs for kernel evaluation procedure)
- ▶ 10-fold CV cost  $\sim 10 \times O\left(\frac{1}{10} n \frac{9}{10} n\right) = O(n^2)$  **KFLOPs**

## Finding image of $\mathbb{R}$ -MRP is by fast look-ups

---

### Algorithm 7: PointWiseImage( $\rho, x$ )

---

**input** :  $\rho$  with box  $\mathbf{x}_\rho$ , the root node of  $\mathbb{R}$ -MRP  $f$  with RP  $s$ , and a point  $x \in \mathbf{x}_\rho$ .  
**output** : Return  $f_{\eta(x)}$  at the leaf node  $\eta(x)$  that is associated with the box  $\mathbf{x}_{\eta(x)}$  containing  $x$ .

```

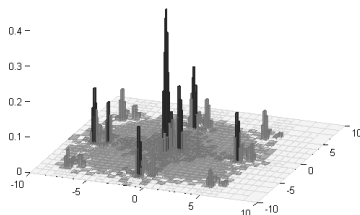
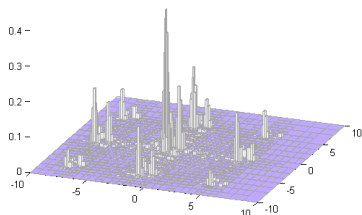
if IsLeaf( $\rho$ ) then
  |   return  $f_\rho$ 
end
else
  |   if  $x \in \mathbf{x}_{\rho R}$  then
  |   |   PointWiseImage( $\rho R, x$ )
  |   end
  |   else
  |   |   PointWiseImage( $\rho L, x$ )
  |   end
end

```

---

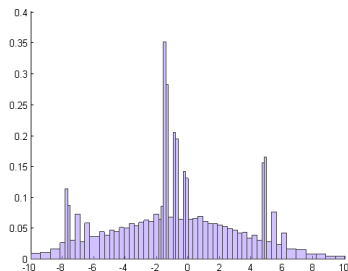
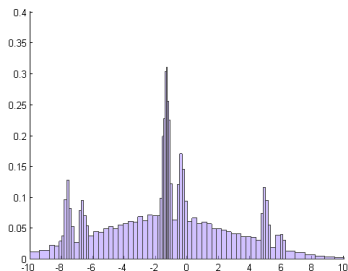
- ▶ Cost of KDE image  $\sim O(n)$  **KFLOPs** (FLOPs for kernel evaluation procedure)
- ▶ 10-fold CV cost  $\sim 10 \times O\left(\frac{1}{10} n \frac{9}{10} n\right) = O(n^2)$  **KFLOPs**
- ▶ But using  $\mathbb{R}$ -MRP approximation to KDE requires  $10 \times O\left(\frac{1}{10} n \lg\left(\frac{9}{10} n\right)\right) = O(n \lg(n))$  **tree-look-ups**

# Coverage, Marginal & Slice Operators of $\mathbb{R}$ -MRP



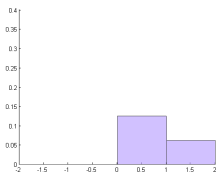
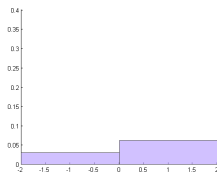
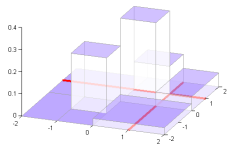
$\mathbb{R}$ -MRP approximation to Levy density and its coverage regions with  $\alpha = 0.9$  (light gray),  $\alpha = 0.5$  (dark gray) and  $\alpha = 0.1$  (black)

# Coverage, Marginal & Slice Operators of $\mathbb{R}$ -MRP



Marginal densities  $f^{\{1\}}(x_1)$  and  $f^{\{2\}}(x_2)$  along each coordinate of  $\mathbb{R}$ -MRP approximation

# Coverage, Marginal & Slice Operators of $\mathbb{R}$ -MRP



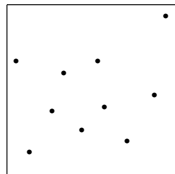
The slices of a simple  $\mathbb{R}$ -MRP in 2D

## Statistical Regular Pavings (SRPs)

- ▶ Extended from the RP;
- ▶ Caches recursively computable statistics at each box or node as data falls through;
- ▶ These statistics include:
  - ▶ the sample count;
  - ▶ the sample mean vector;
  - ▶ the sample variance-covariance matrix;
  - ▶ and the volume of the box.

Caching the sample count in each node (or box).

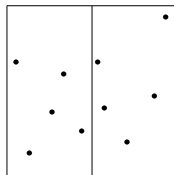
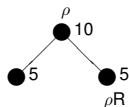
$\rho$   
● 10



# Statistical Regular Pavings (SRPs)

- ▶ Extended from the RP;
- ▶ Caches recursively computable statistics at each box or node as data falls through;
- ▶ These statistics include:
  - ▶ the sample count;
  - ▶ the sample mean vector;
  - ▶ the sample variance-covariance matrix;
  - ▶ and the volume of the box.

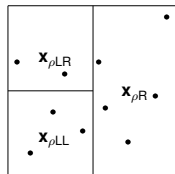
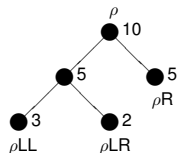
Caching the sample count in each node (or box).



# Statistical Regular Pavings (SRPs)

- ▶ Extended from the RP;
- ▶ Caches recursively computable statistics at each box or node as data falls through;
- ▶ These statistics include:
  - ▶ the sample count;
  - ▶ the sample mean vector;
  - ▶ the sample variance-covariance matrix;
  - ▶ and the volume of the box.

Caching the sample count in each node (or box).



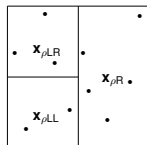
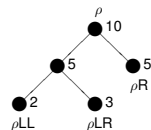
## SRPs as Adaptive Histograms

SRP estimate of  $f$  from random vectors  $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} f$  is

$$f_{n,\dot{s}}(x) = \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{1}(x_i \in \mathbf{x}(x))}{\text{vol}(\mathbf{x}(x))},$$

$\mathbf{x}(x) \in \ell(\dot{s})$  is the leaf box containing  $x$  with volume  $\text{vol}(\mathbf{x}(x))$

Figure : A SRP as a histogram estimate.



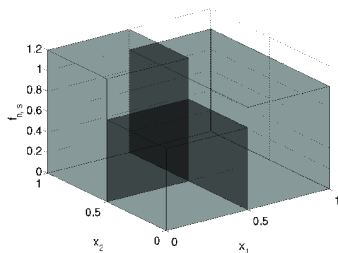
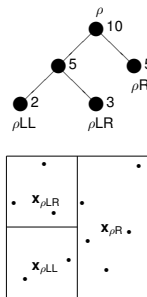
## SRPs as Adaptive Histograms

SRP estimate of  $f$  from random vectors  $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} f$  is

$$f_{n,\dot{s}}(x) = \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{1}(x_i \in \mathbf{x}(x))}{\text{vol}(\mathbf{x}(x))},$$

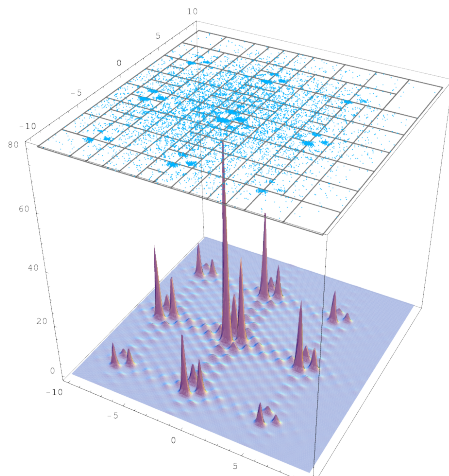
$\mathbf{x}(x) \in \ell(\dot{s})$  is the leaf box containing  $x$  with volume  $\text{vol}(\mathbf{x}(x))$

Figure : A SRP as a histogram estimate.



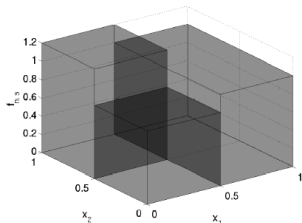
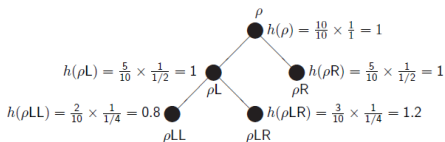
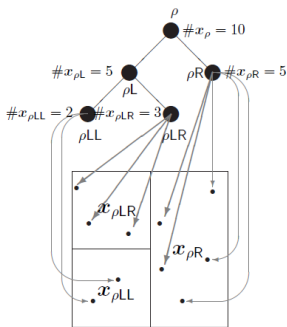
# Nonparametric Density Estimation – Recap

Problem: Take **samples** from an unknown density  $f$  and consistently reconstruct  $f$



# Nonparametric Density Estimation – Recap

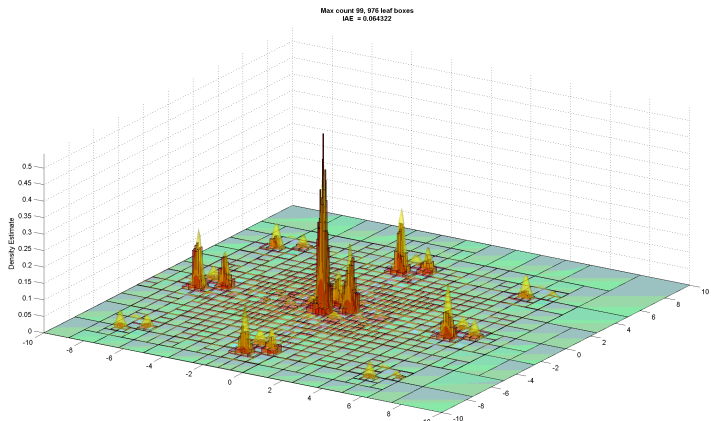
Approach: Use **statistical regular paving** to get  **$\mathbb{R}$ -MRP data-adaptive histogram**



# Nonparametric Density Estimation – Recap

Solution:  $\mathbb{R}$ -MRP histogram averaging allows us to produce a consistent Bayesian estimate of the density (up to 10 dimensions)

(Teng, Harlow, Lee and S., *ACM Trans. Mod. & Comp. Sim.*, 2013)

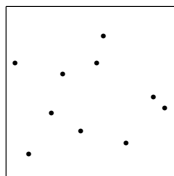


# A Prioritized Queue based Algorithm (for $L_1$ Consistent Initialization)

## Algorithm `SplitMostCounts`

As data arrives, order the leaf boxes of the SRP so that the leaf box with **the most number of points** will be chosen for the next bisection.

$\rho$   
● 10



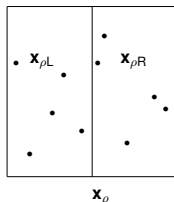
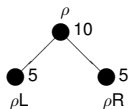
$\mathbf{x}_\rho$

# A Prioritized Queue based Algorithm (for $L_1$ Consistent Initialization)

## Algorithm `SplitMostCounts`

As data arrives, order the leaf boxes of the SRP so that the leaf box with **the most number of points** will be chosen for the next bisection.

Split the root box.

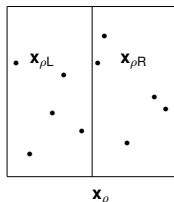
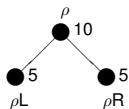


# A Prioritized Queue based Algorithm (for $L_1$ Consistent Initialization)

## Algorithm `SplitMostCounts`

As data arrives, order the leaf boxes of the SRP so that the leaf box with **the most number of points** will be chosen for the next bisection.

Two or more boxes with the most number of points?

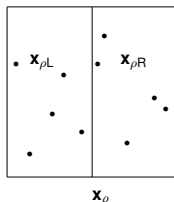
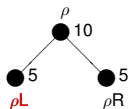


# A Prioritized Queue based Algorithm (for $L_1$ Consistent Initialization)

## Algorithm `SplitMostCounts`

As data arrives, order the leaf boxes of the SRP so that the leaf box with **the most number of points** will be chosen for the next bisection.

Break such ties by randomising the next bisection.

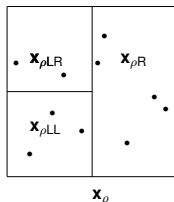
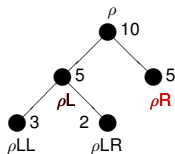


# A Prioritized Queue based Algorithm (for $L_1$ Consistent Initialization)

## Algorithm `SplitMostCounts`

As data arrives, order the leaf boxes of the SRP so that the leaf box with **the most number of points** will be chosen for the next bisection.

Bisect until each box has  $\leq \bar{k}_n$  points (let  $\bar{k}_n = 3$  here).

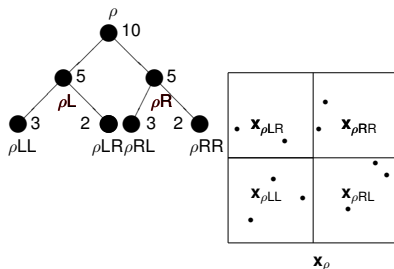


# A Prioritized Queue based Algorithm (for $L_1$ Consistent Initialization)

## Algorithm `SplitMostCounts`

As data arrives, order the leaf boxes of the SRP so that the leaf box with **the most number of points** will be chosen for the next bisection.

Final state



# The SplitMostCounts Algorithm

**Input:** (i) data:  $\mathbf{x}_1, \dots, \mathbf{x}_n \subseteq \mathbb{R}^d$ ; (ii) root box:  $\mathbf{x}_\rho$  // optional;  
 (iii) padding to handle pulsed data:  $\psi \geq 0$  // optional;  
 (iv) S.E.B. max:  $\bar{k}_n$ ; (v) maximum partition size:  $\bar{m}_n$ .  
**Output:** histogram estimate  $f_{n,s}$ .

initialize  $i \leftarrow 1$ ;  $\mathbf{s} \leftarrow \mathbf{x}_\rho + \psi$ ;

repeat until

$\#\mathbf{x}_{\rho v} \leq \bar{k}_n$  for each  $\mathbf{x}_{\rho v} \in \ell(\mathbf{s})$  and  $i \leq \bar{m}_n$  //

$\ell(\mathbf{s}) = \{\text{leaf boxes}\}$

$\mathbf{x}_{\rho v} \leftarrow \text{Uniform}(\hat{\ell}(\mathbf{s}))$  // randomized PQ on leaf boxes

$\mathbf{s} \leftarrow \text{bisect}(\mathbf{s}, \mathbf{x}_{\rho v})$  // bisect leaf box  $\mathbf{x}_{\rho v}$  of  $\mathbf{s}$

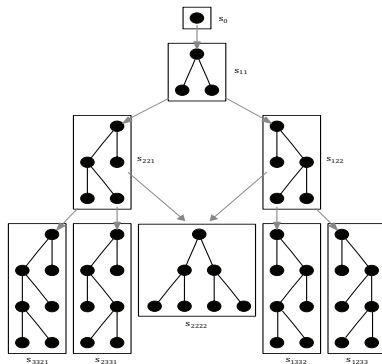
recursively update counts in  $\mathbf{s}$ ;

$i \leftarrow i + 1$ ;

return  $f_{n,s}$ .

# Transition Diagram of Randomized PQ Markov chain

Let  $\mathcal{S}_i$  be the set of all RPs of  $\mathbf{x}_\rho$  made of  $i$  splits and for  $i, j \in \mathbb{N}$  with  $i \leq j$ , let  $\mathcal{S}_{i:j}$  be the set of RPs with  $k$  splits,  $i \leq k \leq j$ .



All possible RP partitions in  $\mathcal{S}_{0:4}$ .

## Proposition: $L_1$ -Consistency of Histogram Estimates from `SplitMostCounts`

Let  $X_1, X_2, \dots$  be independent and identical random vectors in  $\mathbb{R}^d$  whose common distribution  $\mu$  has a non-atomic density  $f$ , i.e.,  $f \ll \lambda^d$ . Let  $\{S_n(i)\}_{i=0}^j$  on  $\mathbb{S}_{0:\infty}$  be the Markov chain formed using `SplitMostCounts` with terminal state  $\dot{s}$  and histogram estimate  $f_{n,\dot{s}}$  over the collection of partitions  $\mathcal{L}_n$ .

As  $n \rightarrow \infty$ , if  $\bar{k}_n \rightarrow \infty$ ,  $n^{-1}\bar{k}_n \rightarrow 0$ ,  $\bar{m}_n \geq n/\bar{k}_n$ , and  $\bar{m}_n/n \rightarrow 0$  then the density estimate  $f_{n,\dot{s}}$  is strongly consistent in  $L_1$ , i.e.

$$\int |f(x) - f_{n,\dot{s}}(x)| dx \rightarrow 0 \text{ with probability 1.}$$

## Proof Sketch

We will assume that  $\bar{k}_n \rightarrow \infty$ ,  $n^{-1}\bar{k}_n \rightarrow 0$ ,  $\bar{m}_n \geq n/\bar{k}_n$ , and  $\bar{m}_n/n \rightarrow 0$ , as  $n \rightarrow \infty$ , and show that the three conditions:

- (a)  $n^{-1}m(\mathcal{L}_n) \rightarrow 0$ ,
- (b)  $n^{-1} \log \Delta_n^*(\mathcal{L}_n) \rightarrow 0$ , and
- (c)  $\mu(\mathbf{x} : \text{diam}(\mathbf{x}(x)) > \gamma) \rightarrow 0$  with probability 1 for every  $\gamma > 0$ ,

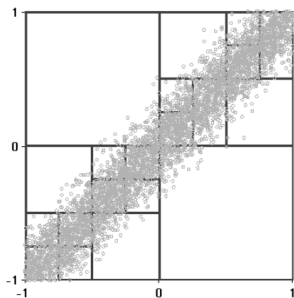
are satisfied. Then by Theorem 1 of Lugosi and Nobel, 1996 our density estimate  $f_{n,\hat{s}}$  is strongly consistent in  $L_1$ .

These conditions mean:

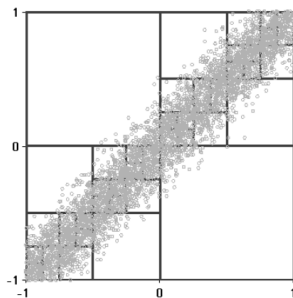
- (a) sub-linear growth of the number of leaf boxes
- (b) sub-exponential growth of a combinatorial complexity measure of the growth of the partition
- (c) shrinking leaf boxes in the partition

## Complementary PQ to “carve out” Support

`SplitMostCounts` uses priority =  $\mu_n(\mathbf{x}_{\rho V})$ .



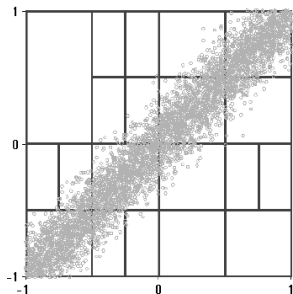
(a) 20 leaves.



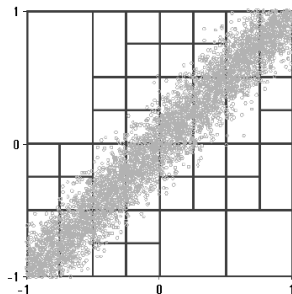
(b) 40 leaves.

## Complementary PQ to “carve out” Support

SupportCarver uses priority =  $(1 - \mu_n(\mathbf{x}_{\rho V}))\text{vol}(\mathbf{x}_{\rho V})$ .



(a) 20 leaves.

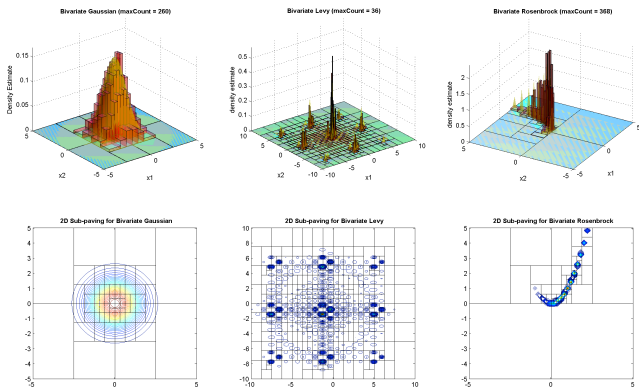


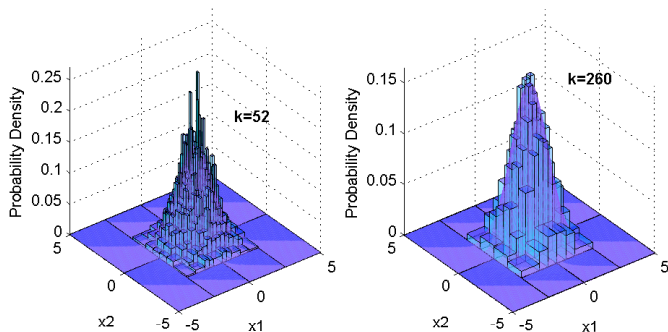
(b) 40 leaves.

Necessary to use SupportCarver for high-dimensional structured densities before using SplitMostCounts

# Some Examples

**Figure :** Histogram density estimates their corresponding pavings for the bivariate Gaussian, Levy and Rosenbrock densities.



Choice of  $k_n$ 

**Figure :** Two histogram density estimates for the standard bivariate gaussian density with different choices of  $k_n$ . The histogram is under-smoothed when  $k_n$  is relatively smaller than  $n$  and over-smoothed when  $k_n$  is relatively larger.

## Adding and Averaging SRPs

Adding  $m$  SRP histogram density estimates

$$\begin{aligned}\sum_{i=1}^m f_{n,s^{(i)}} &= f_{n,s^{(1)}} + f_{n,s^{(2)}} + f_{n,s^{(3)}} + \dots + f_{n,s^{(m)}} \\ &= \left( \left( \left( f_{n,s^{(1)}} + f_{n,s^{(2)}} \right) + f_{n,s^{(3)}} \right) + \dots + f_{n,s^{(m)}} \right) .\end{aligned}$$

## Adding and Averaging SRPs

### Adding $m$ SRP histogram density estimates

$$\begin{aligned}\sum_{i=1}^m f_{n,s^{(i)}} &= f_{n,s^{(1)}} + f_{n,s^{(2)}} + f_{n,s^{(3)}} + \dots + f_{n,s^{(m)}} \\ &= \left( \left( \left( f_{n,s^{(1)}} + f_{n,s^{(2)}} \right) + f_{n,s^{(3)}} \right) + \dots + f_{n,s^{(m)}} \right) .\end{aligned}$$

Averaging  $m$  SRP histogram density estimates recursively yields the sample mean SRP histogram

$$\bar{f}_{n,m} = \frac{1}{m} \sum_{i=1}^m f_{n,s^{(i)}} .$$

## Posterior Distribution over Histograms in $\mathbb{S}_{0:\infty}$

- ▶ Let  $\hat{f}_s$  be a histogram with partition  $\ell(s)$  given by the leaves of RP  $s$  with  $k$  splits and  $k + 1$  leaves in  $\mathbb{S}_k$

## Posterior Distribution over Histograms in $\mathbb{S}_{0:\infty}$

- ▶ Let  $\hat{f}_s$  be a histogram with partition  $\ell(s)$  given by the leaves of RP  $s$  with  $k$  splits and  $k + 1$  leaves in  $\mathbb{S}_k$
- ▶ Then for this partition, the most likely histogram estimate is

$$\hat{f}_s(x; \text{data}) = \frac{1}{n} \hat{f}_s(x; X_{1:n}) = \sum_{i=1}^n \frac{\mathbf{1}(x_i \in \mathbf{x}(x))}{\text{vol}(\mathbf{x}(x))}$$

## Posterior Distribution over Histograms in $\mathbb{S}_{0:\infty}$

- ▶ Let  $\hat{f}_s$  be a histogram with partition  $\ell(s)$  given by the leaves of RP  $s$  with  $k$  splits and  $k + 1$  leaves in  $\mathbb{S}_k$
- ▶ Then for this partition, the most likely histogram estimate is

$$\hat{f}_s(x; \text{data}) = \frac{1}{n} \hat{f}_s(x; X_{1:n}) = \sum_{i=1}^n \frac{\mathbf{1}(x_i \in \mathbf{x}(x))}{\text{vol}(\mathbf{x}(x))}$$

- ▶ Let the prior probability be  $P(s) \propto \frac{1}{C_k^2}$ ,  $s \in \mathbb{S}_{0:\infty}$

## Posterior Distribution over Histograms in $\mathbb{S}_{0:\infty}$

- ▶ Let  $\hat{f}_s$  be a histogram with partition  $\ell(s)$  given by the leaves of RP  $s$  with  $k$  splits and  $k + 1$  leaves in  $\mathbb{S}_k$
- ▶ Then for this partition, the most likely histogram estimate is

$$\hat{f}_s(x; \text{data}) = \frac{1}{n} \hat{f}_s(x; X_{1:n}) = \sum_{i=1}^n \frac{\mathbf{1}(x_i \in \mathbf{x}(x))}{\text{vol}(\mathbf{x}(x))}$$

- ▶ Let the prior probability be  $P(s) \propto \frac{1}{C_k^2}$ ,  $s \in \mathbb{S}_{0:\infty}$
- ▶ Then the posterior density of histogram  $\hat{f}_s$  with  $k$  splits is

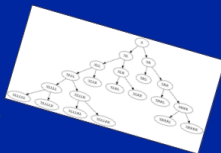
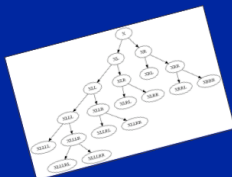
$$P(\hat{f}_s | X_{1:n}) \propto P(X_{1:n} | s) P(s) = \prod_{\mathbf{x}_{\rho\nu} \in \ell(s)} \left( \frac{\#\mathbf{x}_{\rho\nu}}{n \text{vol}(\mathbf{x}_{\rho\nu})} \right)^{n_{\mathbf{x}_{\rho\nu}}} \frac{1}{C_k^2}$$

# Metropolis-Hastings Algorithm

- ▶ Use a proposal density  $q(s'|s^{(i)})$  which depends on current state  $s^{(i)}$ , to generate a new proposed state  $s'$
- ▶ We propose uniformly at random to split a leaf or merge a cherry of current SRP state  $s^{(i)}$
- ▶ **Repeat**
  - ▶ **Draw**  $u \sim U(0, 1)$
  - ▶ **If**  $u < \frac{P(\hat{f}_{s'}|X_{1:n})}{P(\hat{f}_{s^{(i)}}|X_{1:n})} \frac{q(s^{(i)}|s')}{q(s'|s^{(i)})}$  **then**  $s^{(i+1)} \leftarrow s'$
  - ▶ **else**  $s^{(i+1)} \leftarrow s^{(i)}$
- ▶ With a “long enough” burn-in time, this Markov chain will be at the desired stationary distribution  $P(\hat{f}_s|X_{1:n})$  over  $\mathbb{S}_{0:\infty}$

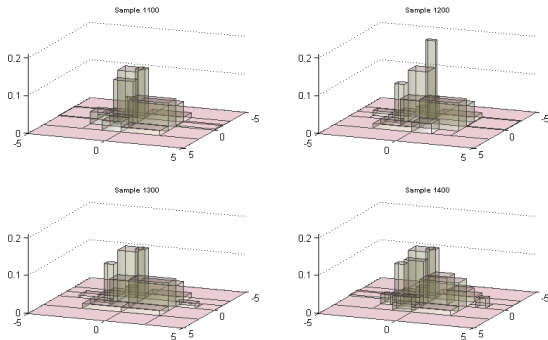
# Metropolis-Hastings Algorithm

- Start from some initial state  $m^0$
- Burn-in: run until initial state is 'forgotten'
- States after burn-in are sample histograms



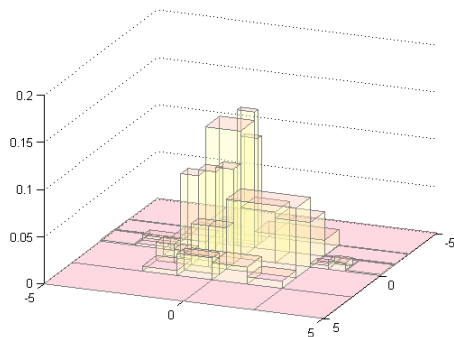
etc...

# Histogram Estimates - Standard Bivariate Gaussian



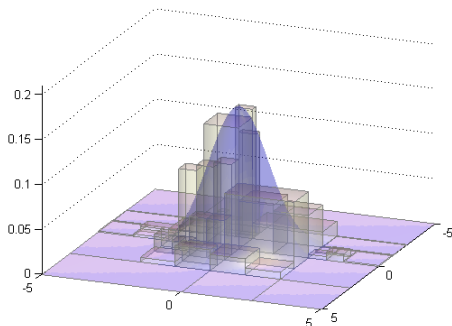
Four sample histograms

# Histogram Estimates - Standard Bivariate Gaussian



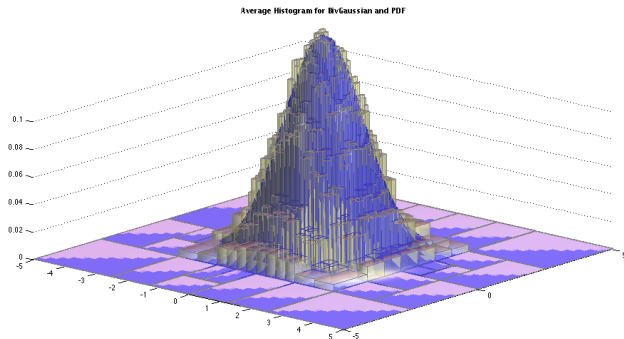
Average of the four sampled histograms

# Histogram Estimates - Standard Bivariate Gaussian



Average of the four sampled histograms with Gaussian PDF

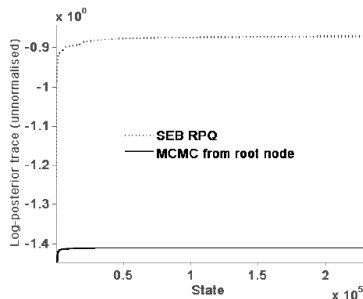
## Histogram Estimates - Standard Bivariate Gaussian



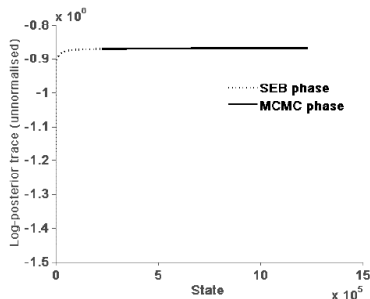
A much better estimate

# Combining Randomized PQ with MCMC

Log-posterior traces of SEB RPQ Vs. MCMC started from root node



(a) Initial SEB phase compared with MCMC from root node.

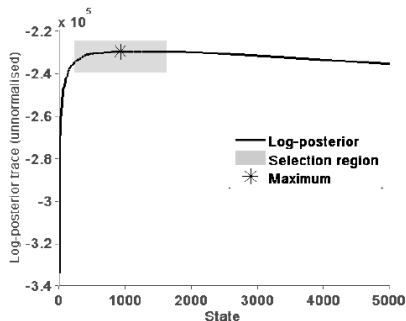


(b) Combined log-posterior trace to  $t = 1,000,000$ .

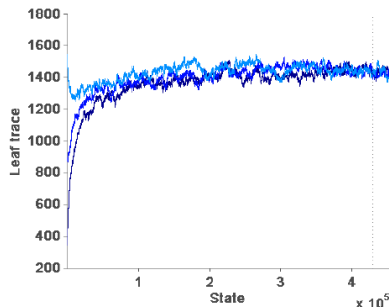
(data drawn from 6D Gaussian Density) – Initialize from highest log-posterior states visited by RPQ

# Combining Randomized PQ with MCMC

Multiple MCMC chains started from high log-posterior region



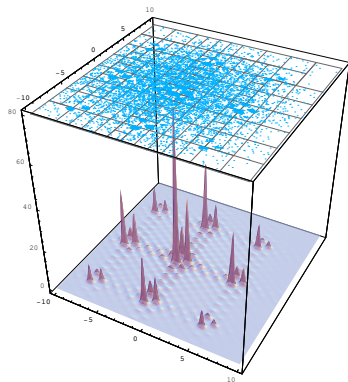
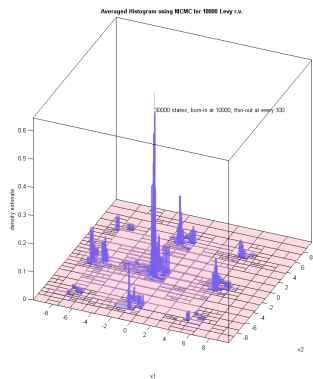
(a) Selection region.



(b) Leaf trace for three chains.

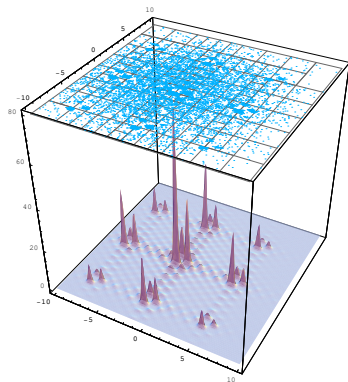
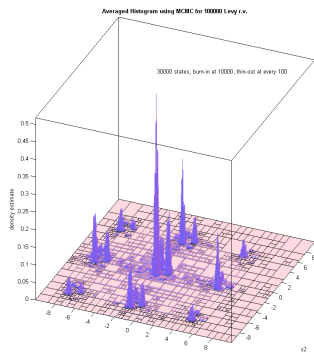
(data drawn from mixture of two 3D Gaussian Densities)

## Histogram Estimates - Bivariate Levy Density



Data points = 10000, Number of states = 30000, Burn-in = 10000,  
Thin-out = 100, Averaged over 201 states, Time taken = 14.16s

## Histogram Estimates - Bivariate Levy Density



Data points = 100000, Number of states = 30000, Burn-in = 10000,  
Thin-out = 100, Averaged over 201 states, Time taken = 50.59s

# Simulations for MCMC from root box

MIAE (std. err.) for  $n$  samples from uniform density in various dimensions (CPU Times  $< O(1\text{minute})$ ).

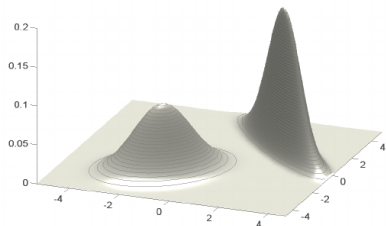
$n$	1D	2D	10D	100D	1000D
$10^2$	0.1112 (0.0707)	0.1425 (0.0882)	0.1170 (0.0723)	0.0958 (0.0605)	0.1111 (0.0524)
$10^3$	0.0366 (0.0192)	0.0363 (0.0219)	0.0442 (0.0275)	0.0413 (0.0196)	0.0305 (0.0195)
$10^4$	0.0164 (0.0095)	0.0124 (0.0073)	0.0115 (0.0070)	0.0111 (0.0083)	0.0089 (0.0065)
$10^5$	0.0041 (0.0020)	0.0040 (0.0026)	0.0041 (0.0028)	0.0050 (0.0030)	0.0043 (0.0025)
$10^6$	0.0011 (0.0005)	0.0016 (0.0007)	0.0010 (0.0006)	0.0012 (0.0001)	0.0010 (0.0004)
$10^7$	0.0004 (0.0003)	0.0003 (0.0002)	0.0003 (0.0002)	0.0002 (0.0001)	-
$10^8$	0.0001 (0.0009)	0.0002 (0.0002)	0.0001 (0.0001)	-	-

Simulations for MCMC and `SplitMostCounts` PQ

MIAE (std. err.) for  $n$  samples from approximated 1D-, 2D- and 10D-Gaussian densities, and 2D- and 10D-Rosenbrock densities ( $L_1$ -minimal Simple function approximation in  $\mathbb{S}_\lambda$ ).

$\lambda$	$n$	Standard Gaussian densities			Rosenbrock densities	
		1D	2D	10D	2D	10D
$10^2$	$10^2$	0.2665 (0.0415)	0.4856 (0.0491)	0.1192 (0.0662)	0.5089 (0.0924)	0.0323 (0.0511)
	$10^3$	0.1390 (0.0192)	0.2558 (0.0127)	0.0543 (0.0172)	0.1712 (0.0224)	0.0095 (0.0191)
	$10^4$	0.0620 (0.0047)	0.0992 (0.0067)	0.0382 (0.0036)	0.0498 (0.0081)	0.0025 (0.0050)
	$10^5$	0.0262 (0.0016)	0.0279 (0.0019)	0.0259 (0.0017)	0.0143 (0.0025)	0.0009 (0.0015)
	$10^6$	0.0099 (0.0008)	0.0086 (0.0006)	0.0073 (0.0009)	0.0045 (0.0005)	0.0004 (0.0005)
	$10^7$	0.0026 (0.0002)	0.0027 (0.0003)	0.0025 (0.0004)	0.0017 (0.0010)	0.0001 (0.0003)
	$10^3$	$10^2$	0.2946 (0.0678)	0.6046 (0.1299)	0.1702 (0.0907)	1.0027 (0.0437)
$10^3$		0.1418 (0.0226)	0.2973 (0.0174)	0.0739 (0.0183)	0.4747 (0.0191)	0.0039 (0.0075)
$10^4$		0.0648 (0.0052)	0.1586 (0.0067)	0.0555 (0.0045)	0.2139 (0.0054)	0.0013 (0.0028)
$10^5$		0.0292 (0.0014)	0.0768 (0.0016)	0.0295 (0.0020)	0.0789 (0.0023)	0.0004 (0.0006)
$10^6$		0.0136 (0.0006)	0.0297 (0.0006)	0.0108 (0.0005)	0.0267 (0.0058)	0.0001 (0.0002)
$10^7$		0.0061 (0.0002)	0.0091 (0.0003)	0.0045 (0.0003)	0.0082 (0.0011)	0.0001 (0.0002)
$10^4$		$10^2$	0.2864 (0.0487)	0.5508 (0.0590)	0.5210 (0.0799)	1.1391 (0.0545)
	$10^3$	0.1380 (0.0152)	0.3301 (0.0120)	0.2719 (0.0251)	0.6018 (0.0139)	0.0791 (0.0223)
	$10^4$	0.0664 (0.0062)	0.1736 (0.0038)	0.1157 (0.0047)	0.3163 (0.0047)	0.0391 (0.0041)
	$10^5$	0.0293 (0.0017)	0.0957 (0.0014)	0.0870 (0.0014)	0.1691 (0.0053)	0.0209 (0.0021)
	$10^6$	0.0138 (0.0005)	0.0495 (0.0005)	0.0788 (0.0009)	0.0882 (0.0048)	0.0123 (0.0012)
	$10^7$	0.0063 (0.0001)	0.0244 (0.0008)	0.0563 (0.0018)	0.0479 (0.0057)	0.0096 (0.0017)

## KDE (diagonal badwidth) Vs. SRP MCMC

Figure B.2: Density II,  $d = 2$ .

When  $d = 2$  Density II is a mixture of two bivariate Normal densities and is the same as Density A studied in [Zhang et al. \(2006\)](#):

$$\mu_a = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \Sigma_a = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}, \quad \mu_b = \begin{pmatrix} -1.5 \\ -1.5 \end{pmatrix}, \quad \Sigma_b = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}.$$

# KDE (diagonal badwidth) Vs. SRP MCMC

Density II is a mixture of two multivariate Normal densities for  $x \in \mathbb{R}^d$ . Density II has high correlation between data coordinates and high bimodality:

$$f_{II}(x | \mu_a, \Sigma_a, \mu_b, \Sigma_b) = \frac{1}{2} \varphi(x | \mu_a, \Sigma_a) + \frac{1}{2} \varphi(x | \mu_b, \Sigma_b),$$

where  $\varphi(x | \mu, \Sigma)$  is the multivariate Normal density with mean  $\mu \in \mathbb{R}^d$  and  $d \times d$  variance-covariance matrix  $\Sigma$ , and

$$\mu_a = \begin{pmatrix} 2.0 \\ \vdots \\ 2.0 \end{pmatrix}, \quad \Sigma_a = \begin{pmatrix} \sigma_a(x_1, x_1) & \sigma_a(x_1, x_2) & \cdots & \sigma_a(x_1, x_d) \\ \sigma_a(x_2, x_1) & \sigma_a(x_2, x_2) & \cdots & \sigma_a(x_2, x_d) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_a(x_d, x_1) & \sigma_a(x_d, x_2) & \cdots & \sigma_a(x_d, x_d) \end{pmatrix},$$

$$\mu_b = \begin{pmatrix} -1.5 \\ \vdots \\ -1.5 \end{pmatrix}, \quad \Sigma_b = \begin{pmatrix} \sigma_b(x_1, x_1) & \sigma_b(x_1, x_2) & \cdots & \sigma_b(x_1, x_d) \\ \sigma_b(x_2, x_1) & \sigma_b(x_2, x_2) & \cdots & \sigma_b(x_2, x_d) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_b(x_d, x_1) & \sigma_b(x_d, x_2) & \cdots & \sigma_b(x_d, x_d) \end{pmatrix},$$

and

$$\sigma_a(x_i, x_j) = \begin{cases} 1 & \text{if } i = j, \\ -0.9^{|i-j|} & \text{if } i \neq j, \end{cases}, \quad \sigma_b(x_i, x_j) = \begin{cases} 1 & \text{if } i = j, \\ 0.3^{|i-j|} & \text{if } i \neq j, \end{cases}.$$

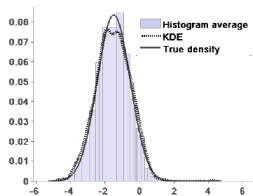
## KDE (diagonal bandwidth) Vs. SRP MCMC

Table 7.2: Estimated errors for KDE and averaged SRP histogram RMRP.

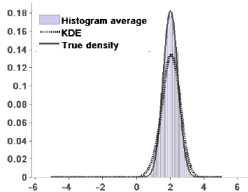
	$\hat{d}_{KL}$	$\hat{L}_1$ error	Time (s)		Leaves	
			min.	max.	min.	max.
<i>2-d</i>						
KDE ( $n_K = 2,000$ )	0.04	0.20	5,000	7,200	<i>n/a</i>	
Averaged RMRP histogram						
$n = 10,000$	0.06	0.22	2	13	811	902
$n = 50,000$	0.03	0.15	15	2,168	1,546	1,719
<i>3-d</i>						
KDE ( $n_K = 2,000$ )	0.13	0.35	5,600	7,200	<i>n/a</i>	
Averaged RMRP histogram						
$n = 10,000$	0.24	0.41	21	451	1,573	1,718
$n = 50,000$	0.12	0.30	295	27,832	3,507	3,783
<i>4-d</i>						
KDE ( $n_K = 2,000$ )	0.25	0.51	7,200	8,050	<i>n/a</i>	
Averaged RMRP histogram						
$n = 50,000$	0.32	0.47	2,524	53,190	6,241	6,570
$n = 100,000$	0.25	0.42	10,382	82,684	9,431	9,775
<i>5-d</i>						
KDE ( $n_K = 2,000$ )	0.41	0.66	7,350	8,880	<i>n/a</i>	
Averaged RMRP histogram						
$n = 50,000$	0.65	0.67	28,841	277,071	9,342	9,803
$n = 100,000$	0.53	0.60	24,244	399,016	15,160	15,563

└ Examples - good, bad and ugly

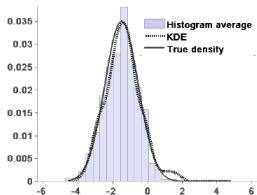
# KDE (diagonal badwidth) Vs. SRP MCMC



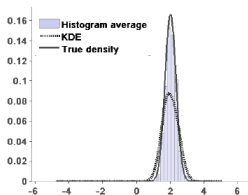
(a)  $d = 2, x_2 = -1.5, n = 50,000$ .



(b)  $d = 2, x_1 = 2.0, n = 50,000$ .



(c)  $d = 3, x_2 = x_3 = -1.5, n = 50,000$ .



(d)  $d = 3, x_1 = x_3 = 2.0, n = 50,000$ .

Figure 7.5: Density II, KDE and averaged RMRP histogram slice.

# Anomaly Detection in Graph Time Series (joint with Carey E. Priebe)

Consider the following two block-specific edge probability matrices:

$$B = \begin{bmatrix} 0.42 & 0.42 \\ 0.42 & 0.50 \end{bmatrix}, \quad B' = \begin{bmatrix} 0.42 & 0.42 \\ 0.42 & 0.40 \end{bmatrix}$$

and the following two prior block probabilities:

$$\pi = (0.6, 0.4) \quad \text{and} \quad \pi' = (0.4, 0.6)$$

In both anomaly simulation scenarios,

- ▶ the initial burst is drawn from  $SBM(B, \pi, n)$
- ▶ and all eight non-anomalous bursts are embedded by  $SBM(B, \pi, m)$ .
- ▶ The 5-th anomalous burst is embedded:
  - ▶ in Scenario B5 by  $SBM(B, \pi, m)$
  - ▶ in Scenario C5 by  $SBM(B, \pi, m)$ .

# Anomaly Detection in Graph Time Series (joint with Carey E. Priebe)

We use  $\mathbb{R}$ -MRP based (10-fold CV) “prior selection”  
 $\pi(s) \propto \exp(-t \times \#leaves)$  to estimate densities from a  
 low-dimensional point-cloud obtained from the Eigen decomposition  
 of the adjacency matrix of each graph.

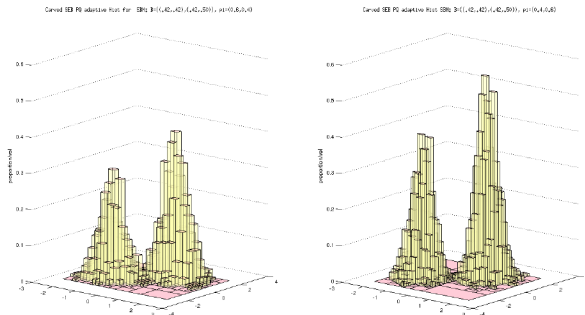


Fig. 1. Density estimates of the embedding of  $SBM(B, \pi, n = 3000)$  and  $SBM(B, \pi', n = 3000)$ .

# Anomaly Detection in Graph Time Series (joint with Carey E. Priebe)

We use  $\mathbb{R}$ -MRP based  $L_1$  computations between all graphs.

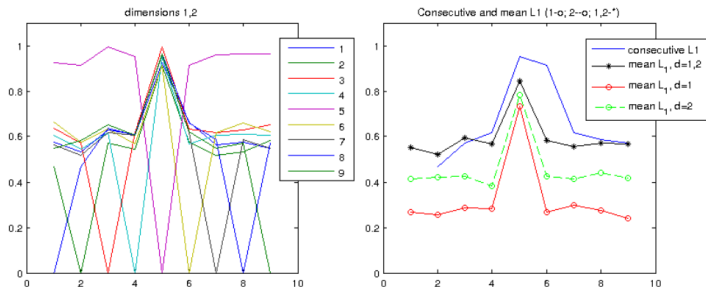
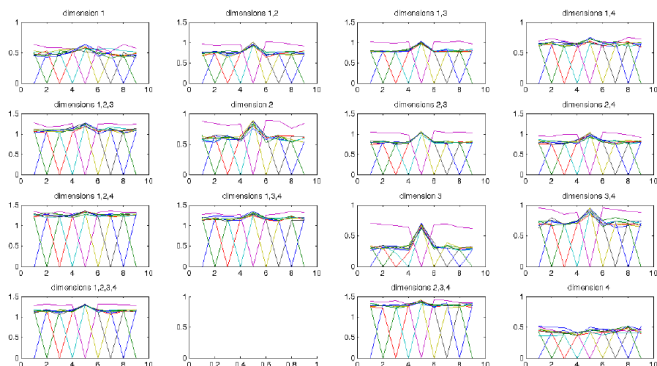


Fig. 2. Pairwise  $L_1$  distances (left) and mean and consecutive  $L_1$  distances (right) between the nine bursts (dimensions 1 and 2) in Scenario C5 with  $n = 5000$ ,  $d = K = 2$ ,  $m = 500$ .

# Anomaly Detection in Graph Time Series (joint with Carey E. Priebe)

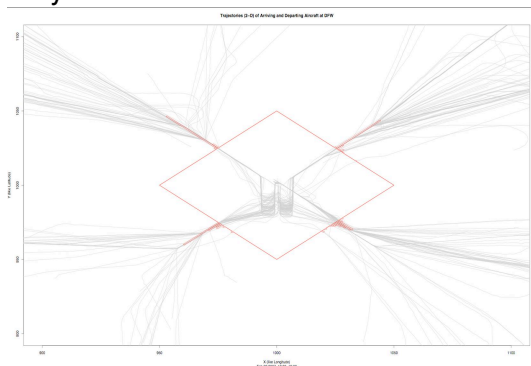
We use  $\mathbb{R}$ -MRP based  $L_1$  computations between all marginal densities of each joint density (4 blocks).



# Air Traffic “Arithmetic” → dynamic air-space configuration

(G. Teng, K. Kuhn and RS, *J. Aerospace Comput., Inf. & Com.*, 9:1, 14–25, 2012.)

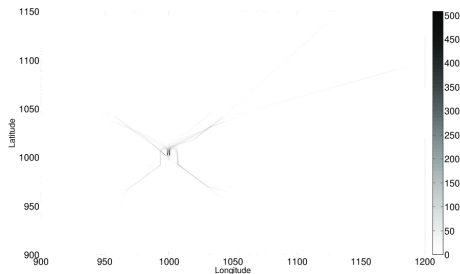
## On a Good Day



# Air Traffic “Arithmetic” → dynamic air-space configuration

(G. Teng, K. Kuhn and RS, *J. Aerospace Comput., Inf. & Com.*, 9:1, 14–25, 2012.)

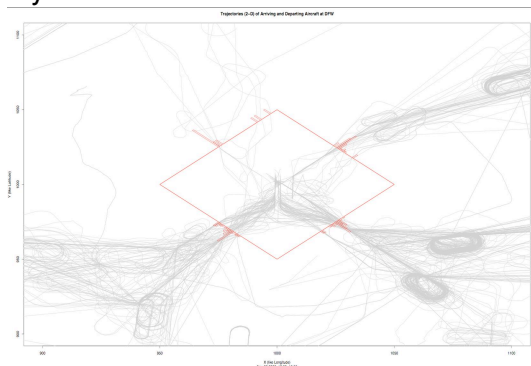
## $\mathbb{Z}_+$ -MRP On a Good Day



# Air Traffic “Arithmetic” → dynamic air-space configuration

(G. Teng, K. Kuhn and RS, *J. Aerospace Comput., Inf. & Com.*, 9:1, 14–25, 2012.)

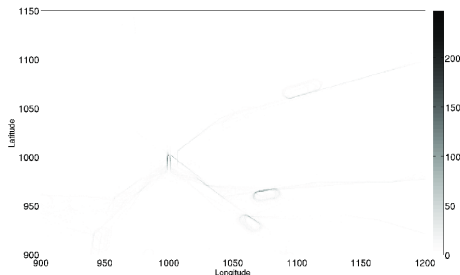
## On a Bad Day



# Air Traffic “Arithmetic” → dynamic air-space configuration

(G. Teng, K. Kuhn and RS, *J. Aerospace Comput., Inf. & Com.*, 9:1, 14–25, 2012.)

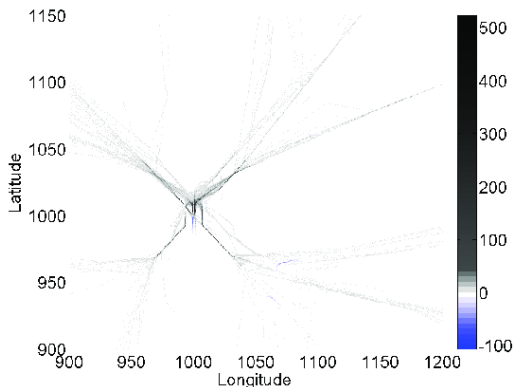
## $\mathbb{Z}_+$ -MRP On a Bad Day



# Air Traffic “Arithmetic” → dynamic air-space configuration

(G. Teng, K. Kuhn and RS, *J. Aerospace Comput., Inf. & Com.*, 9:1, 14–25, 2012.)

$\mathbb{Z}_+$ -MRP pattern for Good Day – Bad Day



## Conclusions

- ▶ Statistical Regular Paving (SRP) is a sufficient statistical data-structure for density estimation and many decisions in massive IID experiments.

# Conclusions

- ▶ Statistical Regular Paving (SRP) is a sufficient statistical data-structure for density estimation and many decisions in massive IID experiments.
- ▶ We can quickly grow or prune the SRP tree data-adaptively

## Conclusions

- ▶ Statistical Regular Paving (SRP) is a sufficient statistical data-structure for density estimation and many decisions in massive IID experiments.
- ▶ We can quickly grow or prune the SRP tree data-adaptively
- ▶ Arithmetic can be efficiently extended to SRPs through  $\mathbb{R}$ -MRPs.

# Conclusions

- ▶ Statistical Regular Paving (SRP) is a sufficient statistical data-structure for density estimation and many decisions in massive IID experiments.
- ▶ We can quickly grow or prune the SRP tree data-adaptively
- ▶ Arithmetic can be efficiently extended to SRPs through  $\mathbb{R}$ -MRPs.
- ▶ Combining PQ-based ( $L_1$ -consistent) initialization + Bayesian MCMC is powerful NFL: MCMC convergence issues exist!

## Conclusions

- ▶ Statistical Regular Paving (SRP) is a sufficient statistical data-structure for density estimation and many decisions in massive IID experiments.
- ▶ We can quickly grow or prune the SRP tree data-adaptively
- ▶ Arithmetic can be efficiently extended to SRPs through  $\mathbb{R}$ -MRPs.
- ▶ Combining PQ-based ( $L_1$ -consistent) initialization + Bayesian MCMC is powerful NFL: MCMC convergence issues exist!
- ▶ Further decisions can be made with appropriate  $\mathbb{R}$ -MRP *arithmetic* (regression, anomaly detection, RPABC+AABC, etc.).

## References

Devroye, L., and Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer.

Devroye, L. and Lugosi, G. (2004). Bin width selection in multivariate histograms by the combinatorial method. *TEST* **13(1)** 129–145.

Jaulin, L., Kieffer, M., Didrit, O. & Walter, E. (2001). *Applied interval analysis*. London: Springer-Verlag.

Lugosi, G. and Nobel, A. (1996). Consistency of data-driven histogram methods for density estimation and classification. *The Annals of Statistics* **24** 687–706.

Harlow, J., Sainudiin, R. and Tucker, W., (2012) *Mapped Regular Pavings, Reliable Computing*, vol. 16, pp. 252-282.

Sainudiin, R., Teng G., Harlow, J., and Lee D., (2013) *Posterior expectation of regularly paved random histograms, ACM Transactions on Modeling and Computer Simulations* 23, 1, Article 6, 20 pages.

Teng G., Kuhn, K. and Sainudiin, R.,(2012) Statistical regular pavings to analyze massive data of aircraft trajectories, *Journal of Aerospace Computing, Information, and Communication*, Vol. 9, No. 1, pp. 14-25.

## Acknowledgements

- ▶ RS's external consulting revenues from the New Zealand Ministry of Tourism
- ▶ WT's Swedish Research Council Grant 2008-7510 that enabled RS's visits to Uppsala in 2006, 2009, 2012
- ▶ Erskine grant from University of Canterbury that enabled WT's visit to Christchurch in 2011 & 2014
- ▶ University of Canterbury MSc Scholarship to JH.
- ▶ Marie Curie International Research Staff Exchange Scheme (IRSES), 2014 - 2017

Thank you!