Cross-identification au CDS

François-Xavier Pineau[1], Thomas Boch[1], Sébastien Derrière[1], Gilles Landais[1], André Schaaff[1], Noémie Wali[2]

[1]Observatoire Astronomique de Strasbourg, Université de Strasbourg, CNRS
[2]Université de Technologie de Belfort-Montbéliard (UTBM)

Réunion annuelle de l'Action Spécifique Observatoires Virtuels France
Institut d'Astrophysique de Paris, le 15 mars 2016

Sommaire

X-Match within the VO

Using VO tools

- No dedicated VO protocol
- Topcat
  - Use the VO Registry and download a VOTable
  - Perform one Cone Search by row
    - good for VizieR's stats...
    - ...but saturates the server when multi-threaded!
- Topcat / Aladin
  - Use the VO Registry and download VOTables
  - Perform the x-match locally
    - Tycho2 vs 2MASS
    - $\Rightarrow$ Download a lot of data
    - $\rightsquigarrow$ `java.lang.OutOfMemoryError`
    - $\Rightarrow$ User must slice the sky, ...
    - but no simple VO protocol allowing full-sky download by non-overlapping regions

X-Match within the VO

Using VO tools

- No dedicated VO protocol
- Topcat
  - Use the VO Registry and download a VOTable
  - Perform one Cone Search by row
    - ★ good for VizieR's stats...
    - ★ ...but saturates the server when multi-threaded!
- Topcat / Aladin
  - Use the VO Registry and download VOTables
  - Perform the x-match locally
    - ★ Tycho2 vs 2MASS
    - ★ ⇒ Download a lot of data
    - ★ ⇝ `java.lang.OutOfMemoryError`
    - ★ ⇒ User must slice the sky, ...
    - ★ but no simple VO protocol allowing full-sky download by non-overlapping regions

X-Match within the VO

Using VO tools

- No dedicated VO protocol
- Topcat
  - Use the VO Registry and download a VOTable
  - Perform one Cone Search by row
    - ★ good for VizieR's stats...
    - ★ ...but saturates the server when multi-threaded!
- Topcat / Aladin
  - Use the VO Registry and download VOTables
  - Perform the x-match locally
    - ★ Tycho2 vs 2MASS
    - ★ $\Rightarrow$ Download a lot of data
    - ★ $\rightsquigarrow$ `java.lang.OutOfMemoryError`
    - ★ $\Rightarrow$ User must slice the sky, ...
    - ★ but no simple VO protocol allowing full-sky download by non-overlapping regions

X-Match within the VO

TAP for small/medium x-matches

- Full VO x-match: TAP is the only way
- TAP (Table Access Protocol)
  - ADQL: Astronomical Data Query Language
  - UWS: Universal Worker Service
- Example

```
SELECT TOP 1000 *
FROM twomass AS a
JOIN tycho2  AS b
ON 1=CONTAINS(POINT('ICRS', a.ra, a.dec),
              CIRCLE('ICRS', b.ra, b.dec, 5./3600.))
```

- ESA GAIA access: TAP, TAP, TAP!

X-Match within the VO

TAP for small/medium x-matches

- Full VO x-match: TAP is the only way
- TAP (Table Access Protocol)
  - ADQL: Astronomical Data Query Language
  - UWS: Universal Worker Service
- Example

```
SELECT TOP 1000 *
FROM twomass AS a
JOIN tycho2  AS b
ON 1=CONTAINS(POINT('ICRS', a.ra, a.dec),
              CIRCLE('ICRS', b.ra, b.dec, 5./3600.))
```

- ESA GAIA access: TAP, TAP, TAP!

X-Match using TAP VizieR

TAP for small/medium x-matches

- CDS TAP Service **benchmark** (Gilles Landais)
  - **No output (simple `count(*)`)**
  - Before optimization (Q3C / H3C)
    - ★ Hipparcos Main (118 k) vs 2MASS (470 M): 14 min
    - ★ Tycho2 (2 M) vs 2MASS (470 M): 48 min
    - ★ 2 large catalogues: $>$ 1 day
  - After optimization (H3C)
    - ★ Hipparcos Main (118 k) vs 2MASS (470 M): 6.5 min
    - ★ Tycho2 (2 M) vs 2MASS (470 M): 11.5 min
    - ★ GSC ACT (25 M) vs 2MASS (470 M): 1 h 30 min
- Pro: SQL flexibility
- Limit: all known TAP implementations on top of a SGBD
  - PostgreSQL (CDS, GAVO, ESAC): no multi-threading
  - Possible performances issues
  - SGBD fitted for large catalogues x-match?

X-Match using TAP VizieR

TAP for small/medium x-matches

- CDS TAP Service **benchmark** (Gilles Landais)
  - **No output (simple `count(*)`)**
  - Before optimization (Q3C / H3C)
    - ★ Hipparcos Main (118 k) vs 2MASS (470 M): 14 min
    - ★ Tycho2 (2 M) vs 2MASS (470 M): 48 min
    - ★ 2 large catalogues: $>$ 1 day
  - After optimization (H3C)
    - ★ Hipparcos Main (118 k) vs 2MASS (470 M): 6.5 min
    - ★ Tycho2 (2 M) vs 2MASS (470 M): 11.5 min
    - ★ GSC ACT (25 M) vs 2MASS (470 M): 1 h 30 min
- Pro: SQL flexibility
- Limit: all known TAP implementations on top of a SGBD
  - PostgreSQL (CDS, GAVO, ESAC): no multi-threading
  - Possible performances issues
  - SGBD fitted for large catalogues x-match?

The CDS X-Match service

Web interface

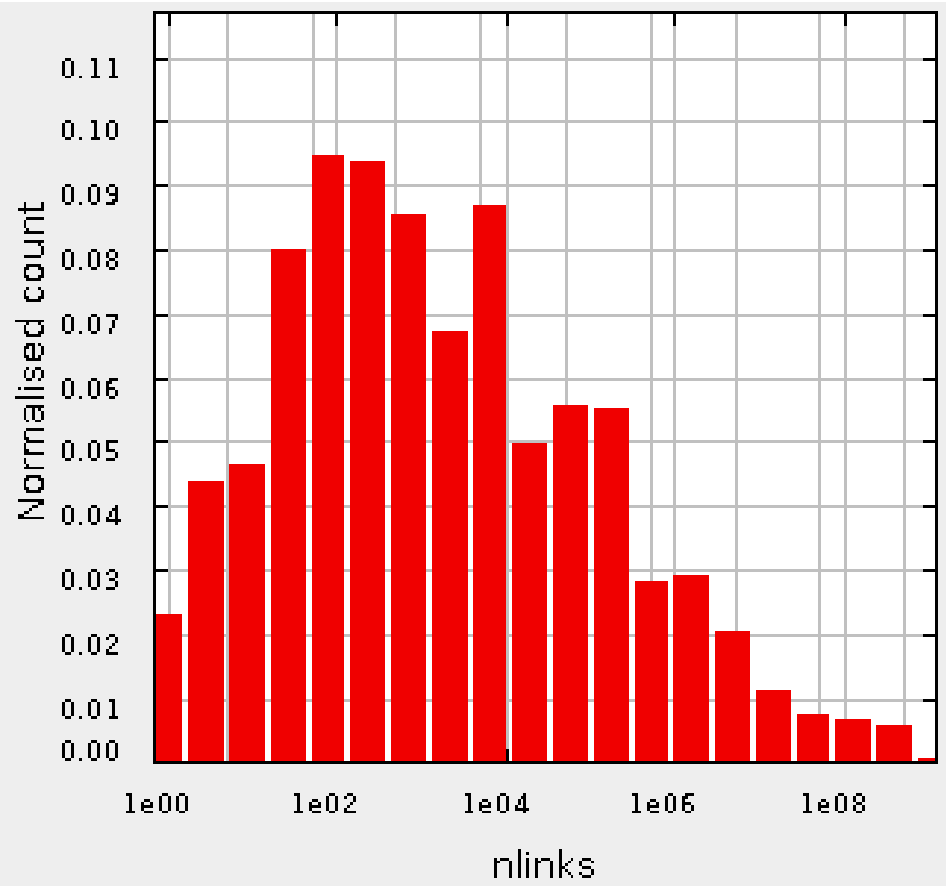
- Released in 2011
- General purpose service (any table having positions) ⇒ no probabilities
- **UWS** (Grégory Mantelet's library)
- **Very efficient** basic xmatch designed for the biggest available catalogues
  - 2 catalogues at the same time
  - **Simple** (few options)
  - E.g. SDSS DR9 vs 2MASS at 2" done in 15 min
    - ★ xmatching: 5 min (50 M links)
    - ★ building result file: 10 min (14 GB)
    - ★ running on a **single** server
- >20 G links computed in 2015, e.g.
  - NOMAD (1.1 G) vs USNOA2 (0.5 G)
  - 750 M links, 117 GB, 1h28

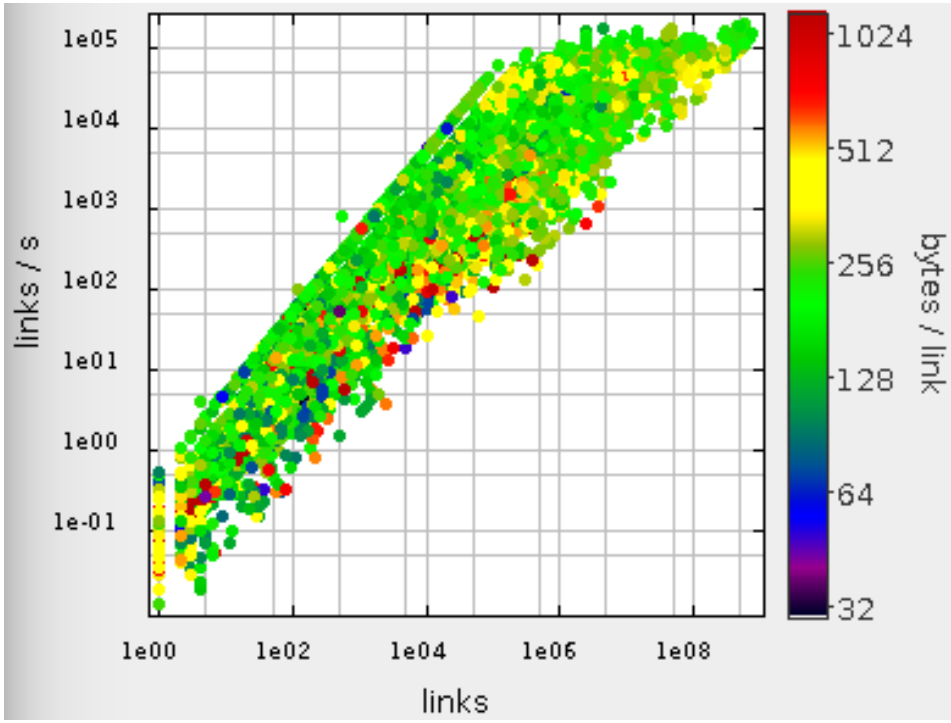
The CDS X-Match service

Web interface

- Released in 2011
- General purpose service (any table having positions) ⇒ no probabilities
- **UWS** (Grégory Mantelet's library)
- **Very efficient** basic xmatch designed for the biggest available catalogues
  - 2 catalogues at the same time
  - **Simple** (few options)
  - E.g. SDSS DR9 vs 2MASS at 2" done in 15 min
    - ★ xmatching: 5 min (50 M links)
    - ★ building result file: 10 min (14 GB)
    - ★ running on a **single** server
- >20 G links computed in 2015, e.g.
  - NOMAD (1.1 G) vs USNOA2 (0.5 G)
  - 750 M links, 117 GB, 1h28

The CDS X-Match service

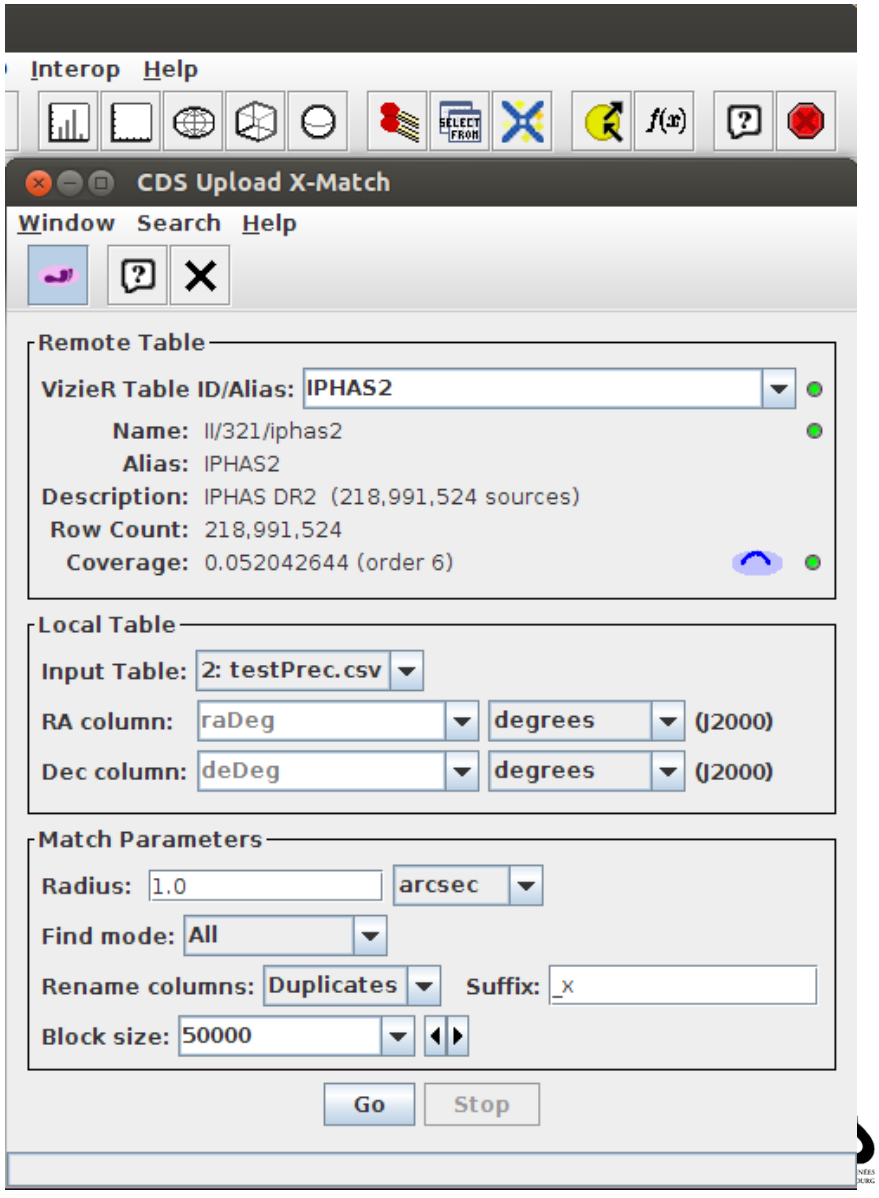




The CDS X-Match service

HTTP API

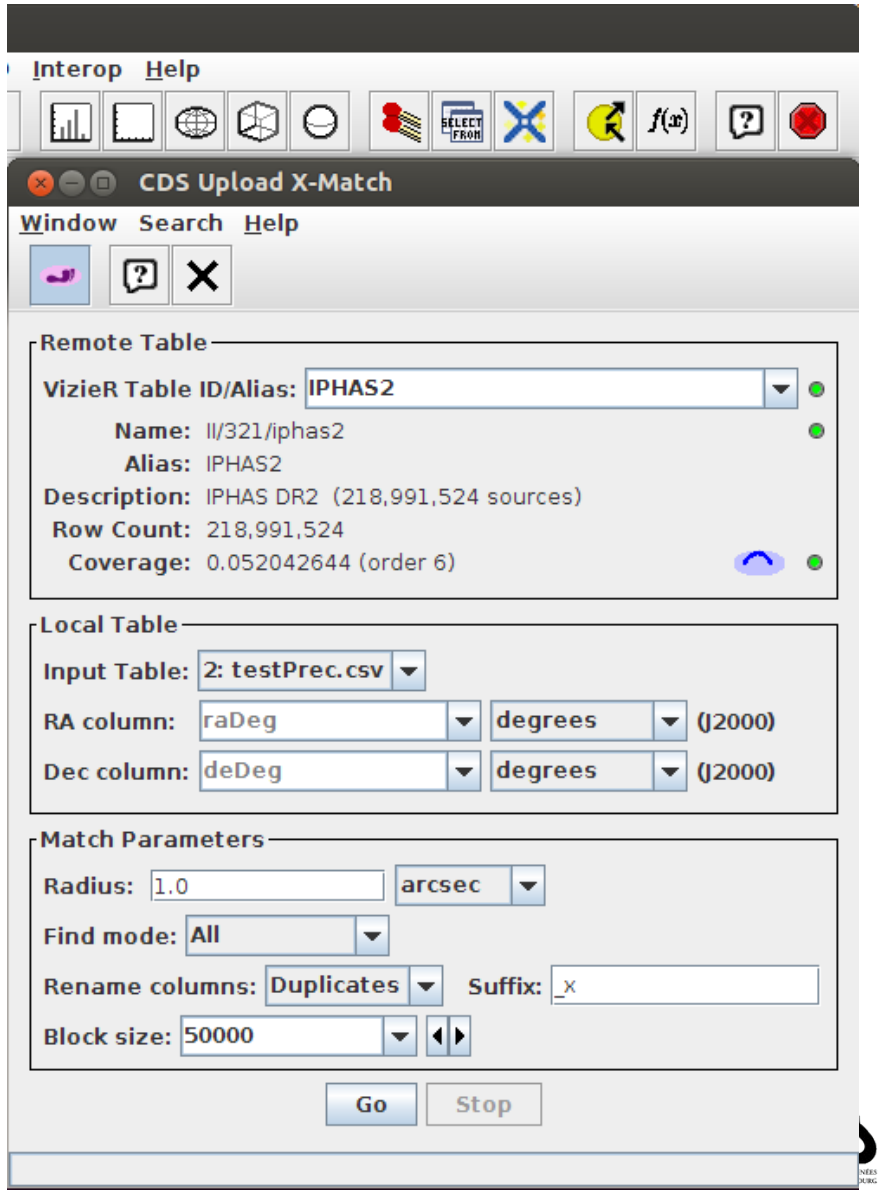
- Realeased in 2013
- Uses **DALI**
- Synchronous programmatic access
- Available from
  - Web Browser, wget, curl, ...
  - TOPCAT, Portal MAST, Astropy
- Output limited to 2 M rows
- 2015: 3 G sources submitted (8 M / day)



The CDS X-Match service

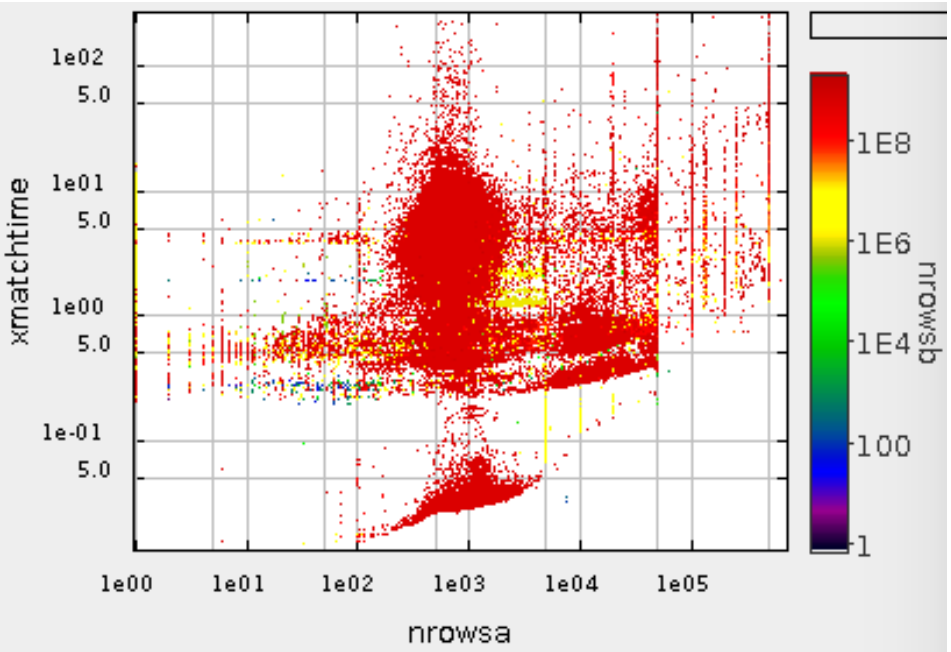
HTTP API

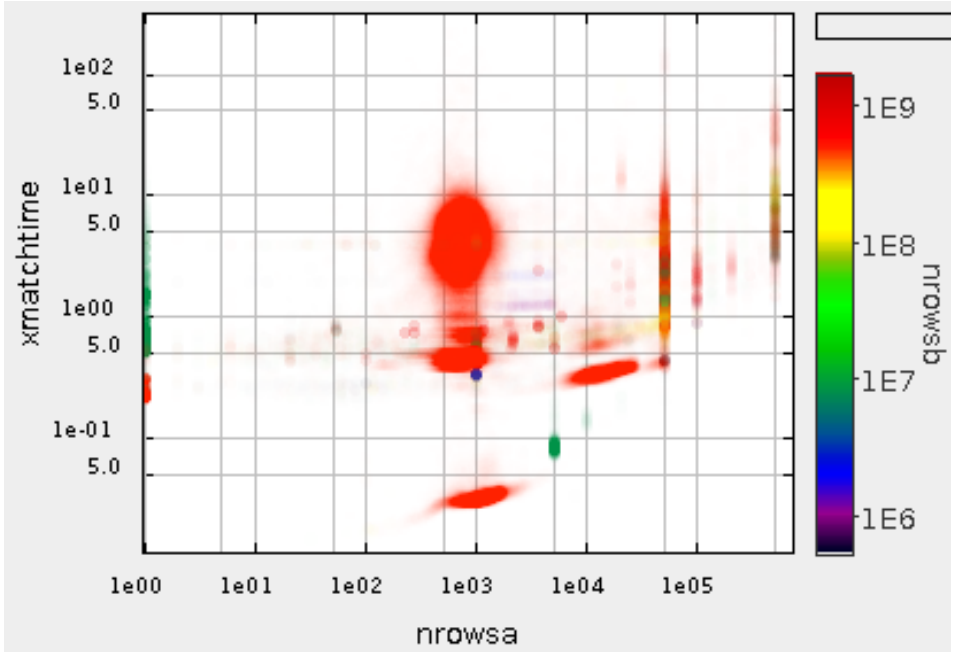
- Realeased in 2013
- Uses **DALI**
- Synchronous programmatic access
- Available from
  - Web Browser, wget, curl, ...
  - TOPCAT, Portal MAST, Astropy
- Output limited to 2 M rows
- 2015: 3 G sources submitted (8 M / day)



The CDS X-Match service

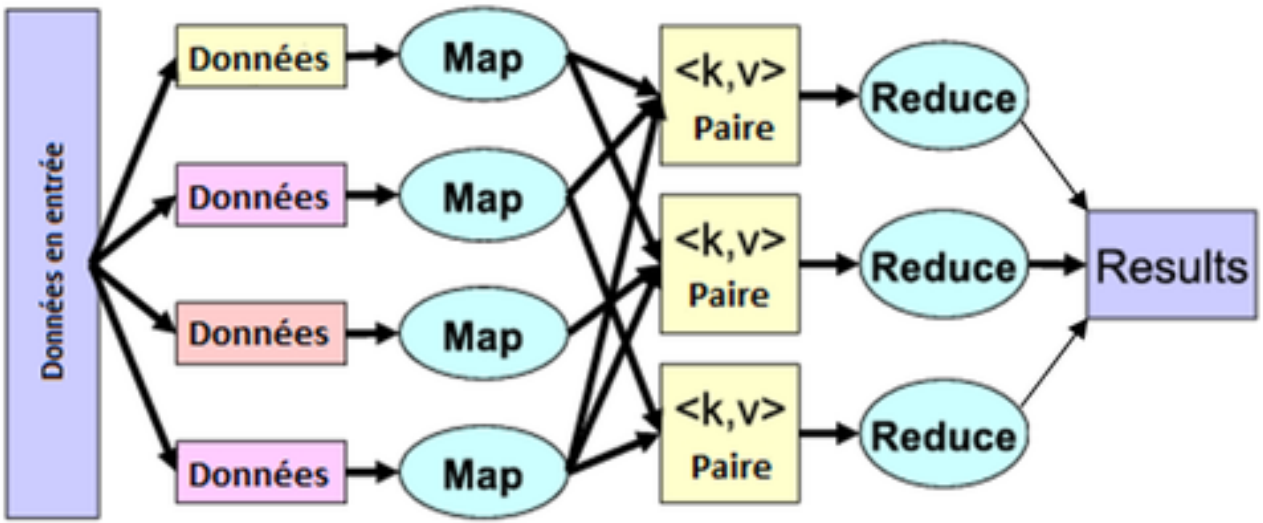
HTTP API





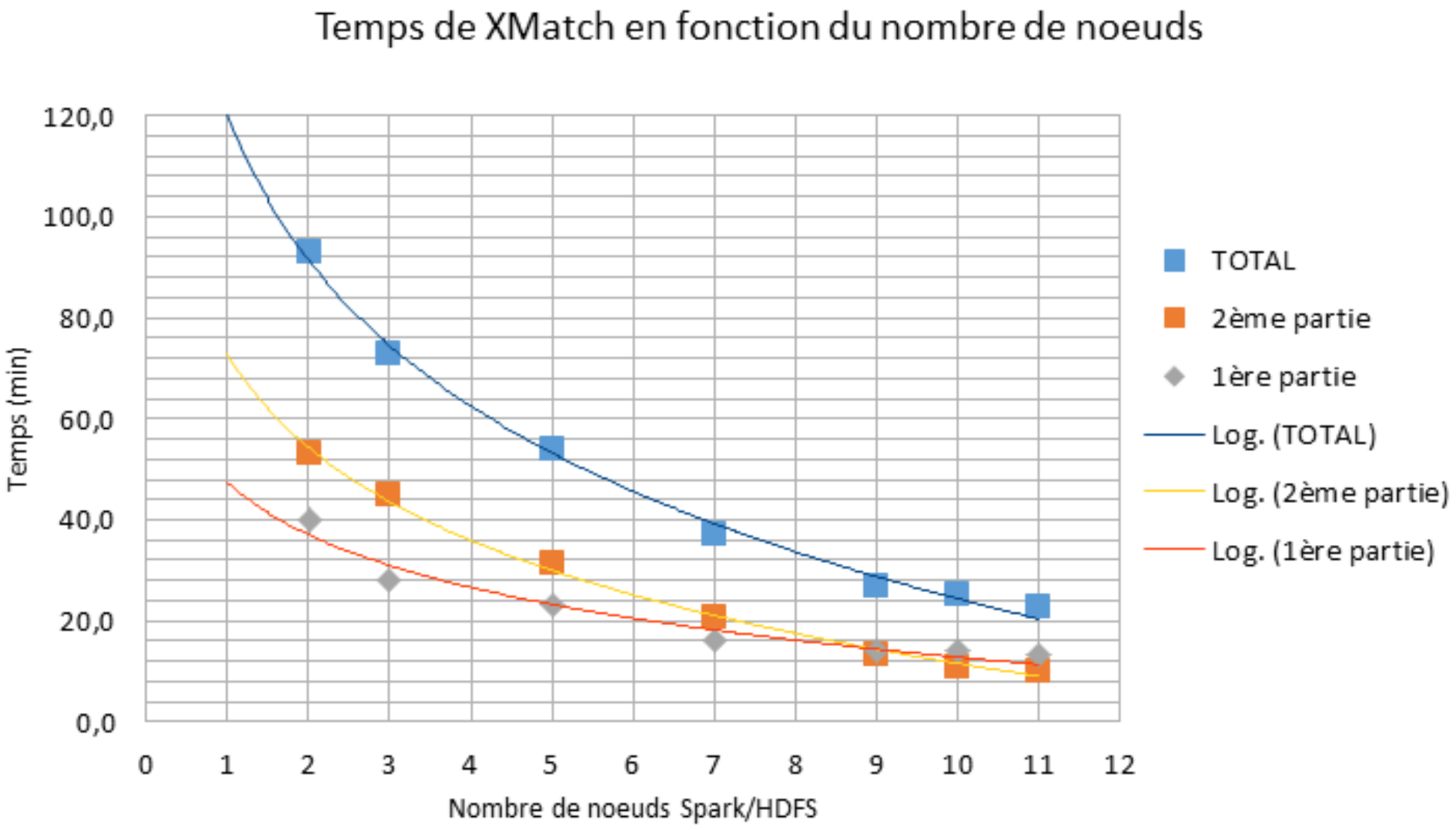
Test of Big Data techno

- N. Wali, A. Schaaff, F.-X. Pineau
- HADOOP / SPARK: MapReduce
- Test on SDSS DR9 / 2MASS
- Bottleneck: shuffle
  - Co-partitioning OK
  - But no way to co-locate data!! (no "block affinity groups")



Credit: Wikipedia

Test of Big Data techno



Sommaire

From x-match to χ-match and x-id

- Existing tools: simple fixed radius x-match
- ARCHES (FP7): we would like identifications
  - very tricky / complex!
  - 2 (possibly) independant parts
    - astrometric part (take into account positional errors)
    - photometric part (not adressed here!)
    - $proba_{id} \propto prior \times likelihood_{astrom} \times likelihood_{phot}$
    - $proba_{id} \propto proba_{astrom} \times likelihood_{phot}$
  - SEDs $\Rightarrow$ multi-catalogue x-match

Candidate selection

Method

- Steps to probabilistic positional x-match
  - Make simplifying assumptions
  - Select candidates: select and group together sources possibly being various detections of a same real source
    - ⋆ Need for a selection criterion
  - Make hypothesis: are the sources really from a same real sources or from different real sources?
  - For each hypothesis:
    - ⋆ derive the associated *likelihood*
    - ⋆ derive the associated *prior*
  - Compute astrometry based probabilities

Candidate selection

Simplifying assumptions

- Radical simplifying assumptions:
  - No proper motions
  - No blending
  - No clustering (density of sources = Poisson law)
  - No systematic offsets
  - You can trust positional uncertainties provided in catalogues

Candidate selection

Selection criterion

How to select a group of n sources from n distinct catalogues as possibly being various observations of a same actual source?

- *Statistical hypothesis testing*
  - H_0 (null hypothesis): all n sources are from the same real source
  - $H_1 = \bar{H_0}$ (alternative hypothesis): at least one source (out of n) is spurious
- User input: γ, the probability to accept H_0 while it is true
  - γ (I call it completeness) is called *true negative* rate
  - we usually fix $\gamma = 0.9973$ (99.73%, value of the 3σ rule in 1 dimensional pb)
  - $\Leftrightarrow$ fixing the **type I error** $= 0.027\% =$ proba to reject null hypothesis while it is true
  - we (theoretically) miss 27/10 000 real association
- **The criterion used is simply a χ^2 test of $2(n-1)$ degrees of freedom**

Candidate selection

Selection criterion

How to select a group of n sources from n distinct catalogues as possibly being various observations of a same actual source?

- *Statistical hypothesis testing*
  - H_0 (null hypothesis): all n sources are from the same real source
  - $H_1 = \bar{H_0}$ (alternative hypothesis): at least one source (out of n) is spurious
- User input: γ, the probability to accept H_0 while it is true
  - γ (I call it completeness) is called *true negative* rate
  - we usually fix $\gamma = 0.9973$ (99.73%, value of the 3σ rule in 1 dimensional pb)
  - $\Leftrightarrow$ fixing the **type I error** $= 0.027\% =$ proba to reject null hypothesis while it is true
  - we (theoretically) miss 27/10 000 real association
- **The criterion used is simply a χ^2 test of $2(n-1)$ degrees of freedom**

Candidate selection

Selection criterion

How to select a group of n sources from n distinct catalogues as possibly being various observations of a same actual source?

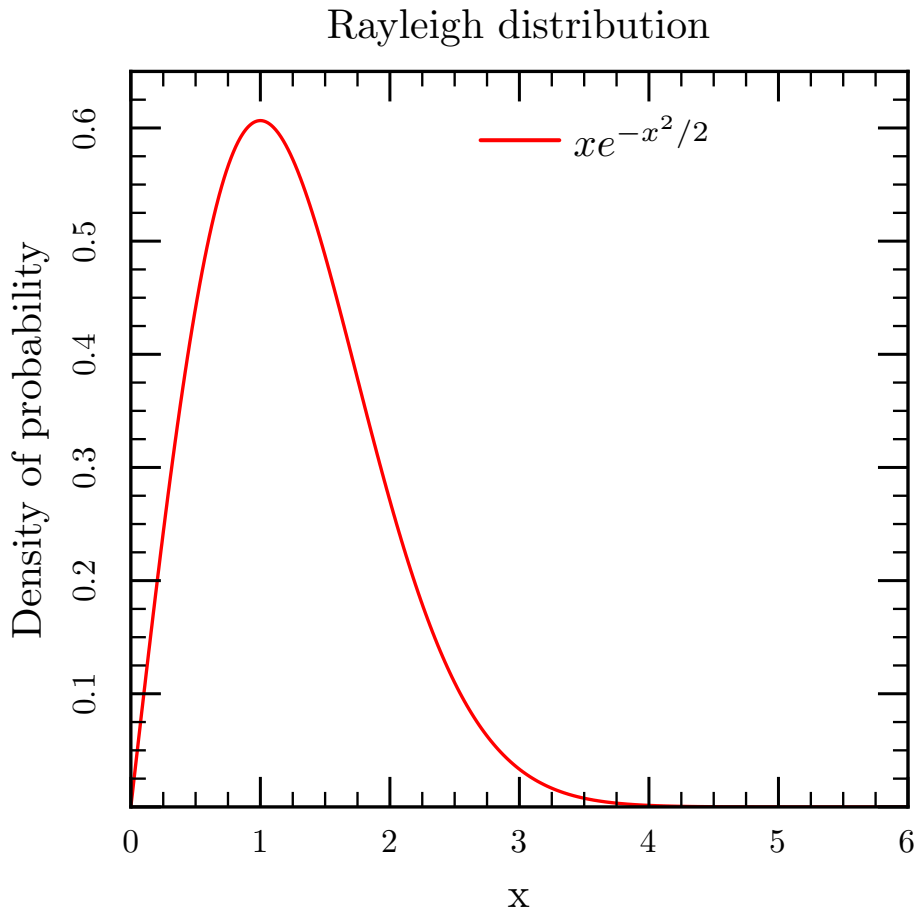
- *Statistical hypothesis testing*
  - H_0 (null hypothesis): all n sources are from the same real source
  - $H_1 = \bar{H_0}$ (alternative hypothesis): at least one source (out of n) is spurious
- User input: γ, the probability to accept H_0 while it is true
  - γ (I call it completeness) is called *true negative* rate
  - we usually fix $\gamma = 0.9973$ (99.73%, value of the 3σ rule in 1 dimensional pb)
  - $\Leftrightarrow$ fixing the **type I error** $= 0.027\% =$ proba to reject null hypothesis while it is true
  - we (theoretically) miss 27/10 000 real association
- **The criterion used is simply a χ^2 test of $2(n-1)$ degrees of freedom**

Candidate selection

Classical 2 catalogues case

- For real associations, i.e. when H_0 is true
  - The distribution of normalized distances is a Rayleigh distribution of scale $\sigma = 1$

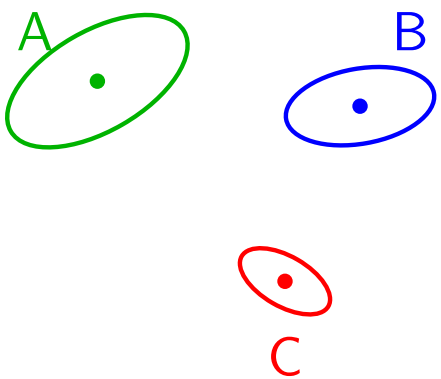
$$r \overset{H_0}{\sim} Rayleigh$$



Candidate selection

Iterative form

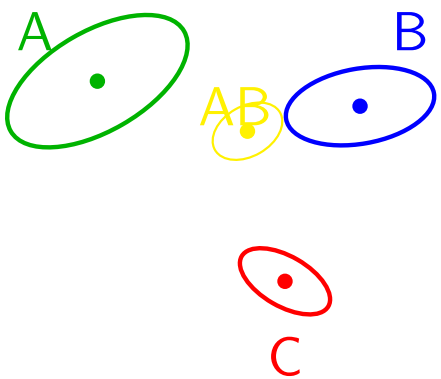
- $\chi_{dof=2(n-1)} \Leftrightarrow (n-1)\chi_{dof=2}$
- $\Rightarrow$ we can perform (n-1) successives and iteratives xmatches
- At each step, we use for next position and next error ellipse
  - the weighted mean position of the previous xmatch (MLE)
  - the error on the weighted mean position of the previous xmatch
- Result independant of the xmatch order (for INNER JOIN!)



Candidate selection

Iterative form

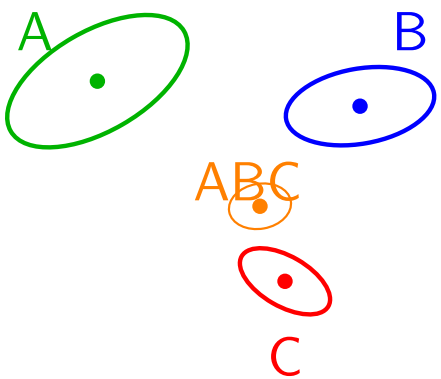
- $\chi_{dof=2(n-1)} \Leftrightarrow (n-1)\chi_{dof=2}$
- $\Rightarrow$ we can perform (n-1) successives and iteratives xmatches
- At each step, we use for next position and next error ellipse
  - the weighted mean position of the previous xmatch (MLE)
  - the error on the weighted mean position of the previous xmatch
- Result independant of the xmatch order (for INNER JOIN!)



Candidate selection

Iterative form

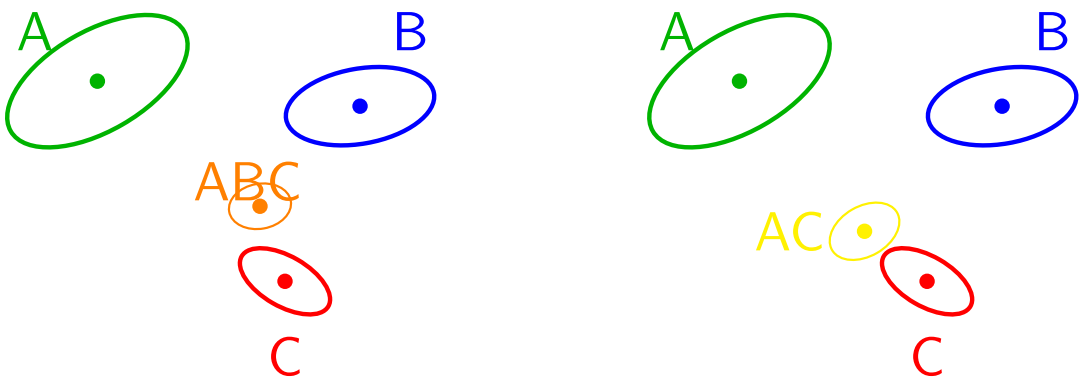
- $\chi_{dof=2(n-1)} \Leftrightarrow (n-1)\chi_{dof=2}$
- $\Rightarrow$ we can perform (n-1) successives and iteratives xmatches
- At each step, we use for next position and next error ellipse
  - the weighted mean position of the previous xmatch (MLE)
  - the error on the weighted mean position of the previous xmatch
- Result independant of the xmatch order (for INNER JOIN!)



Candidate selection

Iterative form

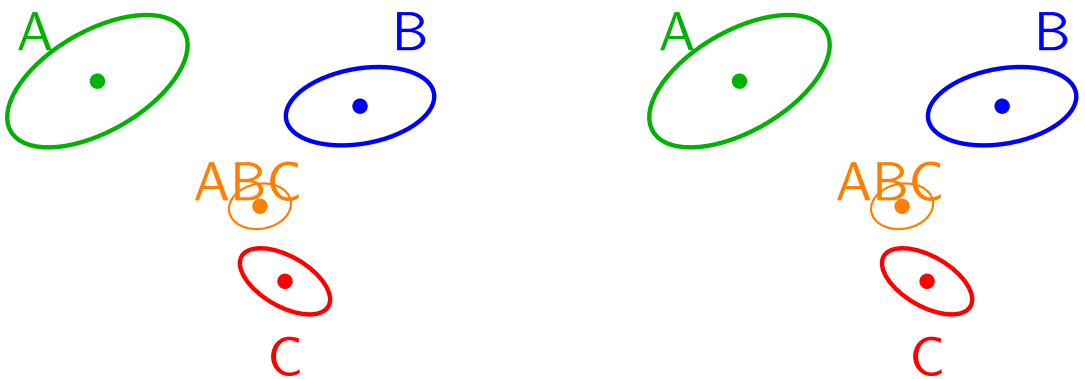
- $\chi_{dof=2(n-1)} \Leftrightarrow (n-1)\chi_{dof=2}$
- $\Rightarrow$ we can perform (n-1) successives and iteratives xmatches
- At each step, we use for next position and next error ellipse
  - the weighted mean position of the previous xmatch (MLE)
  - the error on the weighted mean position of the previous xmatch
- Result independant of the xmatch order (for INNER JOIN!)



Candidate selection

Iterative form

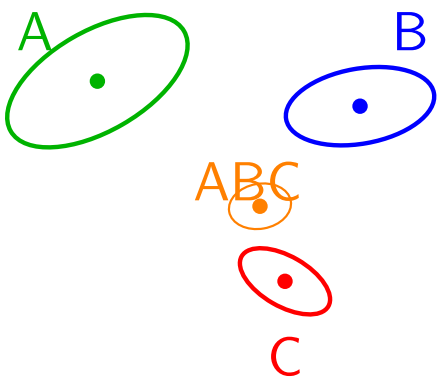
- $\chi_{dof=2(n-1)} \Leftrightarrow (n-1)\chi_{dof=2}$
- $\Rightarrow$ we can perform (n-1) successives and iteratives xmatches
- At each step, we use for next position and next error ellipse
  - the weighted mean position of the previous xmatch (MLE)
  - the error on the weighted mean position of the previous xmatch
- Result independant of the xmatch order (for INNER JOIN!)

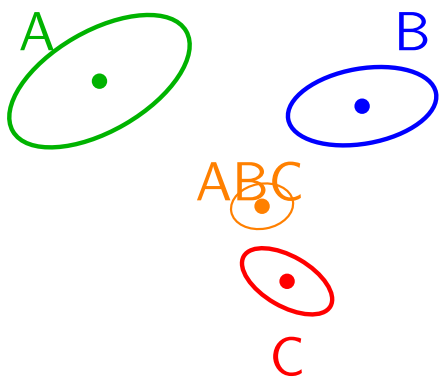


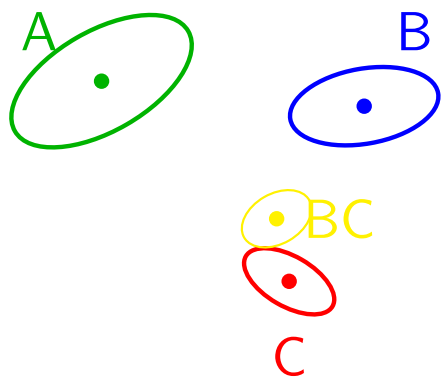
Candidate selection

Iterative form

- $\chi_{dof=2(n-1)} \Leftrightarrow (n-1)\chi_{dof=2}$
- $\Rightarrow$ we can perform (n-1) successives and iteratives xmatches
- At each step, we use for next position and next error ellipse
  - the weighted mean position of the previous xmatch (MLE)
  - the error on the weighted mean position of the previous xmatch
- Result independant of the xmatch order (for INNER JOIN!)



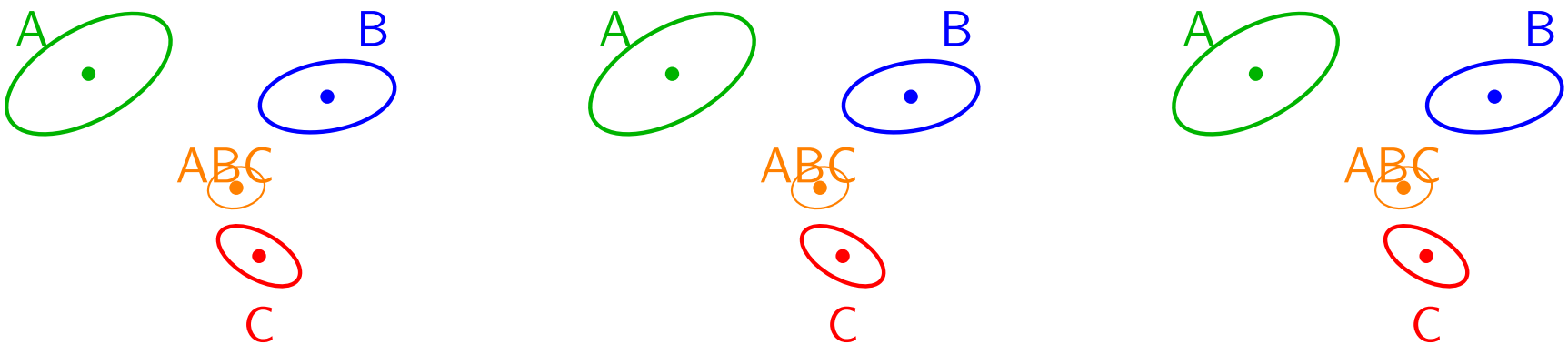




Candidate selection

Iterative form

- $\chi_{dof=2(n-1)} \Leftrightarrow (n-1)\chi_{dof=2}$
- $\Rightarrow$ we can perform (n-1) successives and iteratives xmatches
- At each step, we use for next position and next error ellipse
  - the weighted mean position of the previous xmatch (MLE)
  - the error on the weighted mean position of the previous xmatch
- Result independant of the xmatch order (for INNER JOIN!)



Making hypotheses

For 2 and 3 catalogues

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - AB (H_0)
    - A_B

Making hypotheses

For 2 and 3 catalogues

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - AB (H_0)
    - A_B

- For 3 catalogues
  - 5 hypothesis

Making hypotheses

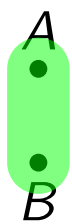
For 2 and 3 catalogues

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - AB (H_0)
    - A_B



- For 3 catalogues
  - 5 hypothesis

Making hypotheses

For 2 and 3 catalogues

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - ★ AB (H_0)
    - ★ A_B

- For 3 catalogues
  - 5 hypothesis
    - ★ ABC (H_0)
    - ★ A_BC
    - ★ AB_C
    - ★ AC_B
    - ★ A_B_C

Making hypotheses

For 2 and 3 catalogues

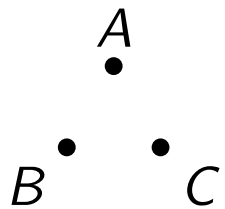
To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - ⋆ AB (H_0)
    - ⋆ A_B

- For 3 catalogues
  - 5 hypothesis
    - ⋆ ABC (H_0)
    - ⋆ AB_C
    - ⋆ AC_B
    - ⋆ A_BC
    - ⋆ A_B_C



Making hypotheses

For 2 and 3 catalogues

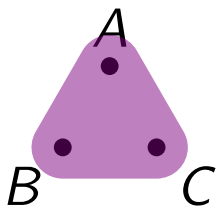
To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - AB (H_0)
    - A_B

- For 3 catalogues
  - 5 hypothesis
    - ABC (H_0)
    - AB_C
    - AC_B
    - A_BC
    - A_B_C



Making hypotheses

For 2 and 3 catalogues

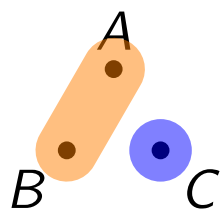
To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - ★ AB (H_0)
    - ★ A_B

- For 3 catalogues
  - 5 hypothesis
    - ★ ABC (H_0)
    - ★ AB_C
    - ★ AC_B
    - ★ A_BC
    - ★ A_B_C



Making hypotheses

For 2 and 3 catalogues

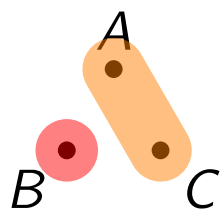
To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - ★ AB (H_0)
    - ★ A_B

- For 3 catalogues
  - 5 hypothesis
    - ★ ABC (H_0)
    - ★ AB_C
    - ★ AC_B
    - ★ A_BC
    - ★ A_B_C



Making hypotheses

For 2 and 3 catalogues

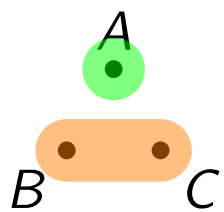
To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - ★ AB (H_0)
    - ★ A_B

- For 3 catalogues
  - 5 hypothesis
    - ★ ABC (H_0)
    - ★ AB_C
    - ★ AC_B
    - ★ A_BC
    - ★ A_B_C



Making hypotheses

For 2 and 3 catalogues

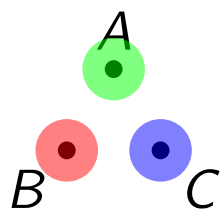
To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^{n} p(H_i) = 1$$

- For 2 catalogues
  - 2 hypothesis
    - ★ AB (H_0)
    - ★ A_B

- For 3 catalogues
  - 5 hypothesis
    - ★ ABC (H_0)
    - ★ AB_C
    - ★ AC_B
    - ★ A_BC
    - ★ A_B_C



Making hypotheses

For n-catalogues

- We generalised for n catalogues
- The number of hypothesis to be tested is given by the BELL number

Table : Values of the seven first BELL numbers

n	2	3	4	5	6	7
B_n	2	5	15	52	203	877

  - n number of catalogues
  - n=5 catalogues $\rightsquigarrow$ 52 probabilities to be computed

- $\Rightarrow$ Combinatorial explosion!

Making hypotheses

For n-catalogues

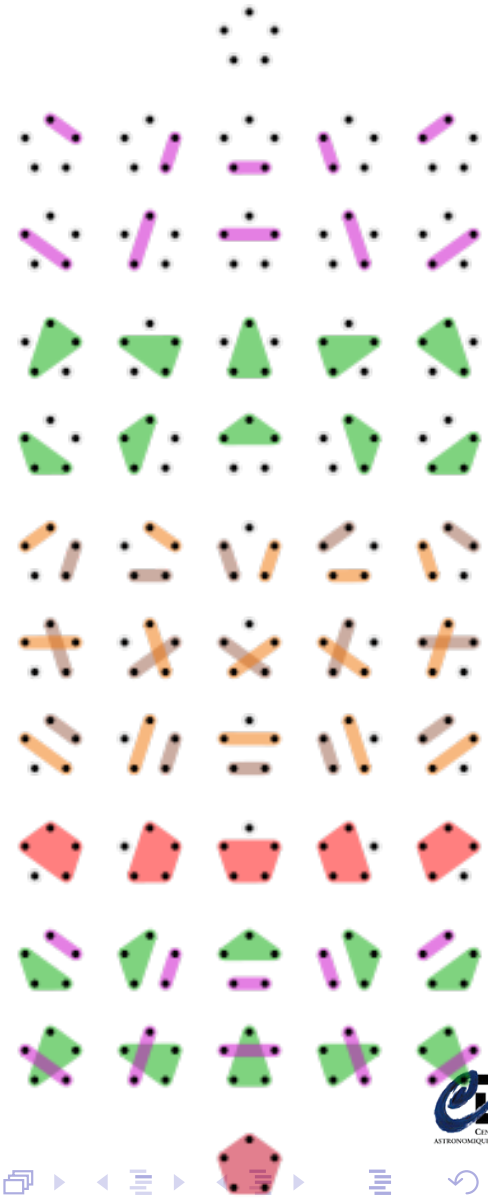
- We generalised for n catalogues
- The number of hypothesis to be tested is given by the BELL number

Table : Values of the seven first BELL numbers

n	2	3	4	5	6	7
B_n	2	5	15	52	203	877

  - n number of catalogues
  - n=5 catalogues $\rightsquigarrow$ 52 probabilities to be computed

- $\Rightarrow$ Combinatorial explosion!



Credits: wikipedia

Making hypotheses

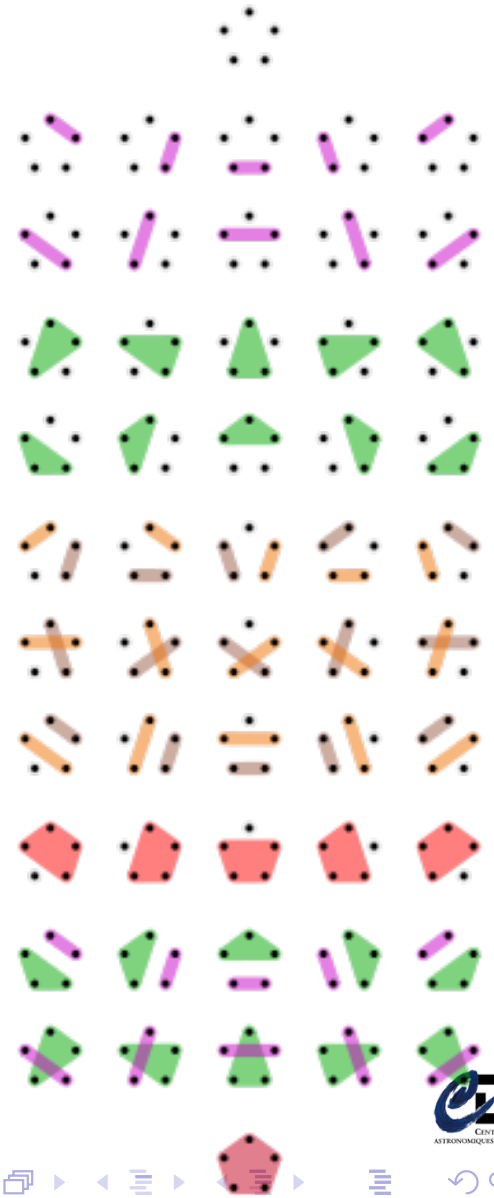
For n-catalogues

- We generalised for n catalogues
- The number of hypothesis to be tested is given by the BELL number

Table : Values of the seven first BELL numbers

n	2	3	4	5	6	7
B_n	2	5	15	52	203	877

  - n number of catalogues
  - n=5 catalogues $\rightsquigarrow$ 52 probabilities to be computed
- $\Rightarrow$ Combinatorial explosion!

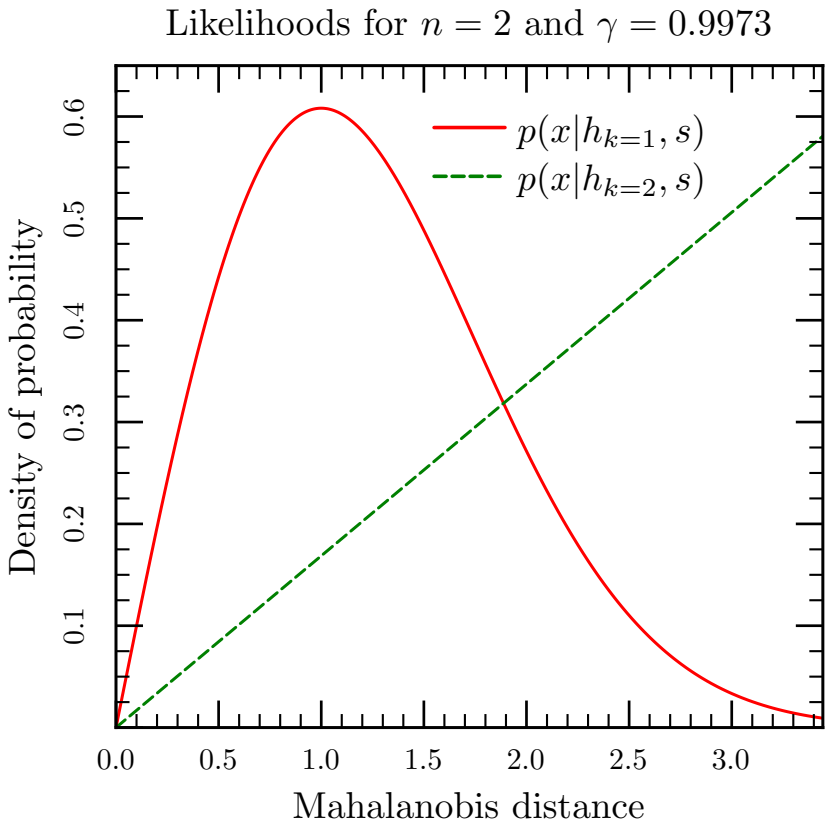


Credits: wikipedia

Likelihoods

For 2 and 3 catalogues

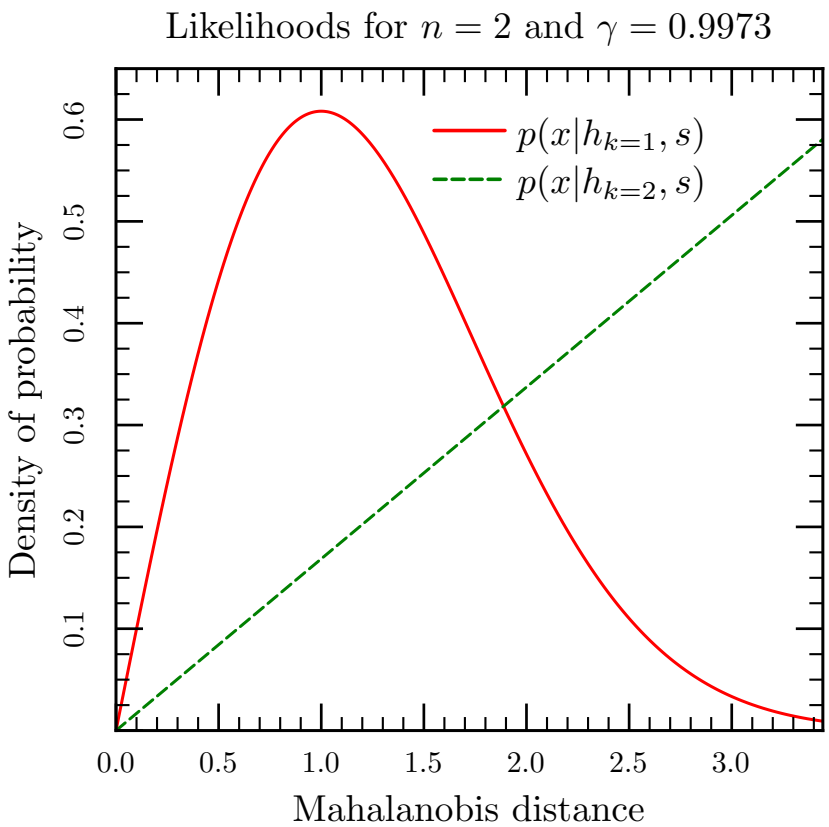
- Likelihood depends only on the nyumber of “actual” source

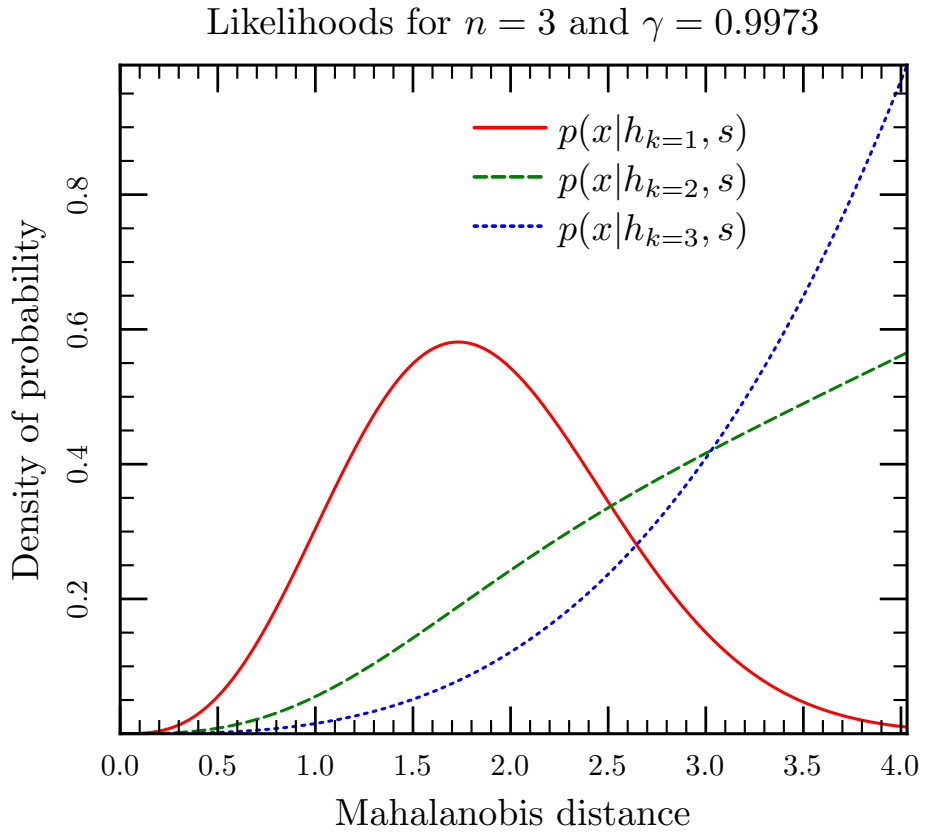


Likelihoods

For 2 and 3 catalogues

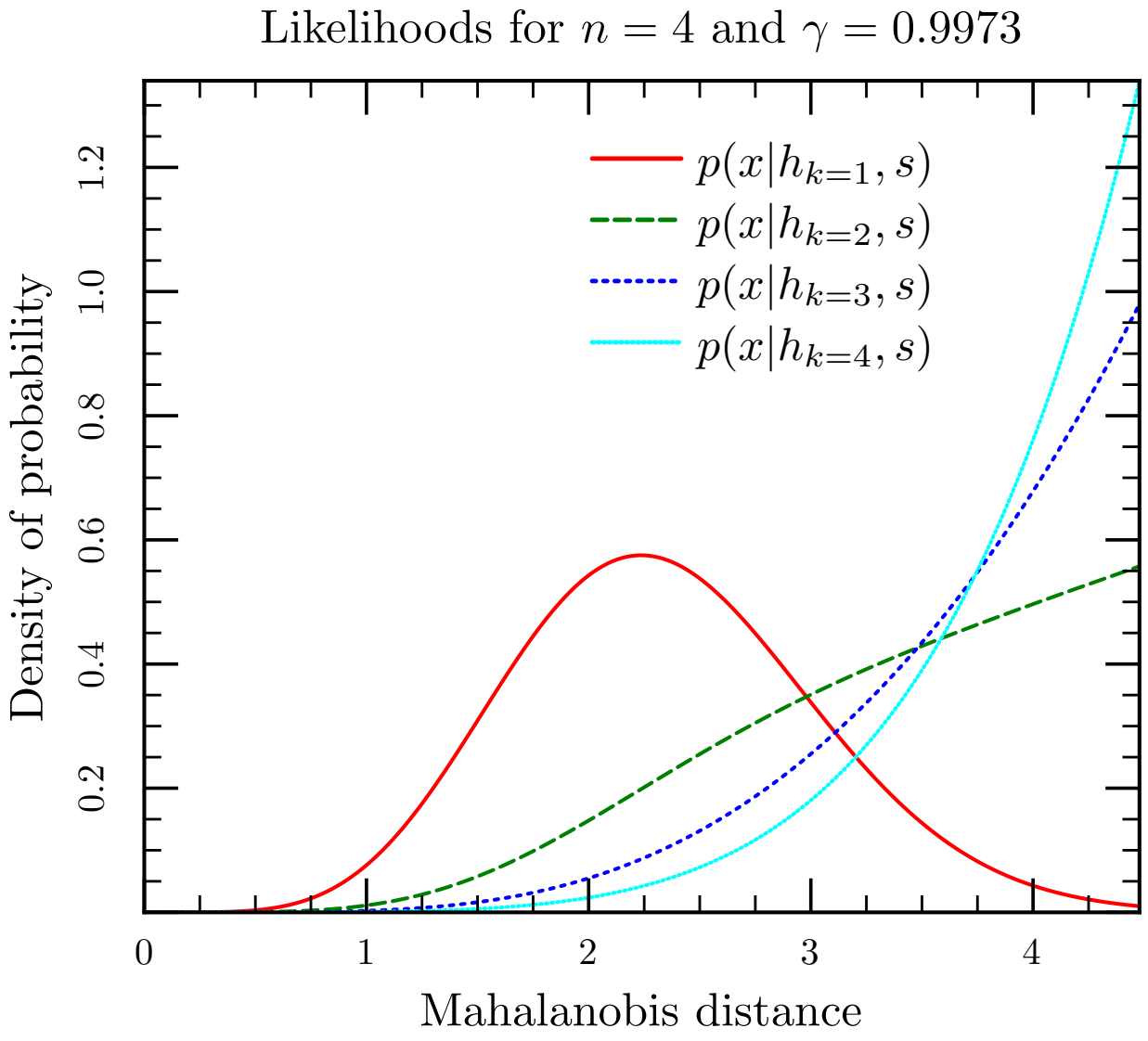
- Likelihood depends only on the nyumber of “actual” source





Likelihoods

For 4 catalogues

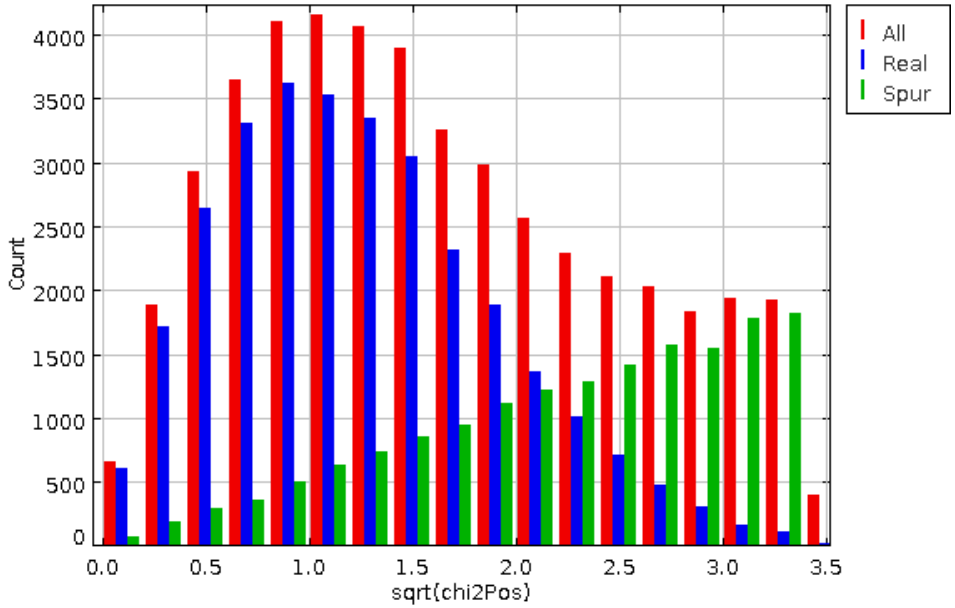


Bayesian probabilities

For 2 catalogues

- Blue curve: $n_{H_0} \times p(x|H_0)$
- Green curve: $n_{H_1} \times p(x|H_1)$
- Red curve = blue + green
- For an association of given x:

$$p(H_0|x) = \frac{\text{Blue curve}(x)}{\text{Red curve}(x)}$$



- Bayes formula:

$$p(H_0|x) = \frac{p(H_0)p(x|H_0)}{p(H_0)p(x|H_0) + p(H_1)p(x|H_1)}$$

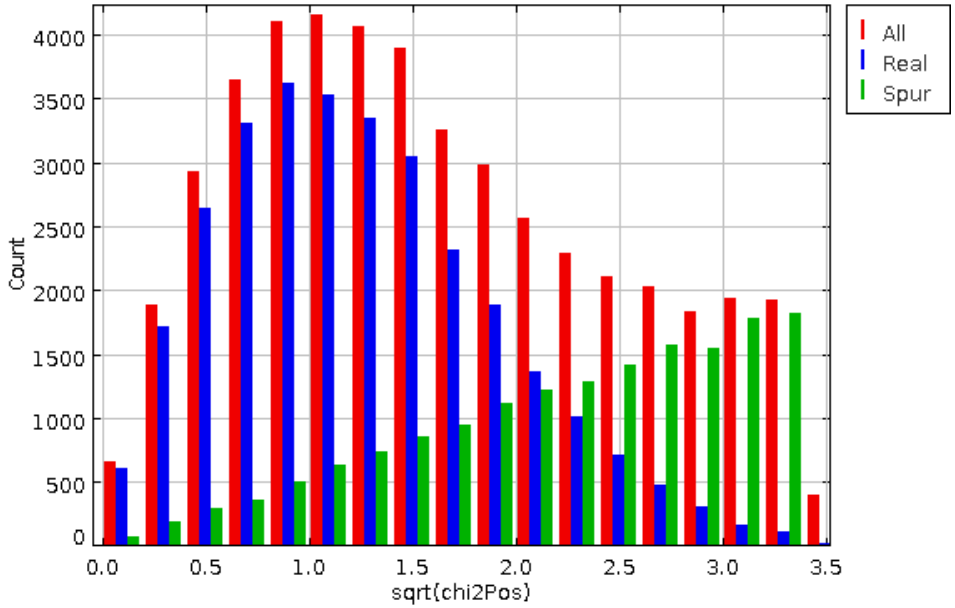
- Here priors $p(H_0) = n_{H_0}/n_{tot}$ and $p(H_1) = n_{H_1}/n_{tot}$

Bayesian probabilities

For 2 catalogues

- Blue curve: $n_{H_0} \times p(x|H_0)$
- Green curve: $n_{H_1} \times p(x|H_1)$
- Red curve = blue + green
- For an association of given x:

$$p(H_0|x) = \frac{\text{Blue curve}(x)}{\text{Red curve}(x)}$$



- Bayes formula:

$$p(H_0|x) = \frac{p(H_0)p(x|H_0)}{p(H_0)p(x|H_0) + p(H_1)p(x|H_1)}$$

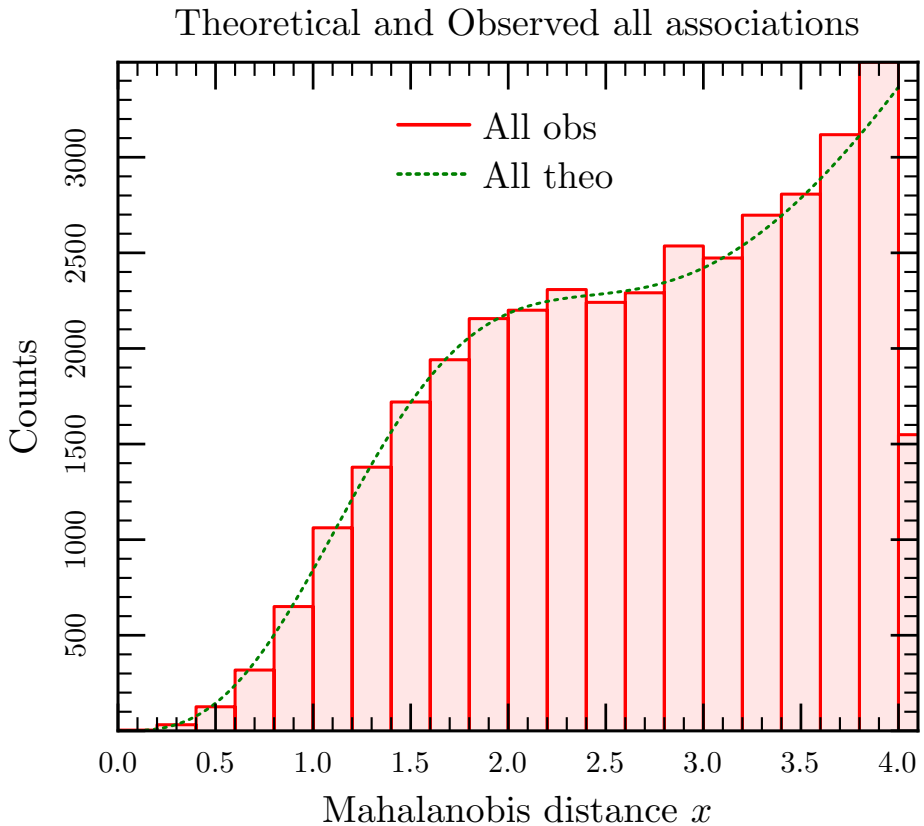
- Here priors $p(H_0) = n_{H_0}/n_{tot}$ and $p(H_1) = n_{H_1}/n_{tot}$

Priors from geometrical considerations

- Common surface area of n catalogues: Ω
- Region of accceptance of the χ^2 test:
  - 2 catalogues case: ellipse
  - n catalogues case: $2(n-1)$-dimensional ellipsoïd
- $\hat{n}_{spur} \propto$ mean volume of $2(n-1)$-dimensional ellipsoïd / Ω^{n-1}
- $\Rightarrow$ a way to define priors

Test on synthetic catalogues

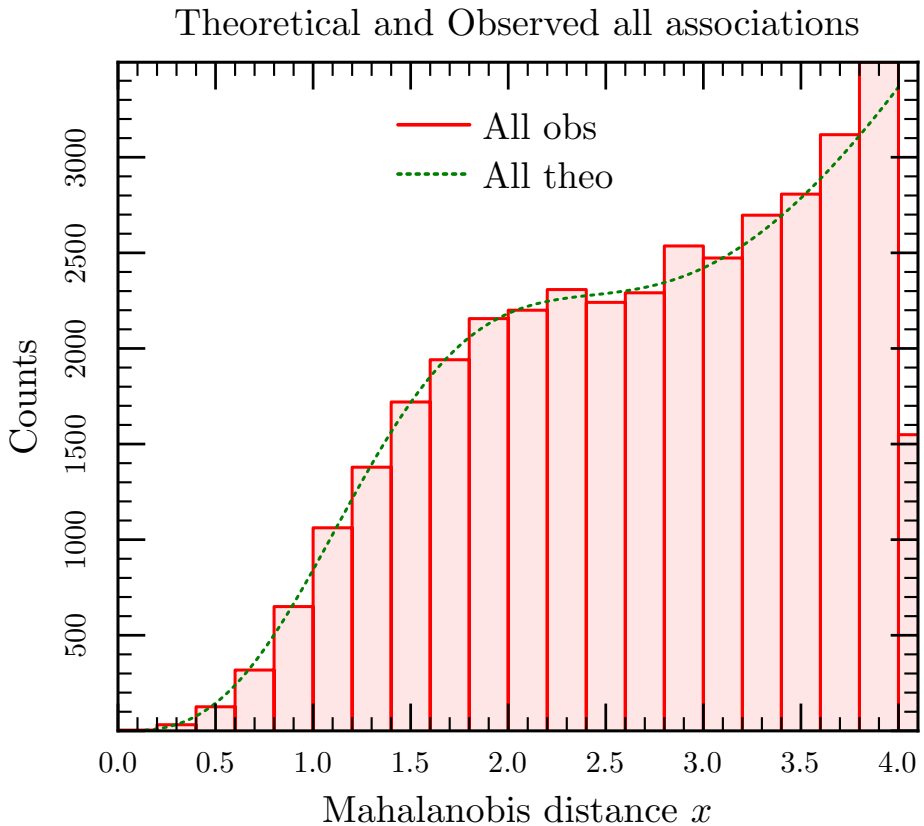
$n_A = 40\,000$ $n_B = 20\,000$ $n_C = 35\,000$
$n_{AB} = 6\,000$ $n_{AC} = 12\,000$ $n_{BC} = 18\,000$
$n_{ABC} = 10\,000$ Cone radius$= 0.42°$
Err A: cte 0.4"; Err B: $\mathcal{N}(0.75", 0.1")$; Err C: linear from 0.8 to 1.2"

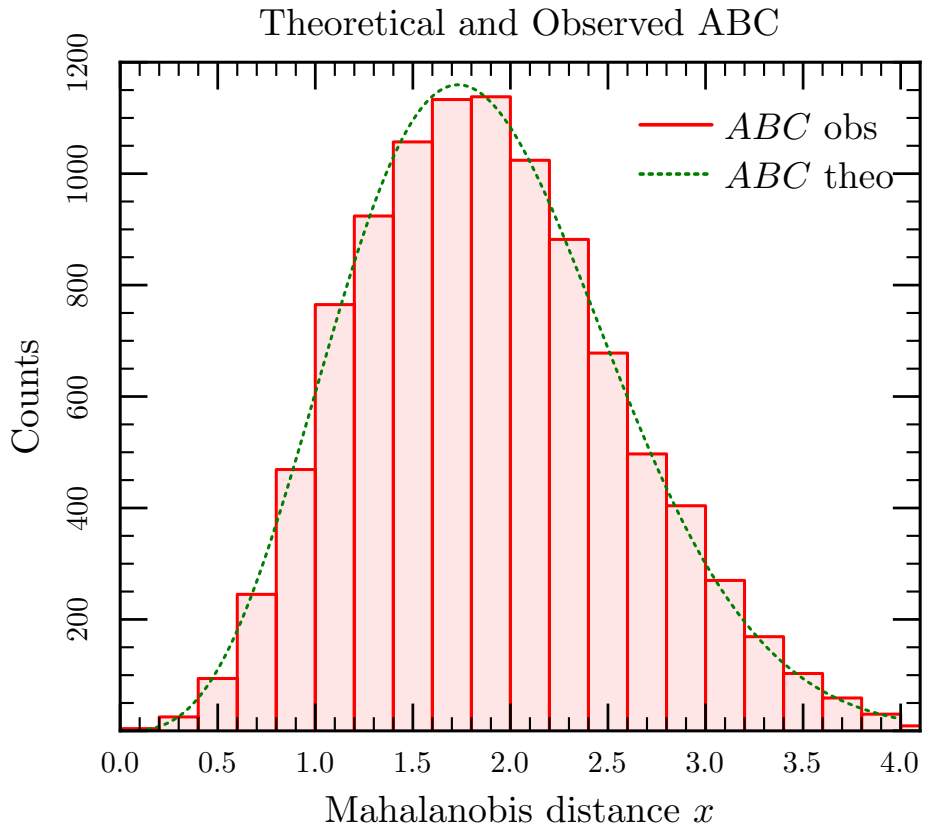


Green curve computed from geometrical considerations only!

Test on synthetic catalogues

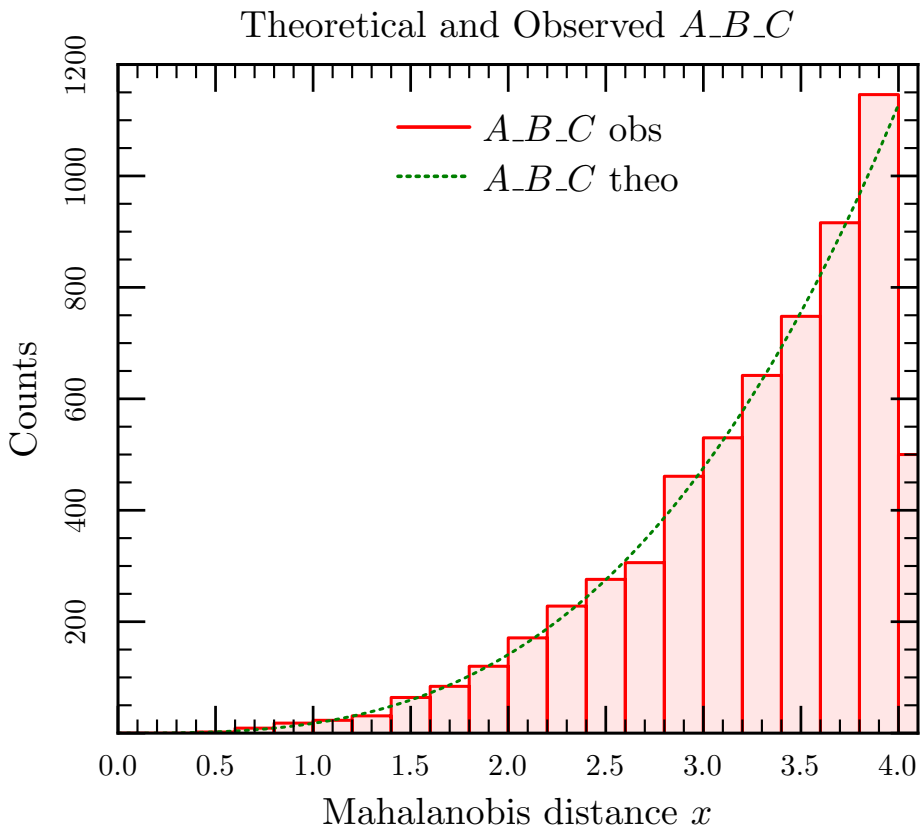
$n_A = 40\,000$ $n_B = 20\,000$ $n_C = 35\,000$
$n_{AB} = 6\,000$ $n_{AC} = 12\,000$ $n_{BC} = 18\,000$
$n_{ABC} = 10\,000$ Cone radius$= 0.42°$
Err A: cte 0.4"; Err B: $\mathcal{N}(0.75", 0.1")$; Err C: linear from 0.8 to 1.2"

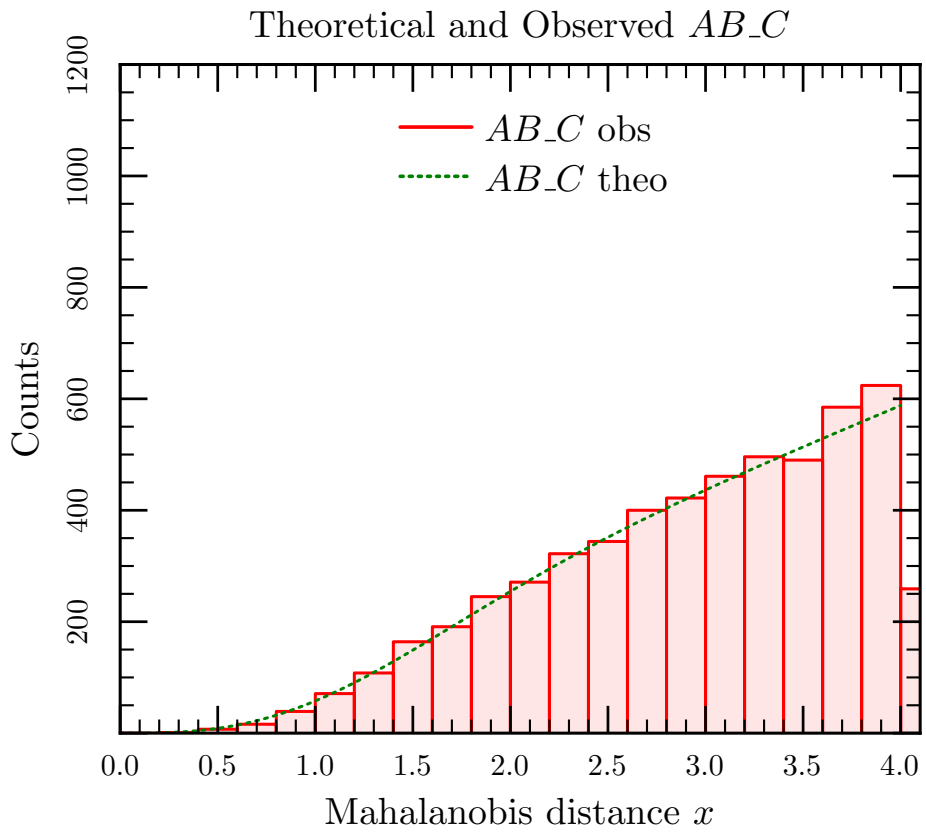




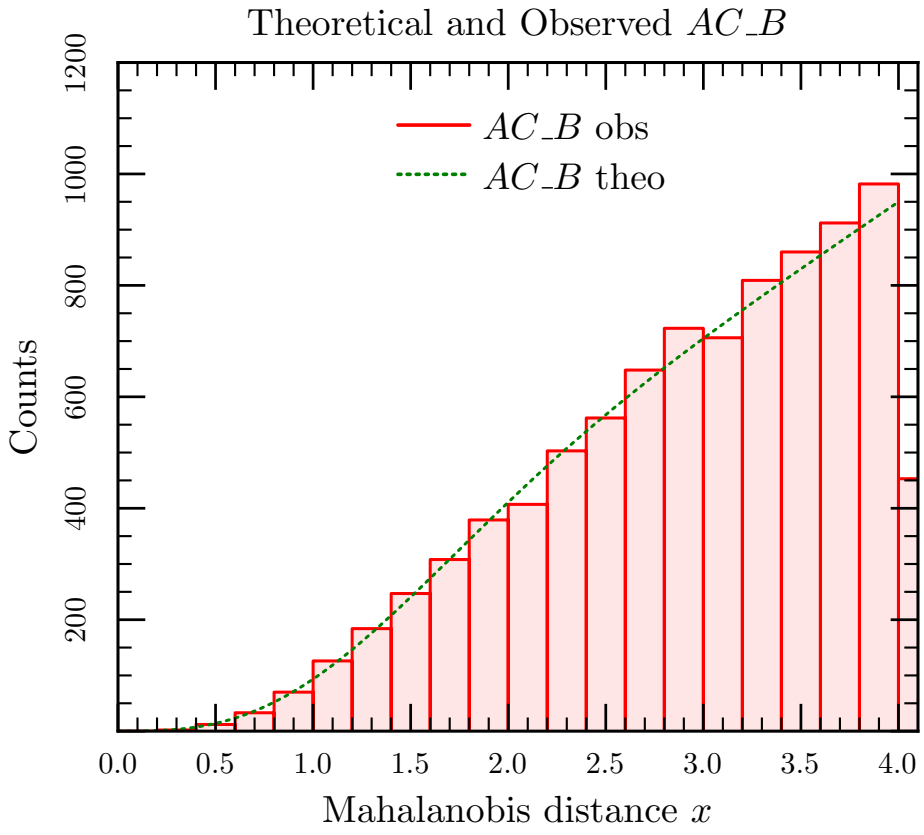
Green curve computed from geometrical considerations only!

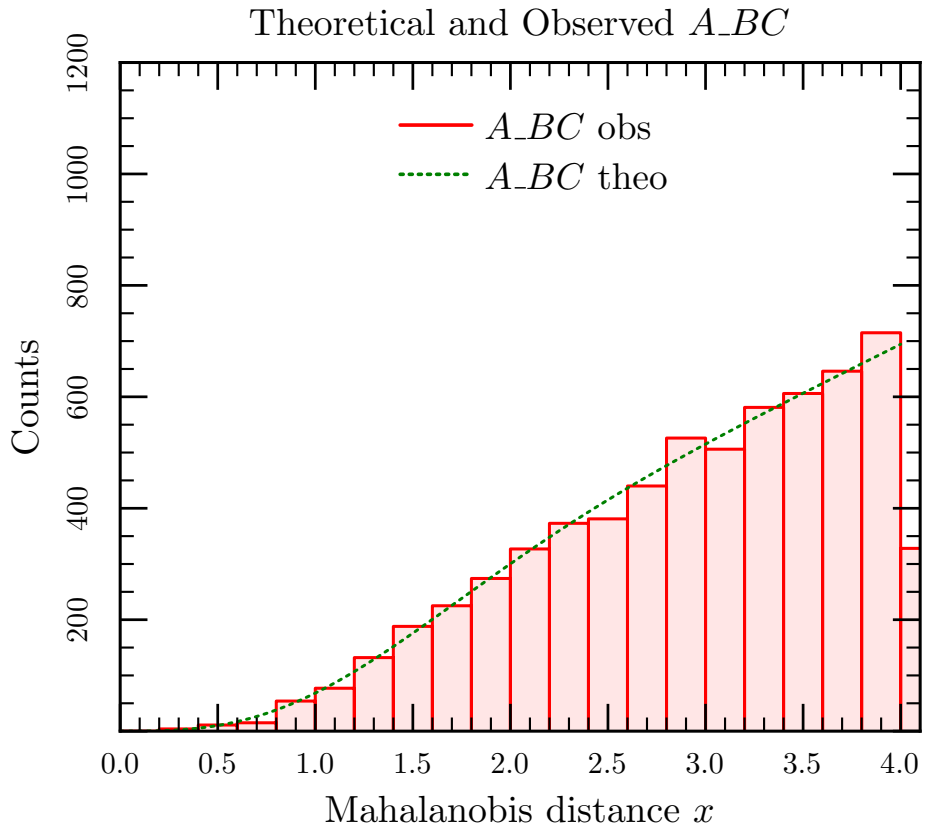
Test on synthetic catalogues





Test on synthetic catalogues





Features: various xmatch algorithms

Algorithm	param	#tbl	prop.mot.	index struct.
chi2 (χ^2)	proba	2	l[1], r[2], b[3]	M/TM-tree

XMatches are chainable: 1 χ^2 xmatch of 4 tables = 3 χ^2 xmatches of 2 tables!

4 to 11 joins ($LIRF\bar{L}\bar{I}\bar{R}L'I'R'F'$) are supported according to the algorithm.

- [1] l: left table contains extended objects or proper motions;
- [2] r: right table contains extended objects or proper motions;
- [3] b: both left and right tables contain extended objects or proper motion.
- [1] mec: Minimum Enclosing Cone

Features: various xmatch algorithms

Algorithm	param	#tbl	prop.mot.	index struct.
chi2 (χ^2)	proba	2	l[1], r[2], b[3]	M/TM-tree
proba2_v*x*	proba	2	no (?)	M-tree
proba3_v*x*	proba	3	no (?)	M-tree
proba4_v*x*	proba	4	no (?)	M-tree
probaN_v*x*	proba	n	no (?)	M-tree

XMatches are chainable: 1 χ^2 xmatch of 4 tables = 3 χ^2 xmatches of 2 tables!

4 to 11 joins ($LIRFL\bar{I}\bar{R}L'I'R'F'$) are supported according to the algorithm.

- [1] l: left table contains extended objects or proper motions;
- [2] r: right table contains extended objects or proper motions;
- [3] b: both left and right tables contain extended objects or proper motion.
- [1] mec: Minimum Enclosing Cone

Features: various xmatch algorithms

Algorithm	param	#tbl	prop.mot.	index struct.
chi2 (χ^2)	proba	2	l[1], r[2], b[3]	M/TM-tree
proba2_v*x*	proba	2	no (?)	M-tree
proba3_v*x*	proba	3	no (?)	M-tree
proba4_v*x*	proba	4	no (?)	M-tree
probaN_v*x*	proba	n	no (?)	M-tree
knn	k+dist	2	r, b	kd/M/TM-tree
cone	dist	2	l, r, b	kd/M/TM-tree

XMatches are chainable: 1 χ^2 xmatch of 4 tables = 3 χ^2 xmatches of 2 tables!

4 to 11 joins ($LIRF\bar{L}\bar{I}\bar{R}L'I'R'F'$) are supported according to the algorithm.

- [1] l: left table contains extended objects or proper motions;
- [2] r: right table contains extended objects or proper motions;
- [3] b: both left and right tables contain extended objects or proper motion.
- [1] mec: Minimum Enclosing Cone

Features: various xmatch algorithms

Algorithm	param	#tbl	prop.mot.	index struct.
chi2 (χ^2)	proba	2	l[1], r[2], b[3]	M/TM-tree
proba2_v*x*	proba	2	no (?)	M-tree
proba3_v*x*	proba	3	no (?)	M-tree
proba4_v*x*	proba	4	no (?)	M-tree
probaN_v*x*	proba	n	no (?)	M-tree
knn	k+dist	2	r, b	kd/M/TM-tree
cone	dist	2	l, r, b	kd/M/TM-tree
mec[1]	dist	n	no (?)	kd/M-tree

XMatches are chainable: 1 χ^2 xmatch of 4 tables = 3 χ^2 xmatches of 2 tables!

4 to 11 joins ($LIRF\bar{L}\bar{I}\bar{R}L'I'R'F'$) are supported according to the algorithm.

- [1] l: left table contains extended objects or proper motions;
- [2] r: right table contains extended objects or proper motions;
- [3] b: both left and right tables contain extended objects or proper motion.
- [1] mec: Minimum Enclosing Cone

Features: various xmatch algorithms

Algorithm	param	#tbl	prop.mot.	index struct.
chi2 (χ^2)	proba	2	l[1], r[2], b[3]	M/TM-tree
proba2_v*x*	proba	2	no (?)	M-tree
proba3_v*x*	proba	3	no (?)	M-tree
proba4_v*x*	proba	4	no (?)	M-tree
probaN_v*x*	proba	n	no (?)	M-tree
knn	k+dist	2	r, b	kd/M/TM-tree
cone	dist	2	l, r, b	kd/M/TM-tree
mec[1]	dist	n	no (?)	kd/M-tree
ext_l[1]	r	2	no	M-tree
ext_r[2]	r	2	no	M-tree
ext_b[3]	r	2	no	M-tree

XMatches are chainable: 1 χ^2 xmatch of 4 tables = 3 χ^2 xmatches of 2 tables!

4 to 11 joins ($LIRF\bar{L}\bar{I}\bar{R}L'I'R'F'$) are supported according to the algorithm.

- [1] l: left table contains extended objects or proper motions;
- [2] r: right table contains extended objects or proper motions;
- [3] b: both left and right tables contain extended objects or proper motion.
- [1] mec: Minimum Enclosing Cone

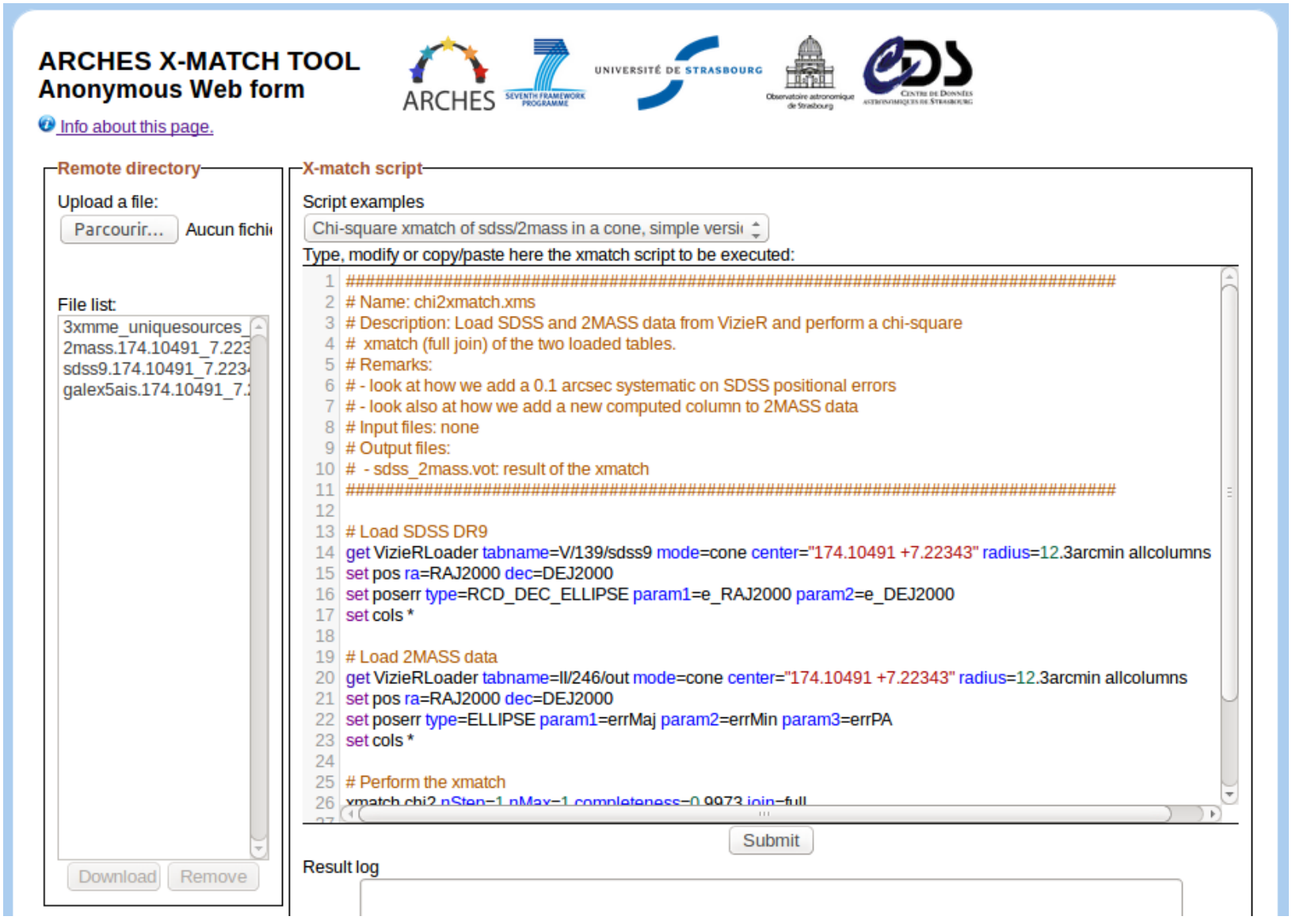
Features: various xmatch algorithms

Algorithm	param	#tbl	prop.mot.	index struct.
chi2 (χ^2)	proba	2	l[1], r[2], b[3]	M/TM-tree
proba2_v*x*	proba	2	no (?)	M-tree
proba3_v*x*	proba	3	no (?)	M-tree
proba4_v*x*	proba	4	no (?)	M-tree
probaN_v*x*	proba	n	no (?)	M-tree
knn	k+dist	2	r, b	kd/M/TM-tree
cone	dist	2	l, r, b	kd/M/TM-tree
mec[1]	dist	n	no (?)	kd/M-tree
ext_l[1]	r	2	no	M-tree
ext_r[2]	r	2	no	M-tree
ext_b[3]	r	2	no	M-tree
...	...	...	...	...

XMatches are chainable: 1 χ^2 xmatch of 4 tables = 3 χ^2 xmatches of 2 tables!
4 to 11 joins ($LIRF\bar{L}\bar{I}\bar{R}L'I'R'F'$) are supported according to the algorithm.

- [1] l: left table contains extended objects or proper motions;
- [2] r: right table contains extended objects or proper motions;
- [3] b: both left and right tables contain extended objects or proper motion.
- [1] mec: Minimum Enclosing Cone

ARCHES X-match tool Web interface



http://serendib.unistra.fr/ARCHESWebService/index.html

Final words

- We plan to integrate the ARCHES engine in the CDS XMatch Service
- Multi-catalogue χ-matches based on INNER JOINS can be performed iteratively
- No more true for e.g. LEFT JOINS (result depends on x-match order)
- Multi-catalogue x-match feasible in ADQL?
- A (evolving) standard describing how to compute probabilities (possibly several ways)??