Posterior Expectation & Minimum Distance Estimation over Adaptive Histograms from randomized Priority Queues on Statistical Regular Pavings

Raazesh Sainudiin

joint work with: Jennifer Harlow, Dominic Lee and Gloria Teng

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

October 2, 2012,

Dept. of Mathematics, University of Oslo, Oslo, October 2, 2012

Massive Metric Data Streams – Introduction

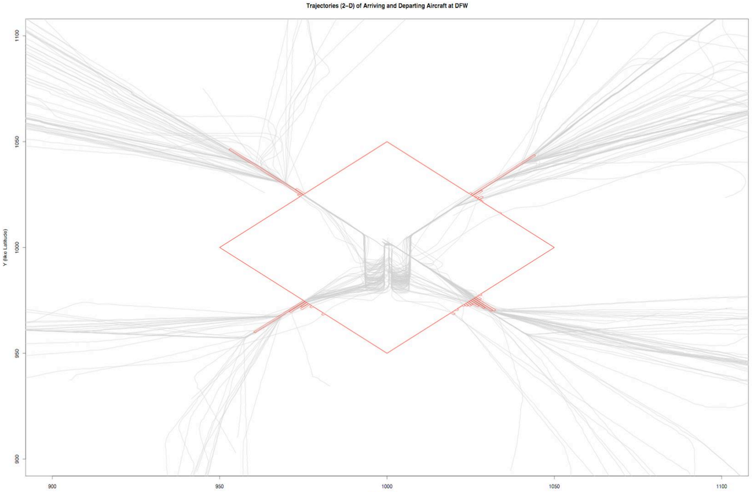
- A massive metric data stream is:

$$\ldots, X_{-3}, X_{-2}, X_{-1}, X_0, \mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \mathbf{X_n}, X_{n+1}, \ldots \sim F, \quad X_i \in \mathbb{R}^d.$$

- 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 Lipschitz $\mathbb{M}$-valued maps: $\{g : \mathbb{R}^d \to \mathbb{M}\}$ – functional estimation when $\mathbb{M} = \mathbb{R}$

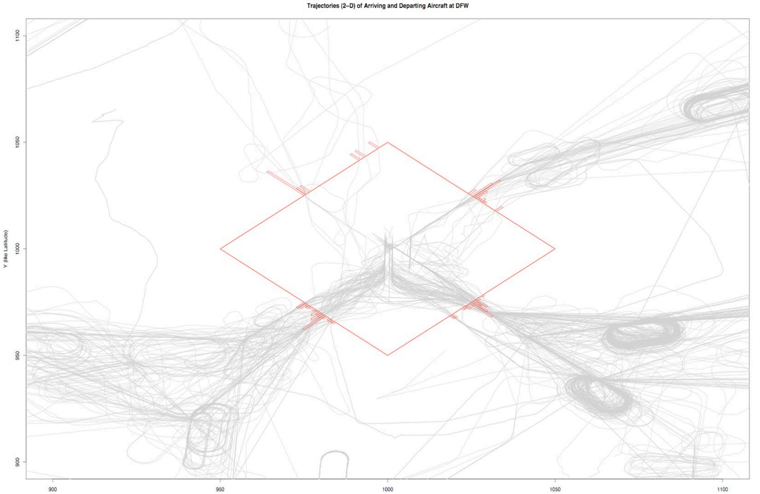
Massive Metric Data Streams – Air Traffic Example

On a Sunny Day



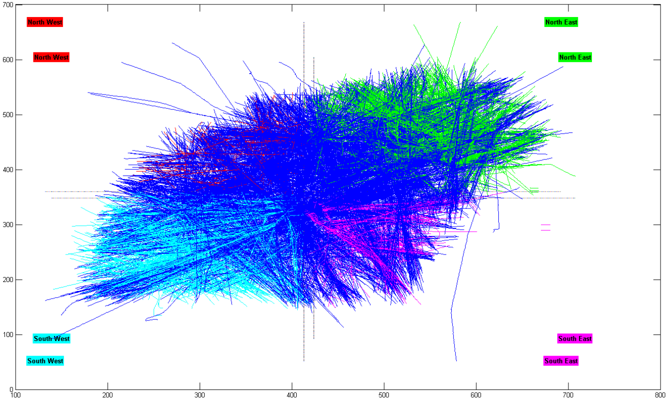
Massive Metric Data Streams – Air Traffic Example

On a Rainy Day



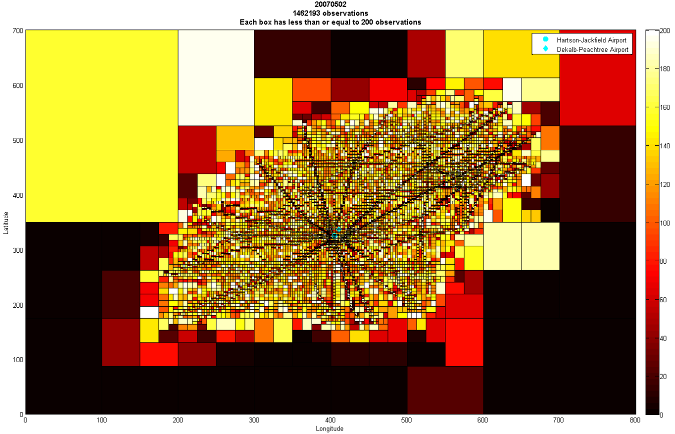
Massive Metric Data Streams – Air Traffic Example

We want to make sense of trajectories like these



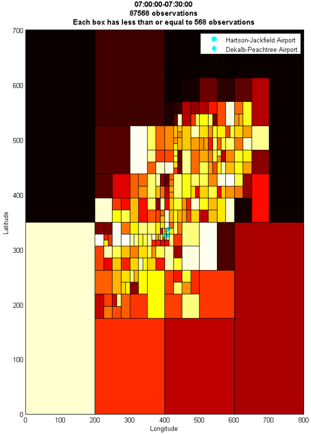
Massive Metric Data Streams – Air Traffic Example

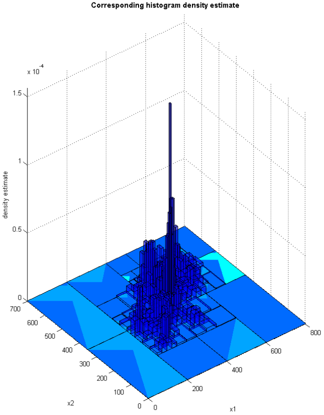
using a picture like this



Massive Metric Data Streams – Air Traffic Example

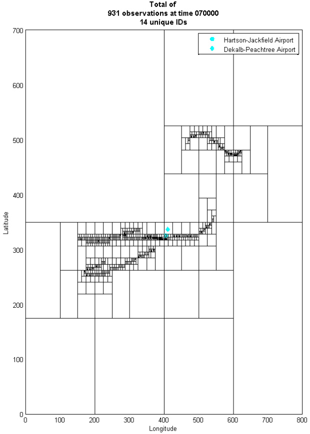
A Histogram Estimate of Air-traffic between 0700 – 0730 hours

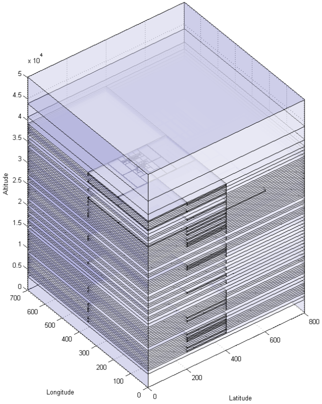




Massive Metric Data Streams – Air Traffic Example

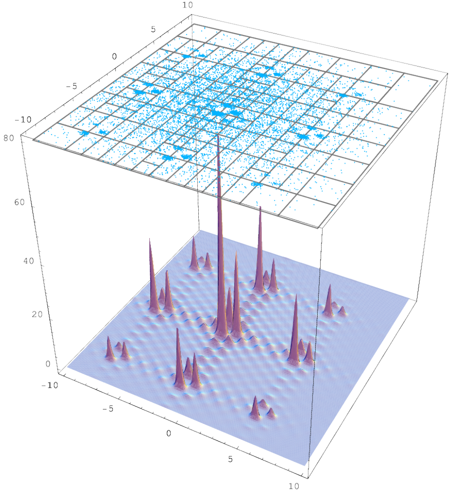
Add the pavings of 14 flight trajectories like this





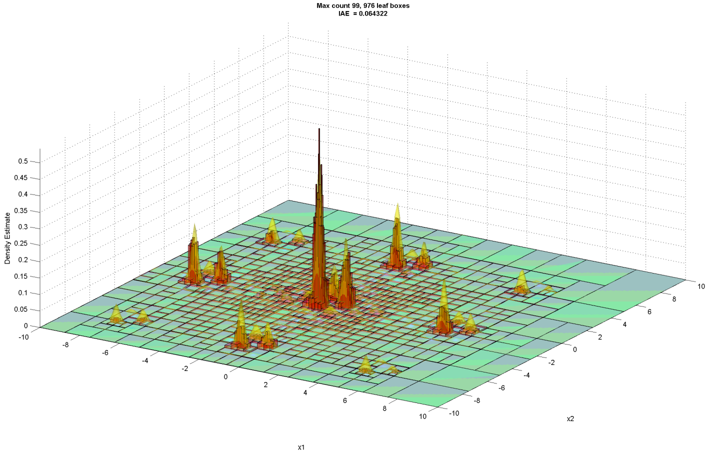
Massive Metric Data Streams – Synthetic Examples

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:

$$\mathbf{x} = [\underline{x}_1, \overline{x}_1] \times [\underline{x}_2, \overline{x}_2] \times \ldots \times [\underline{x}_d, \overline{x}_d] \, , \; \underline{x}_i \leq \overline{x}_i \quad .$$

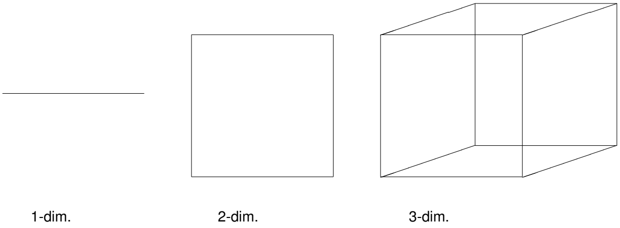


Figure : Boxes in 1D, 2D, and 3D.

Binary Tree Representation

These boxes can be represented by ordered binary trees, a.k.a.:

- *plane binary trees* of enumerative combinatorics
- *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:

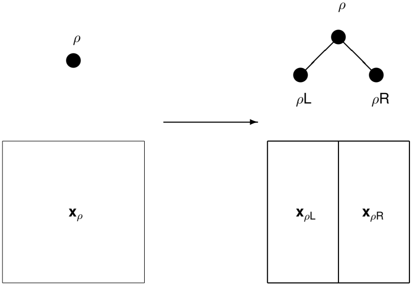
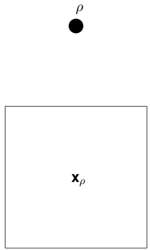


Figure : Bisecting a box or its equivalent node.

Regular Pavings (RPs)

- A sequence of bisections of boxes;
- Start from the root box;
- Along the first widest dimension.

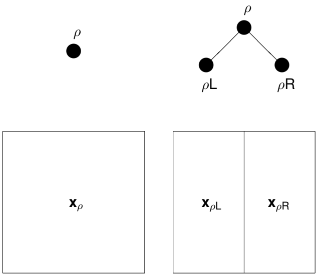
A sequence of bisections on root box **x** to get a 4-leaved RP.



Regular Pavings (RPs)

- A sequence of bisections of boxes;
- Start from the root box;
- Along the first widest dimension.

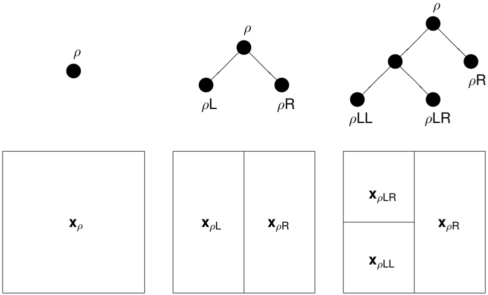
A sequence of bisections on root box **x** to get a 4-leaved RP.



Regular Pavings (RPs)

- A sequence of bisections of boxes;
- Start from the root box;
- Along the first widest dimension.

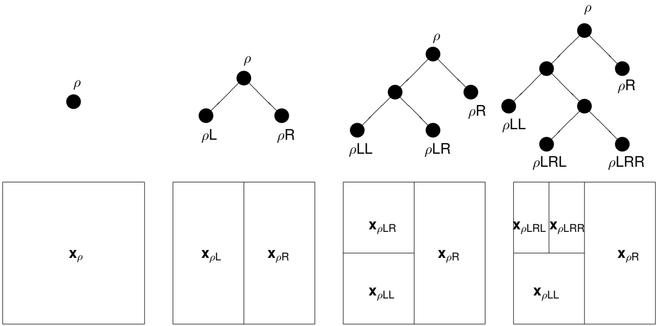
A sequence of bisections on root box **x** to get a 4-leaved RP.



Regular Pavings (RPs)

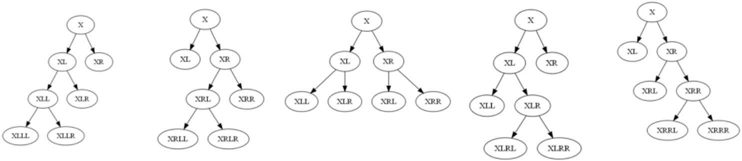
- A sequence of bisections of boxes;
- Start from the root box;
- Along the first widest dimension.

A sequence of bisections on root box **x** to get a 4-leaved RP.

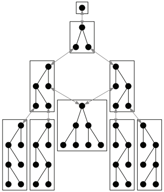


State Space of Regular Pavings

Leaf-depth encoded RPs

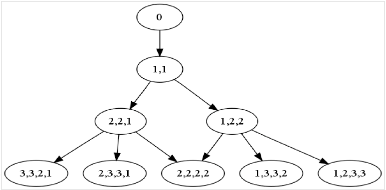


(3, 3, 2, 1) (1, 3, 3, 2) (2, 2, 2, 2) (2, 3, 3, 1) (1, 2, 3, 3)



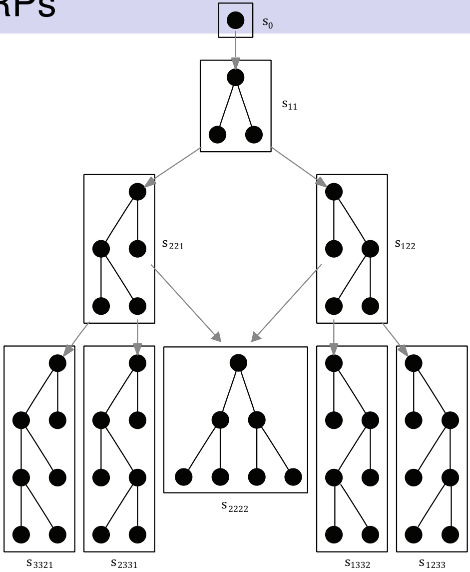
C_k RPs wit k splits

$$
\begin{aligned}
C_0 &= 1 \\
C_1 &= 1 \\
C_2 &= 2 \\
C_3 &= 5 \\
C_4 &= 14 \\
C_5 &= 42 \\
C_k &= \frac{(2k)!}{(k+1)!k!}
\end{aligned}
$$



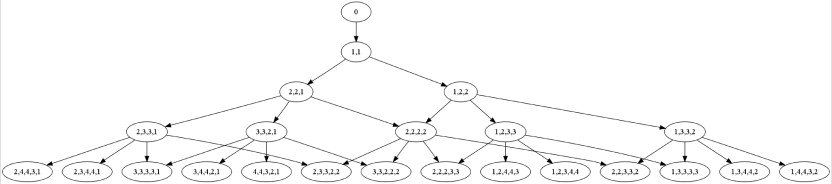
The Space of All Possible RPs

- Let $\mathbb{S}_i$ be the set of all RPs of **x** made of *i* splits and
- Let $\mathbb{S}_{i:j}$ be the set of RPs with *k* splits where $k \in \{i, i+1, \ldots, j\}$
- The space of all RPs is $\mathbb{S}_{0:\infty} := \lim_{i \to \infty} \mathbb{S}_{0:i}$
- RPs are closed under pair-wise union (or overlay) operations (Thompson's Group)
- can get as m_∞-close as desired to any subset of **x**



State Transition Diagram of Regular Pavings

State Transition Diagram of RPs with 0, 1, 2, 3 and 4 splits.

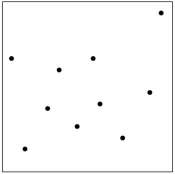


1. The above state space is denoted by $\mathbb{S}_{0:4}$
2. Number of RPs with k splits is the Catalan number C_k
3. There is more than one way to reach a RP by k splits
4. Randomized algorithms here are Markov chains on $\mathbb{S}_{0:\infty}$

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.

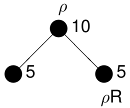
Caching the sample count in each node (or box).

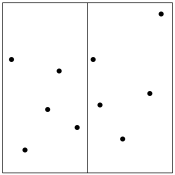


Statistical Regular Pavings (SRPs)

- 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).

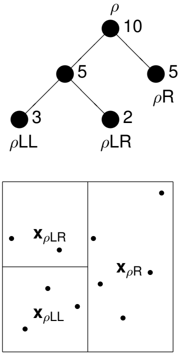




Statistical Regular Pavings (SRPs)

- 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 $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f$ is

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

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

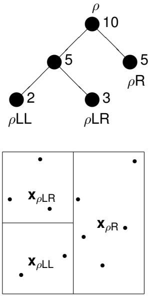


Figure : A SRP as a histogram estimate.

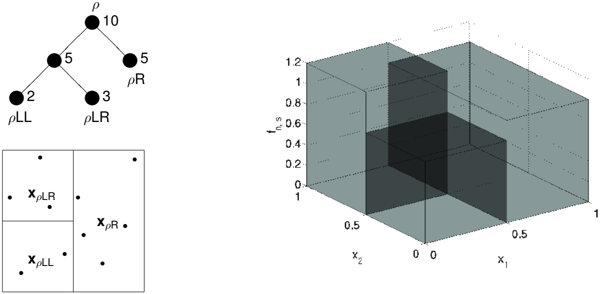
SRPs as Adaptive Histograms

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

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

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

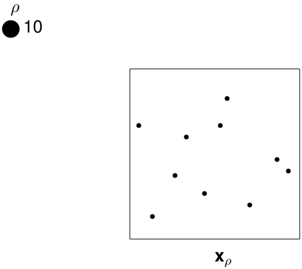
Figure : A SRP as a histogram estimate.



A Prioritized Queue based Algorithm

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.

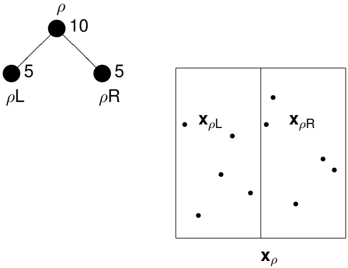


A Prioritized Queue based Algorithm

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.

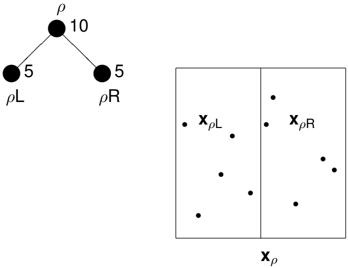


A Prioritized Queue based Algorithm

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?

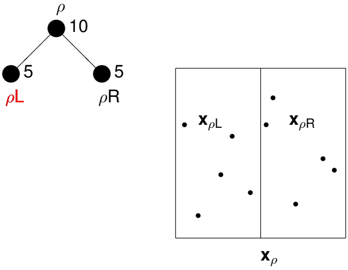


A Prioritized Queue based Algorithm

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.

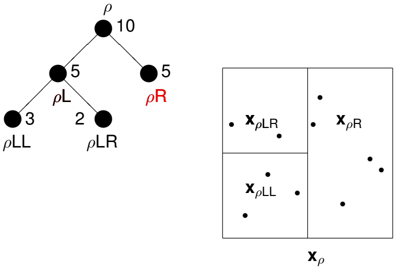


A Prioritized Queue based Algorithm

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 $\leq \overline{k}_n$ points (let $\overline{k}_n = 3$ here).

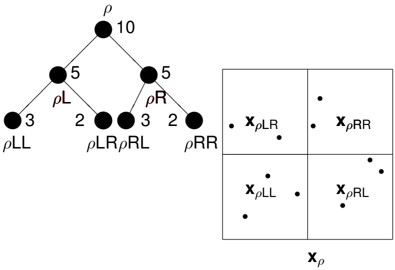


A Prioritized Queue based Algorithm

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



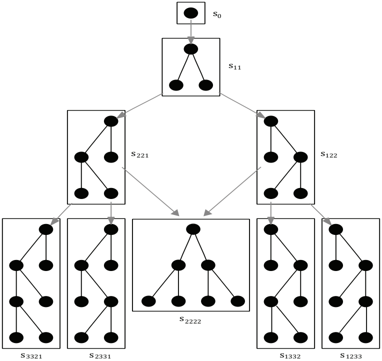
The SplitMostCounts Algorithm

Input: (i) data: $x_1, \ldots, x_n \subseteq \mathbb{R}^d$; (ii) root box: $\mathbf{x}_\rho$ // optional;
(iii) padding to handle pulsed data: $\psi \geq 0$ // optional;
(iv) S.E.B. max: $\overline{k}_n$; (v) maximum partition size: $\overline{m}_n$.
Output: histogram estimate $f_{n,s}$.

initialize $i \leftarrow 1$; $s \leftarrow \mathbf{x}_\rho + \psi$;
repeat until
  $\#\mathbf{x}_{\rho v} \leq \overline{k}_n$ for each $\mathbf{x}_{\rho v} \in \ell(s)$ and $i \leq \overline{m}_n$ //
$\ell(s) = \{$leaf boxes$\}$
  $\mathbf{x}_{\rho v} \leftarrow$ Uniform$(\hat{\ell}(s))$ // randomized PQ on leaf boxes
  $s \leftarrow$ bisect$(s, \mathbf{x}_{\rho v})$ // bisect leaf box $\mathbf{x}_{\rho v}$ of s
  recursively update counts in s;
  $i \leftarrow i + 1$;
return $f_{n,s}$.

Transition Diagram of Randomized PQ Markov chain

Let $\mathbb{S}_i$ 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 $\mathbb{S}_{i:j}$ be the set of RPs with k splits, $i \leq k \leq j$.



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

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

Let $X_1, X_2, \ldots$ be independent and identical random vectors in $\mathbb{R}^d$ whose common distribution μ has a non-atomic density f, i.e., $f \ll \lambda^d$. Let $\{S_n(i)\}_{i=0}^{\dot{I}}$ 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 $n \to \infty$, if $\overline{k}_n \to \infty$, $n^{-1}\overline{k}_n \to 0$, $\overline{m}_n \geq n/\overline{k}_n$, and $\overline{m}_n/n \to 0$ then the density estimate $f_{n,\dot{s}}$ is strongly consistent in L_1, i.e.

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

Proof Sketch

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

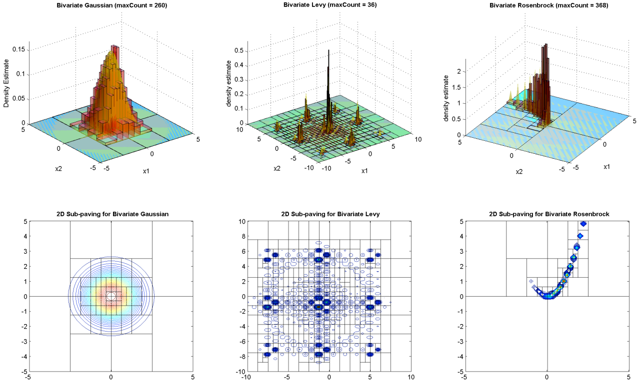
(a) $n^{-1} m(\mathcal{L}_n) \to 0$,
(b) $n^{-1} \log \Delta_n^*(\mathcal{L}_n) \to 0$, and
(c) $\mu(x : \mathrm{diam}(\mathbf{x}(x)) > \gamma) \to 0$ with probability 1 for every $\gamma > 0$,

are satisfied. Then by Theorem 1 of Lugosi and Nobel, 1996 our density estimate $f_{n,\dot{s}}$ 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

Some Examples

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



Choice of k_n

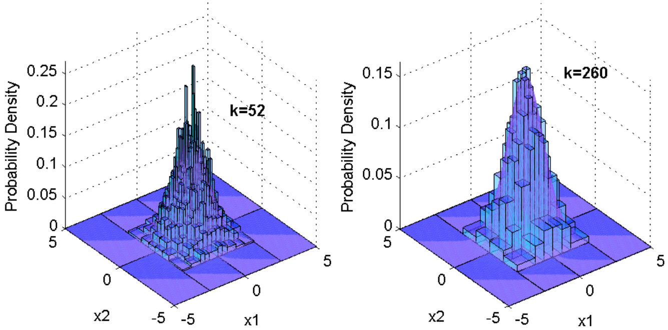
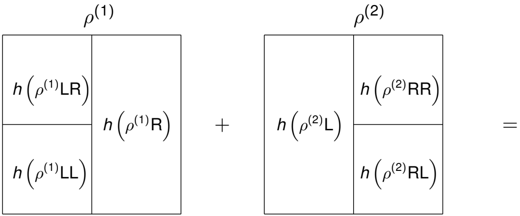


Figure : Two histogram density estimates for the standard bivariate gaussian density with different choices of k_n. The histogram is under-smoothed when k_n is relatively smaller than n and over-smoothed when k_n is relatively larger.

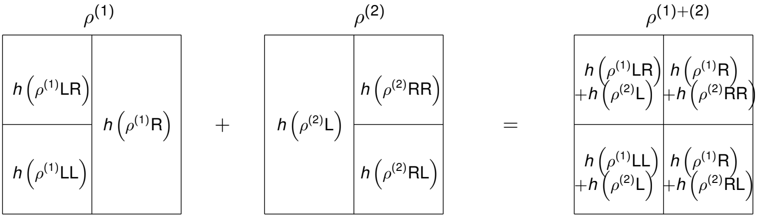
Adding and Averaging SRPs

Do a minimal union (or overlay) operation of $s^{(1)}$ and $s^{(2)}$ with root nodes $\rho^{(1)}$ and $\rho^{(2)}$, and add the value $h(\rho v)$ for each leaf box ρv:



Adding and Averaging SRPs

Do a minimal union (or overlay) operation of $s^{(1)}$ and $s^{(2)}$ with root nodes $\rho^{(1)}$ and $\rho^{(2)}$, and add the value $h(\rho v)$ for each leaf box ρv:



Adding and Averaging SRPs

Adding m SRP histogram density estimates

$$\sum_{i=1}^{m} f_{n,s^{(i)}} = f_{n,s^{(1)}} + f_{n,s^{(2)}} + f_{n,s^{(3)}} + \ldots + f_{n,s^{(m)}}$$

$$= \left(\left(\left(f_{n,s^{(1)}} + f_{n,s^{(2)}} \right) + f_{n,s^{(3)}} \right) + \ldots + f_{n,s^{(m)}} \right) \quad .$$

Adding and Averaging SRPs

Adding m SRP histogram density estimates

$$\sum_{i=1}^{m} f_{n,s^{(i)}} = f_{n,s^{(1)}} + f_{n,s^{(2)}} + f_{n,s^{(3)}} + \ldots + f_{n,s^{(m)}}$$
$$= \left(\left(\left(f_{n,s^{(1)}} + f_{n,s^{(2)}}\right) + f_{n,s^{(3)}}\right) + \ldots + f_{n,s^{(m)}}\right) \quad .$$

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

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

An Example

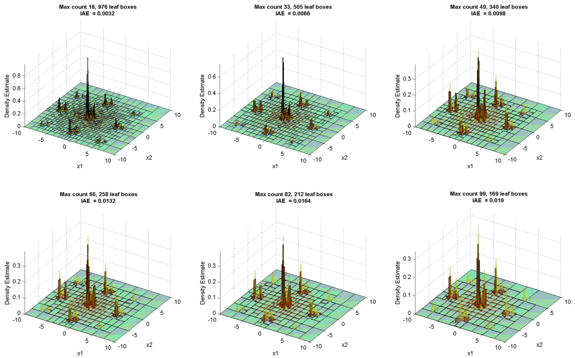


Figure : Histogram density estimates of the bivariate Levy using different values of $\overline{k}_n$.

An Example

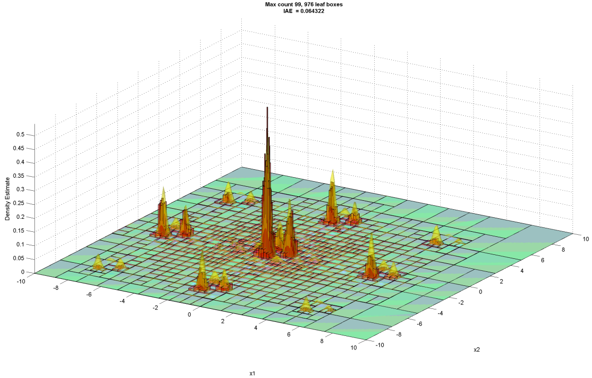


Figure : The averaged histogram density estimate.

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

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

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

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

- 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

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

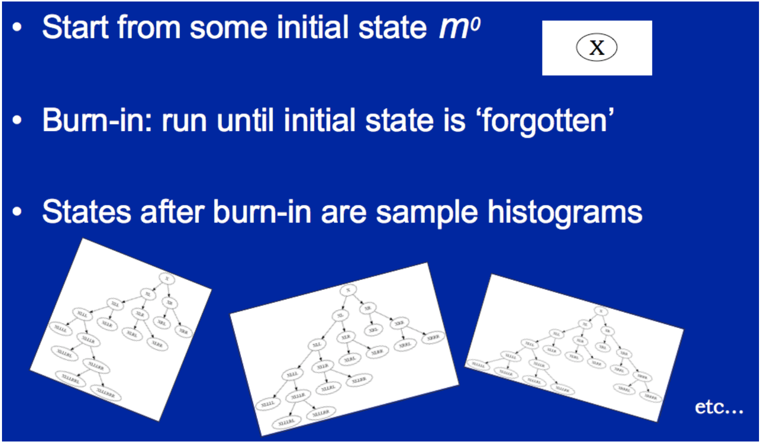
- Let the prior probability be $P(s) \propto \frac{1}{C_k^2}$, $s \in \mathbb{S}_{0:\infty}$
- Then the posterior density of histogram $\hat{f}_s$ with k splits is

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

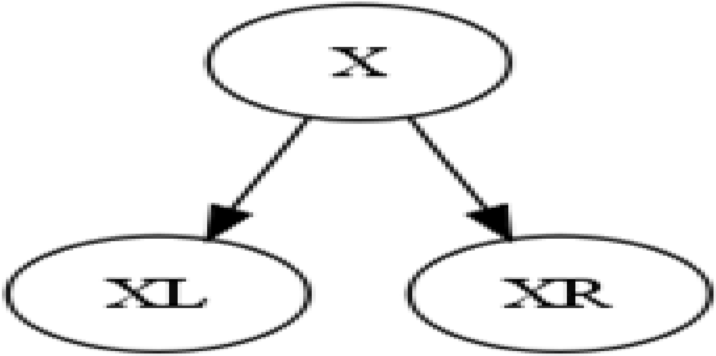
Metropolis-Hastings Algorithm

- Use a proposal density $q(s^{prime}|s^{(i)})$ 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** $u < \frac{P(\hat{f}_{s'}|X_{1:n})}{P(\hat{f}_{s^{(i)}}|X_{1:n})} \frac{q(s^{(i)}|s')}{q(s^{(prime)}|s^{(i)})}$ **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 $\mathbb{S}_{0:\infty}$

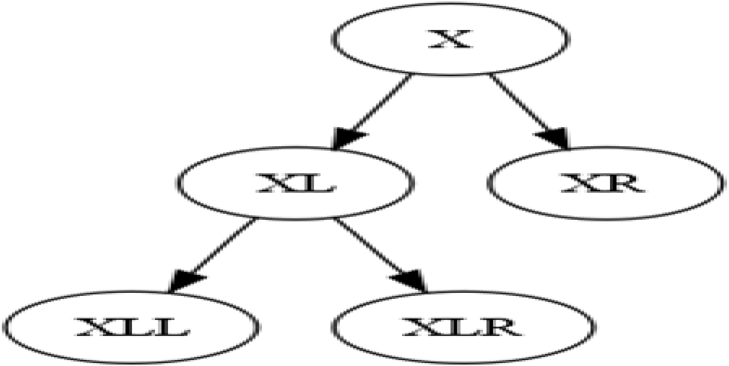
Metropolis-Hastings Algorithm



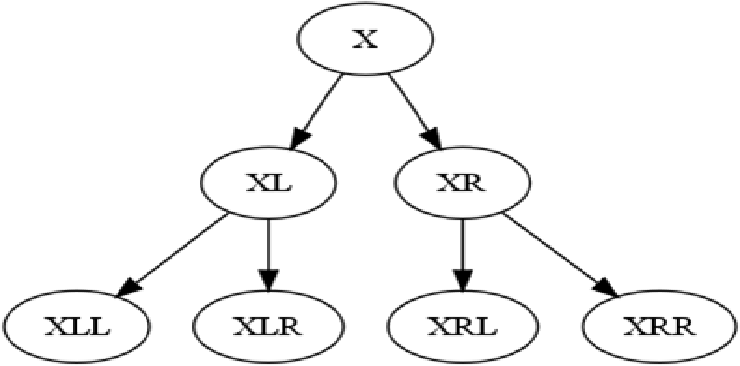
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



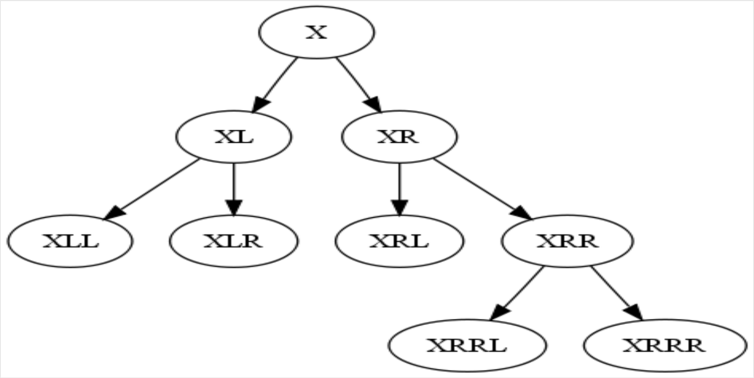
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



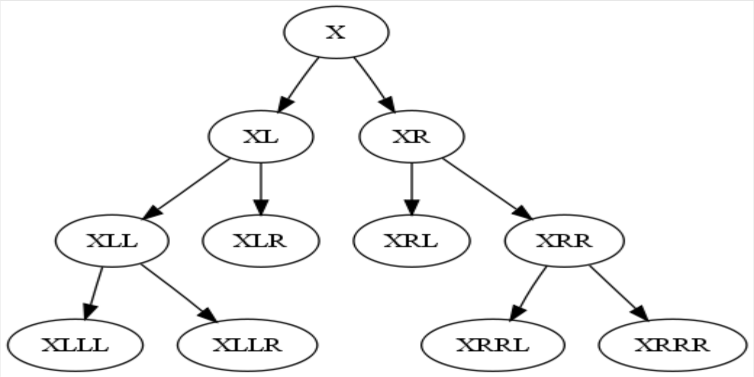
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



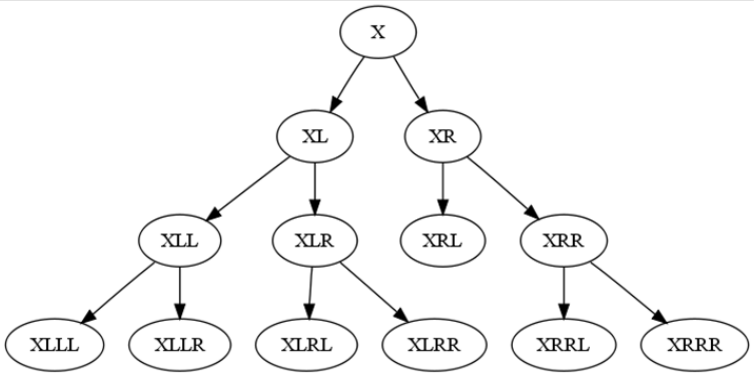
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



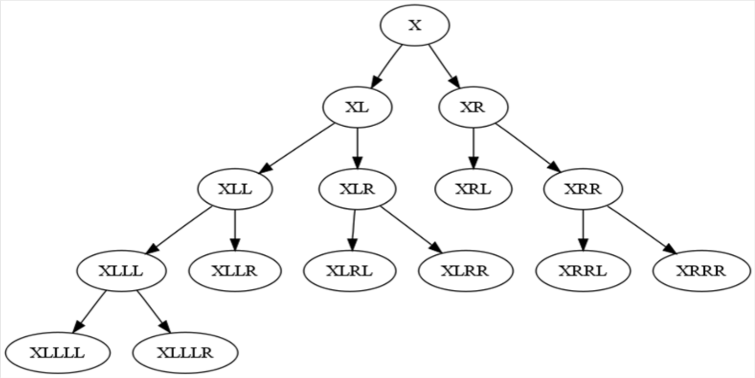
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



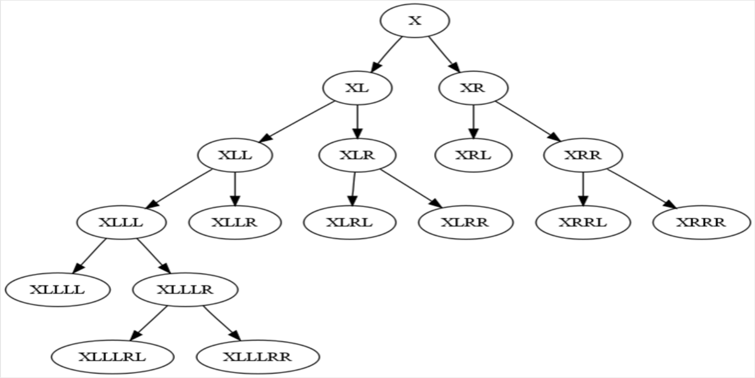
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



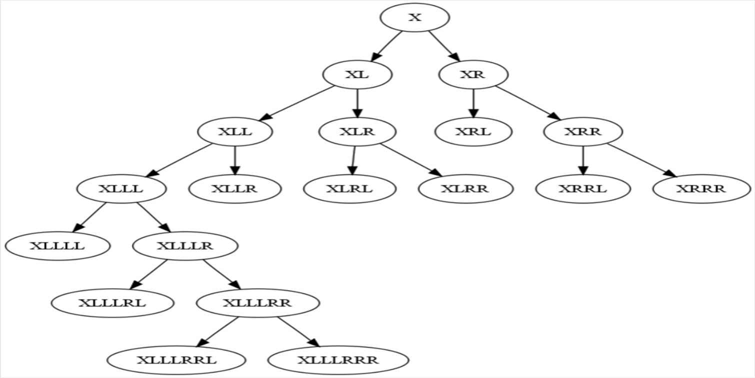
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



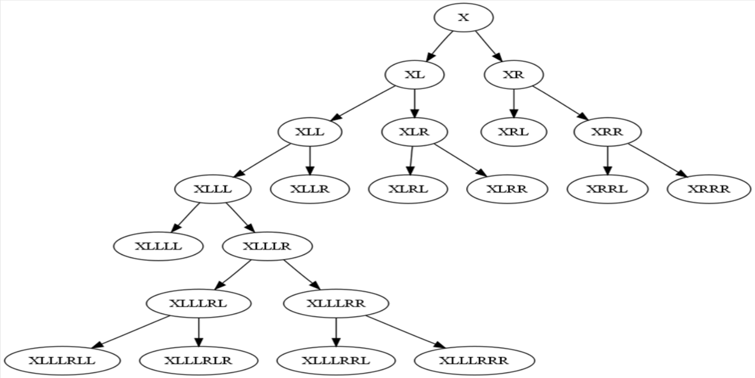
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



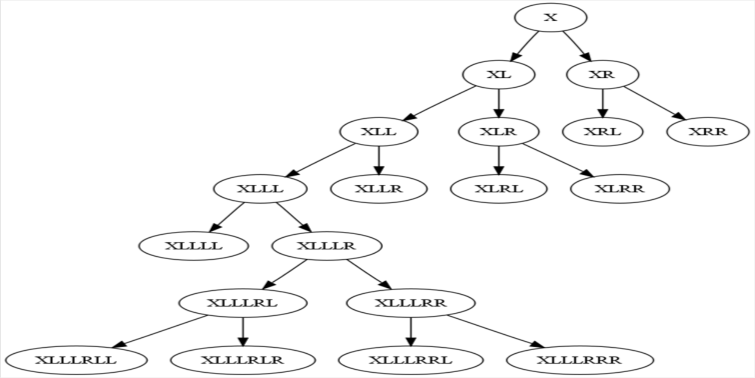
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



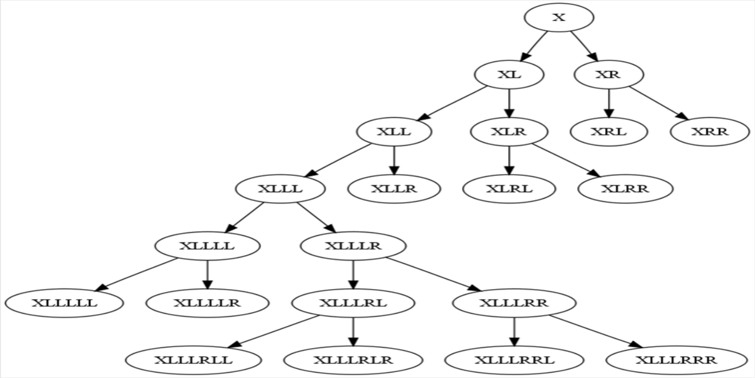
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



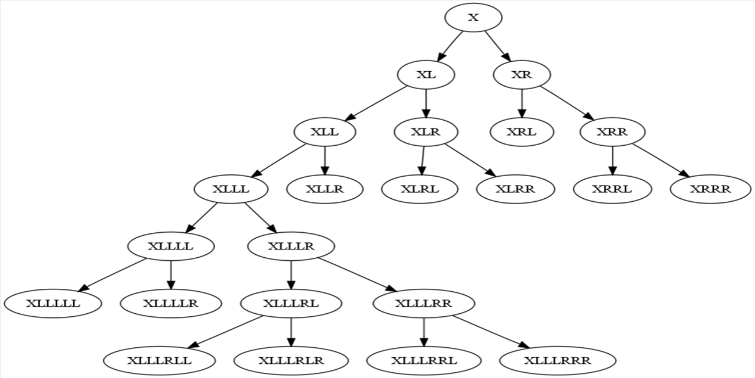
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



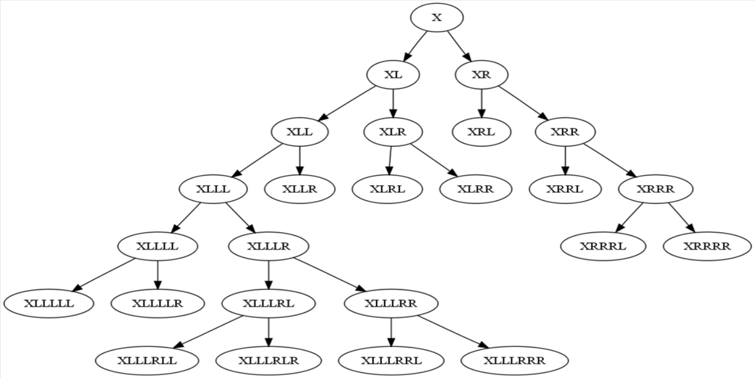
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



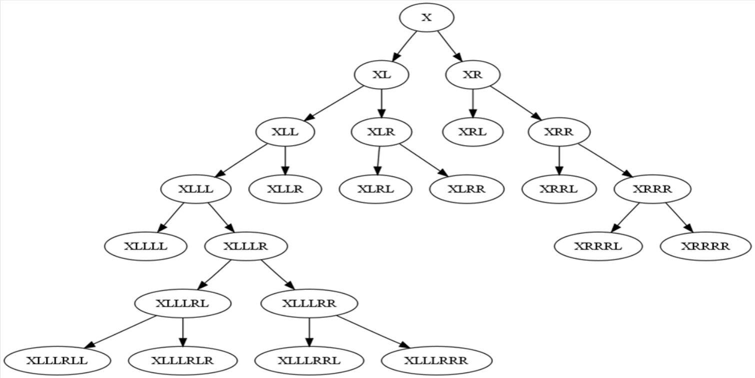
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



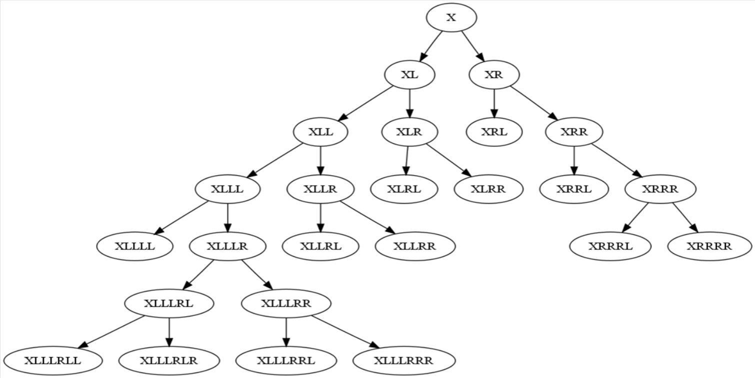
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



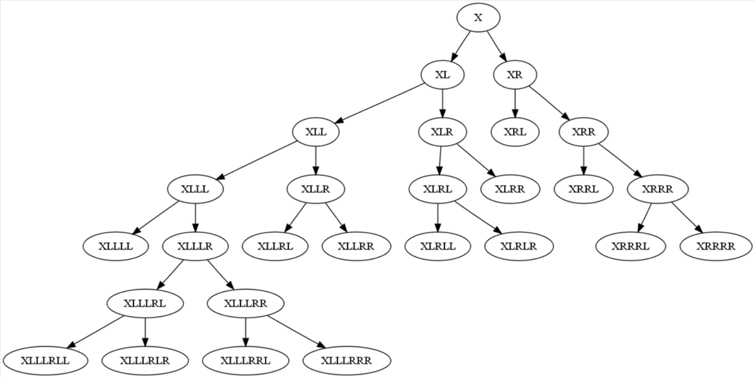
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



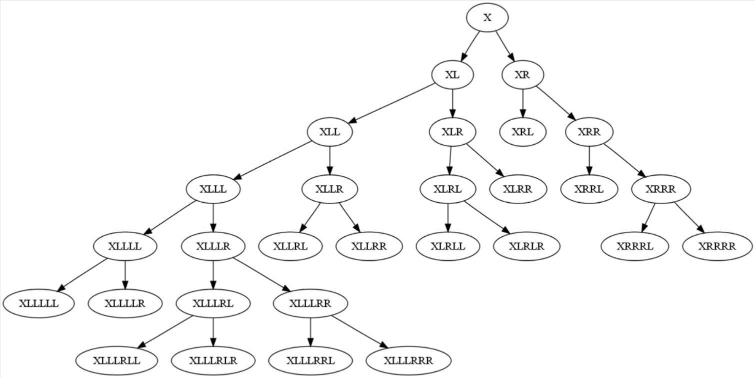
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



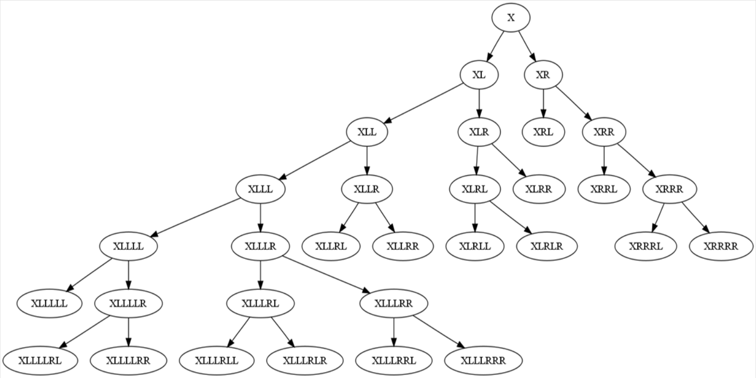
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



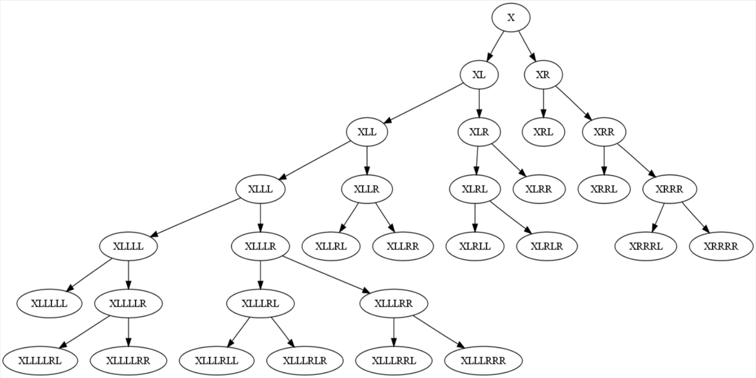
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



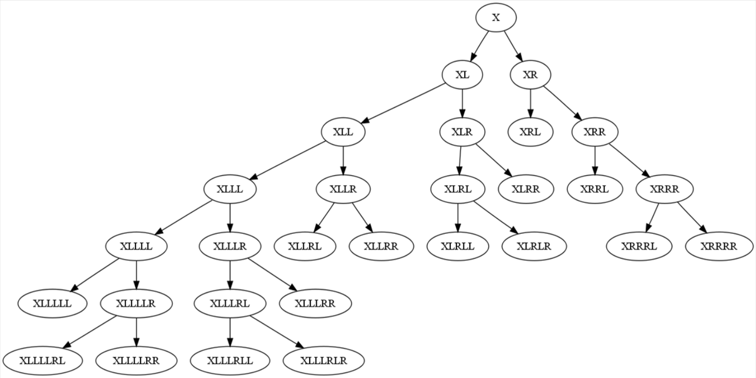
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



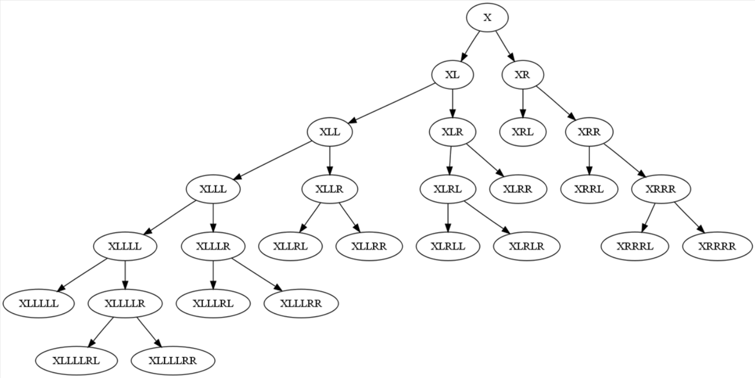
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



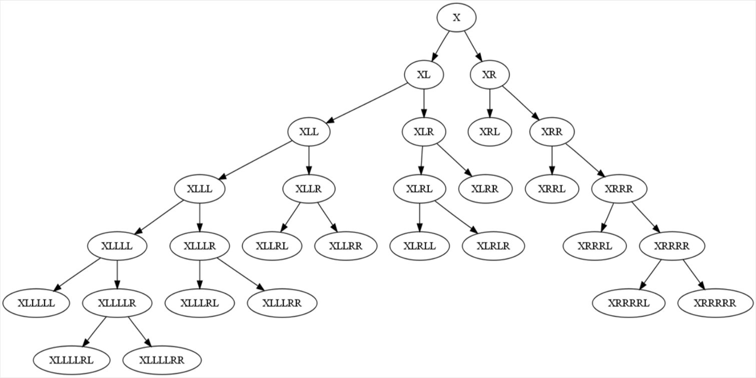
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



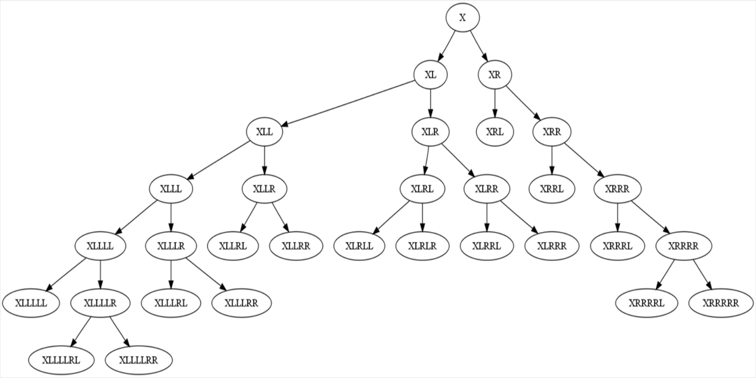
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



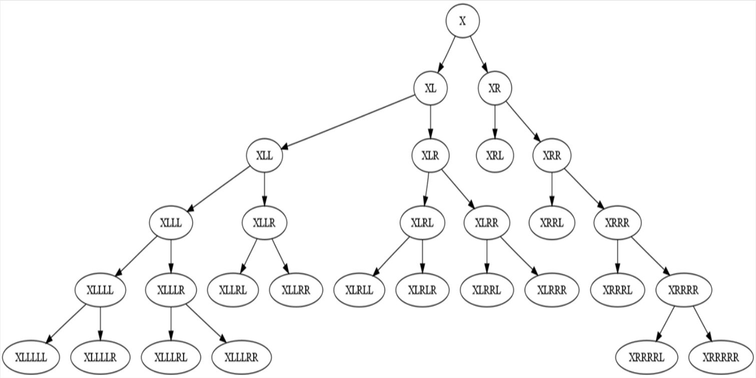
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



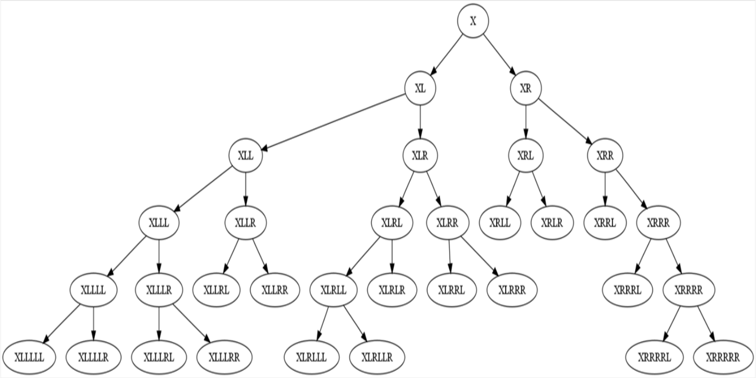
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



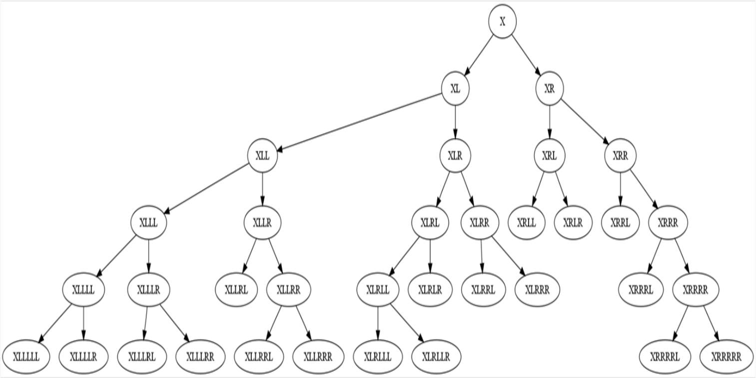
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



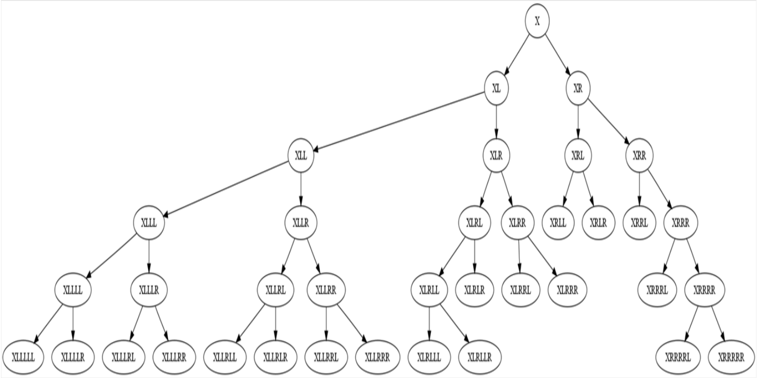
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



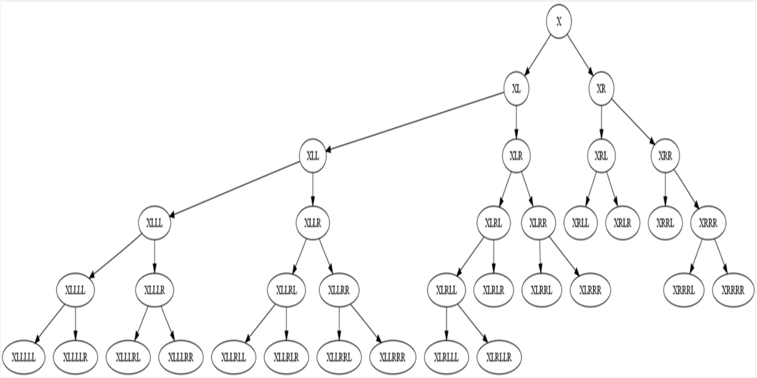
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



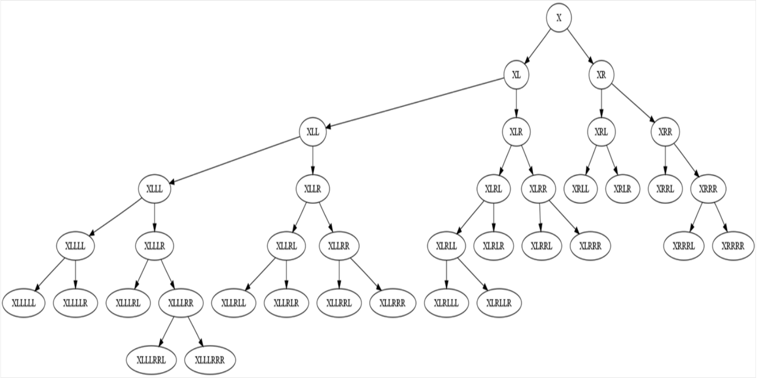
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



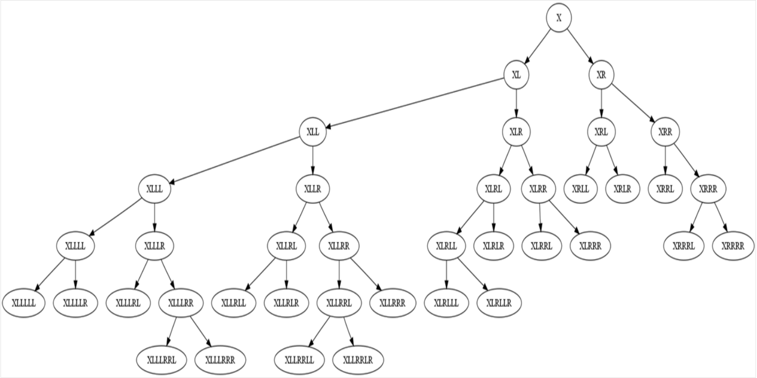
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



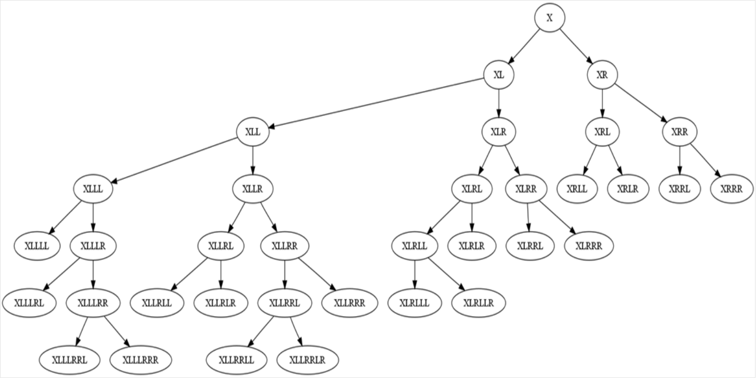
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



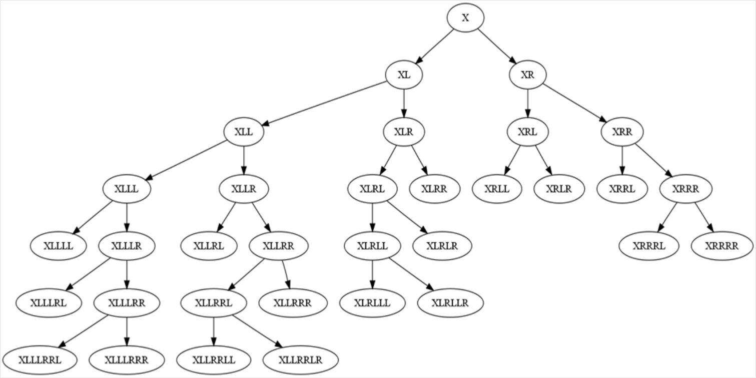
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



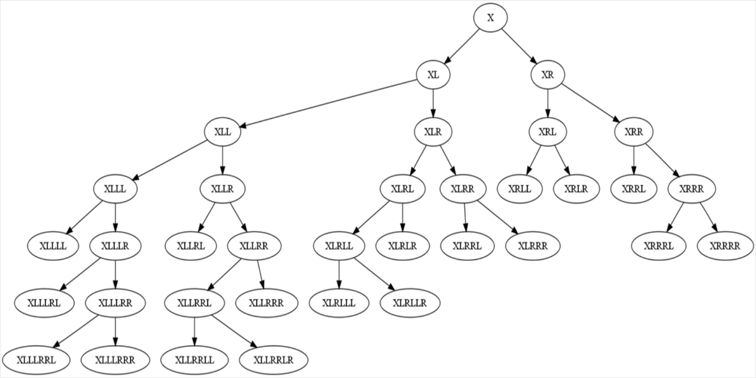
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



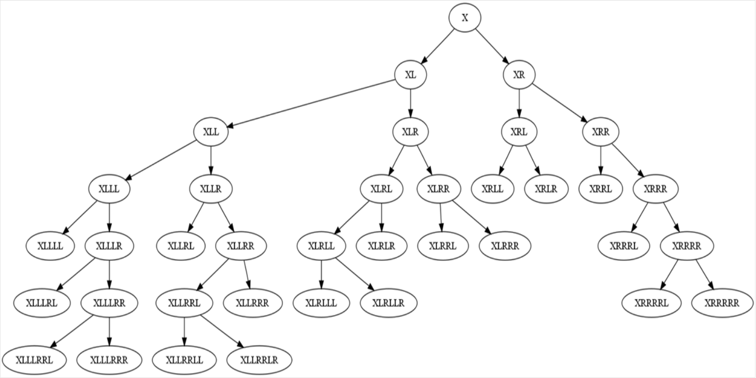
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



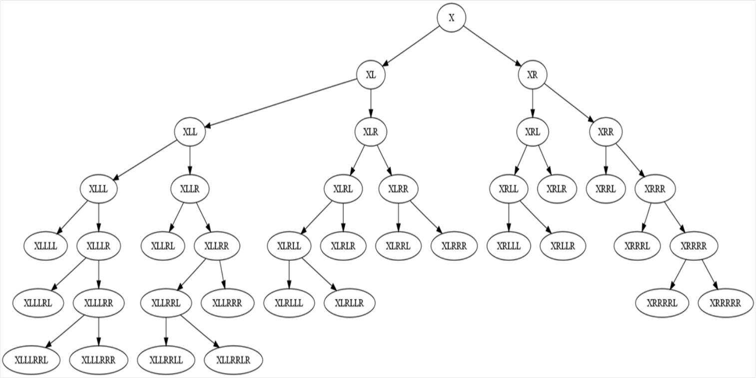
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



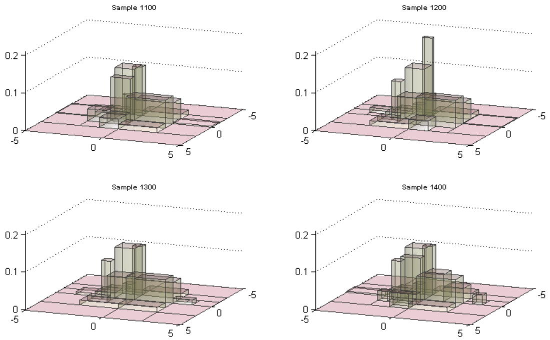
Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$



Monte Carlo Markov Chain over Histograms in $\mathbb{S}_{0:\infty}$

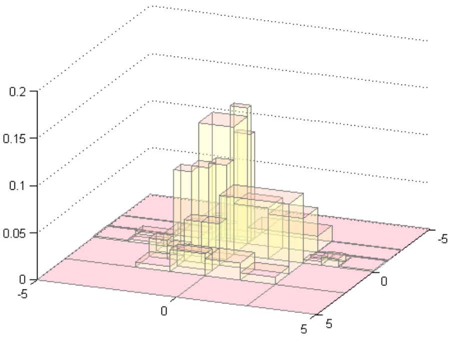


Histogram Estimates - Standard Bivariate Gaussian



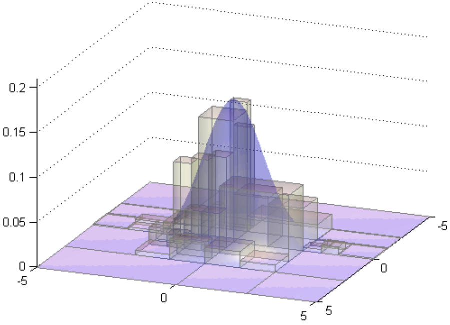
Four sample histograms

Histogram Estimates - Standard Bivariate Gaussian



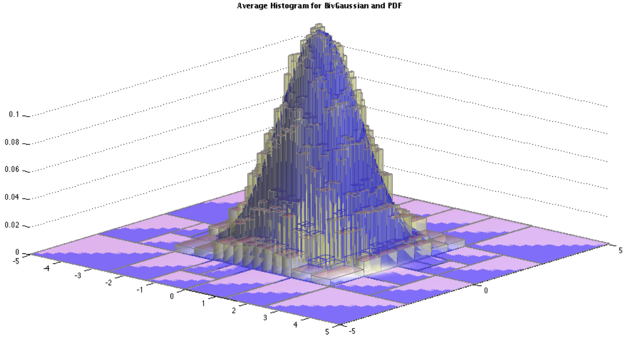
Average of the four sampled histograms

Histogram Estimates - Standard Bivariate Gaussian



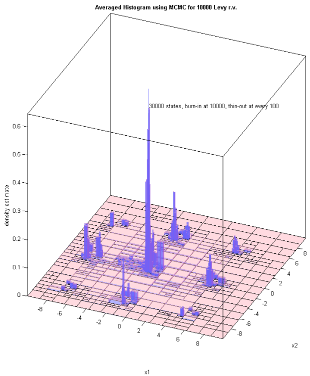
Average of the four sampled histograms with Gaussian PDF

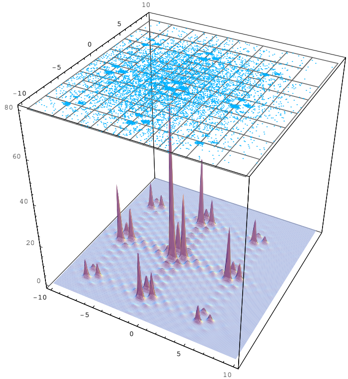
Histogram Estimates - Standard Bivariate Gaussian



A much better estimate

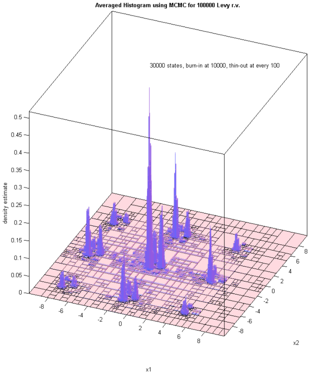
Histogram Estimates - Bivariate Levy Density

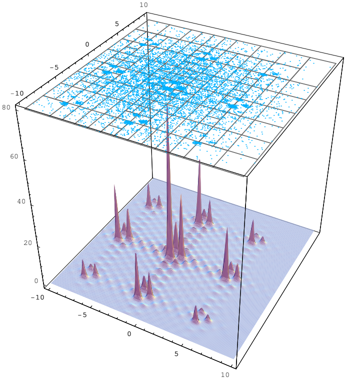




Data points = 10000, Number of states = 30000, Burn-in = 10000, Thin-out = 100, Averaged over 201 states, Time taken = 14.16s

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 and `SplitMostCounts` PQ

MIAE (std. err.) for *n* samples from uniform density in various dimensions (CPU Times < O(1minute)).

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 $\mathbb{S}_\Lambda$).

Λ	n	Standard Gaussian densities 1D	Standard Gaussian densities 2D	Standard Gaussian densities 10D	Rosenbrock densities 2D	Rosenbrock densities 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)	0.0323 (0.0511)
	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)	0.1941 (0.0820)
	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)

Minimum Distance Estimation (MDE)

- Let θ be the current number of splits in `SplitMostCounts`.
- Let $f_{n,\theta}$ be the histogram estimate with corresponding SRP $s \in \mathbb{S}_\theta$, $\int f_{n,\theta} = 1$.
- Denote Θ as the set of the number of splits such that $\Theta := \{0, \ldots, \overline{m}_n - 1\}$ where $\overline{m}_n - 1$ is the maximum number of splits allowed.

Minimum Distance Estimation (MDE)

- Let θ be the current number of splits in `SplitMostCounts`.
- Let $f_{n,\theta}$ be the histogram estimate with corresponding SRP $s \in \mathbb{S}_\theta$, $\int f_{n,\theta} = 1$.
- Denote Θ as the set of the number of splits such that $\Theta := \{0, \ldots, \overline{m}_n - 1\}$ where $\overline{m}_n - 1$ is the maximum number of splits allowed.

The goal is to select the optimal estimate amongst the $|\Theta|$ candidates, $f_{n,\theta}, \theta \in \Theta$ by using a hold-out method proposed by Devroye and Lugosi, 2004 for minimum distance estimation (MDE).

Minimum Distance Estimation (MDE)

Let $0 < \varphi < 1/2$. Given n points, use $n - \varphi n$ points as the training set and the remaining φn points as the validation set.

Minimum Distance Estimation (MDE)

Let $0 < \varphi < 1/2$. Given n points, use $n - \varphi n$ points as the training set and the remaining φn points as the validation set.

Scheffé set

For a pair $(\theta, \theta'), \theta, \theta' \in \Theta, \theta \neq \theta'$, the Scheffé set is

$$A_{\theta,\theta'} := A\left(f_{n-\varphi n,\theta}, f_{n-\varphi n,\theta'}\right) = \left\{x : f_{n-\varphi n,\theta}(x) > f_{n-\varphi n,\theta'}(x)\right\} \quad .$$

Minimum Distance Estimation (MDE)

Let $0 < \varphi < 1/2$. Given n points, use $n - \varphi n$ points as the training set and the remaining φn points as the validation set.

Scheffé set

For a pair $(\theta, \theta'), \theta, \theta' \in \Theta, \theta \neq \theta'$, the Scheffé set is

$$A_{\theta,\theta'} := A\left(f_{n-\varphi n,\theta}, f_{n-\varphi n,\theta'}\right) = \left\{x : f_{n-\varphi n,\theta}(x) > f_{n-\varphi n,\theta'}(x)\right\} \quad .$$

Yatracos class

The class of all sets of the form $A_{\theta,\theta'}$:

$$\mathcal{A}_\Theta := \left\{\left\{x : f_{n-\varphi n,\theta}(x) > f_{n-\varphi n,\theta'}(x)\right\} : \theta, \theta' \in \Theta, \theta \neq \theta'\right\} \quad .$$

Minimum Distance Estimation (MDE)

Minimum distance estimate

The minimum distance estimate $f_{n-\varphi n,\theta^*}$ is the density estimate $f_{n-\varphi n,\theta}$ of smallest index θ^* that minimizes

$$\Delta_\theta = \sup_{A \in \mathcal{A}_\Theta} \left| \int_A f_{n-\varphi n,\theta}(A) - \mu_{\varphi n}(A) \right|$$

where $\mu_{\varphi n}$ is the empirical measure of the validation set $X_{n-\varphi n+1}, \ldots, X_n$.

Minimum Distance Estimation (MDE)

Let Θ be the set of the number of splits such that $\Theta := \{0, \ldots, \overline{m}_n - 1\}$ where $\overline{m}_n - 1$ is the maximum number of splits allowed.

Every time a split happens during `SplitMostCounts`,

- Obtain the Yatracos class for the current split θ;

Minimum Distance Estimation (MDE)

Let Θ be the set of the number of splits such that $\Theta := \{0, \ldots, \overline{m}_n - 1\}$ where $\overline{m}_n - 1$ is the maximum number of splits allowed.
Every time a split happens during `SplitMostCounts`,

- Obtain the Yatracos class for the current split θ;
- Compute $\Delta_\theta = \sup_{A \in \mathcal{A}_\Theta} \left| \int_A f_{n-\varphi n,\theta}(A) - \mu_{\varphi n}(A) \right|$ for each $\theta \in \Theta$.

Minimum Distance Estimation (MDE)

Let Θ be the set of the number of splits such that $\Theta := \{0, \ldots, \overline{m}_n - 1\}$ where $\overline{m}_n - 1$ is the maximum number of splits allowed.
Every time a split happens during `SplitMostCounts`,

- Obtain the Yatracos class for the current split θ;
- Compute $\Delta_\theta = \sup_{A \in \mathcal{A}_\Theta} \left| \int_A f_{n-\varphi n,\theta}(A) - \mu_{\varphi n}(A) \right|$ for each $\theta \in \Theta$.

The candidate estimate that minimizes Δ_θ is the minimum distance estimate.

Setting up MDE

- Need to track statistics for both the training and validation data - recursively computable statistics for validation data;

Setting up MDE

- Need to track statistics for both the training and validation data - recursively computable statistics for validation data;
- Get the Yatracos class for the current Θ - use the RP as a collator (via non-minimal union) to track and compare the histogram estimate at each leaf box of each candidate.

Recursively Computable Statistics for Validation Data

- The training data $\mathcal{T} := \{x_1, \ldots, x_{n-\varphi n}\}$ drive the randomized priority queue RPQ to form an SRP s.
- The validation data $\mathcal{V} := \{x_{n-\varphi n+1}, \ldots, x_n\}$ trickle through s and stay in the boxes of s that contain the data.

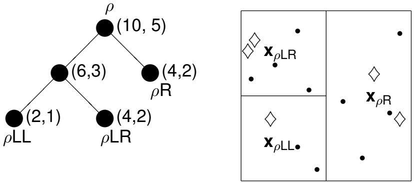


Figure : An SRP s with training ($\bullet$) and validation data ($\diamond$) and their respective sample counts ($\#\mathbf{x}_{\rho v}, \check{\#}\mathbf{x}_{\rho v}$) that are updated recursively as data fall through the nodes of s.

Recursively Computable Statistics for Validation Data

MDE requires the histogram estimate from the training data and the empirical mass of the validation data:

Histogram estimate obtained from the set of training data

$$f_{n-\varphi n}(\rho v) = \frac{\#\mathbf{x}_{\rho v}}{n \cdot \mathrm{vol}\left(\mathbf{x}_{\rho v}\right)} \quad .$$

The empirical measure of the validation data

$$\mu_{\varphi n}(\mathbf{x}_{\rho v}) = \frac{\check{\#}\mathbf{x}_{\rho v}}{\varphi n} \quad .$$

Recursively Computable Statistics for Validation Data

The training data and the empirical mass of the validation data can be tracked as $(f_{n-\varphi n,\theta}(\rho v), \mu_{\varphi n}(\mathbf{x}_{\rho v}))$ at each leaf node.

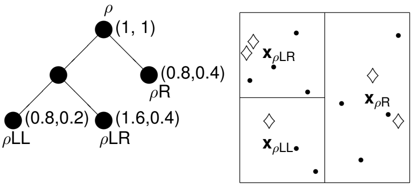


Figure : Tracking the histogram estimate and empirical mass at each node of an SRP s built from a unit square.

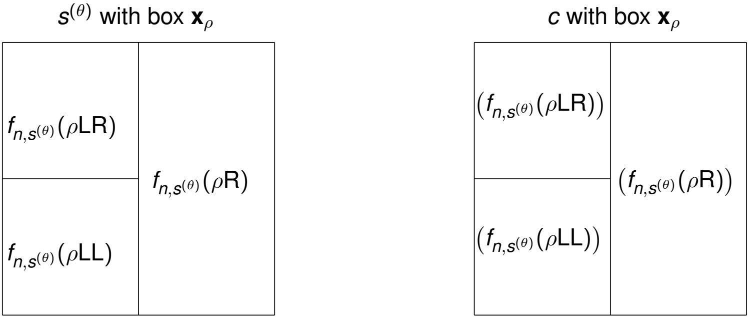
RP as a Collator

- Need to compare the estimates between each SRP $s^{(i)}, 1 < i < \infty$.
- An efficient way to approach this is to collate these histograms onto an *collator regular paving* or *CRP* where the space of CRP is also $\mathbb{S}_{0:\infty}$.
- Consider two SRPs $s^{(\theta)}$ and $s^{(\theta')}$ that have the same mother box and for which the corresponding histogram estimates $f_{n,s^{(\theta)}}$ and $f_{n,s^{(\theta')}}$ are computed.
- By collating the two SRPs we get a CRP c that stores $f_{n,s^{(\theta)}}$ and $f_{n,s^{(\theta')}}$ for each node ρv of c, such that each node ρv has a vector $\boldsymbol{f}_{n,c}(\rho v) := (f_{n,s^{(\theta)}}(\rho v), f_{n,s^{(\theta')}}(\rho v))$.

RP as a Collator

Collating two SRPs $s^{(\theta)}$ and $s^{(\theta')}$ with the same root box $\mathbf{x}_\rho$:

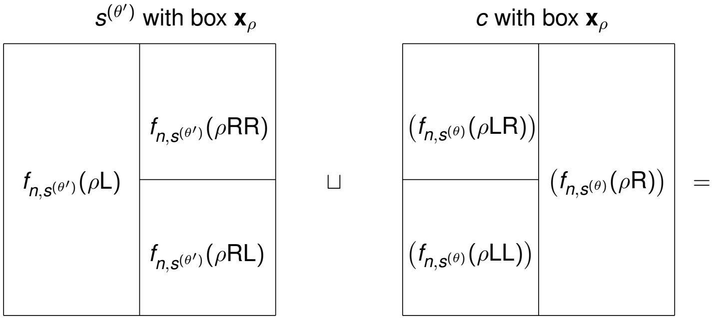
Figure : Make the SRP $s^{(\theta)}$ into a CRP c.



RP as a Collator

Collating two SRPs $s^{(\theta)}$ and $s^{(\theta')}$ with the same root box $\mathbf{x}_\rho$:

Figure : Collate the SRP $s^{(\theta')}$ onto c.



RP as a Collator

Collating two SRPs $s^{(\theta)}$ and $s^{(\theta')}$ with the same root box $\mathbf{x}_\rho$:

Figure : Collate the SRP $s^{(\theta')}$ onto c.

c with box $\mathbf{x}_\rho$

$\begin{pmatrix} f_{n,s^{(\theta)}}(\rho LR) \\ f_{n,s^{(\theta')}}(\rho L) \end{pmatrix}$	$\begin{pmatrix} f_{n,s^{(\theta)}}(\rho R) \\ f_{n,s^{(\theta')}}(\rho RR) \end{pmatrix}$
$\begin{pmatrix} f_{n,s^{(\theta)}}(\rho LL) \\ f_{n,s^{(\theta')}}(\rho L) \end{pmatrix}$	$\begin{pmatrix} f_{n,s^{(\theta)}}(\rho R) \\ f_{n,s^{(\theta')}}(\rho RL) \end{pmatrix}$

Minimum Distance Estimation (MDE): An Example

First candidate: s_0, $\theta = 0$.
Second candidate: $s_{1,1}$, $\theta = 1$.
So $\Theta = \Theta_1 = \{0, 1\}$.
Collate the two candidates:

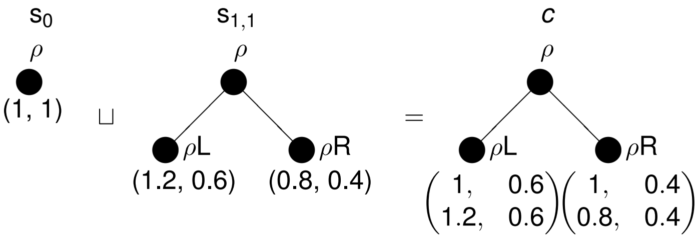
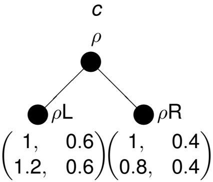


Figure : Collating s_0 and $s_{1,1}$.

Minimum Distance Estimation (MDE): An Example

Compare $f_{n-\varphi n,\theta}, \theta \in \Theta$ for each leaf box $\mathbf{x}_{\rho v} \in \ell(c)$.



At $\mathbf{x}_{\rho \mathrm{L}}$,

$$f_{n-\varphi n,\theta=0}(\mathbf{x}_{\rho \mathrm{L}}) = 1 < f_{n-\varphi n,\theta=1}(\mathbf{x}_{\rho \mathrm{L}}) = 1.2 \quad .$$

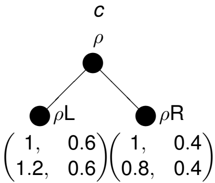
But,

$$f_{n-\varphi n,\theta=1}(\mathbf{x}_{\rho \mathrm{L}}) > f_{n-\varphi n,\theta=0}(\mathbf{x}_{\rho \mathrm{L}}) \quad .$$

Thus $A_{\theta=1,\theta'=0} = \{\mathbf{x}_{\rho \mathrm{L}}\}$ and is in the Yatracos class A_{Θ_1}.

Minimum Distance Estimation (MDE): An Example

Compare $f_{n-\varphi n,\theta}, \theta \in \Theta$ for each leaf box $\mathbf{x}_{\rho v} \in \ell(c)$.



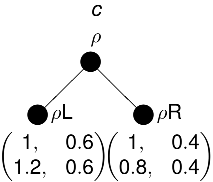
At $\mathbf{x}_{\rho \mathrm{R}}$,

$$f_{n-\varphi n,\theta=0}(\mathbf{x}_{\rho \mathrm{R}}) = 1 > f_{n-\varphi n,\theta=1}(\mathbf{x}_{\rho \mathrm{R}}) = 0.8 \quad .$$

Thus $A_{\theta=0,\theta'=1} = \{\mathbf{x}_{\rho \mathrm{R}}\}$ and is in the Yatracos class A_{Θ_1}.
Finally, we have $\mathcal{A}_{\Theta_1} = \{\mathbf{x}_{\rho \mathrm{L}}, \mathbf{x}_{\rho \mathrm{R}}\}$.

Minimum Distance Estimation (MDE): An Example

Compare $f_{n-\varphi n,\theta}, \theta \in \Theta$ for each leaf box $\mathbf{x}_{\rho v} \in \ell(c)$.



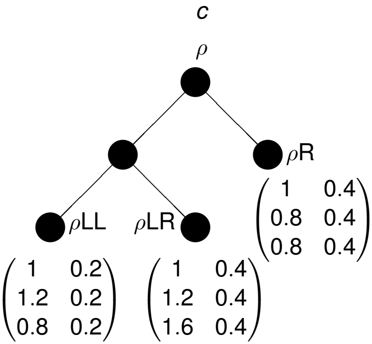
Can also express $\mathcal{A}_{\Theta_1}$ in a matrix as follows:

$$\mathcal{A}_{\Theta_1} = \begin{pmatrix} A_{0,0} & A_{0,1} \\ A_{1,0} & A_{1,1} \end{pmatrix} = \begin{pmatrix} \emptyset & \mathbf{x}_{\rho \mathrm{R}} \\ \mathbf{x}_{\rho \mathrm{L}} & \emptyset \end{pmatrix} . \qquad (1)$$

Note that the diagonal elements are all empty sets because there are no comparisons for the set $\{\theta, \theta'\}$ where $\theta = \theta'$.

Minimum Distance Estimation (MDE): An Example

Make another split at node ρL to produce the nodes ρLL and ρLR and get the SRP s_{221}. Perform collation to get the following collator c:



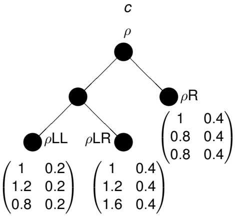
Minimum Distance Estimation (MDE): An Example

Current Θ is now $\Theta = \{0, 1, 2\}$. To update the Yatracos class:

- Denote $\mathbf{x}^*_{\rho v}$ as the leaf box that is being split currently.
- Since only one leaf box is split every time, instead of comparing the estimates $f_{n-\varphi n,\theta}(\mathbf{x}_{\rho v})$ at each leaf box, we need only compare the estimates of its sub-boxes $\{\mathbf{x}^*_{\rho \mathrm{L}}, \mathbf{x}^*_{\rho \mathrm{R}}\}$ to update the Yatracos class.
- Thus we need not check the estimates at all the leaf boxes.

Minimum Distance Estimation (MDE): An Example

Here $\mathbf{x}^*_{\rho v} = \mathbf{x}_{\rho \mathrm{L}}$. Only need to compare the estimates for each θ at the leaf boxes $\mathbf{x}_{\rho \mathrm{LL}}$ and $\mathbf{x}_{\rho \mathrm{LR}}$.



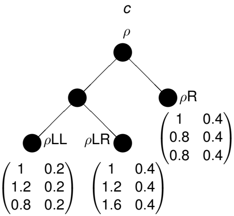
Compare the estimates of the pair $(0, 2)$ at $\mathbf{x}_{\rho \mathrm{LL}}$:

$$f_{n-\varphi n, \theta=0}(\mathbf{x}_{\rho \mathrm{LL}}) = 1 > f_{n-\varphi n, \theta'=2}(\mathbf{x}_{\rho \mathrm{LL}}) = 0.8 \quad .$$

Thus $\mathbf{x}_{\rho \mathrm{LL}}$ will be taken into the set $A_{0,2}$.

Minimum Distance Estimation (MDE): An Example

Here $\mathbf{x}^*_{\rho v} = \mathbf{x}_{\rho \mathrm{L}}$. Only need to compare the estimates for each θ at the leaf boxes $\mathbf{x}_{\rho \mathrm{LL}}$ and $\mathbf{x}_{\rho \mathrm{LR}}$.



From $A_{0,1}$ of Equation 1, we also know that $f_{n-\varphi n,\theta=0}(\mathbf{x}_{\rho \mathrm{R}})$ is larger than $f_{n-\varphi n,\theta=2}(\mathbf{x}_{\rho \mathrm{R}})$. This will also be true for the pair $(0, 2)$. Thus the set $A_{0,2}$ is $\mathbf{x}_{\rho \mathrm{R}} \cup \mathbf{x}_{\rho \mathrm{LL}}$.

Minimum Distance Estimation (MDE): An Example

- If $A_{0,1}$ was a box that was split, no unions will be taken with this box since it is no longer a leaf box.
- In general, at some leaf box $\mathbf{x}_{\rho v}$ for which its estimates are being compared, for any pair $(\theta, \theta'), \theta \neq \theta'$, if $A_{\theta,\theta'} \neq \{\mathbf{x}_{\rho v^*}\}$, where $\mathbf{x}_{\rho v^*}$ was the box being split, we will take the union of $\mathbf{x}_{\rho v}$ with the elements of the set $A_{\theta-1,\theta'}$.
- Besides, since the sub-boxes will have the same estimate as its parent box, it will be redundant to make comparisons for the pair $\{0,1\}$.
- Therefore instead of doing $\binom{3}{2}$ comparisons, we now only need $\binom{2}{1}$ comparisons, i.e. comparing the estimates at the pairs $\{0,2\}$ and $\{1,2\}$.

Minimum Distance Estimation (MDE): An Example

Continue the comparisons for the other sub-box $\mathbf{x}_{\rho \mathrm{LR}}$ and for all remaining comparison pairs.
The final Yatracos class $\mathcal{A}_{\Theta_2}$ is:

$$\mathcal{A}_{\Theta_2} = \begin{pmatrix} A_{0,0} & A_{0,1} & A_{0,2} \\ A_{1,0} & A_{1,1} & A_{1,2} \\ A_{2,0} & A_{2,1} & A_{2,2} \end{pmatrix} = \begin{pmatrix} \emptyset & \mathbf{x}_{\rho \mathrm{R}} & \mathbf{x}_{\rho \mathrm{R}} \cup \mathbf{x}_{\rho \mathrm{LL}} \\ \mathbf{x}_{\rho \mathrm{L}} & \emptyset & \mathbf{x}_{\rho \mathrm{LL}} \\ \mathbf{x}_{\rho \mathrm{LR}} & \mathbf{x}_{\rho \mathrm{LR}} & \emptyset \end{pmatrix} .$$

Minimum Distance Estimation (MDE): An Example

The corresponding Δ_θ values for $\mathcal{A}_{\Theta_1}$ and $\mathcal{A}_{\Theta_2}$:

	$\Delta_\theta = \left| \int_{\mathbf{x}_{\rho v}} f_{n-\varphi n} - \mu_{\varphi n}(\mathbf{x}_{\rho v}) \right|$	
θ	$\Theta_1 = \{0, 1\}$	$\Theta_2 = \{0, 1, 2\}$
$\theta = 0$	0.1	0.15
$\theta = 1$	0	0.1
$\theta = 2$	-	**0**

Table : Table of Δ_θ values.

- Take $\mathcal{A}_{\Theta_2}$ to be the final Yatracos class;
- Observe the column associated with Θ_2;
- The minimum distance estimate $f_{n-\varphi n,\theta^*}$ is the estimate at $\theta = 2$ since $\Delta_{\theta=2}$ is the minimum over all $\theta \in \Theta_2$.

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 - averaging histograms and other *arithmetics* on $\mathbb{S}_{0:\infty}$.

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 - averaging histograms and other *arithmetics* on $\mathbb{S}_{0:\infty}$.
- Smoother posterior mean from MCMC samples on the space of adaptive multi-variate histograms with partitions in $\mathbb{S}_{0:\infty}$. 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 - averaging histograms and other *arithmetics* on $\mathbb{S}_{0:\infty}$.
- Smoother posterior mean from MCMC samples on the space of adaptive multi-variate histograms with partitions in $\mathbb{S}_{0:\infty}$. NFL: MCMC convergence issues exist!
- Higher (1000) dimensional densities can be estimated fast and rough (but L_1-consistent) with the approach especially with `SplitMostCounts` PQ and further decisions can be done with appropriate *mapped RP arithmetic* over $\mathbb{S}_{0:\infty}$.

References

Devroye, L., Gyorfi, L. and Lugosi, G. (1996). *A probabilistic theory of pattern recognition.* New York: Springer.

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.

Sainudiin, R. and York, T. L. (2005). *An Auto-validating Rejection Sampler*. BSCB Dept. Technical Report BU-1661-M, Cornell University, Ithaca, New York.

Tucker, W. (2004). *Auto-validating numerical methods*. Lecture Notes, Uppsala University, Sweden.

Gratitude for Operating Fiscal Environment

2012, NZD 8300, Maths & Stats Dept. research travel grant and College of Engineering Strategic research grant,Univ. of Canterbury, Christchurch, New Zealand

2012, 2-day Visitor, Stochastic Networks Workshop (organizer Ilze Ziedins), Auckland, New Zealand

2011, 2-week Visitor, Advanced Learning and Evolutionary Algorithms Group (ALEA), INRIA, Bordeaux, France

2010, 6-month Sabbatical Award, College of Engineering, U. of Canterbury, Christchurch, New Zealand

2010, 6-month Visiting Scientist, Theoretical Stats. & Maths. Unit, Indian Statistical Institute, Bengaluru, India

2009, 2-week Visitor, Computer-Aided Proofs in Analysis Group (CAPA), Angström Laboratory, Uppsala, Sweden

2009, NZD 8000, External Consulting Revenues from New Zealand Ministry of Tourism

Remark: 8 unsuccessful applications to NZ funds (3 Marsden fast-starts) :(P – due to lack of “track record”...

This work in Computational Mathematics and Statistics is based on: 1,800,101 ASCII characters $\sim$ 21 days of non-stop coding at 1 character per second in GPL licensed *mrs: a C++ class library for statistical set processing* as of Thu Apr 12 08:34:44 NZST 2012 by Bycroft, Harlow, Sainudiin, Teng and York)