The ARCHES cross-correlation tool Hands On session

François-Xavier Pineau[1]

[1]Observatoire Astronomique de Strasbourg, Université de Strasbourg, CNRS

Paris, 1th December, 2015

X-MATCH TOOL TUTORIAL

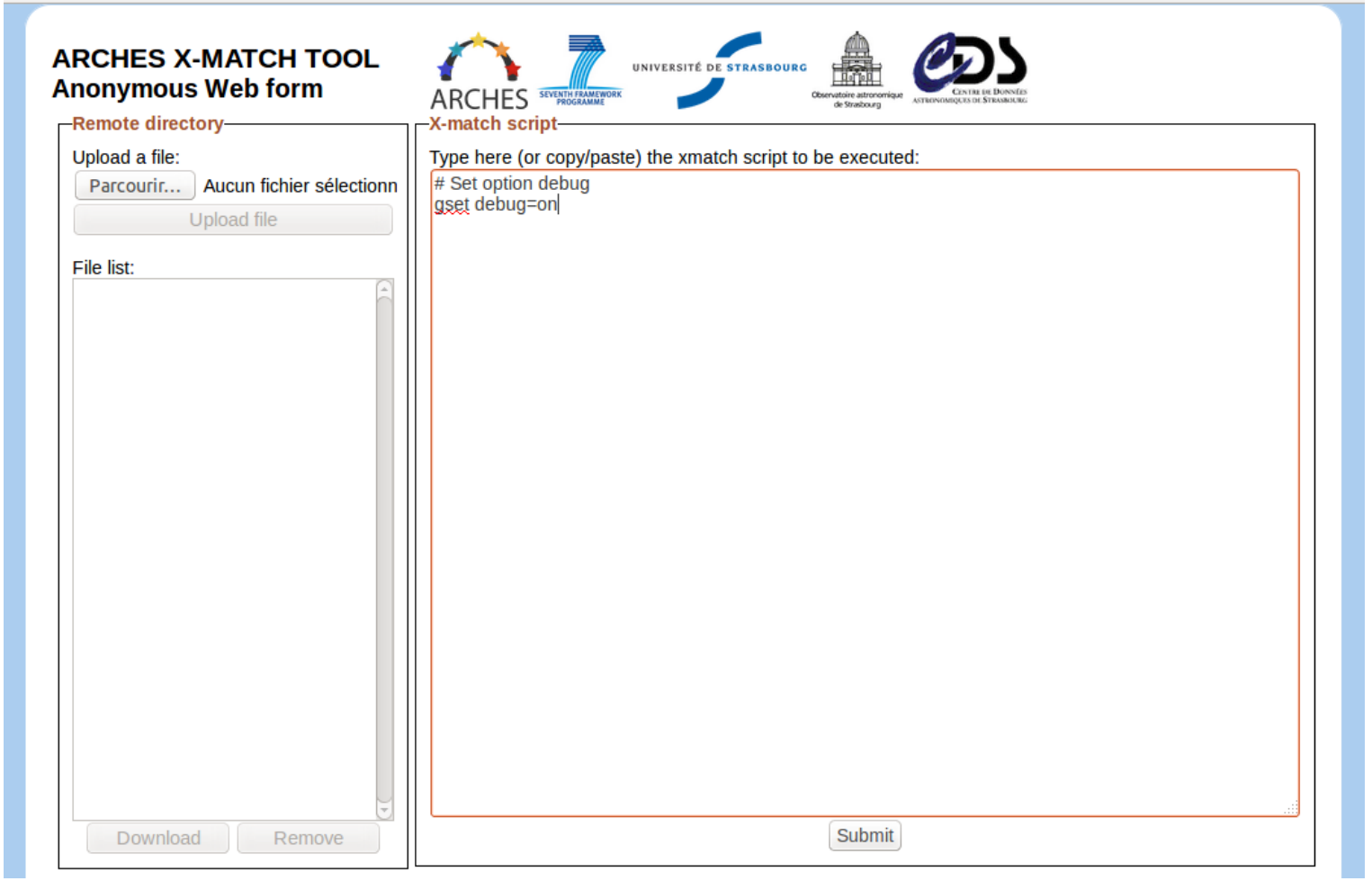
Documentation and setup

- Startup instructions
  - Look at the web page:
    http://serendib.unistra.fr/ARCHESWebService/XMatchARCHES
  - Or:
    - ⋆ Download and run the following *script*:
      http://serendib.unistra.fr/ARCHESWebService/archesxmatch.bash
    - ⋆ Download the example script:
      http://serendib.unistra.fr/ARCHESWebService/example.xms
    - ⋆ Download the documentation:
      http://serendib.unistra.fr/ARCHESWebService/XMatch_soft_doc.pdf
- Informations:
  - Login: anonymous
  - Passwd: anonymous
  - HTTP session last 30 min after last detected activity

X-MATCH TOOL TUTORIAL

Informations

Web interface (in preparation)



X-MATCH TOOL TUTORIAL

Informations

- Service limitations
  - Max 15 jobs at the same time: extra jobs are kicked out;
  - Size of uploaded files limited to 200 MB;
  - 10 min timeout: jobs running for more than 10 minutes are kicked out.
- The tool is deliberately stupid: it does not makes any guess, the user must declare everything
  - e.g. if some positions / positional errors are empty or =0, remove them using the *where* command
- The tool IS NOT ABLE to deal with large All-sky catalogues at once!
- The xmatch region MUST be covered by all catalogues when using *probaXXX* algorithms
- It is the user's responsability to xmatch at once areas having similar properties (sky densities, ...)
- Error messages are not yet user friendly (sorry)

X-MATCH TOOL TUTORIAL

First script submission

- Look at the script *example.xms*: it contains commented commands
  - Look in particular at how data is loaded from VizieR
  - Look at how a systematic is added on SDSS positional errors
- Login using command `./archesxmatch.bash i`
- How many files do you have in your working directory?
- Submit the *example.xms* script
- How many files do you now have in your working directory?
- Download the result file *example.fits* and open it with your favourite tool
- Any questions?

X-MATCH TOOL TUTORIAL

Theoretical Mahalanobis distance histogram

Purpose

- Use the x-match tool to:
  - generate 3 syntetical catalogues;
  - perform χ^2 x-matches of the 3 generated catalogues.
- Check that the number of "real" associations in output is consistent with the input completeness (e.g. $\gamma = 0.9973$)
- Reproduce Fig. 1 using e.g. TOPCAT
- Check the symmetry of the results changing the xmatch order
  - 1 xmatch 2 xmatch 3
  - 1 xmatch 3 xmatch 2
  - ...

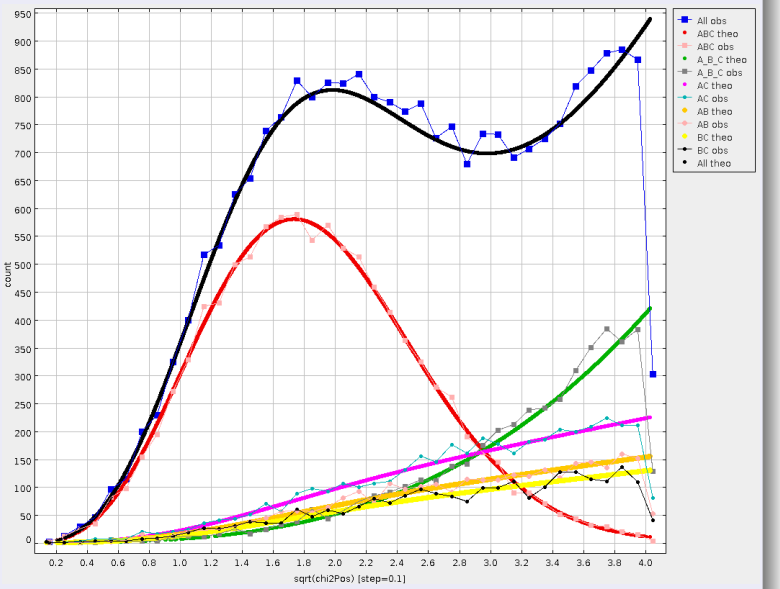


Figure : Mahalanobis distance histogram on simulated data

X-MATCH TOOL TUTORIAL

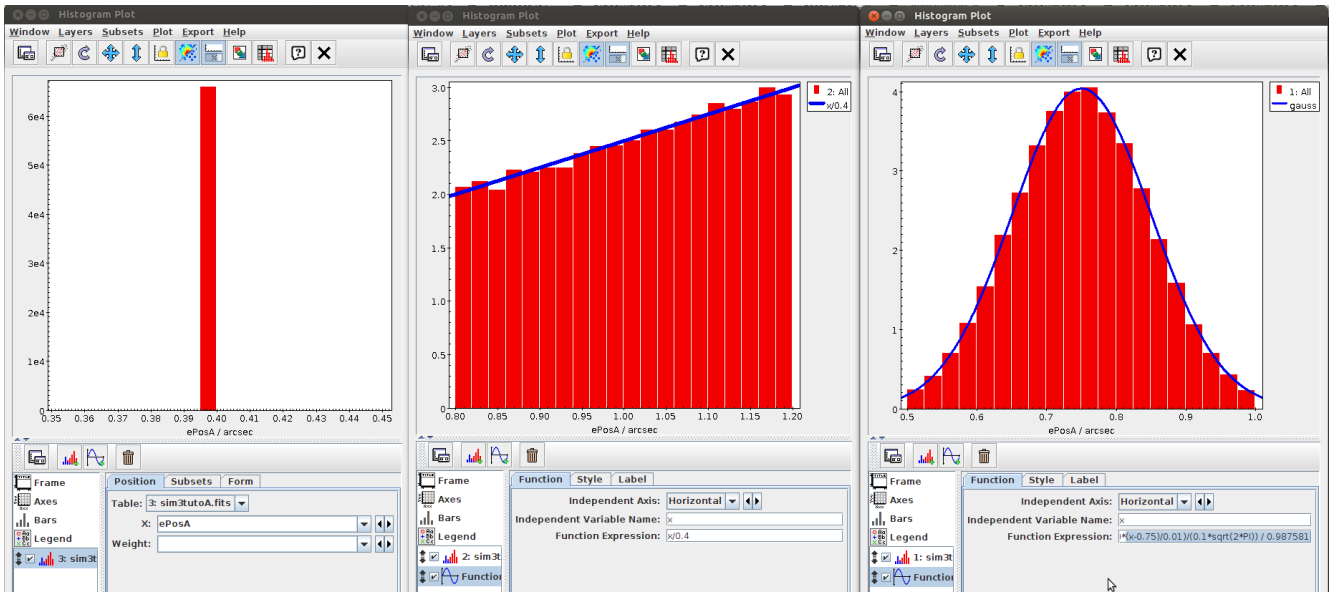
Theoretical Mahalanobis distance histogram

- Step 1: generate 3 syntetical catalogues
  - use command *synthetic*
  - set a cone of $\approx$ 25 arcminutes
  - for catalogue A:
    - ★ set fixed value (e.g. 0.4 arcsec) CIRCULAR positional errors
  - for catalogue B:
    - ★ set CIRCULAR positional errors
    - ★ set error distribution following the function x, $x \in [0.8, 1.2]$
  - for catalogue C:
    - ★ set CIRCULAR positional errors
    - ★ set error distribution following a gaussian function, e.g. $\frac{1}{0.1\sqrt{2\pi}} \exp(-\frac{1}{2}\frac{(x-0.75)^2}{0.1^2})$, $x \in [0.5, 1]$
  - set the number of sources in each possible subset of catalogues, e.g.
    - ★ nA=40 000 nB=20 000 nC=35 000
    - ★ nAB=6 000 nAC=7 000 nBC=8 000 nABC=10 000
  - save the generated files

X-MATCH TOOL TUTORIAL

Theoretical Mahalanobis distance histogram

- Step 2: check the coherence of the generated files, e.g.:
  - #rows in file A = nA + nAB + nAC + nABC; idem for B and C
  - #rows in common file = nA + nB + nC + nAB + nAC + nCB + nABC
  - positional error distributions, e.g.:
    - ★ normalize positional error histograms
    - ★ overplot the error distribution function $f(x)/\int_{x_{min}}^{x_{max}} f(x)\mathrm{d}x$



X-MATCH TOOL TUTORIAL

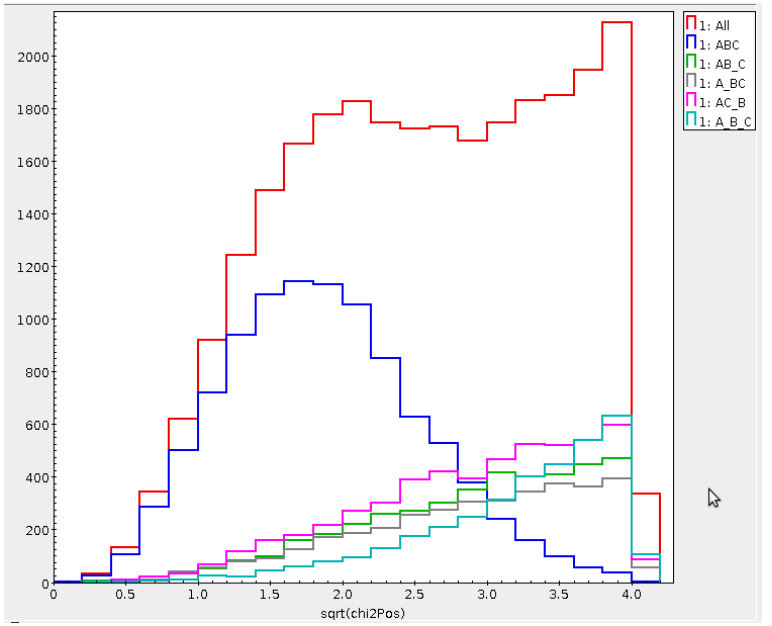
Theoretical Mahalanobis distance histogram

- Step 3: perform a 3 catalogues x-match.
  - Load and set tables using commands:
    - *get*, *set pos*, *set poserr* and *set cols*.
  - Choose a *completeness* $\gamma \in [0, 1]$, e.g. 0.9973
  - Method 1:
    - perform 2 successive χ^2 x-matches
    - you will need to use at least one *merger*
    - the result SHOULD NOT depend on the xmatches order (AxBxC=AxCxB=...)
  - Method 2:
    - peform the x-match at once with e.g. command *xmatch probaN_v1*

X-MATCH TOOL TUTORIAL

Theoretical Mahalanobis distance histogram

- Step 4: check the results
  - Build the 5 components (Views/Rows Subsets in TOPCAT):
    - ★ ABC: *A_id == B_id && B_id == C_id*
    - ★ AB_C: *A_id == B_id && B_id != C_id*
    - ★ A_BC: *A_id != B_id && B_id == C_id*
    - ★ AC_B: *A_id == C_id && B_id != C_id*
    - ★ A_B_C: *A_id != B_id && B_id != C_id && A_id != C_id*
  - Verify #rows = #ABC + #AB_C + #A_BC + #AC_B + #A_B_C
  - Verify the fraction of "real" *ABC* associations recovered $\approx \gamma$
  - For all associations and for the 5 components, plot the histogram of the Mahalanobis distance (or χ-distance, square root of column `chi2Pos`)



X-MATCH TOOL TUTORIAL

Theoretical Mahalanobis distance histogram

- Step 5.1: plot theoretical curves over the Mahalanobis distance histograms
  - ABC histogram: binStep$\times$nABC$\times\chi_{dof=4}(x)$, $\chi_{dof=4}(x) = \frac{x^3}{2}\exp(-\frac{x^2}{2})$
  - ABC normalized histogram (=Likelihood): $\chi_{dof=4}(x)/\gamma$

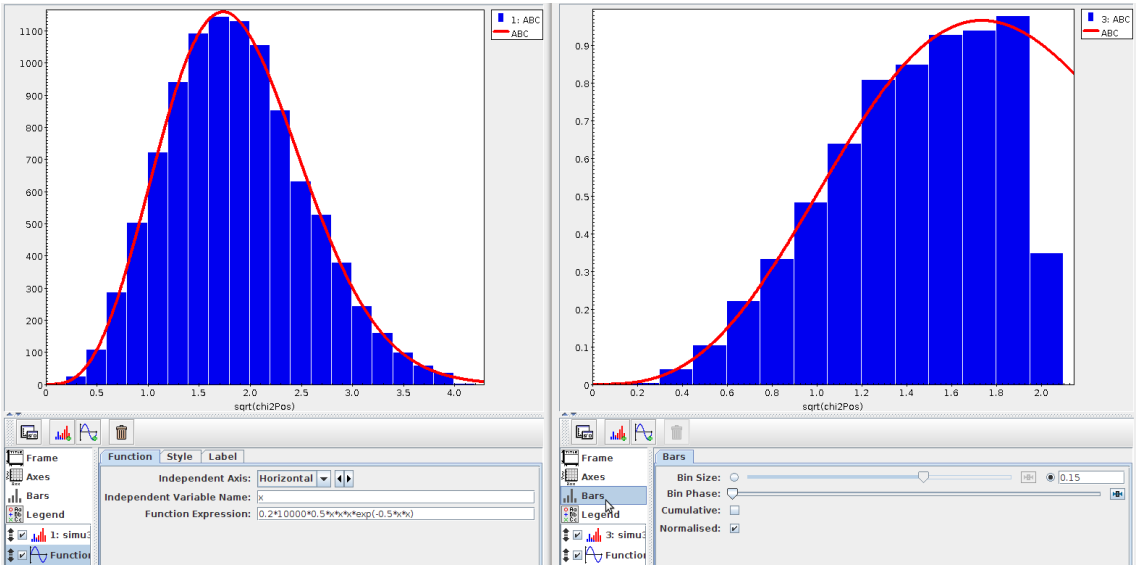
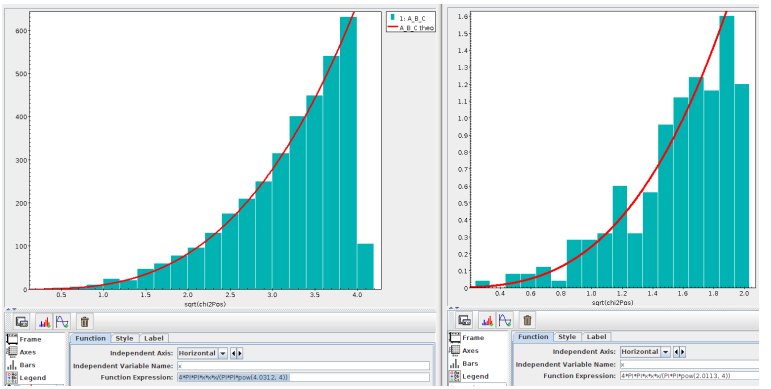


Figure : Left: ABC histogram with $\gamma = 0.9973$; Right: ABC normalized histogram with $\gamma = 0.6$

X-MATCH TOOL TUTORIAL

Theoretical Mahalanobis distance histogram

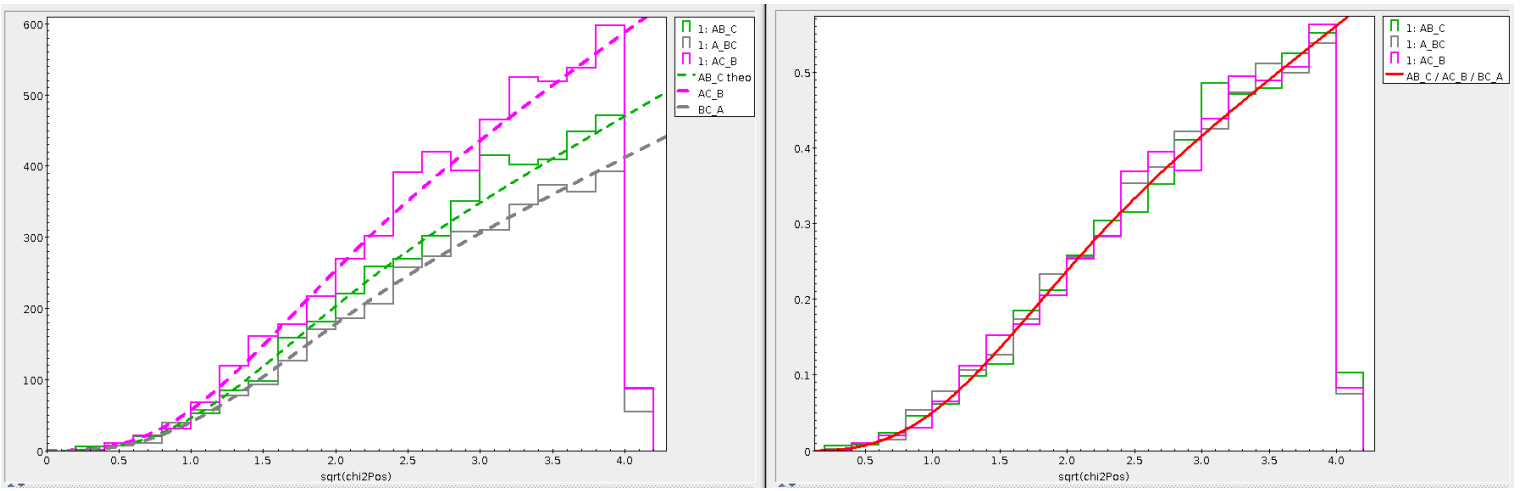
- Step 5.2: plot theoretical curves over the Mahalanobis distance histograms
  - S: Surface area of the xmatched region ($\approx \pi r^2$ here)
  - k: Mahalanobis distance threshold
    - ★ $k = 4.03127$ for $\gamma = 0.9973$ (for 3 catalogues only)
    - ★ use e.g. `www.wolframalpha.com` to solve $\int_0^k \chi_{dof=4}(x)\mathrm{d}x = \gamma$:
      `solve integrate x^3/2*exp(-x^2/2)dx from 0 to k = 0.9973 for k`
  - nTotX: total number of sources in catalogue X
  - A_B_C histogram: binStep$\times$nTotA$\times$nTotB$\times$nTotC$\frac{\sigma_A^2\sigma_B^2+\sigma_A^2\sigma_C^2+\sigma_B^2\sigma_C^2}{S^2} \times 2\pi^2 x^3$
  - ABC normalized histogram (=Likelihood): $4x^3/k^4$



X-MATCH TOOL TUTORIAL

Theoretical Mahalanobis distance histogram

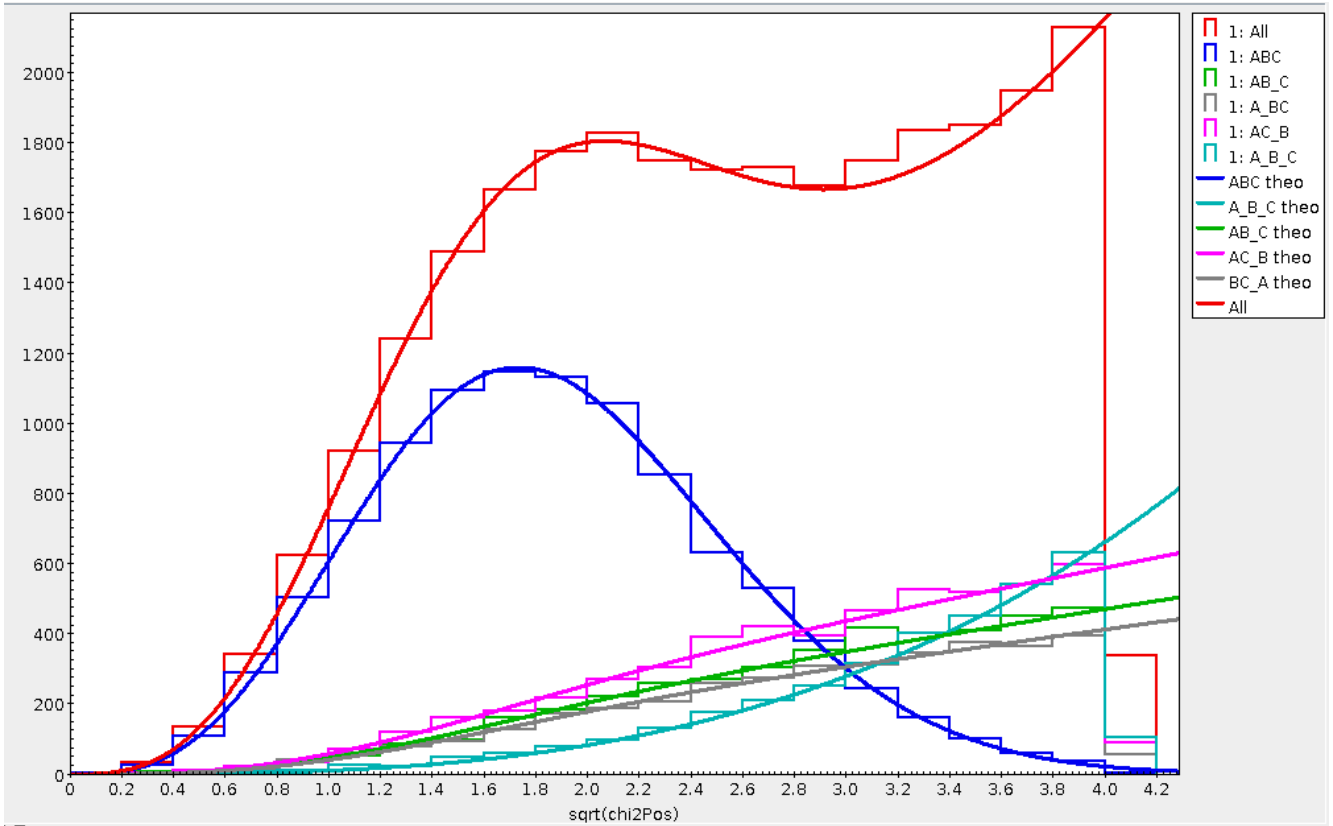
- Step 5.3: plot theoretical curves over the Mahalanobis distance histograms
  - $\sigma^2_{AB\_C} = \frac{\sigma^2_A * \sigma^2_B}{\sigma^2_A + \sigma^2_B} + \sigma^2_C$, similarly for $\sigma^2_{AC\_B}$ and $\sigma^2_{BC\_C}$
  - AB_C histogram: binStep (nABC+nAB) nTotC $(\sigma^2_{AB\_C}/S)2\pi x(1 - \exp(-x^2/2))$, similarly for AC_B and BC_A
  - AB_C/AC_B/BC_A normalized histograms (=Likelihood): $2\pi x(1 - \exp(x^2/2))/(\pi[k^2 - 2(1 - \exp(-k^2/2))])$



X-MATCH TOOL TUTORIAL

Theoretical Mahalanobis distance histogram

- Step 5.4: plot theoretical curves over the Mahalanobis distance histograms
  - Summing the 5 curves, you obtain the curve of all associations



X-MATCH TOOL TUTORIAL

Theoretical Mahalanobis distance histogram

- Closing remarks
  - About probabilities:
    - ⋆ Distributions fitting normalized histograms are likelihoods
    - ⋆ Curves fitting histograms are $\propto$ prior $\times$ likelihood
    - ⋆ proba ABC(x) = curve ABC(x) / curve total(x)
    - ⋆ similarly for proba A_B_C(x), AB_C(x), ...
  - About the tool
    - ⋆ nTotA, nTotB, nTotC are known (= number of rows in each table)
    - ⋆ (nAB+nABC) is estimated from the xmatch of A and B
    - ⋆ similarly for (nAC+nABC) and (nBC+nABC)
    - ⋆ from this plus the xmatch of A with B and C we can estimate nABC
    - ⋆ $\Rightarrow$ we are able to compute all probabilities
    - ⋆ $\Rightarrow$ for *n* catalogues, we have to perform the xmatches for all possible subset of catalogues!