Probabilistic positional association of catalogs of astrophysical sources: the ASPECTS code

(arxiv.org/abs/1209.5361)

Michel Fioc[1,2]

[1]Institut d'astrophysique de Paris

[2]Université Pierre et Marie Curie

IAP, 2013-5-30

The cross-identification problem

Consider two catalogs of astrophysical sources

$$K = \{M_1, \ldots, M_n\} \quad \text{and} \quad K' = \{M'_1, \ldots, M'_{n'}\}$$

defined on a common surface of area S.

How can one decide, just from the positions $c_1, \ldots, c_n$ and $c'_1, \ldots, c'_{n'}$ of K- and K'-sources and from their uncertainties, whether

- M_i is identical to ("is associated with") M'_j, (event $A_{i,\,j}$)

or

- M_i has no counterpart in K'? (event $A_{i,\,0}$)

Probability distribution of the position of a source

Assume the observed positions of M_i and M'_j, $\vec{r}_i$ and $\vec{r}'_j$, are normally distributed around their true positions, $\vec{r}_i^{\,0}$ and $\vec{r}'^{\,0}_j$:

$$\vec{r}_i \sim \mathcal{N}(\vec{r}_i^{\,0}, \Gamma_i) \quad \text{and} \quad \vec{r}'_j \sim \mathcal{N}(\vec{r}'^{\,0}_j, \Gamma'_j).$$

If M_i and M'_j are the same point source, their true positions are identical. Then,

$$\vec{r}_{i,j} = \vec{r}'_j - \vec{r}_i \sim \mathcal{N}(0, \Gamma_{i,j}),$$

where $\Gamma_{i,j} = \Gamma_i + \Gamma'_j$, and Γ_i and Γ'_j are given in the same basis (warning: near the poles, the North of M_i differs from that of M'_j), i.e.

$$P(c_i \mid c'_j \cap A_{i,j}) = \frac{\exp(-\frac{1}{2}\,\vec{r}_{i,j}^{\,\intercal} \cdot \Gamma_{i,j}^{-1} \cdot \vec{r}_{i,j})}{2\pi\sqrt{\det \Gamma_{i,j}}} = \xi_{i,j}.$$

(For extended sources, one may add to $\Gamma_{i,j}$ a term depending on unknown parameters to account for the possible difference between the true positions.)

If M_i has no counterpart, it is randomly distributed in S:

$$P(c_i \mid A_{i,0}) = \frac{1}{S} = \xi_{i,0}.$$

Naïve "answer"

If $\Gamma_i = \sigma^2$ for all K-sources and $\Gamma'_j = \sigma'^2$ for all K'-sources,

1. define $R = \text{some factor} \times \sqrt{\sigma^2 + \sigma'^2}$;
2. if there is a source M'_j in the disk of radius R centered on M_i, M'_j is the counterpart of M_i;
3. if not, M_i has no counterpart.

Weaknesses

- The factor is somewhat arbitrary.
- There may be more than one K'-source in the disk around M_i: which one is the counterpart?
- The higher the density of K'-sources, the more likely that an unrelated one will be close to M_i.
- If positional uncertainties are elliptical, there are different ellipses for each (M_i, M'_j): M'_j may be considered as a counterpart for the ellipse defined by (M_i, M'_j), but not for that defined by (M_i, M'_k).
- Positional uncertainties are not always known.

A probabilistic answer

What is really wanted is the *probability* that M_i is associated with M'_j ($j > 0$), or that M_i has no counterpart in K' ($j = 0$), given the positions of all sources and the uncertainties on these, i.e.

$$P(A_{i,\,j} \mid C \cap C'),$$

where $C = c_1 \cap \ldots \cap c_n$ and $C' = c'_1 \cap \ldots \cap c'_{n'}$.

Unknown parameters

This probability depends on at least one unknown: the *a priori* probability (not knowing C and C') that any K-source has a counterpart in K',

$$f = P\Big(\bigcup_{j>0} A_{i,\,j}\Big) = 1 - P(A_{i,\,0}).$$

(We will also use the *a priori* probability that any K'-source has a counterpart in K,

$$f' = P\Big(\bigcup_{i>0} A_{i,\,j}\Big) = 1 - P(A_{0,\,j}).)$$

$P(A_{i,\,j} \mid C \cap C')$ may also depend on other unknowns, such as the positional uncertainty in one catalog, or the combined uncertainty of both.

Possible assumptions on associations

To compute the probabilities, some model of association must be assumed:

Several-to-one: several K-sources may be associated with the same K'-source, but at most one K'-source is associated to a given K source. More precisely,

$$\begin{cases} \text{for all } M_i\text{, the events } (A_{i,j})_{j\in[\![1,n']\!]} \text{ are exclusive;} \\ \text{for all } M'_j\text{, the events } (A_{i,j})_{i\in[\![1,n]\!]} \text{ are independent.} \end{cases} \qquad (H_{\text{s:o}})$$

Reasonable if the angular resolution is much poorer in K' than in K.

One-to-several: symmetric of several-to-one (K and K' swapped). $(H_{\text{o:s}})$
Appropriate for extended sources looking single at the wavelength of K but breaking up at that of K'.

One-to-one: any K-source has at most one counterpart in K' and reciprocally, i.e.

$$\text{all the events } (A_{i,j})_{i\in[\![1,n]\!],\, j\in[\![1,n']\!]} \text{ are exclusive.} \qquad (H_{\text{o:o}})$$

One has then $fn = f'n'$. Natural assumption for point sources if the angular resolution is high in both K and K'.

Several-to-several: not considered.

In $H_{\text{s:o}}$ and $H_{\text{o:s}}$, catalogs do not play symmetrical roles. Assumption $H_{\text{o:o}}$ is therefore more neutral, but calculations with $H_{\text{s:o}}$ are much simpler and will serve as a guide; they also provide initial values for one-to-one computations.

Three related problems

- Calculate, for each assumption, the probability of association $P(A_{i,\,j} \mid C \cap C')$ that M_i is the same as M'_j ($j > 0$) or that M_i has no counterpart ($j = 0$), given the coordinates of all sources and the unknown parameters.

- Estimate unknown parameters from the data, in particular the *a priori* probability f that any K-source has some counterpart.

- Select the most likely assumption, i.e. the most appropriate association model, given the data.

Computation of $P_{\text{s:o}}(A_{i,\,j} \mid C \cap C')$

$$P(A_{i,\,j} \mid C \cap C') = \frac{P(A_{i,\,j} \cap C \mid C')}{P(C \mid C')} \,.$$

Computation of the denominator

M_i may be associated to M'_j which may be associated to M_k which may be associated to M'_ℓ which may be associated to $M_i \ldots \Rightarrow$ One needs to consider all possible combinations of all the events $A_{k,\,j_k}$ and order them.

Event

$$\bigcap_{k=1}^{n} \bigcup_{j_k=0}^{n'} A_{k,\,j_k} = \bigcup_{j_1=0}^{n'} \bigcup_{j_2=0}^{n'} \cdots \bigcup_{j_n=0}^{n'} \bigcap_{k=1}^{n} A_{k,\,j_k}$$

is certain, so

$$P_{\text{s:o}}(C \mid C') = P_{\text{s:o}}\Big(C \cap \bigcap_{k=1}^{n} \bigcup_{j_k=0}^{n'} A_{k,\,j_k} \mid C'\Big) = \sum_{j_1=0}^{n'} \sum_{j_2=0}^{n'} \cdots \sum_{j_n=0}^{n'} P_{\text{s:o}}\Big(C \cap \bigcap_{k=1}^{n} A_{k,\,j_k} \;\Big|\; C'\Big)$$

$$= \sum_{j_1=0}^{n'} \sum_{j_2=0}^{n'} \cdots \sum_{j_n=0}^{n'} P_{\text{s:o}}\Big(C \;\Big|\; \bigcap_{k=1}^{n} A_{k,\,j_k} \cap C'\Big) P_{\text{s:o}}\Big(\bigcap_{k=1}^{n} A_{k,\,j_k} \;\Big|\; C'\Big).$$

One has

$$P_{\mathrm{s:o}}\Big(C \;\Big|\; \bigcap_{k=1}^{n} A_{k,\,j_k} \cap C'\Big) = \text{cst.} \times \prod_{k=1}^{n} \xi_{k,\,j_k}$$

and

$$P_{\mathrm{s:o}}\Big(\bigcap_{k=1}^{n} A_{k,\,j_k} \;\Big|\; C'\Big) = P_{\mathrm{s:o}}\Big(\bigcap_{k=1}^{n} A_{k,\,j_k}\Big) = \left(\frac{f}{n'}\right)^{q} (1-f)^{n-q},$$

where q is the number of K-sources with a counterpart.

Finally,

$$P_{\mathrm{s:o}}(C \mid C') = \text{cst.} \times \sum_{j_1=0}^{n'} \sum_{j_2=0}^{n'} \cdots \sum_{j_n=0}^{n'} \prod_{k=1}^{n} \zeta_{k,\,j_k}\,,$$

where

$$\zeta_{k,\,0} := (1-f)\xi_{k,\,0} \quad \text{and} \quad \zeta_{k,\,j_k} := \frac{f\xi_{k,\,j_k}}{n'} \text{ if } j_k > 0.$$

Computation of the numerator

Similarly,

$$P_{\mathrm{s:o}}(A_{i,\,j} \cap C \mid C') = \text{cst.} \times \zeta_{i,\,j} \sum_{j_1=0}^{n'} \cdots \sum_{j_{i-1}=0}^{n'} \sum_{j_{i+1}=0}^{n'} \cdots \sum_{j_n=0}^{n'} \prod_{\substack{k=1\\k\neq i}}^{n} \zeta_{k,\,j_k}.$$

Ratio

$P_{\text{s:o}}(C \mid C')$ and $P_{\text{s:o}}(A_{i,\,j} \cap C \mid C')$ may be factorized:

$$\sum_{j_1=0}^{n'} \sum_{j_2=0}^{n'} \cdots \sum_{j_n=0}^{n'} \prod_{k=1}^{n} \zeta_{k,\,j_k} = \prod_{k=1}^{n} \sum_{j_k=0}^{n'} \zeta_{k,\,j_k},$$

so

$$\begin{aligned} P_{\text{s:o}}(A_{i,\,j} \mid C \cap C') &= \frac{\zeta_{i,\,j} \prod_{\substack{k=1 \\ k \neq i}}^{n} \sum_{j_k=0}^{n'} \zeta_{k,\,j_k}}{\prod_{k=1}^{n} \sum_{j_k=0}^{n'} \zeta_{k,\,j_k}} = \frac{\zeta_{i,\,j}}{\sum_{k=0}^{n'} \zeta_{i,\,k}} \\ &= \begin{cases} \dfrac{f\xi_{i,\,j}}{(1-f)n'\xi_{i,\,0} + f\sum_{k=1}^{n'} \xi_{i,\,k}} & \text{if } j > 0, \\ \dfrac{(1-f)n'\xi_{i,\,0}}{(1-f)n'\xi_{i,\,0} + f\sum_{k=1}^{n'} \xi_{i,\,k}} & \text{if } j = 0. \end{cases} \end{aligned}$$

The sums on k may be restricted to sources M'_k close to M_i.

Likelihood and estimation of unknown parameters under $H_{\text{s:o}}$

Maximize the likelihood

$$L = \text{cst.} \times P(C \cap C')$$

to observe all sources at their effective positions. Maximum likelihood estimates $\hat{x}$, $\hat{y}$, etc., of the unknown parameters x, y, etc., are thus obtained by solving

$$\left(\frac{\partial L}{\partial x}\right)_{(x,\, y,\, \ldots)=(\hat{x},\, \hat{y},\, \ldots)} = 0.$$

Under the several-to-one assumption,

$$L_{\text{s:o}} = \text{cst.} \times \prod_{i=1}^{n} \sum_{k=0}^{n'} \zeta_{i,\, k},$$

from which one derives that

$$\frac{\partial \ln L_{\text{s:o}}}{\partial f} = \frac{n(1-f) - \sum_{i=1}^{n} P_{\text{s:o}}(A_{i,\,0} \mid C \cap C')}{f(1-f)},$$

so

$$\hat{f}_{\text{s:o}} = 1 - \frac{1}{n} \sum_{i=1}^{n} \hat{P}_{\text{s:o}}(A_{i,\,0} \mid C \cap C'), \quad \text{where } \hat{P}_{\text{s:o}} = (P_{\text{s:o}})_{f=\hat{f}_{\text{s:o}}}.$$

As $\hat{f}_{\text{s:o}}$ appears on both sides, we calculate it by a back and forth iteration between the l.h.s. and the r.h.s., starting from some arbitrary $f \in [0, 1]$ (this converges very fast).

Theoretical computation of $P_{\mathrm{o:o}}(A_{i,\,j} \mid C \cap C')$

A K'-source associated to a K-source may not be associated to another one, so

$$P_{\mathrm{o:o}}(C \mid C') = \mathrm{cst.} \times \sum_{\substack{j_1=0\\ j_1\notin X_0}}^{n'} \sum_{\substack{j_2=0\\ j_2\notin X_1}}^{n'} \cdots \sum_{\substack{j_n=0\\ j_n\notin X_{n-1}}}^{n'} \prod_{k=1}^{n} \eta_{k,\,j_k}\,,$$

where

$$X_0 := \varnothing, \quad X_k := (X_{k-1} \cup \{j_k\}) \setminus \{0\},$$

$$\eta_{k,\,0} := (1-f)\xi_{k,\,0} \quad \text{and} \quad \eta_{k,\,j_k} := \frac{f\xi_{k,\,j_k}}{n' - \#X_{k-1}} \quad \text{if } j_k > 0.$$

(At depth k in the recursive sum, X_{k-1} is the set of excluded K'-sources: it contains the counterparts associated with $M_1, \ldots, M_{k-1}$, which may therefore not be associated with $M_k, \ldots, M_n$.)

$P_{\mathrm{o:o}}(A_{i,\,j} \mid C \cap C')$ is computed similarly and

$$P_{\mathrm{o:o}}(A_{i,\,j} \mid C \cap C') = \frac{\zeta_{i,\,j} \sum_{\substack{j_1=0\\ j_1\notin X^*_0}}^{n'} \cdots \sum_{\substack{j_{i-1}=0\\ j_{i-1}\notin X^*_{i-2}}}^{n'} \sum_{\substack{j_{i+1}=0\\ j_{i+1}\notin X^*_i}}^{n'} \cdots \sum_{\substack{j_n=0\\ j_n\notin X^*_{n-1}}}^{n'} \prod_{\substack{k=1\\ k\neq i}}^{n} \eta^*_{k,\,j_k}}{\sum_{\substack{j_1=0\\ j_1\notin X_0}}^{n'} \sum_{\substack{j_2=0\\ j_2\notin X_1}}^{n'} \cdots \sum_{\substack{j_n=0\\ j_n\notin X_{n-1}}}^{n'} \prod_{k=1}^{n} \eta_{k,\,j_k}}$$

(the asterisk means that j is also excluded if $j > 0$).

Likelihood and estimation of unknown parameters under $H_{\text{o:o}}$

The recursive sums may not be factorized, so the ratio may not be simplified, contrary to the several-to-one case. Because of the combinatorial explosion of the number of terms, $P_{\text{o:o}}(A_{i,\,j} \mid C \cap C')$ and $L_{\text{o:o}}$ seem impossible to evaluate.

Assume nonetheless that one can compute $P_{\text{o:o}}(A_{i,\,j} \mid C \cap C')$. Then, one can show that one still has

$$\frac{\partial \ln L_{\text{o:o}}}{\partial f} = \frac{n(1-f) - \sum_{i=1}^{n} P_{\text{o:o}}(A_{i,\,0} \mid C \cap C')}{f(1-f)},$$

which gives $\hat{f}_{\text{o:o}}$ by the back and forth iteration described earlier.

Since $L_{\text{o:o}}(f = 0)$ (i.e., when all sources are randomly distributed) is known, $L_{\text{o:o}}$ may also be obtained for any f by integrating $\partial \ln L_{\text{o:o}} / \partial f$.

One can also compute $L_{\text{o:o}}$ like this:

$$L_{\text{o:o}} = \text{cst.} \times \prod_{i=1}^{n} \frac{(1-f)\xi_{i,\,0}}{P_{\text{o:o}}(A_{i,\,0} \mid C \cap C' \cap \bigcap_{k=1}^{i-1} A_{k,\,0})}.$$

Practical computation of $P_{\text{o:o}}(A_{i,\,j} \mid C \cap C')$

A partially true idea:

$P_{\text{o:o}}(A_{i,\,j} \mid C \cap C')$ *depends only on the neighbors of* M_i *and* M'_j.

One may indeed expect that, although the numerator $P_{\text{o:o}}(A_{i,\,j} \cap C \mid C')$ and the denominator $P_{\text{o:o}}(C \mid C')$ depend on distant sources, the effect of these cancels in their *ratio*.

Therefore, order K-sources by increasing distance to M_i and consider the ratio of the recursive sums in the numerator and denominator up to some depth ℓ,

$$p_\ell := \frac{\zeta_{i,\,j} \sum_{\substack{j_2=0 \\ j_2 \notin \widetilde{X}_1^*}}^{n'} \cdots \sum_{\substack{j_\ell=0 \\ j_\ell \notin \widetilde{X}_{\ell-1}^*}}^{n'} \prod_{k=2}^{\ell} \widetilde{\eta}_{k,\,j_k}^*}{\sum_{\substack{j_1=0 \\ j_1 \notin \widetilde{X}_0}}^{n'} \sum_{\substack{j_2=0 \\ j_2 \notin \widetilde{X}_1}}^{n'} \cdots \sum_{\substack{j_\ell=0 \\ j_\ell \notin \widetilde{X}_{\ell-1}}}^{n'} \prod_{k=1}^{\ell} \widetilde{\eta}_{k,\,j_k}}.$$

(The tilde is for the reordering.)

One has

$$p_n = P_{\text{o:o}}(A_{i,\,j} \mid C \cap C').$$

When $\ell \nearrow$, more distant neighbors are progressively included in p_ℓ. The ratio p_ℓ oscillates for small ℓ, then stabilizes for some ℓ_0.

It is therefore tempting to conclude that the sequence (p_ℓ) has converged and to set $P_{\text{o:o}}(A_{i,\,j} \mid C \cap C') = p_{\ell_0}$ (all the more tempting that, when M_i is the only K-source considered, $p_1 = P_{\text{s:o}}(A_{i,\,j} \mid C \cap C')$).

All-sky simulations with *known circular* positional uncertainties

Combined positional uncertainty $\mathring{\sigma} = \sqrt{\sigma^2 + \sigma'^2}$ known $\Rightarrow$ only f must be estimated.

Input values: $f = 1/2$; $\mathring{\sigma} = 10^{-3}$ rad; $n' = 10^5$; $n \in [\![10^3, 10^5]\!]$.

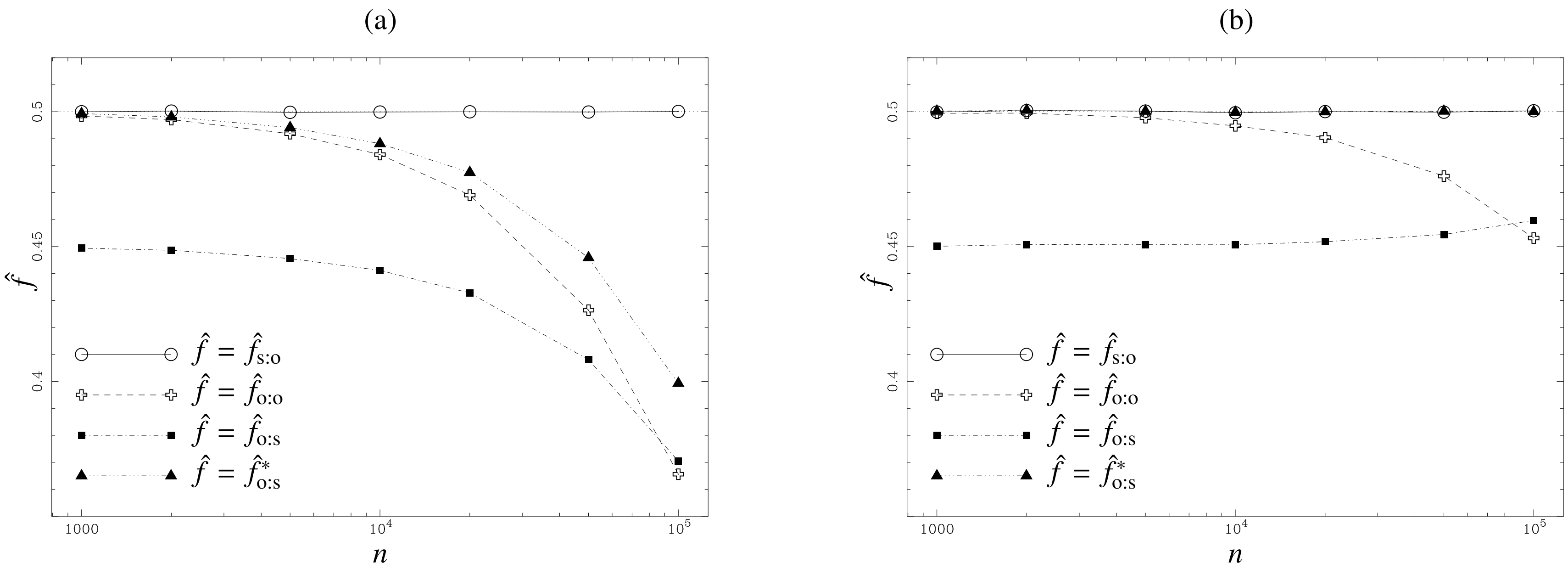

Figure : Mean value of different estimators $\hat{f}$ of f as a function of n.
$\hat{f}^*_{\mathrm{o:s}} = \hat{f}'_{\mathrm{o:s}} n'/n$ is an estimator derived from $\hat{f}'_{\mathrm{o:s}}$, assuming one-to-one associations.
$\hat{f}_{\mathrm{o:s}}$ is not a maximum likelihood estimator.

(a) Several-to-one simulations.

(b) One-to-one simulations ($\hat{f}_{\mathrm{s:o}}$ and $\hat{f}^*_{\mathrm{o:s}}$ overlap).

Simulations with *unknown circular* positional uncertainties

Both f and $\mathring{\sigma}$ must be estimated.
Input values: $f = 1/2$; $\mathring{\sigma} = 10^{-3}$ rad; $n = n' = 2 \times 10^4$.

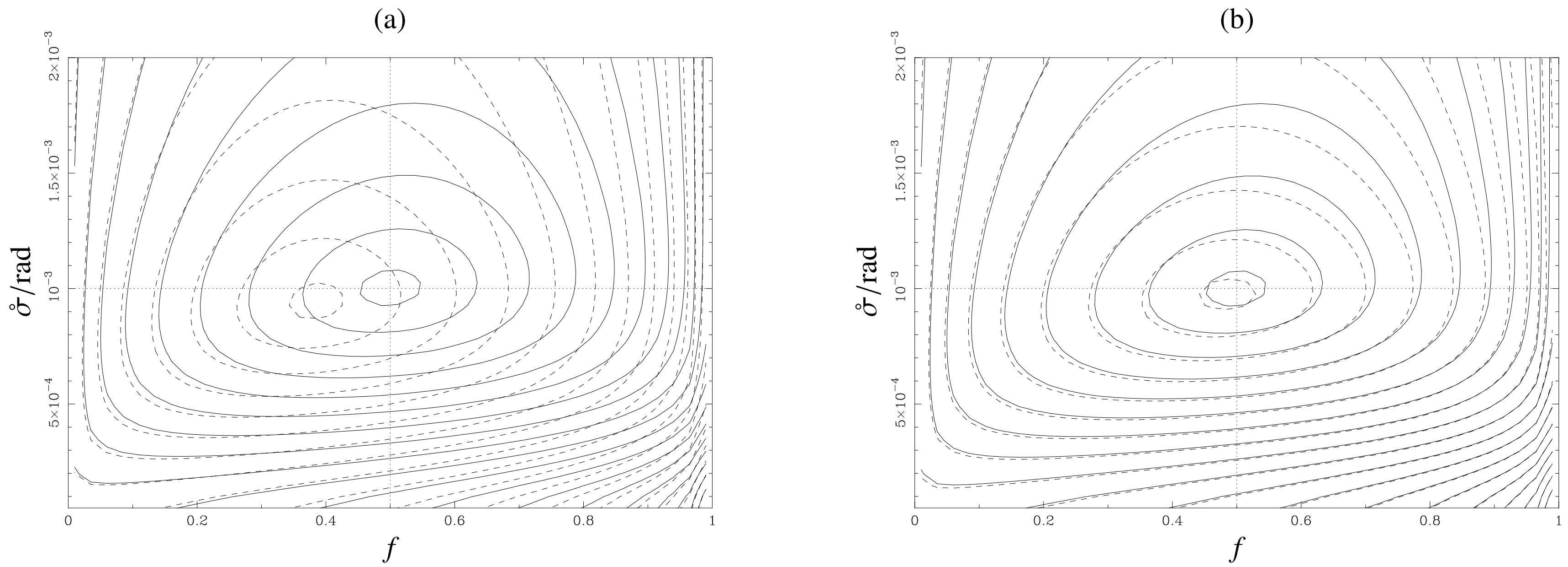

Figure : Contour lines of $L_{\text{s:o}}$ (solid) and $L_{\text{o:o}}$ (dashed). The input values of f and $\mathring{\sigma}$ are indicated by dotted lines.

(a) Several-to-one simulations.

(b) One-to-one simulations.

Comparison of maximum likelihoods

Circular positional uncertainties. Combined positional uncertainty known.
Input values: $f = 1/2$; $\mathring{\sigma} = 10^{-3}$ rad; $n' = 10^5$; $n \in [\![10^3, 10^5]\!]$.

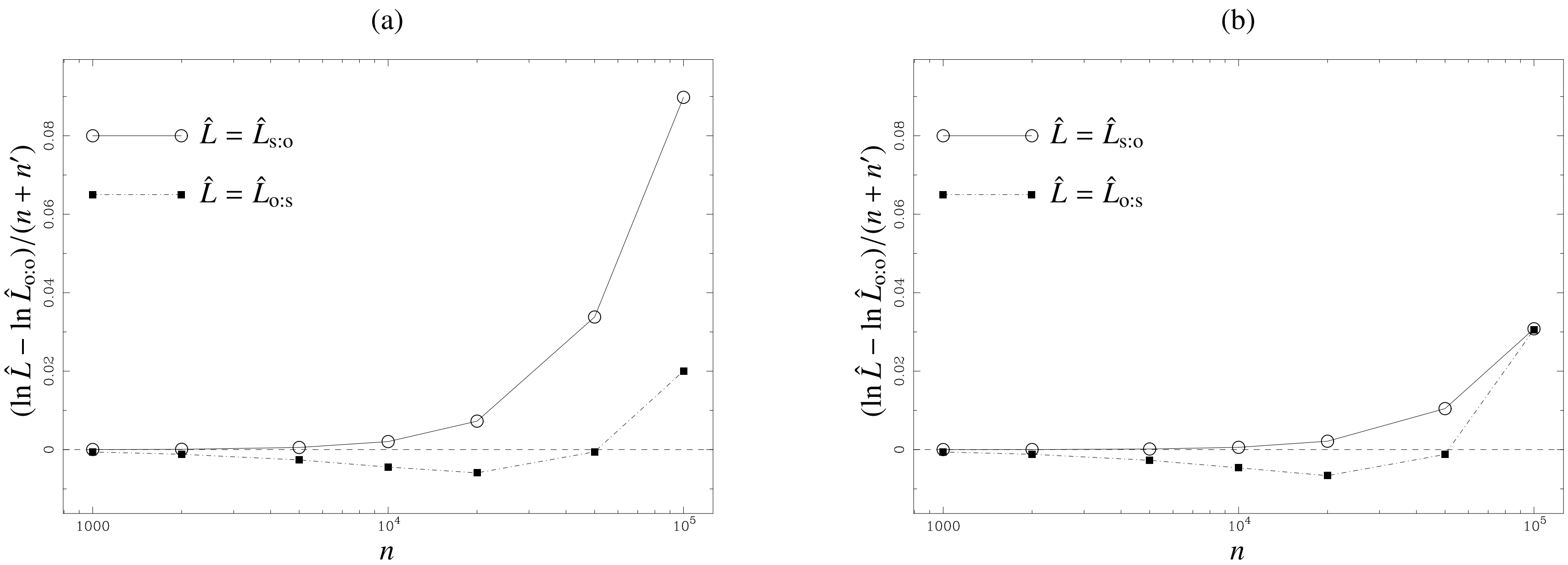

Figure : Normalized mean value of different likelihoods at their maximum as a function of n.

(a) Several-to-one simulations.

(b) One-to-one simulations.

Analysis of simulations

- Several-to-one estimators always provide closer values to f and $\mathring{\sigma}$ than one-to-one estimators, even for one-to-one simulations!
- One-to-one estimators are *statistically inconsistent*: the bias does not tend to 0 when $n \nearrow$.
  Note however that maximum likelihood estimators may be inconsistent in some circumstances. Conditions used to prove their consistency are not applicable here.
- The maximum of the several-to-one likelihood is larger than that of the one-to-one likelihood, even for one-to-one simulations.
- One also has

$$\begin{array}{ccll} \hat{f}_{\mathrm{s:o}} n & > & \hat{f}'_{\mathrm{o:s}} n' & \text{for several-to-one simulations,} \\ & \approx & & \text{for one-to-one simulations,} \\ & < & & \text{for one-to-several simulations.} \end{array}$$

Practical recommendations

The failure of one-to-one estimators and the fact that $\hat{L}_{\mathrm{o:o}} < \hat{L}_{\mathrm{s:o}}$ for one-to-one simulations are embarrassing, but no algorithmic or numerical mistake was found in the numerous tests we made, both analytically and numerically, manually and with Mathematica. The same problem occurs on a circle.

For want of a better solution, we therefore recommended the following to compute $P(A_{i,\,j} \mid C \cap C')$:

- if $\hat{f}_{\mathrm{s:o}} n \gtrnapprox \hat{f}'_{\mathrm{o:s}} n'$, assume $H_{\mathrm{s:o}}$ and use $P_{\mathrm{s:o}}$ with $f = \hat{f}_{\mathrm{s:o}}$;
- if $\hat{f}_{\mathrm{s:o}} n \lnapprox \hat{f}'_{\mathrm{o:s}} n'$, assume $H_{\mathrm{o:s}}$ and use $P_{\mathrm{o:s}}$ with $f' = \hat{f}'_{\mathrm{o:s}}$;
- if $\hat{f}_{\mathrm{s:o}} n \approx \hat{f}'_{\mathrm{o:s}} n'$, assume $H_{\mathrm{o:o}}$ and use $P_{\mathrm{o:o}}$. As $\hat{f}_{\mathrm{s:o}}$ and $\hat{f}'_{\mathrm{o:s}}$ are, respectively, good estimators of f and f' in the case of one-to-one simulations and as one must then have $f n = f' n'$, take
$$f = \sqrt{\hat{f}_{\mathrm{s:o}} \hat{f}^{*}_{\mathrm{o:s}}}\,.$$

The version on arXiv corresponds to this stage.

Reconsideration

After scrutiny of a simple example ($n = n' = 2$), it became clear that *distant sources do matter*, but only by their number, not their exact positions: they only lock some number of counterparts which may not be associated to M_i and its neighbors.

In p_ℓ, n' must be repaced by n'_{eff}, the number of K'-sources that may *effectively* be associated with M_i and its $\ell - 1$ nearest neighbors. One has

$$n'_{\text{eff}} = n' - \sum_{\text{distant } M_k} (1 - P_{\text{o:o}}[A_{k,\,0} \mid \mathrm{C} \cap \mathrm{C}']).$$

For $\ell = n$, $n'_{\text{eff}} = n'$ and one recovers the theoretical result for $P_{\text{o:o}}(A_{i,\,j} \mid \mathrm{C} \cap \mathrm{C}')$.

As $P_{\text{o:o}}$ depends on n'_{eff} which itself depends on $P_{\text{o:o}}$, both may be computed with a back and forth iteration, taking $P_{\text{s:o}}$ as the initial value of $P_{\text{o:o}}$.

What happened with the partially true idea that only neighbors matter is that, after a transient phase where p_ℓ oscillated, a steady state was reached. The ratio p_ℓ had however not converged, but was slowly drifting to $p_n = P_{\text{o:o}}(A_{i,\,j} \mid \mathrm{C} \cap \mathrm{C}')$.

Simulations with *known circular* positional uncertainties (revised)

$\mathring{\sigma}$ known. Only f must be estimated.
Input values: $f = 1/2$; $\mathring{\sigma} = 10^{-3}$ rad; $n' = 10^5$; $n \in [\![10^3, 10^5]\!]$.

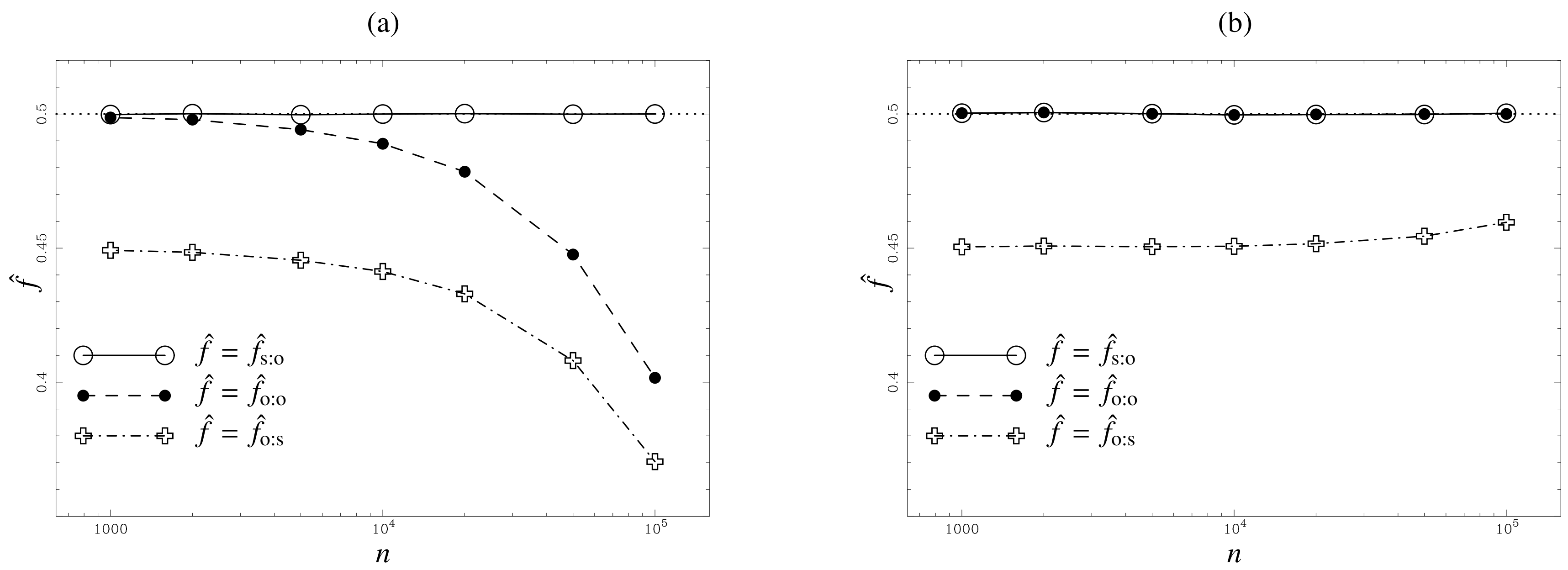

Figure : Mean value of different estimators $\hat{f}$ of f as a function of n.

(a) Several-to-one simulations.

(b) One-to-one simulations.

Simulations with *unknown circular* positional uncertainties (revised)

Both f and $\mathring{\sigma}$ must be estimated.
Input values: $f = 1/2$; $\mathring{\sigma} = 10^{-3}$ rad; $n = n' = 2 \times 10^4$.

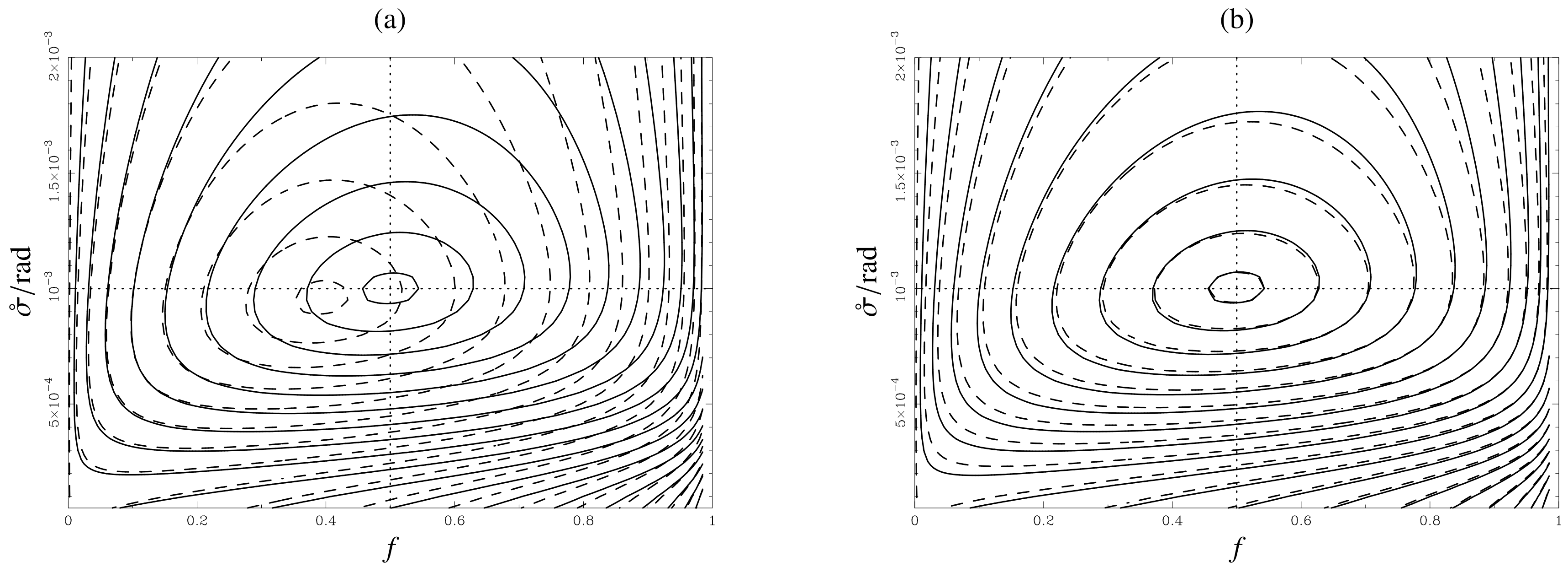

Figure : Contour lines of $L_{\mathrm{s:o}}$ (solid) and $L_{\mathrm{o:o}}$ (dashed). The input values of f and $\mathring{\sigma}$ are indicated by dotted lines.

(a) Several-to-one simulations.

(b) One-to-one simulations.

Simulations with *unknown elliptical* positional uncertainties

Both f and $\mathring{\sigma}$ must be estimated.

Input values: $f = 1/2$; randomly oriented positional uncertainty ellipses with a semi-major axis of 1.5×10^{-3} rad and a semi-minor axis of 0.5×10^{-3} rad; $n = n' = 2 \times 10^4$.

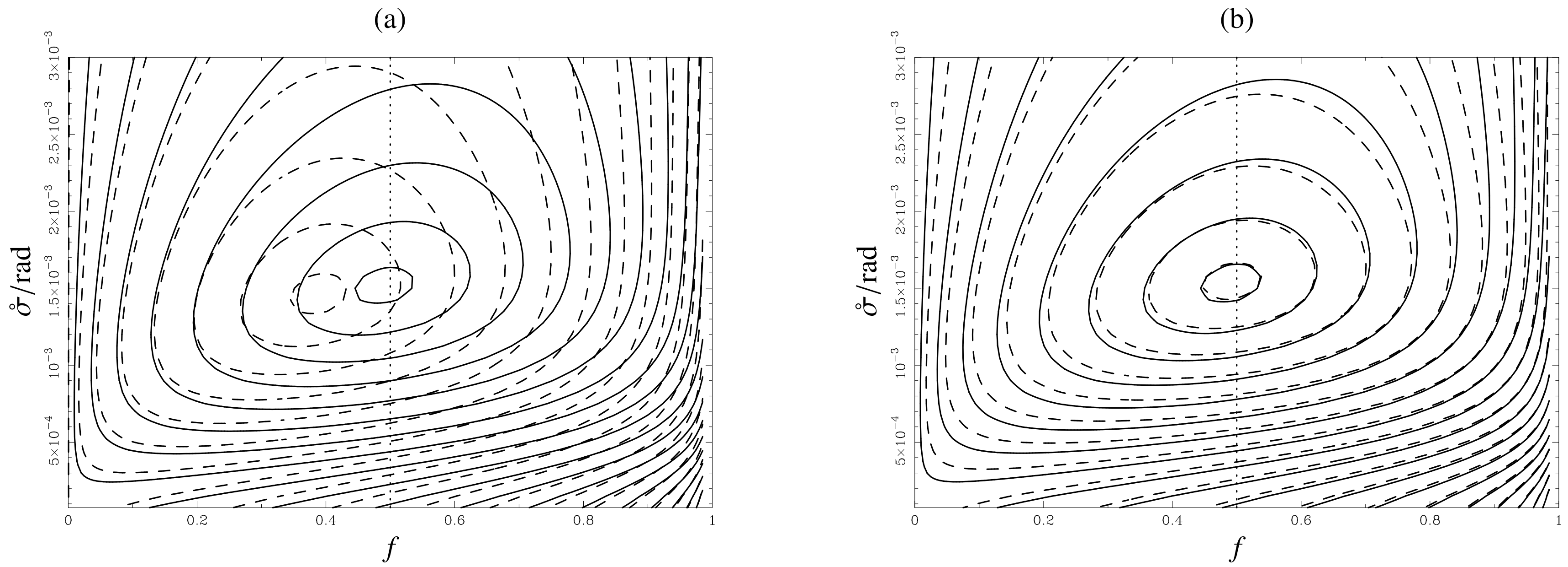

Figure : Contour lines of $L_{\mathrm{s:o}}$ (solid) and $L_{\mathrm{o:o}}$ (dashed). The input value of f is indicated by a dotted line.

(a) Several-to-one simulations.

(b) One-to-one simulations.

Comparison of maximum likelihoods (revised)

Circular positional uncertainties. Combined positional uncertainty known.
Input values: $f = 1/2$; $\mathring{\sigma} = 10^{-3}$ rad; $n' = 10^5$; $n \in [\![10^3, 10^5]\!]$.

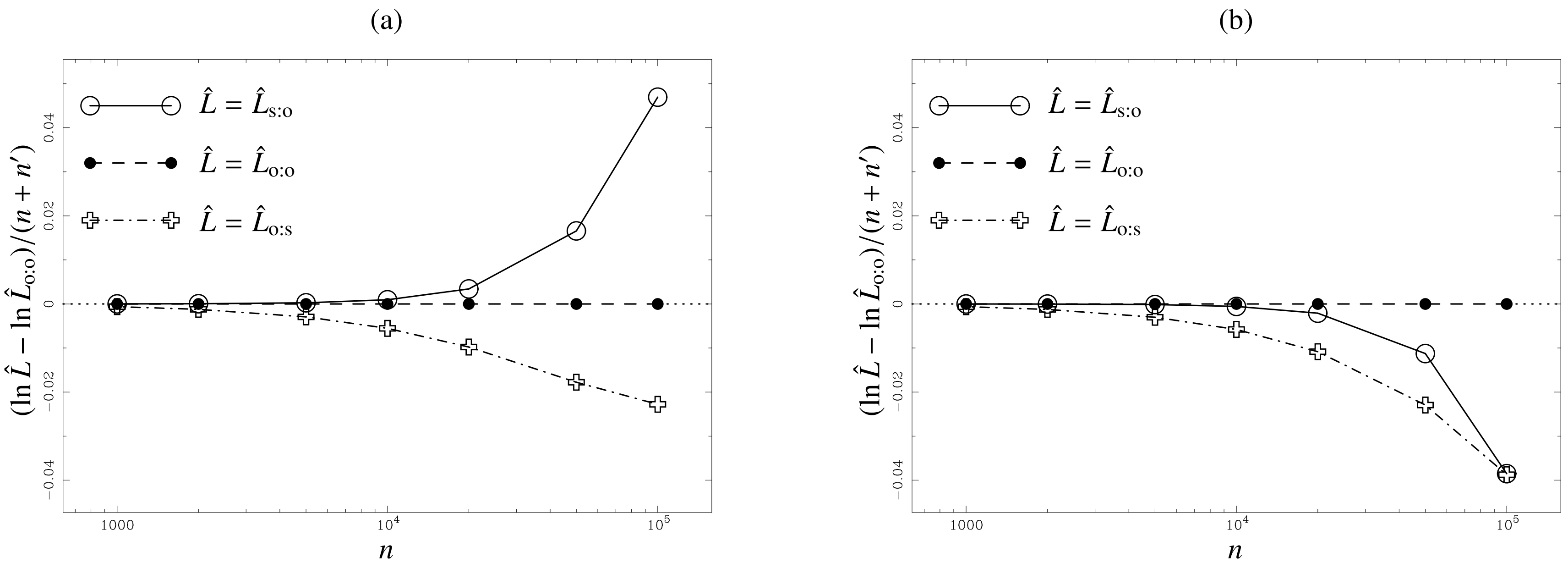

Figure : Normalized mean value of different likelihoods at their maximum as a function of n.

(a) Several-to-one simulations.

(b) One-to-one simulations.

Conclusions

- Several-to-one estimators provide unbiased values of f for several-to-one simulations, but also for one-to-one simulations.
- Revised one-to-one estimators provide unbiased values of f for one-to-one simulations, but not for several-to-one simulations (not a problem).
- The same holds for $\mathring{\sigma}$ if it is unknown.
- These estimators are robust: if $\mathring{\sigma}$ is unknown or if positional uncertainties are elliptical, the right value of f is still recovered.
- For several-to-one simulations, $\hat{L}_{\mathrm{s:o}} > \hat{L}_{\mathrm{o:o}} > \hat{L}_{\mathrm{o:s}}$, as expected.
- For one-to-one simulations, $\hat{L}_{\mathrm{o:o}} > \hat{L}_{\mathrm{s:o}}$ and $\hat{L}_{\mathrm{o:o}} > \hat{L}_{\mathrm{o:s}}$, as expected.

The Aspects code

All these simulations were created and analyzed with the Fortran 95 code Aspects ([aspε]), an acronym for

As*sociation* **P**$^{\textit{ositionnell}}_{\textit{robabilist}}$**E** *de* **ca**T*alogues de* **s***ources.*

Version 1 available at `www2.iap.fr/users/fioc/Aspects/`.
Version 2 in preparation.