Finding the asymptotically optimal Baire distance for multi-channel data

Patrick Erik Bradley

Institute für Photogrammetrie und Fernerkundung
Karlsruhe Institut für Technologie (KIT)

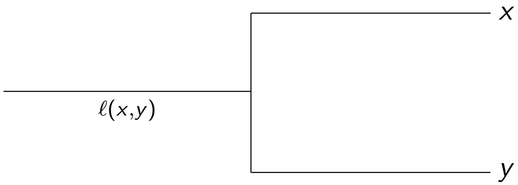
International Workshop on *p*-Adic Methods for Modeling of Complex Systems
Bielefeld
April 15–19, 2013

1. Baire Distance

Let $x, y \in X$ be words over an Alphabet A. The *Baire distance* is

$$d(x, y) = 2^{-\ell(x,y)},$$

$\ell(x, y)$ = length of longest common initial subword:



length = number of letters from A
(with multiple occurrences)

1. Baire Distance

Remark. Basis $\frac{1}{2}$ in Baire distance is arbitrary!

- Replace $\frac{1}{2}$ by any fixed $0 < \epsilon < 1$.

Definition. *ϵ-Baire distance*:

$$d_\epsilon(x, y) = \epsilon^{\ell(x,y)}$$

Observe. The metrics d and d_ϵ are equivalent.

1. Baire Distance

Motivation.

- Baire distance is an ultrametric:

$$d(x, y) \leq \max\{d(x, z), d(x, y)\}$$

- p-adic distance is an ϵ-Baire distance, if (integral) p-adic expansions

$$x = \alpha_0 + \alpha_1 p + \alpha_2 p^2 + \dots$$

  are viewed as words

$$\alpha_0\, \alpha_1\, \alpha_2 \dots$$

  Here, $\epsilon = \frac{1}{p}$.

- Similar for discrete valuation rings, e.g. integer rings of p-adic number fields: Then $\epsilon = \frac{1}{q}$ with $q = p^k$.

2. p-Adic Classification

Observation. Every finite alphabet A embeds into the ring of integers O_K of some p-adic number field as a (possibly incomplete) set of representants of the residue field $O_K/\mathfrak{m}_K$.

Typical example. $K = \mathbb{Q}_p$, $O_K = \mathbb{Z}_p$, $A \subseteq \{0, \ldots, p-1\}$.

Consequence.

- Freedom of choice in prime number p.
- Freedom of choice in p-adic alphabet.

Example.

- No need for large p if alphabet is large.
- Why not use e.g. Teichmüller representatives as alphabet, and exploit their multiplicative structure?

2. *p*-Adic Classification

- *p*-**Adic classification** aims at finding hierarchies inherent in data
- **Task.** Find an embedding of data into a *p*-adic number field

  Then tree structure of data is fixed
- Traditional classification often imposes hierarchy on data

2. p-Adic Classification

Applications.

- 2-Adic image sementation in spectral domain. $A = \{0, 1\}$
  (Benois-Pineau, Khrennikov, Kotovich 2001)

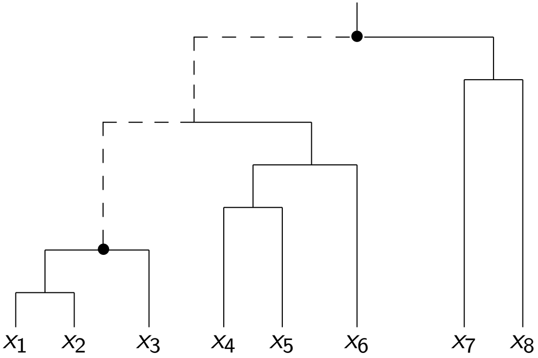
- Decimal number data. $A = \{0, 1, \ldots, 9\}$
  (Contreras & Murtagh 2010)

Usefulness.

- Efficient hierarchical classification
  (Murtagh; Benois-Pineau, Khrennikov, Kotovich)
- Good classification results
  (Contreras & Murtagh).

2. p-Adic Classification

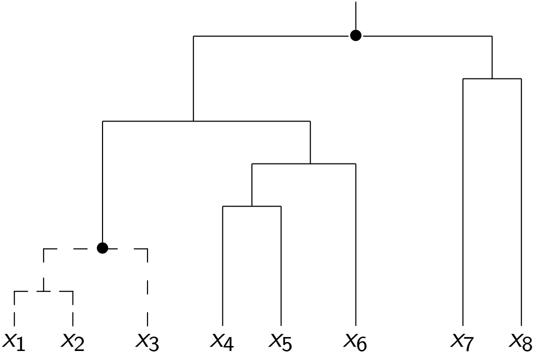
Remark. Set X of words over alphabet A defines a unique *dendrogram* $D(X)$, i.e. tree representation of X:



where path $\bullet - - - \bullet$ represents longest common subword of $\{x_1, x_2, x_3\}$.

2. p-Adic Classification

- Dendrogram $D(X)$ does not depend on ϵ in ϵ-Baire metric d_ϵ
- View nodes of $D(X)$ as *clusters*, and top node as *root*.



- Top half-edge points towards ∞.

2. p-Adic Classification

- *Classification* means:

$$X = \coprod c_i,$$

  i.e. disjoint union of clusters c_i.

- Classification is obtained by a *classification algorithm* with some optimization criteria, using d_ϵ.

  E.g. a splitting algorithm (P.E.B. 2009)

- Optimal classifications depend on ϵ.

3. Optimal Baire distance

- Given data X
- with set $P = \{p_1, \ldots, p_n\}$ of attributes
- assuming possible values $V = \{v_1, \ldots, v_m\}$

By a permutation $\sigma \in S_n$ obtain:

$$p^{\sigma}(x) := p_{\sigma(1)}(x)\, p_{\sigma(2)}(x) \ \ldots \ p_{\sigma(n)}(x),$$

a word with letters in alphabet V.

- σ-ϵ-Baire distance $d_{\epsilon}^{\sigma}(x, y)$.
- Which $\sigma \in S_n$ is most suitable for classification?

3. Optimal Baire distance

- Look at distances between different words:

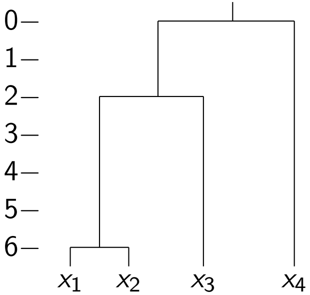
σ	1	2	3	4	5	6	7	8	9
x_1	e	n	i	g	m	a			
x_2	e	n	i	g	m	a	t	i	c
x_3	e	n	g	i	n	e			
x_4	t	r	a	i	n	i	n	g	

- Average σ-ϵ-Baire distance $\times 12$:

$$\begin{aligned} E_\sigma &= \epsilon^6 + \epsilon^2 + \epsilon^0 \\ &+ \epsilon^6 + \epsilon^2 + \epsilon^0 \\ &+ \epsilon^2 + \epsilon^2 + \epsilon^0 \\ &+ \epsilon^0 + \epsilon^0 + \epsilon^0 = 6\epsilon^0 + 4\epsilon^2 + 2\epsilon^6 \end{aligned}$$

3. Optimal Baire distance

- Dendrogram $D(\sigma, X)$:



- 1 more dense cluster $\{x_1, x_2, x_3\}$, 1 singleton $\{x_4\}$

3. Optimal Baire distance

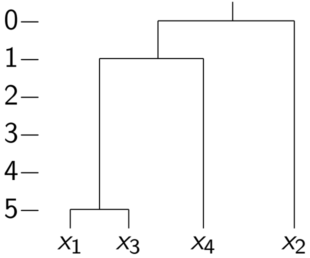
- Now $\tau = (9, 8, 7, 1, 2, 3, 4, 5, 6)$:

τ	9	8	7	1	2	3	4	5	6
x_1				e	n	i	g	m	a
x_2	c	i	t	e	n	i	g	m	a
x_3				e	n	g	i	n	e
x_4		g	n	t	r	a	i	n	i

$$\begin{aligned} E_\tau &= \epsilon^0 + \epsilon^5 + \epsilon^1 \\ &+ \epsilon^0 + \epsilon^0 + \epsilon^0 \\ &+ \epsilon^5 + \epsilon^0 + \epsilon^1 \\ &+ \epsilon^1 + \epsilon^0 + \epsilon^1 = 6\epsilon^0 + 4\epsilon^1 + 2\epsilon^5 \end{aligned}$$

3. Optimal Baire distance

- Dendrogram $D(\tau, X)$:



- 1 less dense cluster $\{x_1, x_3, x_4\}$, 1 singleton $\{x_2\}$

3. Optimal Baire distance

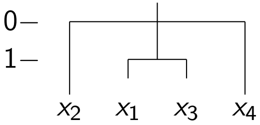
- Now $\rho = (8, 3, 1, 2, 4, 5, 6, 7, 9)$:

ρ	8	3	1	2	4	5	6	7	9
x_1		i	e	n	g	m	a		
x_2	i	i	e	n	g	m	a	t	c
x_3		g	e	n	i	n	e		
x_4	g	a	t	r	i	n	i	n	

$$\begin{aligned} E_\rho = &\ \epsilon^0 + \epsilon^1 + \epsilon^0 \\ &+ \epsilon^0 + \epsilon^0 + \epsilon^0 \\ &+ \epsilon^1 + \epsilon^0 + \epsilon^0 \\ &+ \epsilon^0 + \epsilon^0 + \epsilon^0 = 10\epsilon^0 + 2\epsilon^1 \end{aligned}$$

3. Optimal Baire distance

- Dendrogram $D(\rho, X)$:



- 1 less dense cluster $\{x_1, x_3\}$, 2 singletons $\{x_2\}$ and $\{x_4\}$

3. Optimal Baire distance

- Low average σ-ϵ-Baire distance leads to lots of common initial features. $\Rightarrow$ high density clusters, few singletons
- $E_\sigma = 6\epsilon^0 + 4\epsilon^2 + 2\epsilon^6$
- Highest average σ-ϵ-Baire distance has lots of branching at highest levels in the hierarchy, high number of singletons,
- $E_\rho = 10\epsilon^0 + 2\epsilon^1$

3. Optimal Baire distance

Task. Find a permutation $\sigma \in S_n$ such that

$$E^\sigma(\epsilon, X) = \sum_{x,y \in X} d_\epsilon^\sigma(x, y)$$

is minimal.

- optimal Baire distance d_ϵ^σ.

Remark. Optimal σ depends on ϵ.

Problem. $|S_n| = n!$ is quite large for large n.

3. Optimal Baire distance

- Δ = combinatorial $n-1$-simplex with corners $N = \{1, \ldots, n\}$
- The faces $F(N)$ are the power set 2^N
- $F_i(N) = \{x \in F(n) \mid |x| = i + 1\}$ are the i-faces of Δ

3. Optimal Baire distance

The graph Γ_Δ

- Vertices: the faces $F(N)$
- Edges: pairs (v, v') with $v \in F_i(N)$, $v' \in F_{i+1}(N)$, and $v \subseteq v'$
- Use

$$c \colon 2^N \to \mathbb{N},$$
$$I \mapsto |\{(x_1, x_2) \in X^2 \mid x_1 \neq x_2 \text{ and } \forall i \in I \colon i(x_1) = i(x_2)\}|$$

- Vertex weights:

$$w(v) = c(v)$$

- Edge weights:

$$w(e) = w(v) - w(v'),$$

where $e = (v, v')$

3. Optimal Baire distance

Lemma. $w(e) \geq 0$.

- Γ_Δ is a directed acyclic graph with initial vertex $v_\emptyset$ and terminal vertex v_N.
- An injective path $\gamma\colon v_\emptyset \rightsquigarrow v_J$ in Γ_Δ has ϵ-length

$$\ell_\epsilon(\gamma) = \sum_{\mu=0}^{\nu-1} w(e_\mu)\epsilon^\mu,$$

where $\gamma = (e_0, \dots, e_{\nu-1})$ (sequence of edges)

3. Optimal Baire distance

Definition. Permutation $\sigma \in S_n$ is *compatible* with $\gamma \colon v_\emptyset \rightsquigarrow v_J$, if

$$\{\sigma(i)\} = J_i \setminus J_{i-1},$$

where γ travels through the sequence of sets $J_0 = \emptyset, \ldots, J_\nu = J$.

Lemma. If σ is compatible with γ, then

$$E^\sigma(\epsilon) - \ell_\epsilon(\gamma) = c(N)\epsilon^n,$$

i.e. does not depend on γ.

Corollary. Dijkstra's shortest path algorithm on Γ_Δ finds the global minima for $E^\sigma(\epsilon)$ with any given $\epsilon \in (0,1)$.

Problem. Size of Γ_Δ makes Corollary impractical.

3. Optimal Baire distance

- Gradient descent

Theorem. Gradient descent is of run-time complexity at most $O(n^2 \cdot |X^2|)$.

Proof.

- In first step, there are n choices for edges.
- After n steps, the permutations are found.
- Finding minimal edge in step ν is of complexity $O(\nu)$.
- Complexity of computing $w(e)$ is at most $O(|X^2|)$

Hence, the upper bound. □

3. Optimal Baire distance

- Gradient descent yields only local minimum.
- For global minimum, take all minimal edges.

Algorithm 1. Input. Γ_Δ and weights w.

Start. $V_0 := \{v_\emptyset\}$, $E_0 := \{\text{all edges of } \Gamma_\Delta\}$

Step 1. $E_1 := \{e \in E_0 \mid o(e) \in V_0 \text{ and } w(e) \text{ smallest}\}$,
$V_1 := \{t(e) \mid e \in E_1\}$.

Step ν. $E_\nu := \{e \in E_{\nu-1} \mid o(e) \in V_{\nu-1} \text{ and } w(e) \text{ smallest}\}$,
$V_\nu := \{t(e) \mid e \in E_\nu\}$.

Output. All paths $\gamma\colon v_\emptyset \rightsquigarrow v_N$ with smallest sum of weights.

3. Optimal Baire distance

Theorem. There is a constant $C \in (0, 1)$ such that Algorithm 1 finds a global minimum for $E^{\sigma}(\epsilon)$, whenever $0 < \epsilon < C$.

3. Optimal Baire distance

Notation.

- Let L be a list of sets, $I := \bigcup_{J \in L} J$.
- Then $c_L(J) := c(I \cup J)$. Observe: $c_{(\emptyset)}(J) = c(J)$

Algorithm 2. Input X, N.

1. Step. $L := (\emptyset)$ (ordered list)
   1.1 Compute $M_1^L := \{i \mid c := c_L(i) \text{ is largest}\}$. If $c > 0$, continue.
   1.2 Choose $i \in M_1^L$. Find a $j \in M_1^L \setminus \{i\}$ s.t. $c_L(i,j) = c$. If found, then find a $k \in M_1^L \setminus \{i,j\}$ s.t. $c_L(i,j,k) = c$. Etc. Obtain $I_1 \subseteq M_1^L$.
   1.3 If $I_1 \neq M_1^L$, then Step 1.2 with $M_1^L := M_1^L \setminus I_1$.
   1.4 Obtain $M_1^L = I_1 \cup \cdots \cup I_r$ (disjoint union) s.t. $c_L(I_\rho) = c$.
   1.5 $\mathcal{I}_1^L := \{I_\rho \mid |I_\rho| \text{ is largest}\}$.

ν. Step^L. $\forall I_\rho \in \mathcal{I}_{n-1}$ do
   1.1 $L := \text{append } L$ with I_ρ
   1.2 Do as in previous step
   1.3 Obtain $\mathcal{I}_\nu^L$

3. Optimal Baire distance

Output. A set of sequences $L = (\emptyset, I_1, \ldots, I_\nu)$ with disjoint $I_\mu \in \mathcal{I}^{(I_1,\ldots,I_\mu)}$ of cardinality i_μ.

- The permutations corresponding to L are all $\sigma \in S_n$ such that

$$\begin{aligned} \sigma(\{1,\ldots,i_1\}) &= I_1 \\ \sigma(\{i_1+1,\ldots,i_1+i_2\}) &= I_2 \\ &\vdots \\ \sigma(\{i_1+\cdots+i_{\nu-1}+1,\ldots,i_1+\ldots,i_\nu\}) &= I_\nu \end{aligned}$$

- $I := \bigcup_{J\in L} J \Rightarrow \sigma(\{i_\nu+1,\ldots,n\}) = N \setminus I$

- $\forall\, j \in N \setminus I \colon c_I(j) = 0$,
  i.e. coincidences occur here only on the diagonal

3. Optimal Baire distance

Theorem. Algorithm 2 computes all σ such that $E^{\sigma}(t) \in \mathbb{N}[t]$ is lexicographically minimal.

3. Optimal Baire distance

Example.

σ	1	2	3	4	5	6	7	8	9
x_1	e	n	i	g	m	a			
x_2	e	n	i	g	m	a	t	i	c
x_3	e	n	g	i	n	e			
x_4	t	r	a	i	n	i	n	g	

- Step 1. $c_{(\emptyset)}(1) = c_{(\emptyset)}(2) = c_{(\emptyset)}(9) = 6$. $M_1^{(\emptyset)} = \{1, 2, 9\}$
  $c_{(\emptyset)}(1, 2) = 6$, $c_{(\emptyset)}(1, 2, 9) = 2 \Rightarrow I_1 = \{1, 2\}$, $I_2 = \{9\}$
  $\mathcal{I}_1^{(\emptyset)} = \{I_1\}$

3. Optimal Baire distance

Example.

σ	1	2	3	4	5	6	7	8	9
x_1	e	n	i	g	m	a			
x_2	e	n	i	g	m	a	t	i	c
x_3	e	n	g	i	n	e			
x_4	t	r	a	i	n	i	n	g	

- $I_1 = \{1, 2\}$
- Step 2. $c_{(\emptyset, I_1)}(3) = \cdots = c_{(\emptyset, I_1)}(9) = 2$. $M_2^{(\emptyset, I_1)} = \{3, \ldots, 9\}$
  $c_{(\emptyset, I_1)}(3, 4) = c_{(\emptyset, I_1)}(3, 4, 5) = c_{(\emptyset, I_1)}(3, 4, 5, 6) = 2$,
  $c_{(\emptyset, I_1)}(3, 4, 5, 6, 7) = 0 \Rightarrow I_2 = \{3, 4, 5, 6\}$,
  $|M_2^{(\emptyset, I_1)} \setminus I_2| = 3 < |I_2|$. $\mathcal{I}_2^{(\emptyset, I_1)} = \{I_2\}$

3. Optimal Baire distance

Example.

σ	1	2	3	4	5	6	7	8	9
x_1	e	n	i	g	m	a			
x_2	e	n	i	g	m	a	t	i	c
x_3	e	n	g	i	n	e			
x_4	t	r	a	i	n	i	n	g	

- $I_1 = \{1, 2\}$, $I_2 = \{3, 4, 5, 6\}$
- Step 3. $c_{(\emptyset, I_1, I_2)}(7) = c_{(\emptyset, I_1, I_2)}(8) = c_{(\emptyset, I_1, I_2)}(9) = 0.$

3. Optimal Baire distance

Example.

σ	1	2	3	4	5	6	7	8	9
x_1	e	n	i	g	m	a			
x_2	e	n	i	g	m	a	t	i	c
x_3	e	n	g	i	n	e			
x_4	t	r	a	i	n	i	n	g	

- Output. $L = \{\emptyset, I_1, I_2\}$
- $I_1 = \{1,2\}$, $I_2 = \{3,4,5,6\}$, $N \setminus (I_1 \cup I_2) = \{7,8,9\}$
- Any optimal σ permutes first $\{1,2\}$, then $\{3,4,5,6\}$, then $\{7,8,9\}$
- $E^{\sigma}(t) = 6 + 4t^2 + 2t^6$

3. Optimal Baire distance

Complexity.

Worst case.

σ	1	2	$\dots$	n
x_1	$\times$			
x_2		$\times$		
$\vdots$			$\ddots$	
x_n				$\times$

- $c(I)$ is constant on all sets I of constant cardinality
- Hence, all permutations of N are computed.
- This example is pathologic: No preferred permutations!

3. Optimal Baire distance

Expected complexity.

- In general, $c(I) = 0$ for $|I|$ large can be expected.
- The set $\{I \subseteq N \mid c(I) > 0\}$ is in general sparse.
- This should make Algorithm 3 practical in general.

4. Application

1. **Hyperspectral data** (with Andreas Braun)
   - AVIRIS Indian Pines dataset: 145×145 pixels with 220 spectral channels
   - Coincidences due to signal vs. Coincidences due to noise $\Rightarrow$ PCA yields six components explaining 99.66% of total variance
   - Reduce to six variables
   - Find optimal permutations in S_6 at different resolutions
   - Incorporate these into a multi-Baire-kernel SVM
   - Results are comparable to a Linear SVM, more complete results in most classes for multi-Baire-kernel SVM

4. Application

Byzantine Chant.

- single melody
- drone (Ison)

Skope of Chant.

- Enhance a poetic liturgical text
- Highlight important words
- Bring mind and heart to contents of text
- Audible Icon

4. Application

Byzantine Chant.

Idea of classification steps.

- Highlight *emphasized* syllables (they occur in important words)
- Identify Cadenzas
- Cadenzas are intermediate (end of thought) or final (end of phrase)
- Bring phrases of different lengths to uniform length by inserting blanks
- The unemphasized syllables (somehow) lead to the pitch of the next emphasized syllable
- "somehow" depends on number of syllables and musical distance to target pitch
- Melody usually stays on one pitch, or moves one step. Larger jumps occur on occasion.

4. Application

Paschal Canon, Tone 1.

It is the dáy of Resurréction! Let us be rádiant, o ye péoples!

Final Cadenza: Páscha, the Lórd's Páscha!

Alaska:

For from death to lífe, and from earth to héaven has Christ our God brought us

Boston:

For Christ God hath brought us from death unto lífe, and from earth unto héaven

Final Cadenza: as we sing the triúmphal hymn.

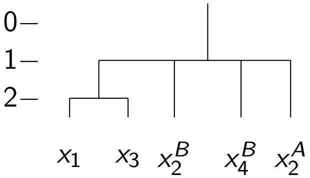
Refrain. Christ is risen from the dead.

http://www.musicarussica.com/compact_discs/i-099
http://www.musicarussica.com/compact_discs/i-083

4. Application

Question. $\exists$ optimal σ which brings highlighted syllables first, in order of occurence, then other syllables in some order?

- Phrases without final cadenzas $^B = \mathrm{Boston}$, $^A = \mathrm{Alaska}$



- Final Cadenzas: