# Cross-Matching with the Chandra Source Catalog

*Arnold Rots* *2025-02-27*

## Table of Contents

## 1. Scope

This note provides the recipe for calculating cross-match probabilities for CSC sources, using the algorithm from Budavári & Szalay (2008) and Heinis, Budavári, & Szalay (2009) that was also used for the CSC1-SDSS cross-match, which, in turn, was used for the absolute astrometric error determination for Release 1 by Rots & Budavári (2011). See also Budavári & Loredo (2015). We will discuss multi (i.e., more than two) catalog matching, but will focus on matching catalog pairs. We will also provide provisions for situations where the spatial resolutions are significantly different and assign classifications using a set of six match classes.

## 2. Context

The basic initial assumption is that we have two sets of sources, $\mathcal{L}$ and $\mathcal{M}$, containing $N_L$ sources from $\mathcal{L}$ that are contained in the intersection with the coverage of $\mathcal{M}$ and $N_M$ sources from $\mathcal{M}$ that are contained in the intersection with the coverage of $\mathcal{L}$. In practice, we shall assume that, although the coverages may be different, they are subject to the requirement that one is wholly contained in the other.

Each source $i$ is associated with an elliptical position error, characterized by semi-major axis $a_i$ (in radians), semi-minor axis $b_i$ (in radians), and position angle $\phi_i$. We assume that the error distribution is Gaussian. Each source is also characterized by an elliptical Gaussian raw size; for point sources, this raw size is the PSF. The Gaussian sigmas are obtained by multiplying 95% confidence errors by 0.4085, or by multiplying 90% confidence errors by 0.4660. 1.7σ corresponds to approximately 90% confidence.

There are three distinct steps involved in determining match probabilities: calculate a Bayes Factor for each match candidate tuple; calculate likelihoods for each tuple, on the basis of the Bayes Factors in the set; determine the likelihood threshold for accepting matches and assign a match probability and/or classification.

The derivation of CSC raw sizes and PSFs is described in Section 7.

In the following we do, in principle, not impose any restrictions on the number of catalogs involved in the match, but the current implementation is limited to two catalogs. In this memo bold face lower case letters with a single overbar denote vectors, bold face upper case letters with a double overbar denote matrices. [I could not figure out how to do the underlining correctly in MS Word equations]

# 3. Bayes Factors

## 3.1. Reference

Xmatch's algorithm for calculating Bayes Factors is based on Eqn (16) in Budavári and Szalay (2008):

$$B_{ij} = \frac{2}{\sigma_i^2 + \sigma_j^2} \cdot \exp\left(-\frac{\psi_{ij}^2}{2\left(\sigma_i^2 + \sigma_j^2\right)}\right)$$

(1)

This the Bayes factor for sources *i* and *j*, having circular errors $\sigma_i$ and $\sigma_j$, respectively, and separated by $\psi_{ij}$. This is the case for matching two catalogs with sources that have circular errors.

## 3.2. Elliptical Errors and more than Two Catalogs

To generalize this for *n* catalogs with sources that have elliptical errors, we turn to inverse covariance matrices.

The covariance matrix $\bar{\bar{\boldsymbol{C}}}_\iota$ representing the error ellipse $(a_i, b_i, \phi_i = 0)$ for source $i$, as defined in Section 2, is:

$$\bar{\bar{\boldsymbol{C}}}_\iota = \begin{bmatrix} a_i^2 & 0 \\ 0 & b_i^2 \end{bmatrix}$$

(2)

For reference, the inverse of the symmetric $2 \times 2$ matrix $\begin{bmatrix} p & r \\ r & q \end{bmatrix}$ is: $\frac{1}{pq-r^2}\begin{bmatrix} q & -r \\ -r & p \end{bmatrix}$.

For a source *i* with elliptical error semi-major axis $a_i$, semi-minor axis $b_i$, and semi-major axis position angle with respect to north $\varphi_i$ the inverse covariance matrix is:

$$\bar{\bar{\boldsymbol{C}}}_\iota^{\,-1} = \frac{1}{a_i^2 b_i^2}\begin{bmatrix} a_i^2 \sin^2\phi_i + b_i^2 \cos^2\phi_i & \left(b_i^2 - a_i^2\right)\sin\phi_i \cos\phi_i \\ \left(b_i^2 - a_i^2\right)\sin\phi_i \cos\phi_i & a_i^2 \cos^2\phi_i + b_i^2 \sin^2\phi_i \end{bmatrix}$$

(3)

Its determinant is $\frac{1}{a_i^2 b_i^2}$.

We define for an *n*-tuple of sources the matrix $\bar{\bar{\boldsymbol{K}}}$ (effectively the covariance matrix of the combined tuple) through its inverse, as the sum of the inverse covariance matrices:

$$\bar{\bar{\boldsymbol{K}}}^{-1} = \sum_{i=1}^{n} \bar{\bar{\boldsymbol{C}}}_\iota^{\,-1}$$

(4)

First we determine the tangent point to be used for the *n*-tuple:

$$\alpha_0 = \frac{\sum_{i=1}^{n} \left| \overline{\overline{\boldsymbol{C}}}_{\boldsymbol{\imath}}^{-1} \right| \alpha_i}{\sum_{i=1}^{n} \left| \overline{\overline{\boldsymbol{C}}}_{\boldsymbol{\imath}}^{-1} \right|}$$

$$\delta_0 = \frac{\sum_{i=1}^{n} \left| \overline{\overline{\boldsymbol{C}}}_{\boldsymbol{\imath}}^{-1} \right| \delta_i}{\sum_{i=1}^{n} \left| \overline{\overline{\boldsymbol{C}}}_{\boldsymbol{\imath}}^{-1} \right|}$$

(5)

This is equivalent to taking a mean position by weighing the individual positions by the inverse of the area of their error ellipses. From this point forward we will define the source position vectors $\overline{\boldsymbol{x}_{\boldsymbol{\imath}}} = \begin{bmatrix} x_{i,1} \\ x_{i,2} \end{bmatrix}$ in the tangent plane with respect to the tangent point $(\alpha_0, \delta_0)$.

$$x_{i,1} = \frac{-\cos\delta_i \sin(\alpha_i - \alpha_0)}{\sin\delta_i \sin\delta_0 + \cos\delta_i \cos\delta_0 \cos(\alpha_i - \alpha_0)}$$

$$x_{i,2} = \frac{\sin\delta_i \cos\delta_0 - \cos\delta_i \sin\delta_0 \cos(\alpha_i - \alpha_0)}{\sin\delta_i \sin\delta_0 + \cos\delta_i \cos\delta_0 \cos(\alpha_i - \alpha_0)}$$

(6)

The vectors $\overline{\boldsymbol{y}}$ (the mean position of the tuple) and $\overline{\boldsymbol{u}}$ become:

$$\overline{\overline{\boldsymbol{K}}}^{-1} \cdot \overline{\boldsymbol{y}} = \sum_{i=1}^{n} \overline{\overline{\boldsymbol{C}}}_{\boldsymbol{\imath}}^{-1} \cdot \overline{\boldsymbol{x}}_{\boldsymbol{\imath}} = \overline{\boldsymbol{u}}$$

(7)

If the $k'_{ij}$ are the elements of matrix $\overline{\overline{\boldsymbol{K}}}^{-1}$, then:

$$\overline{\boldsymbol{y}}^T \cdot \overline{\overline{\boldsymbol{K}}}^{-1} \cdot \overline{\boldsymbol{y}} = \frac{k'_{22}u_1^2 - 2k'_{12}u_1u_2 + k'_{11}u_2^2}{k'_{11}k'_{22} - {k'_{12}}^2} = \overline{\boldsymbol{u}}^T \cdot \overline{\overline{\boldsymbol{K}}} \cdot \overline{\boldsymbol{u}}$$

(8)

The Bayes Factor for the tuple becomes:

$$B = 2^{n-1} \frac{\sqrt{|\overline{\overline{\boldsymbol{K}}}|}}{\prod_{i=1}^{n} \sqrt{|\overline{\overline{\boldsymbol{C}}}_\iota|}} \exp\left\{\frac{1}{2}\left(\overline{\boldsymbol{y}}^T \cdot \overline{\overline{\boldsymbol{K}}}^{-1} \cdot \overline{\boldsymbol{y}} - \sum_{i=1}^{n} \overline{\boldsymbol{x}_\iota}^T \cdot \overline{\overline{\boldsymbol{C}}}_\iota^{-1} \cdot \overline{\boldsymbol{x}_\iota}\right)\right\}$$

$$= 2^{n-1} \frac{\sqrt{|\overline{\overline{\boldsymbol{K}}}|}}{\prod_{i=1}^{2} \sqrt{|\overline{\overline{\boldsymbol{C}}}_\iota|}} \exp\left\{\frac{1}{2}\left(\overline{\boldsymbol{u}}^T \cdot \overline{\overline{\boldsymbol{K}}} \cdot \overline{\boldsymbol{u}} - \sum_{i=1}^{n} \overline{\boldsymbol{x}_\iota}^T \cdot \overline{\overline{\boldsymbol{C}}}_\iota^{-1} \cdot \overline{\boldsymbol{x}_\iota}\right)\right\}$$

( 9 )

Where:

$$\overline{\boldsymbol{u}} = \sum_{i=1}^{n} \overline{\overline{\boldsymbol{C}}}_\iota^{-1} \cdot \overline{\boldsymbol{x}}_\iota$$

( 10 )

The effective σ of the tuple becomes:

$$\sigma^2 = \frac{\sqrt{\left|\overline{\overline{\boldsymbol{K}^{-1}}}\right|}}{\sqrt{\prod_{i=1}^{n} \left|\overline{\overline{\boldsymbol{C}_1}^{-1}}\right|}}$$

( 11 )

## 3.3.Covariance Matrices for Two Catalogs with Elliptical Errors

For two sources, *i* and *j*, respectively in catalogs 1 and 2, the effective inverse covariance matrix of the match is:

$$\overline{\overline{\boldsymbol{K}}}^{-1} = \overline{\overline{\boldsymbol{C}}}_\iota^{-1} + \overline{\overline{\boldsymbol{C}}}_J^{-1}$$

( 12 )

The tangent point becomes:

$$\alpha_0 = \frac{\left|\overline{\overline{\boldsymbol{C}}}_\iota^{-1}\right| \alpha_i + \left|\overline{\overline{\boldsymbol{C}}}_J^{-1}\right| \alpha_j}{\left|\overline{\overline{\boldsymbol{C}}}_\iota^{-1}\right| + \left|\overline{\overline{\boldsymbol{C}_2}}^{-1}\right|}$$

$$\delta_0 = \frac{\left|\overline{\overline{\boldsymbol{C}}}_\iota^{-1}\right| \delta_i + \left|\overline{\overline{\boldsymbol{C}}}_J^{-1}\right| \delta_j}{\left|\overline{\overline{\boldsymbol{C}}}_\iota^{-1}\right| + \left|\overline{\overline{\boldsymbol{C}}}_J^{-1}\right|}$$

( 13 )

The Bayes Factor becomes:

$$B_{12} = 2 \cdot \frac{\sqrt{|\bar{\bar{\boldsymbol{K}}}|}}{\sqrt{|\bar{\bar{\boldsymbol{C}}}_\iota| \cdot |\bar{\bar{\boldsymbol{C}}}_J|}} \exp\left\{\frac{1}{2}\left(\bar{\boldsymbol{y}}^T \cdot \bar{\bar{\boldsymbol{K}}}^{-1} \cdot \bar{\boldsymbol{y}} - (\bar{\boldsymbol{x}}_\iota{}^T \cdot \bar{\bar{\boldsymbol{C}}}_\iota{}^{-1} \cdot \overline{\boldsymbol{x}_\iota} + \bar{\boldsymbol{x}}_J{}^T \cdot \bar{\bar{\boldsymbol{C}}}_J{}^{-1} \cdot \overline{\boldsymbol{x}_J})\right)\right\}$$

$$= 2 \cdot \frac{\sqrt{\left|\overline{\overline{\boldsymbol{C}_\iota{}^{-1}}}\right| \cdot \left|\overline{\overline{\boldsymbol{C}_J{}^{-1}}}\right|}}{\sqrt{\left|\overline{\overline{\boldsymbol{K}^{-1}}}\right|}} \exp\left\{\frac{1}{2}\left(\bar{\boldsymbol{u}}^T \cdot \overline{\overline{\boldsymbol{K}}} \cdot \bar{\boldsymbol{u}} - (\bar{\boldsymbol{x}}_\iota{}^T \cdot \bar{\bar{\boldsymbol{C}}}_\iota{}^{-1} \cdot \overline{\boldsymbol{x}_\iota} + \bar{\boldsymbol{x}}_J{}^T \cdot \bar{\bar{\boldsymbol{C}}}_J{}^{-1} \cdot \overline{\boldsymbol{x}_J})\right)\right\}$$

( 14 )

Where:

$$\bar{\boldsymbol{u}} = \bar{\bar{\boldsymbol{C}}}_\iota{}^{-1} \cdot \bar{\boldsymbol{x}}_\iota + \bar{\bar{\boldsymbol{C}}}_J{}^{-1} \cdot \bar{\boldsymbol{x}}_J$$

( 15 )

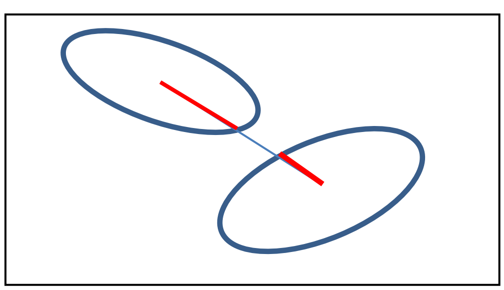
and the effective sigma of the match:

$$\sigma^2 = \frac{\sqrt{\left|\overline{\overline{\boldsymbol{K}^{-1}}}\right|}}{\sqrt{\left|\overline{\overline{\boldsymbol{C}_\iota{}^{-1}}}\right| \cdot \left|\overline{\overline{\boldsymbol{C}_J{}^{-1}}}\right|}}$$

( 16 )

### 3.4. Alternative Calculation of Sigmas for Two Catalogs with Elliptical Errors

When dealing with two catalogs with elliptical errors it is essential that the directionality of the ellipses and the connecting vector be retained. This information is lost in the calculation of σ in Eqn ( 16 ). Keeping that in mind there is an alternative option. Returning to Eqn ( 1 ), we can use for the σ values the radii of the error ellipses where the vector connecting the two sources intersects with those ellipses; these are the red segments in the diagram below.



The length of the vector connecting the two sources is:

$$\psi_{ij} = \arccos\left(\sin\delta_i \cdot \sin\delta_j + \cos\delta_i \cdot \cos\delta_j \cdot \cos(\alpha_i - \alpha_j)\right)$$

( 17 )

The position angle of this vector is:

$$\theta_{ij} \approx \arcsin\left(\frac{\sin(\alpha_i - \alpha_j)}{\sin\psi_{ij}} \cos\left(0.5(\delta_i + \delta_j)\right)\right)$$

( 18 )

Strictly speaking, the individual angles $\theta_i$ and $\theta_j$ are:

$$\theta_{i,j} = \arcsin\left(\frac{\sin(\alpha_i - \alpha_j) \cdot \cos\delta_{i,j}}{\sin\psi_{ij}}\right)$$

( 19 )

But use of Eqn ( 18 ) is perfectly adequate.

The resultant values for σ are:

$$\sigma_i^2(j) = \frac{a_i^2 b_i^2}{a_i^2 \sin^2(\phi_i - \theta_{ij}) + b_i^2 \cos^2(\phi_i - \theta_{ij})})$$

( 20 )

where $a_i$, $a_j$, $b_i$, and $b_j$ are the semi axes and $\phi_i$ and $\phi_j$ the position angles of the ellipses for sources *i* and *j*, respectively. σ is then calculated as:

$$\sigma^2 = \sigma_i^2(j) + \sigma_j^2(i)$$

( 21 )

We shall use this value for calculating the normalized separation of the two sources, $\psi/\sigma$.

### 3.5.Remarks

For practical reasons Bayes Factors are only evaluated for source pairs with a separation less than 10 times the sum of their raw size semi axes. In addition, any pairs with Bayes Factors less than $10^2$ are ignored in the evaluation of likelihoods.

## 4. Cross-Match Probabilities

Next, we convert the Bayes Factors to probabilities.

Let there be $m$ tuples drawn from $n$ catalogs, each with, respectively, $N_1, N_2, \ldots, N_n$ sources, where the numbers $N_i$ are scaled to the entire sky (4π radians).

## 4.1. Initial Prior

We start out with the initial prior $P_0(0)$:

$$P_0(0) = \frac{\min(N_1, N_2, \dots, N_n)}{\prod_{i=1}^{n} N_i}$$

( 22 )

## 4.2. Probabilities and Iteration

Now we iterate on the prior. Calculate for iteration $k$ and each tuple $i$ its posterior match probability $p_i$:

$$p_i(k) = \left(1 + \frac{1 - P_0(k)}{B_i \cdot P_0(k)}\right)^{-1}$$

( 23 )

Update $P_0$ to a new prior:

$$P_0(k+1) = \frac{\sum_{i=1}^{m} p_i(k)}{\prod_{i=1}^{n} N_i}$$

( 24 )

Remember that both $\sum_{i=1}^{m} p_i(k)$ and the $N_i$ need to be scaled to the entire sky.

And iterate until:

$$\frac{P_0(k+1) - P_0(k)}{P_0(k+1)} < \ 10^{-3}$$

( 25 )

It is prudent to limit the maximum number of iterations to something like 20: if all Bayes Factors are very small this will not converge while no harm will be done if the iterations are terminated.

## 4.3.Priors for Matching Two Catalogs

Matching only a pair of catalogs simplifies the equations in the previous sections slightly.

Let catalog 1's area of coverage be $A_1$ and let its sample contain $N_1$ sources; $A_2$ and $N_2$ for catalog 2. Also, let the area of overlap be $A_0$.

Scale the number of sources in both catalogs to the area of overlap:

$$N_i' = \frac{A_0}{A_i} N_i$$

( 26 )

Then the initial prior is:

$$P_0(0) = \frac{\min(N_1', N_2')}{N_1' \cdot N_2'} \cdot \frac{A_0}{S}$$

( 27 )

Where, if $A_0$ is expressed in square arcminutes, $= 4 \cdot (60 \cdot 180)^2/\pi = 466{,}560{,}000/\pi$ .

The posterior match probability for source pair *j* is:

$$p_j(k) = \left(1 + \frac{1 - P_0(k)}{B_j \cdot P_0(k)}\right)^{-1}$$

( 28 )

Iterate on the prior:

$$P_0(k+1) = \frac{\sum_{j=1}^{m} p_j(k)}{N_1' \cdot N_2'} \cdot \frac{A_0}{S}$$

( 29 )

## 4.4. Probability Thresholds

Budavári & Szalay (2008), in Section 5.3, propose a self-consistent mechanism for determining the threshold that match probabilities have to meet in order to be considered accepted. On that basis we have adopted the following criterion for acceptance.

For a given set of $m$ $n$-tuples the iteration on the probabilities for the individual $n$-tuples is derived from the Bayes Factors and a prior that involves the sum of probabilities and the source densities. The issue here is that the source densities introduce a scaling of the probabilities as derived from the BFs (that's why the $P$ versus $\log(BF)$ curves may never be identical) and that for assigning matches we want to apply a uniform thresholding criterion. Here is the recipe:

Assume we have a list of $m$ source $n$-tuples with probability $p[i]$ and a sum $S_P = \sum_{i=1}^{m} p[i]$ .
Note that we count array elements as 1-relative for clarity.

1. If $S_P < 0.2$ reject all matches; else:
2. Sort the list according to decreasing $p[i]$
3. Set $k = S_P$ (truncate)
4. Set the threshold for these $m$ $n$-tuples to $P = s \cdot p[k]$
   For practical reasons it is advisable to require $P$ to be no less than 0.4
5. Accept all $n$-tuples in the list with $p[i] > P$

$s$ is still an input parameter. Tests have shown that good results are obtained with $s = 0.9$.

## 5. Multi-Set Matching

*This section has not been updated, since work on extending the code to more than two catalogs has not been sufficiently developed. It should be ignored for now.*

## 6. Spatial Resolution Issues

Cross-matching has a dark side to it, as Tom Loredo has commented, and this is particularly apparent in situations where the spatial resolution varies widely between the catalogs to be matched – and this is unfortunately applicable *par excellence* to the *Chandra* Source Catalog. Consider two sources separated by a few arcseconds observed on-axis (where they are easily resolved) *versus* 10 or 15 arcminutes off-axis (where there is no chance of resolving them). In such a situation the position errors may all be fairly small and essentially irrelevant: the deciding factor for matches is really the size of the PSF.

The situation is even more complicated when matching catalogs are derived from data obtained in vastly different parts of the electromagnetic spectrum, as there is no guarantee that closely spaced detections actually originate from the same physical object. And, in addition, there is the possibility that detections of a certain source may be contaminated by radiation from another.

For this reason we are also taking the raw (i.e., detected) sizes of sources into consideration. In the case of point sources that raw size is the PSF.

## 7. PSF Parameters

For the CSC, the parameters of the ECF 90% PSF ellipses are available from the Region files. For single-Obi stacks this does not present a problem and one can just take the values and scale the semi axes down by 0.4660. For multi-Obi stacks one needs to calculate a composite PSF where each Obi is assigned a weight proportional to the number of counts it contributes to the total number detected in the source. If there are $c_i$ photons detected in Obi $i$ of $n$ Obis:

$$w_i = \frac{c_i}{\sum_{j=1}^{n} c_j} \tag{30}$$

The quick solution is to derive the PSF parameters from the composite PSF's covariance matrix $\bar{\bar{\boldsymbol{K}}}$, where its inverse is calculated as before:

$$\bar{\bar{\boldsymbol{K}}}^{-1} = \sum_{i=1}^{n} w_i \bar{\bar{\boldsymbol{C}}}_i^{\,-1} \tag{31}$$

Ideally, though, one should construct a composite PSF by taking the weighted average of the PSF images and determine the 1-σ ellipse of this image directly.

## 8. Implementation of Cross Matching Pairs of Catalogs

In this section we will restrict the discussion to the matching of two catalogs. Consequently, Bayes Factors are calculated following Section 3.3, Eqn. ( 14 ).

In the final implementation we decided to run the cross match probabilities twice: once using the Bayes Factors based on the error ellipses (set 1) and a second run using the largest BF for a given tuple based on either the error ellipses or the raw size ellipses (but remembering which was used; set 2). In both sets we use the self-consistency criterion (Section 4.4) and a minimum probability value of 0.4 to decide which matches to accept. Match classes are assigned according to the following decision tree:

1. All matches with $\frac{\psi}{\sigma_1} \geq 3.4$ or $\frac{\psi}{\sigma_2} \geq 3.4$ (i.e., beyond 99.7% confidence) are rejected out of hand, where $\psi$ is the separation of the pair and $\sigma_i$ the composite uncertainty in the sets 1 and 2 as defined in Eqn ( 16 ).
2. All unambiguous matches from set 1 (based on error ellipses) with $\frac{\psi}{\sigma_1} \leq 1.7$ (i.e., within the 76% confidence region) are considered definite.
3. All unambiguous matches from set 1 with $\frac{\psi}{\sigma_1} > 1.7$ are considered likely.
4. Ambiguous matches in set 1: all matches from set 1 are considered. At this point good matches (definite or likely) are identified in cases where additional ambiguous matches are clearly inferior: where the next highest probability $< (p - 0.5)^2$, *p* being the highest probability.
5. For the remaining ambiguous cases in set 1 with a good match (*p* > 0.9) and a somewhat inferior match, we consider it a good match (definite or likely), but potentially contaminated, based on separation or normalized separation.
6. Turning to set 2, if there is a unique match in that set based on error ellipses (but missing in set 1) we accept it as a likely match. This is something that can happen in sparse fields.
7. Finally, remaining unique matches from set 2 based on raw size and with $\frac{\psi}{\sigma_2} < 1.7$ are classified as raw.

### Table 1. Cross Match Codes and Normalized Separation

| Code | Assessment | Restrictions for the matched pair | Restrictions for additional matched pairs |
|---|---|---|---|
| d | Definite | $\frac{\psi}{\sigma_1} \leq 1.7$ | None |
| l | Likely | $1.7 < \frac{\psi}{\sigma_1} < 3.4$ or $\frac{\psi}{\sigma_2} \leq 3.4$ | None |
| c | Definite, but potentially contaminated | $\frac{\psi}{\sigma_1} \leq 1.7$ | $\frac{\psi}{\sigma_1} \leq 3.4$ |
| k | Likely, but potentially contaminated | $1.7 < \frac{\psi}{\sigma_1} < 3.4$ | $\frac{\psi}{\sigma_1} \leq 3.4$ |
| r | Raw size match | $\frac{\psi}{\sigma_2} \leq 1.7$ | None |
| a | Ambiguous | $\frac{\psi}{\sigma_{1,2}} \leq 3.4$ | $\frac{\psi}{\sigma_{1,2}} \leq 3.4$ |

# 9. Very Large Numbers of Catalogs

*This section has not been updated, since work on extending the code for more than two catalogs has not been sufficiently developed. It should be ignored for now.*

# 10. References

Budavári, T., & Loredo, T. J. 2015, Ann. Rev. Stat. Appl. 2015.2, 113

Budavári, T., & Szalay, A. 2008, ApJ 679, 301

Heinis, S., Budavári, T., & Szalay, A. 2009, ApJ 705, 739

Rots, A. H., & Budavári, T. 2011, ApJS 192, 8

# 11. Appendix – Implementation Details

## 11.1. Running xMatch

The name of the executable is CSCxmatchPairsmixF. This may well change.

Xmatch should be run separately for each CSC Field: a field of view region that is defined as the union of the fields of view of a set of spatially connected observations.

There are five options (all are optional):

| | |
|---|---|
| -i | Input file path [stdin] |
| -plim | Scaling factor for limiting probability [0.90] |
| -pplim | Minimum probability for allowing unambiguous match [0.40] |
| -prtall | Print all considered matches to stdout (not recommended for production) |
| -pntsrc | Only use point sources/stars in the cross match |
| -U | Display input options |

The input file contains two lines with the paths to the input data files.

## 11.2. Input Data File Formats

The input data files begin with a header record containing four parameters:

Catalog name
Catalog type
Field name
Area covered by the data file in square arcminutes

The catalog type may have one of four different values (not case-sensitive):

CHANDRA (also handles XMM and eRASS)
SDSS (also handles 2MASS)
WISE

GAIA
PANSTARRS

All other values will default to CHANDRA.

The header line is followed by an unspecified number of one-line records, containing (tab-separated):

Source name (less than 32 characters)
Designation (less than 32 characters) [Omitted for cat types PANNSTARS, SDSS, CHANDRA]
RA (decimal degrees)
Dec (decimal degrees)
Error ellipse semi-major axis (arcsec)
Error ellipse semi-minor axis (arcsec)
Error ellipse position angle of major axis from north, through east (degrees)
Raw size ellipse semi-major axis (arcsec)
Raw size ellipse semi-minor axis (arcsec)
Raw size ellipse position angle of major axis from north, through east (degrees)
Source type (1 character) [Only for types SDSS (S or G) and Chandra (P or X)]

For WISE, SDSS, and GAIA the error axes are RA and Dec, respectively.

For GAIA the error axes are in mas.

For WISE the error axes are FWHM.

For GAIA the first two parameters are switched.

### 11.3. Output File Formats

There are, aside from stdout, four output files:

Unique matches: one, tab-delimited, record per matched pair
Ambiguous matches: one, tab-delimited, record per matched pair based on error ellipses
Ambiguous raw matches: one, tab-delimited, record per matched pair based on raw size ellipses
Sources with ambiguous matches in catalog 0
Sources with ambiguous matches in catalog 1

The uniquely matched pair records, as well as the ambiguous match records contain:

Field ID
Catalog 0: Source ID
Source type
RA
Dec
Error ellipse semi-major axis (arcsec)
Error ellipse semi-minor axis (arcsec)

Error ellipse position angle of major axis from north, through east (degrees)
Raw size ellipse semi-major axis (arcsec)
Raw size ellipse semi-minor axis (arcsec)
Raw size ellipse position angle of major axis from north, through east (degrees)
Source type (1 character)

Catalog 1: Source ID
Source type
RA
Dec
Error ellipse semi-major axis (arcsec)
Error ellipse semi-minor axis (arcsec)
Error ellipse position angle of major axis from north, through east (degrees)
Raw size ellipse semi-major axis (arcsec)
Raw size ellipse semi-minor axis (arcsec)
Raw size ellipse position angle of major axis from north, through east (degrees)
Source type (1 character)

BF type (*e* or *r*)
Match class (*d, l, c, k, r*, or *a*; see Table 1)
Match probability
Distance ψ between the two members of the pair (arcsec)
Normalized separation ψ/σ *(beware, this is based on the error ellipse match σ for match classes d, l, c, and k; but on the raw size match σ for class r)*

Raw size matches are only considered for sources that do not have a validated match based on error ellipses. Ambiguous matches for these cases are recorded in a separate *AmbiguousXmatchListRaw* file. Consequently, ambiguous raw size matches are not recorded for sources involved in validated matches based on error ellipses.

The ambiguous match records in the per-catalog file contain for the ambiguously matched source:

Source ID
Source type

And then for each matched source from the other catalog:

Source ID
Source type
BF type
Set 2 probability
Set 1 probability
Source separation ψ (arcsec)
ψ/σ