MaxEnt 2008

São Paulo, 6-11 July 2008

Michel Petitjean

CEA/DSV/iBiTec-S/SB2SM (CNRS URA 2096)
91191 Gif-sur-Yvette Cedex, France.

michel.petitjean@cea.fr

http://petitjeanmichel.free.fr/itoweb.petitjean.html

AN ASYMMETRY COEFFICIENT FOR MULTIVARIATE DISTRIBUTIONS

PHYSICAL SYSTEMS

Some symmetric physical systems having degenerated energy levels and offering a continuous separation of its energy levels induced by, or inducing a symmetry breaking, may be such that symmetry could itself offer continuous variations.

We must treat symmetry as a measurable quantity.

SYMMETRY // SKEWNESS // CHIRALITY

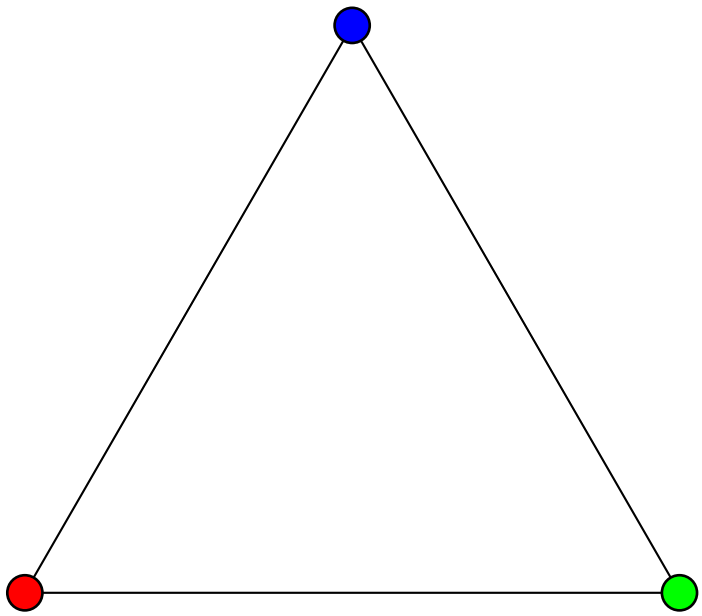
SKEWNESS: degree of asymmetry of a distribution

Asymmetry coefficients exist:
therefore, symmetry is measurable !!

CHIRALITY: lack of mirror symmetry
May apply to objects and distributions having COLORS

Chirality is measurable, too.
In fact, an asymmetric univariate distribution is CHIRAL
(reflection through a point is related to indirect symmetry)

Geometric chirality measures are NOT related with physical light-matter interactions. However, optical rotatory power and circular dichroism are revealed in chiral media.



THIS EQUILATERAL TRIANGLE IS CHIRAL ...
... IF WE CAN SEE THE COLORS AT THE VERTICES.

IF WE CAN'T, IT IS ACHIRAL.

GENERAL THEORY. Part I: COLORED MIXTURES

Colors cannot be handled in the euclidean space

1. We consider a probability space: (C, A, P)

C: space of colors (e.g. $C = \{red, green, blue\}$)
It is possible to have an infinite number of colors.

A: σ-algebra defined on C

P: a probability measure on (C, A)

2. We consider the measurable space: $(C \times R^d, A \otimes B)$
B: Borel σ-algebra of R^d

3. We define a mapping Φ from C on (R^d, B):
To each color c is associated a d-variate distribution $\tilde{P}_c = \Phi(c)$.
The value of the distribution function of $\tilde{P}_c$ at $x \in R^d$ is $\tilde{F}(x|c)$

4. We consider a random variable (K, X) taking values in $(C \times R^d, A \otimes B)$, with marginal distribution function F in R^d such that:

$$F(x) = \int_{c \in C} \tilde{F}(x|c)P(dc)$$

X is called a colored mixture,
and its distribution F is a colored mixture of distributions.

When K is a.s. constant, it is equivalent to consider that there is only one color in C, and there is no essential difference between X and an ordinary random vector.

GENERAL THEORY. Part II: THE COLORED MIXTURE MODEL

We consider two random variables (K_1, X_1) **and** (K_2, X_2) **on** $(C \times R^d, A \otimes B)$**,** X_1 **and** X_2 **being two colored mixtures.**

Joint distribution of (K_1, K_2)**:** P_{12}

We have a couple of mappings (Φ_1, Φ_2)**, thus for each couple of colors** (c_1, c_2) **we have a couple of** d**-variate distributions:**

$$(\tilde{P}_{1c_1}, \tilde{P}_{2c_2}) = (\Phi_1(c_1), \Phi_2(c_2))$$

Joint distribution of $(\tilde{P}_{1c_1}, \tilde{P}_{2c_2})$**:** $\tilde{W}$

Joint distribution function of (X_1, X_2):

$$W(x_1, x_2) = \int_{c_1 \in C} \int_{c_2 \in C} \tilde{W}(x_1, x_2 | c_1, c_2) P_{12}(dc_1, dc_2)$$

ADDITIONAL ASSUMPTION: $K_1 \overset{a.s.}{=} K_2$

It means that: $P_{12}(dc_1, dc_2) = P(dc_1)\delta_{[c_2=c_1]} dc_2$
(δ is the Dirac-Delta function)

and then: $W(x_1, x_2) = \int_{c \in C} \tilde{W}(x_1, x_2 | c) P(dc)$

EXAMPLE 1

$C = \{red, green\}$

K_1, K_2: $\Pr(red) = 1/2$, $\Pr(green) = 1/2$

In this example we consider a.s. constant random vectors.

Colored mixture X_1 (a_1 and b_1 are distinct constants in R^d):

$\tilde{P}_{1,red}$: $\Pr(X_1 = a_1 | red) = 1$
$\tilde{P}_{1,green}$: $\Pr(X_1 = b_1 | green) = 1$
Distribution of X_1: $\Pr(X_1 = a_1) = \Pr(X_1 = b_1) = 1/2$

Colored mixture X_2 (a_2 and b_2 are distinct constants in R^d):

$\tilde{P}_{2,red}$: $\Pr(X_2 = a_2 | red) = 1$
$\tilde{P}_{2,green}$: $\Pr(X_2 = b_2 | green) = 1$
Distribution of X_2: $\Pr(X_2 = a_2) = \Pr(X_2 = b_2) = 1/2$

We have a two-step process:

(1) We get **one** color c from $K_1 \overset{a.s.}{=} K_2$
(2) We get the distributions $\tilde{P}_{1c}$ and $\tilde{P}_{2c}$ from c:
$\tilde{P}_{1,red}$ and $\tilde{P}_{2,red}$ when $c = red$
$\tilde{P}_{1,green}$ and $\tilde{P}_{2,green}$ when $c = green$

In general, the colored mixtures CANNOT be independant. The set of joint distributions of X_1 and X_2 is constrained by the link in the space of colors.

Here, there is only one possible distribution of (X_1, X_2):

$\Pr(X_1 = a_1, X_2 = a_2) = 1/2$
$\Pr(X_1 = b_1, X_2 = b_2) = 1/2$
$\Pr(X_1 = a_1, X_2 = b_2) = 0$
$\Pr(X_1 = b_1, X_2 = a_2) = 0$

X_1 and X_2 are not independant: they are correlated!

EXAMPLE 2 (generalizes example 1)

We assume:
(a) The mixing distribution of the colors is discrete and finite:
there are k colors
(b) All mixed distributions are discrete and finite
(c) For each color, the two discrete marginals are distributed over an equal number of values n_c ($c = 1, \ldots, k$)
(d) For each color, the two discrete marginals are uniform
(e) The full marginals X_1 and X_2 are uniformly distributed

It is proved that (X_1, X_2) **has** $\prod_{c=1}^{c=k} n_c$ **possible joint distributions.**

We have modelized the situation where two set of n points are each partitioned into k groups of n_c points, $c = 1, \ldots, k$, each pair of groups being associated to a color.

Each of these pairs of groups is such that the two subsets of n_c points offer $n_c!$ possible pairwise correspondances.

---

1 correspondence $\leftrightarrow$ 1 joint distribution.
(permutation matrix) / n = joint distrib. probability matrix.

$k = n$ colors: two groups of n *"discernable points or particles"*
We have two groups of n points pairwise associated.
(e.g.: regression in the plane: values are pairwise associated)

$k = 1$ color: two groups of n *"indiscernable points or particles"*
We have two groups of n points under free correspondence.
There are $n!$ possible correspondances.

SIMILARITY STUDIES

Remark:
Measuring symmetry or chirality is measuring self-similarity

We need a distance between colored mixtures
i.e. we need a probability metric able to $<<$ see $>>$ the colors

---

The **$\mathbf{L^2}$-Wasserstein distance** is a probability metric between distributions of random vectors
(appears in the Monge-Kantorovich transportation problem):

$D^2 = Inf_{\{W\}} E[(X_1 - X_2)' \cdot (X_1 - X_2)]$
$\{W\}$: set of all joint distributions of (X_1, X_2).

The **COLORED $\mathbf{L^2}$-Wasserstein distance** is a probability metric between distributions of **COLORED MIXTURES**:

$D^2 = Inf_{\{W\}} E[(X_1 - X_2)' \cdot (X_1 - X_2)]$
$\{W\}$: set of all joint distributions of (X_1, X_2).

Here $\{W\}$ is a subset of all joint distributions of the couple of random vectors (X_1, X_2) when there are no colors.

Reminder: the link in the space of colors induces constraints

$\{W\}$ is shown to be not empty.

SIMILARITY STUDIES EXAMPLES

SAMPLES / LEAST SQUARES METHODS

- Procrustes methods:
  optimal superposition of two groups of n points in R^d.
  under affine transformation, or isometry, or rotation, etc.

- RMS alignment/superposition (chemistry, biochemistry):
  as above, but pure rotation only (most time in R^3)

The Procrustes and RMS distances are instances of
the colored L^2-Wasserstein distance.
*(E.g.: case of a **fixed** pairwise correspondence)*

When there is only one color (or no color),
they are also instances of the L^2-Wasserstein distance.
*(case of a **free** pairwise correspondence)*

The minimized distance is a distance between
classes of equivalence of distributions.
E.g., minimizing the distance for rotation means that we consider
the class of distributions images via rotation.

Minimization: analytical solutions are known in several cases.
The optimal rotation is unknown for $d > 3$.

MEASURING CHIRALITY: GENERAL THEORY

We consider a colored mixture X in R^d.
Its inertia T is assumed to be finite and non null.

We consider the colored mixtures $\bar{X}$ distributed as rotated and translated inverted images of X.
In other words, the distributions of X and $\bar{X}$ are images through some indirect isometry, i.e. through composition of some rotation R and translation t and mirror inversion.

Remark: we have the constraints induced by $\mathbf{K} \overset{\mathbf{a.s.}}{=} \bar{\mathbf{K}}$

Definition of the CHIRAL INDEX

$$\chi = \frac{d}{4T} Min_{\{R,t\}} D^2$$

$$D^2 = Inf_{\{W\}} E[(X - \bar{X})' \cdot (X - \bar{X})]$$

$\{W\}$: set of all joint distributions of $(X, \bar{X})$.

Properties

χ depends only on the distribution of (K, X)

χ is insensitive to rotations, translations, inversions, and scaling

χ takes values on $[0;1]$

$\chi = 0$ **IF and ONLY IF** the distribution is **ACHIRAL**

Other properties of the chiral index

The minimisation for translation is reached for $EX = E\bar{X}$
(and the optimal rotation is analytically known in R^2 and in R^3)

$$\chi = \frac{d}{4T} Min_{\{R\}} Inf_{\{W\}} E[(X - \bar{X})' \cdot (X - \bar{X})]$$

$$\chi = \frac{d}{2}[1 - [Sup_{\{R,W\}} \sum_{i=1}^{i=d} c_i]/T]$$

$\{W\}$: set of all joint distributions of $(X, \bar{X})$.
c_i: covariance attached to the axis i ($i = 1 \ldots d$)

When the mixed distributions are all those of a.s. constant vectors:
(i.e. never two of them have the same color)

$\chi = d\lambda_d/T$ (λ_d is the smallest eigenvalue of $Cov(X)$)

Here the maximum $\chi = 1$ is reached when $Cov(X)$ is proportional the the identity matrix.

Case of samples (modelizes a fnite set of n points in R^d)

X: rectangular array of n lines and d columns
A: centering operator: $A = I - \mathbf{1}\mathbf{1}'/n$
- I: identity matrix of size n
- $\mathbf{1}$: vector of size n with all components equal to 1

P: permutation matrix of order n (eqv. to a joint distribution)
Q: arbitrary fixed orthogonal matrix of order n with $det(Q) = -1$

$$\chi = \frac{d}{4nT} Min_{\{P,R\}}[Tr(X - PXQ'R')'A(X - PXQ'R')]$$

”Continuity” property

We would like something like that:
”closer” two distributions are, closer their chiral indices are.

with a weak convergence criterion for distributions,
so that we can get a strong theorem.

NON COLORED case

X_n: random vector with probability distribution P_n
X: random vector with probability distribution P

X_n is a sequence of random vectors converging to X in law

Assumptions:

$E[X'X]$ exists
$E[X_n'X_n] \longrightarrow E[X'X]$
$E[(X - EX)'(X - EX)] \neq 0$

Theorem: $\chi(P_n) \longrightarrow \chi(P)$

Works for samples of a parent population: estimation of $\chi(P)$

COLORED or non colored case: samples

χ is a continuous function of the array X
(any matricial norm works)

THE DIRECT SYMMETRY INDEX

COLORED or non colored case: samples (n equally weighted points)

X: rectangular array of n lines and d columns
A: centering operator: $A = I - \mathbf{1}\mathbf{1}'/n$
  I: identity matrix of size n
  $\mathbf{1}$: vector of size n with all components equal to 1
P: permutation matrix of order n

$$DSI = \frac{1}{2T} Min_{\{P \neq I, R\}}[Tr(X - PXR')'A(X - PXR')]$$

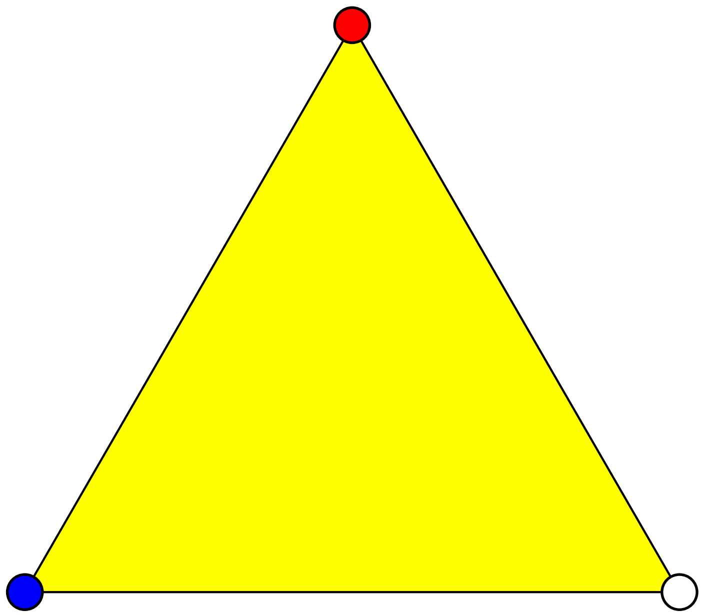
DSI is a continuous function of X, taking values on $[0; 1]$.
It is insensitive to rotations, translations, inversions and scaling.

BUT: cannot be extended to continuous distributions.
(notice the condition $P \neq I$ and its consequences)

It is due to the problem itself,
NOT to the Wasserstein distance

The problem is partly solvable for finite sets of rotations.

Some extremal figures

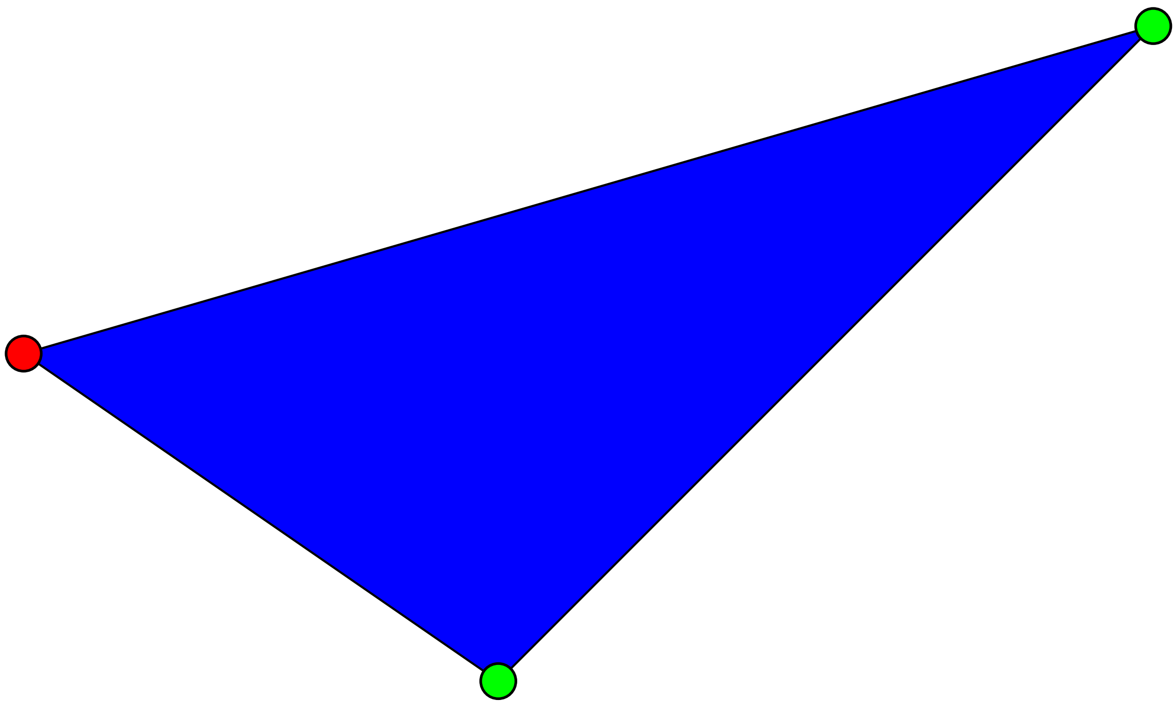


THE MOST CHIRAL TRIANGLE WITH ALL
NON-EQUIVALENT VERTICES IS EQUILATERAL

$$\chi = \mathbf{1}$$

This result generalizes in any dimension: the most chiral simplex with all non-equivalent vertices is regular: $\chi = \mathbf{1}$.

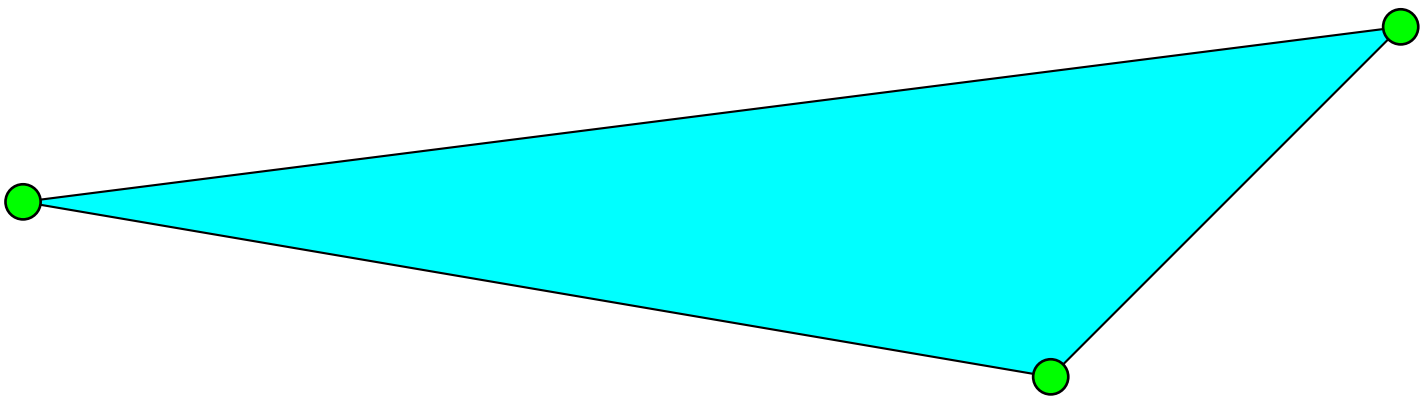
Remark: *only the vertices are considered not the interior, the sides, the faces, etc.*



THE MOST CHIRAL TRIANGLE
WITH 2 EQUIVALENT VERTICES

Distances ratio: $\sqrt{\mathbf{1}-\sqrt{\mathbf{6}}/\mathbf{4}} : \mathbf{1} : \sqrt{\mathbf{1}+\sqrt{\mathbf{6}}/\mathbf{4}}$

$$\chi = \mathbf{1} - \sqrt{\mathbf{2}}/\mathbf{2}$$



THE MOST CHIRAL TRIANGLE
WITH 3 EQUIVALENT VERTICES

(we are no more in the colored case!)

Distances ratio: $\mathbf{1} : \sqrt{\mathbf{4} + \sqrt{\mathbf{15}}} : \sqrt{(\mathbf{5} + \sqrt{\mathbf{15}})/\mathbf{2}}$

$$\chi = \mathbf{1} - \mathbf{2}\sqrt{\mathbf{5}}/\mathbf{5}$$

△

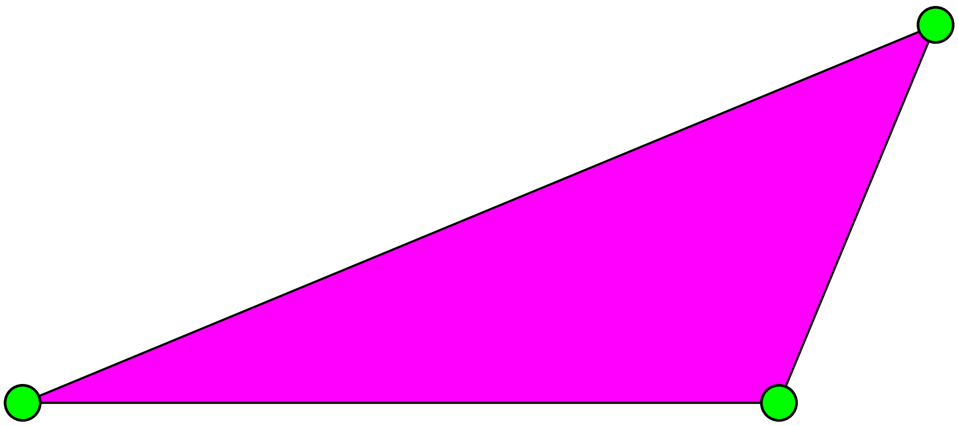
THE UNEQUIVALENCE OF ALL VERTICES PRECLUDES THE EXISTENCE OF ANY DIRECT SYMMETRY:

AT LEAST 2 POINTS SHOULD BE EQUIVALENT.

ONE OF THE MOST DISSYMETRIC TRIANGLES
WITH 2 UNEQUIVALENT VERTICES

Abscissas: $(\mathbf{-1-\sqrt{3}})/\mathbf{2}, (\mathbf{-1+\sqrt{3}})/\mathbf{2}, \mathbf{1}$

THIS DEGENERATE TRIANGLE IS SUCH THAT
$DSI = 1$ IN ANY DIMENSION.



THE MOST DISSYMETRIC TRIANGLE
WITH 3 EQUIVALENT VERTICES

Angles: $\pi/\mathbf{4}, \pi/\mathbf{8}, \mathbf{5}\pi/\mathbf{8}$

$$DSI = \mathbf{1} - \sqrt{\mathbf{2}}/\mathbf{2}$$

$\triangle$

REMARKABLE PROPERTY OF THE 5 EXTREMAL TRIANGLES

The 5 extremal triangles have the following geometric property.

The squared lengths of the sides are equal to three times
the squared distances vertex-barycenter:

$$d^2(p_2, p_3) = 3d^2(p_1, g)$$

$$d^2(p_1, p_2) = 3d^2(p_2, g)$$

$$d^2(p_3, p_1) = 3d^2(p_3, g)$$

$$g = (p_1 + p_2 + p_3)/3$$

CARE:

THE RELATION IS SYMMETRIC FOR TWO POINTS ONLY

ASYMMETRY COEFFICIENT AND MULTIVARIATE SKEWNESS
(no more colors)

Karl Pearson's skewness (1895) is null
for many "asymmetric" distributions.

The chiral index is null $<<$ IF and ONLY IF $>>$
the distribution is indirect-symmetric

Other advantage over multivariate analogs of Pearson's skewness:

the existence of the third-order moments is not required
(the existence of the inertia suffices)

UNIVARIATE CASE

$\chi = (1 + r_{min})/2$

r_{min} is the lower bound of the correlation coefficient between the distribution and itself.

It is shown that r_{min} cannot be positive: $\chi \in [0; 1/2]$

The upper bound is asymptotically reached by the Bernouilli law with parameter $m \longrightarrow 0$ or $m \longrightarrow 1$.

The Bernouilli distribution of parameter m: explicit calculation

$\Pr(X=0)=1-m \qquad \Pr(X=1)=m$

$EX=m \qquad T=Var(X)=m(1-m)$

We take Y distributed as X. The marginals X and Y are known: we parametrize their joint distributions by the quantity q.

$q=\Pr(X=0,Y=0)$

Then we get the set of joint distributions of (X,Y):

$$\begin{array}{ll}
\Pr(X=0,Y=0)=q & q\geq 0 \\
\Pr(X=1,Y=0)=(1-m)-q & q\leq(1-m) \\
\Pr(X=0,Y=1)=(1-m)-q & q\leq(1-m) \\
\Pr(X=1,Y=1)=m-(1-m-q) & q\geq(1-2m)
\end{array}$$

$E(XY)=2m-1+q \qquad Cov(X,Y)=(2m-1+q)-m^2$

$r=[q-(1-m)^2]/m(1-m)$.

We get χ from the minimization of $Cov(X,Y)$

The minimum is reached either for $q=(1-2m)$ or for $q=0$, depending on m.

If $m\in]0;1/2]$ then $r_{min}=-m/(1-m)$ and $\chi=1-1/(2-2m)$

If $m\in[1/2;1[$ then $r_{min}=-(1-m)/m$ and $\chi=1-1/2m$

$\chi=0$ IF and ONLY IF $m=1/2$

THE 3 POINTS SET ON THE REAL LINE

This is the simplest chiral set which can be built:
no color, no weights, $d = 1$, only 3 points, only one parameter.

α is the distance ratio between the two adjacent segments.

The following properties are mandatory for any chirality measure:

(a) It must depend **ONLY** on α
(b) It must be a continuous function of α
(c) It must be null when $\alpha = 1$
(d) It must be null **ONLY** for $\alpha = 1$
(e) It must return the same value for α and $1/\alpha$ (scaling invariance)

The chiral index satisfies to (a)-(e):

$$\chi = (1-\alpha)^2/4(1+\alpha+\alpha^2)$$

Sophisticated multivariate chirality measures and asymmetry coefficients must be first checked against the 3 points sets in order to see whether or not properties (a)-(e) stand.

SAMPLING / SYMMETRY TESTS

Let $x_{i:n}$ ($i = 1, \ldots, n$) be the **ORDERED** sample of size n.

Observed sample mean: $\bar{x}$
Observed standard deviation: σ.

The minimal correlation is reached when
the sample sorted in ascending order is correlated with
the sample sorted in descending order.

$$r_{min} = [\sum_{i=1}^{i=n}(x_{i:n} - \bar{x})(x_{n+1-i:n} - \bar{x})]/n\sigma^2$$

$$\chi_n = (1 + r_{min})/2$$

The chiral index is easily computable on a pocket calculator.

Other expressions of χ from the embedded intervals

From half **rangelengths**: $\chi_n = 1 - [\sum_{i=1}^{i=n}(\frac{x_{i:n} - x_{n+1-i:n}}{2})^2]/(n\sigma^2)$

From **midranges**: $\chi_n = [\sum_{i=1}^{i=n}(\frac{x_{i:n} + x_{n+1-i:n}}{2})^2 - n \cdot \bar{x}^2]/(n\sigma^2)$

The ratio above is: variance of midranges / sample variance

Symmetry tests: asymptotic distributions of χ_n ??
(under normality assumption, or uniformity assumption, or other...)

BIVARIATE DISTRIBUTIONS

Wasserstein distance (colored or not) between the distributions of X and Y, minimized for rotation:

$$D^2 = E[X'X] + E[Y'Y] - 2|G|$$

$$G^2 = (E[X'Y])^2 + (E[X'\Pi Y])^2 \qquad \Pi = \left[\begin{array}{c|c} 0 & -1 \\ \hline 1 & 0 \end{array}\right]$$

X_1 and X_2 are identically distributed in R^2 (joint distributions: W)

$$\bar{X} = EX_1 = EX_2$$

$$T = E(X_1 - \bar{X})'(X_1 - \bar{X}) = E(X_2 - \bar{X})'(X_2 - \bar{X})$$

$$\chi = 1 - Sup_{\{W\}}|\mu_1 - \mu_2|/T$$

$(\mu_1 - \mu_2)$ is the difference between the two eigenvalues of V

$$(\mu_1 - \mu_2)^2 = [Tr(V)]^2 - 4Det(V)$$

$$2V = E[(X_1 - \bar{X})(X_2 - \bar{X})' + (X_2 - \bar{X})(X_1 - \bar{X})']$$

Expression in the complex plane

Complex random variables z_1 and z_2, identically distributed (joint distributions: W)

$$\bar{z} = Ez_1 = Ez_2$$

$$T = E[\|z_1 - \bar{z}\|^2] = E[\|z_2 - \bar{z}\|^2]$$

$$\chi = 1 - Sup_{\{W\}}|E(z_1 - \bar{z})(z_2 - \bar{z})|/T$$

BIVARIATE SAMPLES

X: array of the n observations, n lines and 2 columns

Inertia: $T = Tr(X'AX)/n$

$A = I - \mathbf{11}'/n$ (centering operator)

P: permutation matrix of size n

$\chi = 1 - Max_{\{P\}}|\mu_1 - \mu_2|/nT$

$(\mu_1 - \mu_2)$ is the difference between the two eigenvalues of V

$(\mu_1 - \mu_2)^2 = [Tr(V)]^2 - 4Det(V)$

$V = (AX)'(P + P')(AX)/2$

In the complex plane: $z \in C^n$ contains the n observations

$\chi = 1 - [Max_{\{P\}}(Az)'P(Az)]/nT$

In the non colored case:

Theorem 1: There is an optimal P which is symmetric.
($P' = P$)

Theorem 2: $Sup(\chi) \in [1 - 1/\pi; 1 - 1/2\pi]$
(stands also for continuous distributions)

Conjecture: $Sup(\chi) = 1 - 1/\pi$

Family of sets conjectured to be of maximal chirality:
(asymptotic)

$$Sup(\chi) = 1 - 1/\pi$$

The calculations are easier in the complex plane.

Fix $\epsilon > 0$ then choose even integer $m > 1/\epsilon$.

$\omega = e^{i(2\pi)/(2m)}$ $(\omega^{2m} = 1)$

Select an integer $r > m^4/\epsilon^2$ then
select an even integer $k > r^{m-1}/\epsilon$

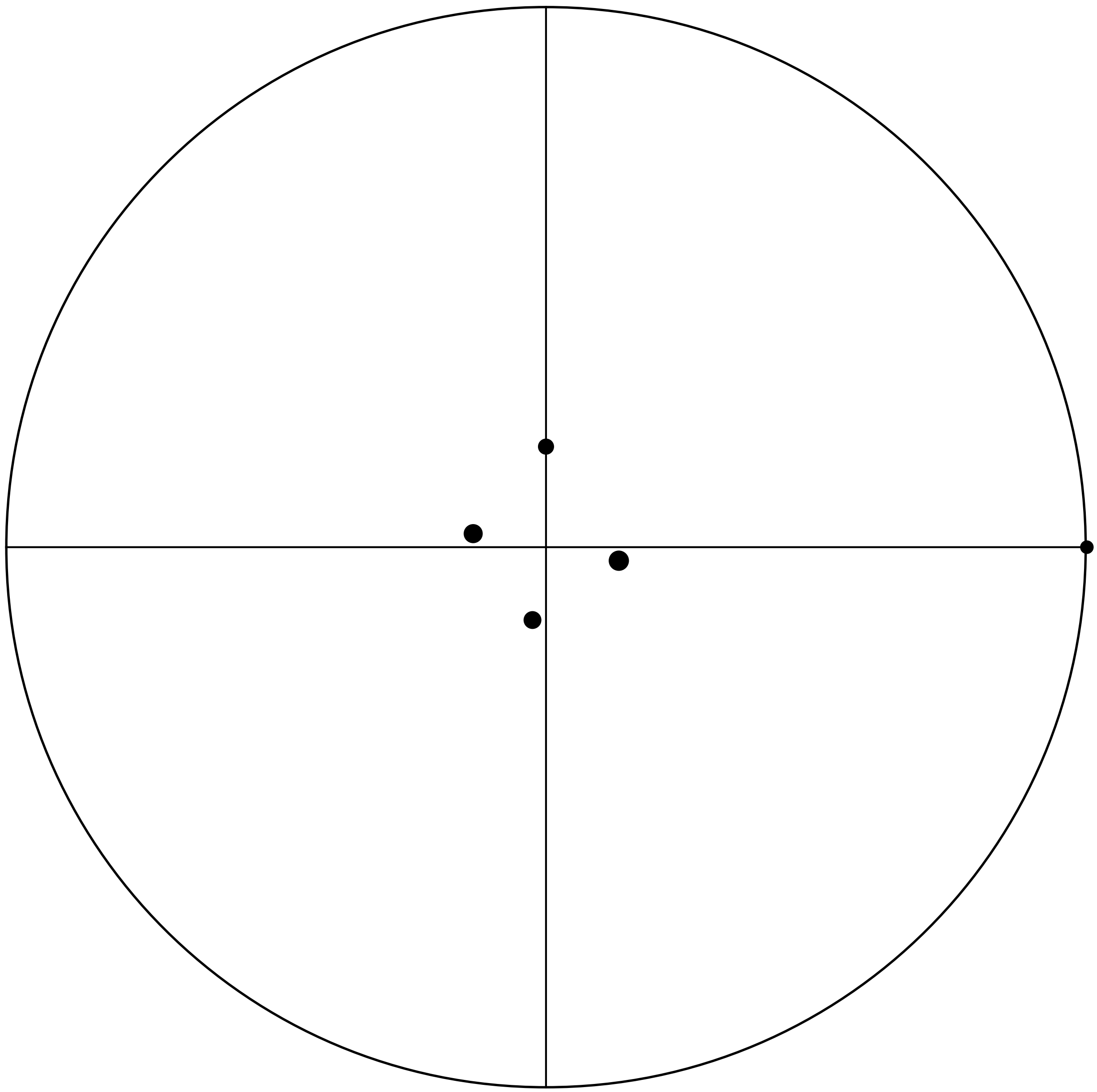
$z \in C^n$ z is a complex vector of $m + 3$ blocks of elements

Each block j ($j = 0..m + 2$), contains identical elements.

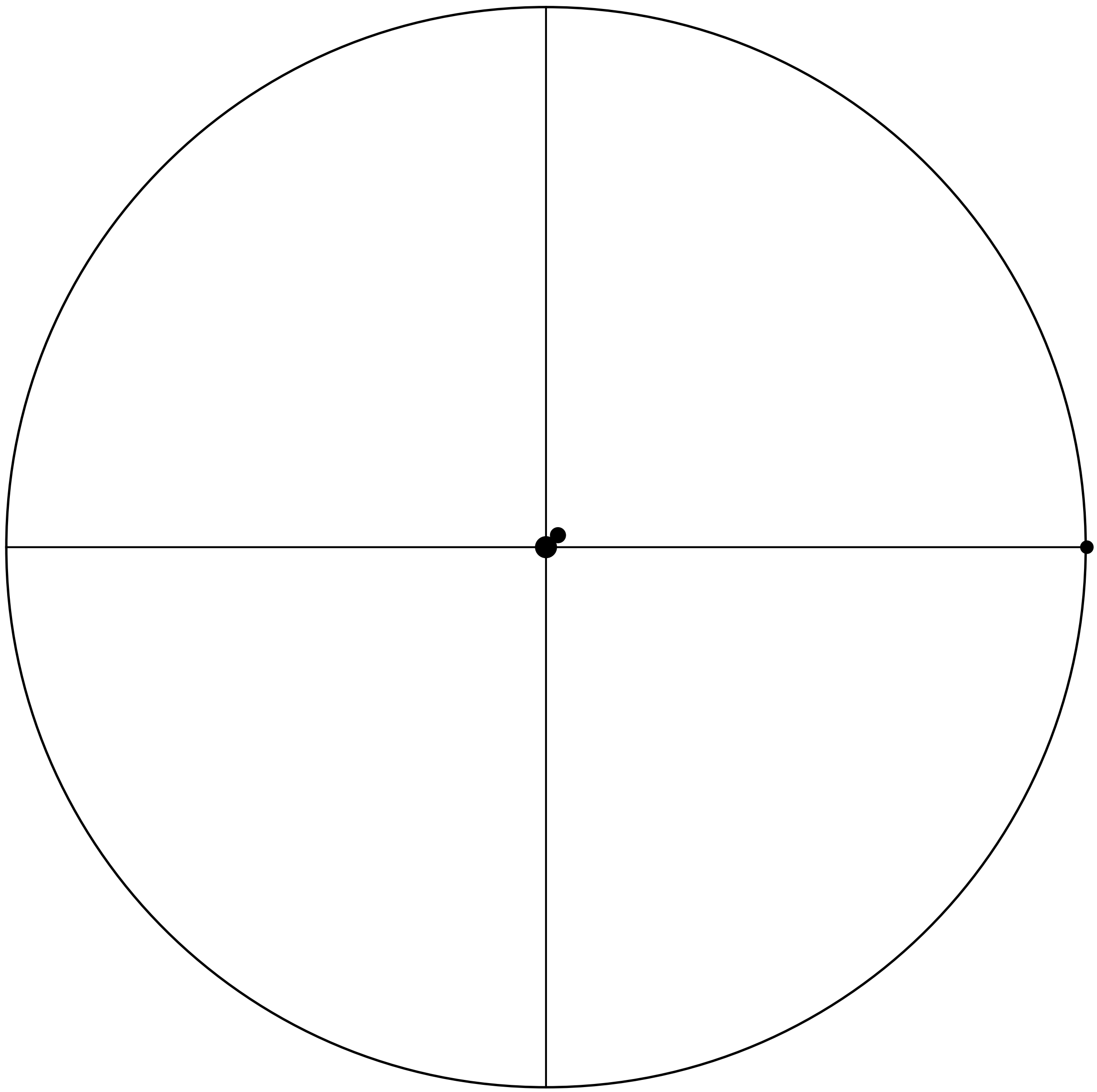
$n = 1 + r + r^2 + \ldots + r^{m-1} + k + \frac{k}{2} + \frac{k}{2}$

$S = \sum_{j=0}^{j=m-1} \omega^j r^{j/2}$ (z is such that $z'\mathbf{1} = 0$ and $z'z = 0$)

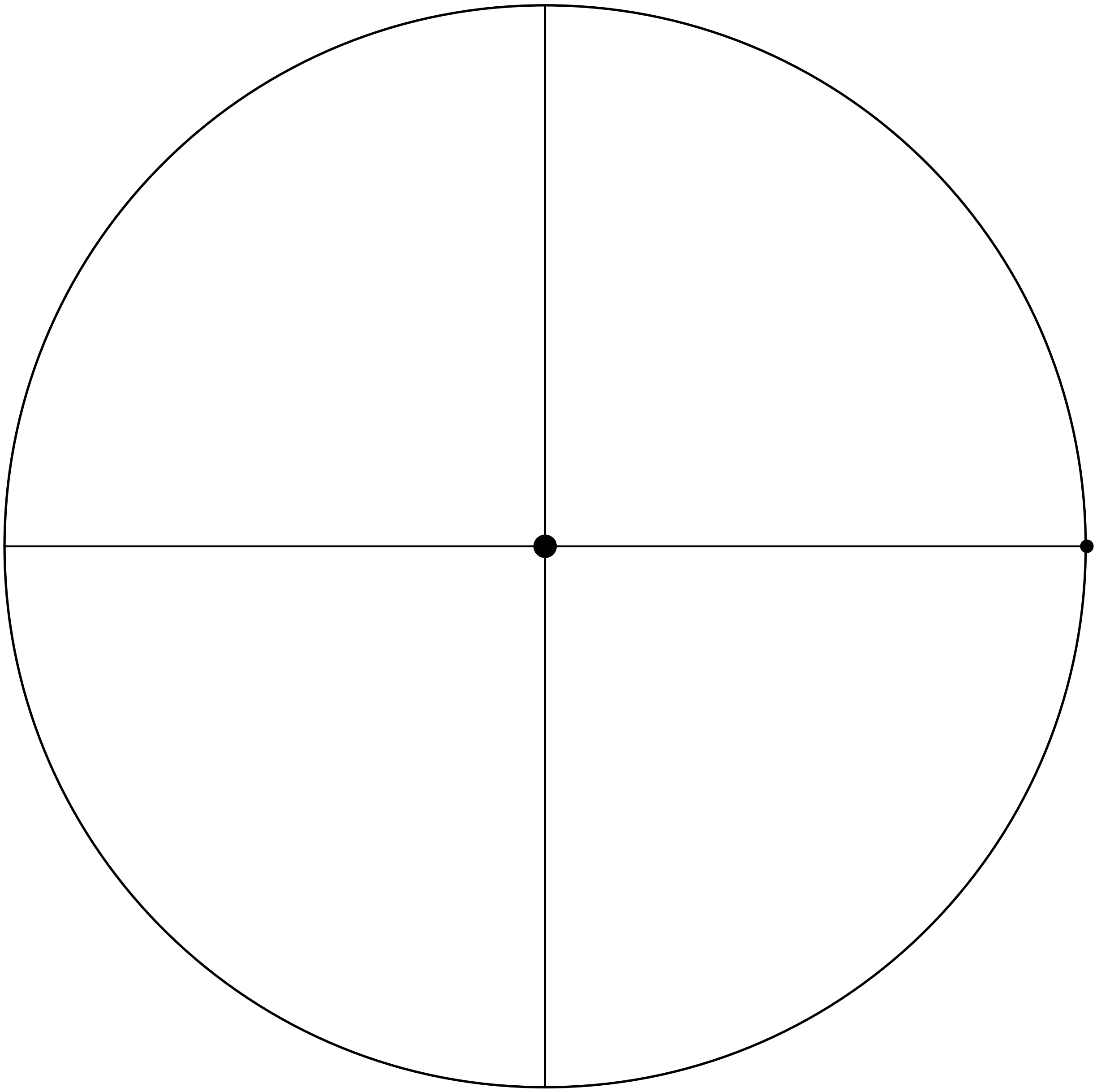
$block$	z_j	$multiplicity$
0	1	1
1	$\omega/r^{1/2}$	r
2	ω^2/r	r^2
$\vdots$	$\vdots$	$\vdots$
j	$\omega^j/r^{j/2}$	r^j
$\vdots$	$\vdots$	$\vdots$
$m-1$	$\omega^{m-1}/r^{(m-1)/2}$	r^{m-1}
m	$-S/k$	k
$m+1$	iS/k	$k/2$
$m+2$	$-iS/k$	$k/2$



ϵ= 0.750
m = 2 ; m+3 = 5 ; r = 29
k = 0.400E+02 ; n = 0.110E+03



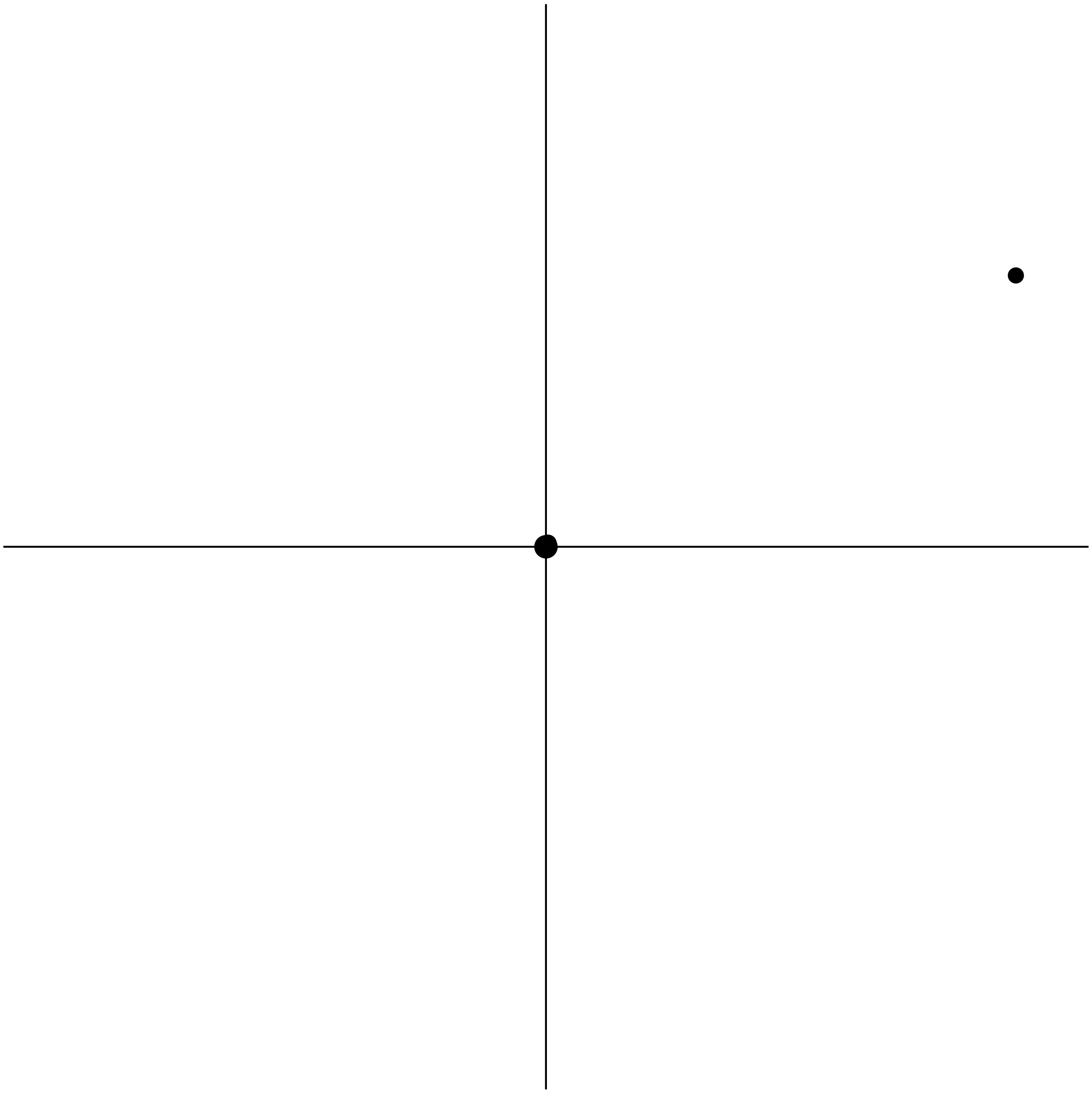
ϵ= 0.500
m = 4 ; m+3 = 7 ; r = 1025
k = 0.215E+10 ; n = 0.539E+10



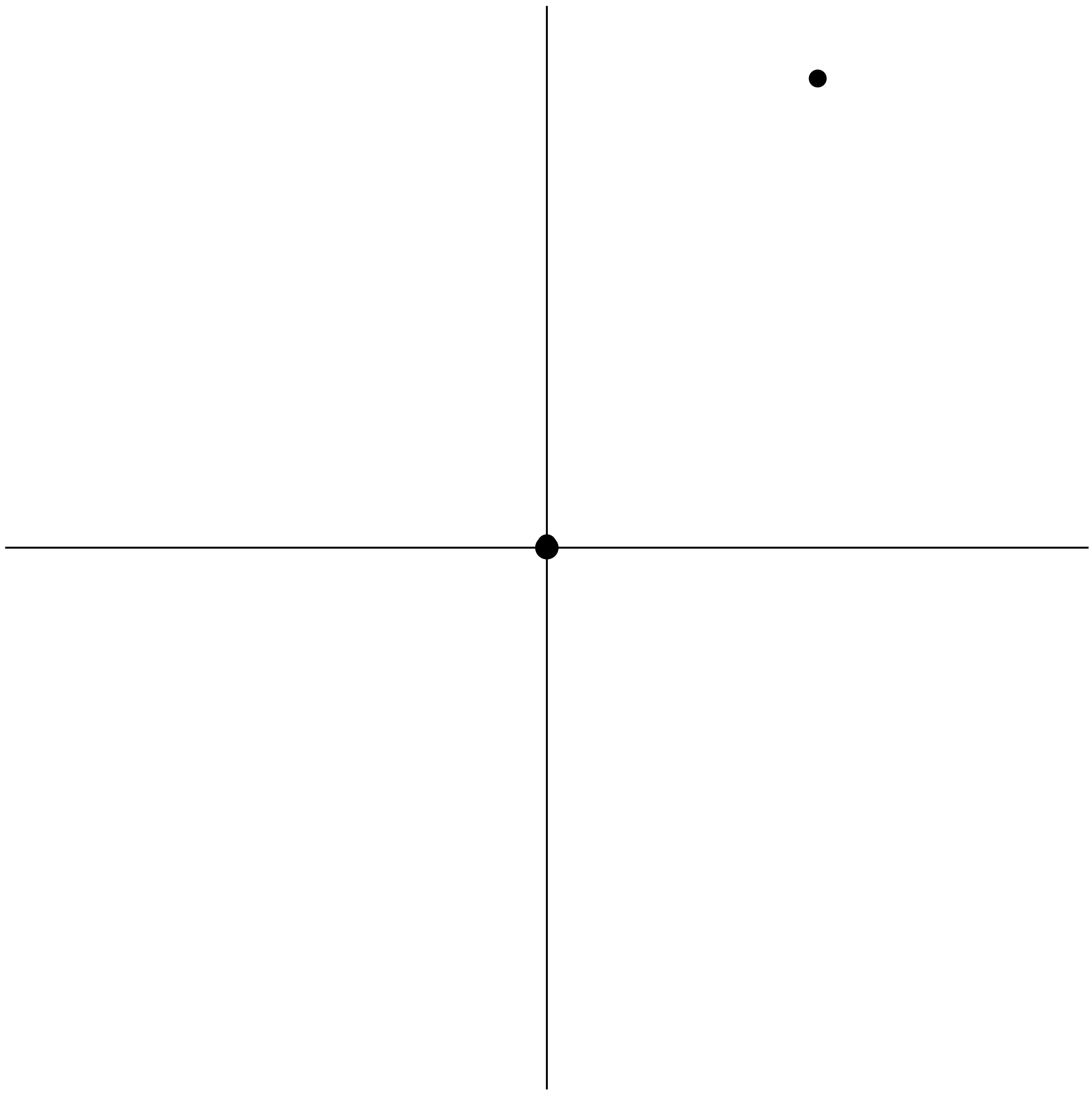
ϵ= **0.250**

m = 6 ; m+3 = 9 ; r = 20737

k = 0.153E+23 ; n = 0.345E+23



ϵ**= 0.250** Deleted points: 1 Scaling: 144.
m = 6 ; m+3 = 9 ; r = 20737
k = 0.153E+23 ; n = 0.345E+23



ϵ**= 0.250** Deleted points: 2 Scaling: 20737.

m = 6 ; m+3 = 9 ; r = 20737

k = 0.153E+23 ; n = 0.345E+23

TRIVARIATE DISTRIBUTIONS

Wasserstein distance (colored or not) between the distributions of X and Y, minimized for rotation:

$$D^2 = E[(X-Y)'(X-Y)] - 2q'Bq$$

q: unit quaternion associated to the largest eigenvalue of B

$$B = \left[\begin{array}{c|c} 0 & E[Y \wedge X] \\ \hline E[Y \wedge X]' & (Z+Z') - I \cdot Tr(Z+Z') \end{array}\right]$$

$$Z = E[YX']$$

Remark: the three components of $E[Y \wedge X]$
are computed from the elements of Z.

Setting X centered and Y distributed as $-X$:

$$\chi = \frac{3}{4T} Inf_{\{W\}} D^2$$

W: joint distribution of (X, Y)

In the non colored case:

Theorem: $Sup(\chi) \in [1/2; 1]$

$Sup(\chi) = ???$

HIGHER DIMENSIONS

The following family X_ε of finite discrete distributions has a chiral index χ_ε tending to $1/2$ when ε tends to zero.

There are $d+1$ weighted points in R^d (simplex).

X_ε: array of the $d+1$ points
M: respective weights of the $d+1$ points

$$X_\varepsilon = \left(\begin{array}{cccc} 0 & 0 & \dots & 0 \\ \hline 1/\varepsilon & 0 & \dots & 0 \\ 0 & 1/\varepsilon^2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1/\varepsilon^d \end{array}\right) \qquad M = \frac{1}{c}\begin{pmatrix} 1 \\ \varepsilon^2 \\ \vdots \\ \varepsilon^{2d} \end{pmatrix} \qquad c = \sum_{i=0}^{i=d} \varepsilon^{2i}$$

This family of discrete distributions is asymptotically isoinertial, i.e. its covariance matrix tend to be proportional to I.

$$Lim_{\varepsilon \longrightarrow 0}(\chi_\varepsilon) = 1/2$$

This is an optimal upper bound for the chiral index when $d=1$, but not for $d=2$.

Calculating this upper bound for any d is an open problem.
(and the optimal rotation is unknown for $d \geq 4$)

Conjectures:
- The uper bound of the chiral index is asymptotically reachable only for isoinertial distributions.
- This upper bound is unreachable for any d

MISCELLANEOUS

Colored sample:
$\chi = \frac{d}{4nT} Min_{\{P,R\}}[Tr(X - PXQ'R')'A(X - PXQ'R')]$

Can be generalized when the n points are the vertices of a **graph**, $\{P\}$ being the set of permutations associated to the **GRAPH AUTOMORPHISMS**.

Examples in chemistry:

The graph of the water molecule H-O-H has three nodes and two edges, and has 2 automorphisms.

The graph of Br-CHF-Cl has 5 nodes and 4 edges, and has only 1 automorphism.
(assuming a regular tetrahedron geometry, we would have $\chi = 1$, and NOT $\chi = 0$).

Generalizing the case of samples of colored mixtures:
Cyclobutane squeletton C4: **there are 8 permutations, not 24**, although there are no colors!
Works with colors, but difficult to generalize to continuous distributions, even without colors.

SOME OTHER OPEN PROBLEMS

How $<<$ idealize $>>$ a quasi-achiral set ?

How measure chirality when the mass is infinite ?
(lattices, infinite helices, etc.)