Inference of cluster distance and geometry from astrometry

Coryn A.L. Bailer-Jones
Max Planck Institute for Astronomy, Heidelberg
Email: calj@mpia.de
2018-01-30

Abstract

Here I outline a method to determine the distance to and size/shape of a stellar cluster given measurements of the parallaxes and sky positions of (some of) its cluster members. I start rather generally, but then proceed to make some simplifying assumptions. This example also serves to illustrate the key steps involved in solving many general parameter inference problems.

Notation convention: I write scalars in lower case, e.g. ϖ, vectors in bold, e.g. rvec, and matrices in upper case, e.g. C. We have a set of N stars, and any set of N quantities I denote using curly brackets, e.g. $\{\varpi\}$ or $\{\mathbf{r}\}$.

1 Problem statement

We have a set of N stars with measured parallaxes $\{\varpi\}$ with covariance ($N \times N$ matrix) C. We also have the position on the the sky of each star, $\boldsymbol{\alpha} = (\alpha, \delta)$, which we assume to have no uncertainty.[1] Let $\mathbf{r}$ be 3D position of the star, the magnitude of which – the true, unknown distance – is r. Note that we are not assessing cluster membership here: we assume all stars are cluster members.

The centre of the cluster is at position $\mathbf{r}_c$. The magnitude of this is the cluster distance, r_c, which is one of the parameters we wish to infer. I assume that the direction to the cluster centre is known. The parameters which describe the intrinsic geometry of the cluster – its size and shape in 3D – are denoted $\boldsymbol{s}$. This could comprise many parameters to describe some complex shape, or it may be as simple as a one-parameter isotropic model.

The *cluster model* describes the spatial distribution of stars in the cluster. It specifies

$$P(\{\mathbf{x}\}\,|\,\boldsymbol{s}) \tag{1}$$

where $\mathbf{x}$ is the 3D position of a star relative to the cluster centre, the magnitude of which is x (see Figure 1). As $\{\boldsymbol{\alpha}\}$ are known (no uncertainty), we can instead write the cluster model in terms of the N scalar quantities $\{x\}$

$$P(\{x\}\,|\,\{\boldsymbol{\alpha}\}, \boldsymbol{s})\ . \tag{2}$$

Assuming further that the stars are distributed independently in the cluster, the above can just be written as the product of N terms of the form $P(x\,|\,\boldsymbol{\alpha}, \boldsymbol{s})$.

[1] This is a reasonable assumption for Gaia astrometry, because the positions are determined to a *much* higher precision than the angular size of the cluster.

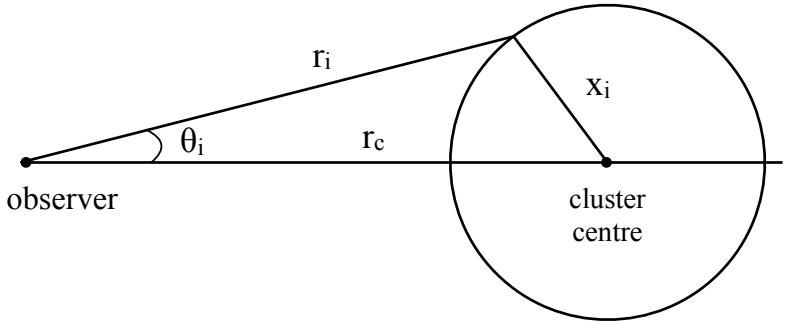

Figure 1: Sketch of a cluster. The distance to the centre of the cluster from us is r_c. An arbitrary star in the cluster has distance r_i from us and x_i from the centre of the cluster.

It follows from Figure 1) that $\mathbf{r} = \mathbf{x} + \mathbf{r}_c$. The *cosine rule* tells us that

$$x^2 = r^2 + r_c^2 - 2rr_c\cos\theta \tag{3}$$

whereby θ, the angular distance we observe between a star and the cluster centre, is determined entirely by $\boldsymbol{\alpha}$.

The *generative model* is the set of equations which predicts the noise-free data from the parameters of the model. These equations here are the cluster model, the cosine rule, and the defintion of parallax, $\varpi_{\rm true} = 1/r$: The cluster model gives us x in terms of the cluster parameters, $\boldsymbol{s}$, and the cosine rule transforms this into r (together with $\boldsymbol{\alpha}$).

The *measurement model* (also known as the *noise model*) describes the probability density function (PDF) over the data (i.e. the measured parallaxes) given the model predictions ($\{1/r\}$). This we can write as

$$P(\{\varpi\}\,|\,\{r\}, C) \tag{4}$$

where C is the covariance between the parallax measurements. The parallax measurement model would generally be expressed as an N-dimensional Gaussian distribution with mean $\{1/r_i\}$ and covariance C. If the parallax measurements are independent, C becomes a diagonal matrix and we can write the measurement model as a product of N 1D Gaussians, each with mean $1/r_i$ and standard deviation $\sigma_{\varpi i}$

$$\prod_{i=1}^{N} P(\varpi_i\,|\,r_i, \sigma_{\varpi_i}) \tag{5}$$

where $\sigma_{\varpi i}$ is the standard uncertainty in the parallax ϖ_i for each star i.

2 Solution

We first identify what we want to find. This is the posterior

$$P(r_c, \boldsymbol{s}\,|\,\{\varpi\}, C, \{\boldsymbol{\alpha}\})\ . \tag{6}$$

We then use *Bayes' theorem* to write this in terms of the likelihood and prior

$$P(r_c, \boldsymbol{s}\,|\,\{\varpi\}, C, \{\boldsymbol{\alpha}\}) = \frac{P(\{\varpi\}\,|\,C, \{\boldsymbol{\alpha}\}, r_c, \boldsymbol{s})\,P(r_c, \boldsymbol{s})}{P(\{\varpi\}\,|\,C, \{\boldsymbol{\alpha}\})} \tag{7}$$

whereby the denominator is independent of the parameters, so is not of explicit interest.[2] Note that I have removed C and $\{\boldsymbol{\alpha}\}$ from the prior, because the prior has nothing to do with the measurement covariance, and I assume the prior to be independent of the direction on the sky.

To proceed, we need to write the likelihood in terms of quantities we know, which are the parallax measurement model (equation 4) and the cluster model (equation 1). The former model involves the variable r and the latter the variable $\mathbf{x}$; these variables are connected via the geometry (Figure 1) so one can be converted into the other. This suggests that we should introduce one of these variables, and here I chose r. As r is unknown – and we don't even want to find the individual distances – we should integrate (marginalize) over it. I therefore write the likelihood using the *marginalization law* as

$$P(\{\varpi\}|C,\{\boldsymbol{\alpha}\},r_c,\boldsymbol{s}) = \int P(\{\varpi\},\{r\}|C,\{\boldsymbol{\alpha}\},r_c,\boldsymbol{s})\,d\{r\} \tag{8}$$

$$= \int P(\{\varpi\}|\{r\},C,\{\boldsymbol{\alpha}\},r_c,\boldsymbol{s})\,P(\{r\}|\,C,\{\boldsymbol{\alpha}\},r_c,\boldsymbol{s})\,d\{r\} \tag{9}$$

$$= \int P(\{\varpi\}|\{r\},C)\,P(\{r\}|\,\{\boldsymbol{\alpha}\},r_c,\boldsymbol{s})\,d\{r\}\,. \tag{10}$$

whereby the integral is N-dimensional (over all N true, unknown distances). In the second line I have used the product rule ($P(A,B) = P(A|B)P(B)$) to split the integrand into two terms. In the third line I have used the concept of *conditional independence* to remove dependency on some variables. In the first term on that line, the distribution of the parallaxes is determined entirely by the true distances and the parallax covariance: This is the parallax measurement model (equation 4), which this is what this term therefore becomes in line three. Once we have the true distances, the angular positions of the stars as well as the cluster distance and size/shape provide no additional information on the distribution of the parallaxes, and so these variables can be dropped in the conditioning. Similarly in the second term on the third line, the distribution of the true distances of the stars in the cluster – given the cluster distance and size/shape – is entirely independent of the parallax covariance, so C drops out of this term. Assuming further that the true positions of the stars in the cluster are independent of one another (given the cluster distance and size/shape), this second in line three can be written as product, i.e.

$$P(\{r\}|\,\{\boldsymbol{\alpha}\},r_c,\boldsymbol{s}) = \prod_{i=1}^{N} P(r_i|\,\boldsymbol{\alpha}_i,r_c,\boldsymbol{s})\,. \tag{11}$$

The position on the sky for a star, $\boldsymbol{\alpha}$, specifies the angle θ in Figure 1. It therefore follows from equation 3 that for fixed $\boldsymbol{\alpha}$ and r_c, x is determined by r. We can therefore transform the above PDF in r into one in x as

$$P(r_i|\,\boldsymbol{\alpha}_i,r_c,\boldsymbol{s})\,dr_i = P(x_i|\,\boldsymbol{\alpha}_i,\boldsymbol{s})\,dx_i \tag{12}$$

for any star i, where I could remove r_c from on the right side of this equation because conditional on knowing $\boldsymbol{\alpha}$ and $\boldsymbol{s}$, the distribution of x – the distance of stars from the centre of the cluster – is independent of how far the cluster is from us. We can then re-write equation 12 as

$$P(r_i|\boldsymbol{\alpha}_i,r_c,\boldsymbol{s}) = P(x_i|\boldsymbol{\alpha}_i,\boldsymbol{s})\left(\frac{\partial x_i}{\partial r_i}\right)_{r_c,\theta_i}\,. \tag{13}$$

[2] We can consider it as a normalization constant of the numerator, and can compute it (if we need it) by integrating the numerator over the whole parameter space.

Note that this involves a *partial* derivative because r_c and θ_i are held constant. This term is the *Jacobian*: it transforms the probability density. We compute it from equation 3:

$$\left(\frac{\partial x}{\partial r}\right)_{r_c,\theta} = \frac{r - r_c\cos\theta}{x} \,. \tag{14}$$

We now substitute equation 14 into 13 and this into equation 11, to write

$$P(\{r\}|\ \{\boldsymbol{\alpha}\}, r_c, \boldsymbol{s}) = \prod_{i=1}^{N} P(x_i|\boldsymbol{\alpha}_i, \boldsymbol{s}) \left(\frac{r_i - r_c\cos\theta_i}{x_i}\right) \,. \tag{15}$$

Note that the first term on the right is the cluster model, equation 2.

This we have now expressed the likelihood in equation 10 in terms of things we know, namely the measurement model (equation 4) and the cluster model (equation 2). Specifying a form for the prior, $P(r_c, \boldsymbol{s})$, we can now in principle compute the posterior (equation 7) by sampling, for example by MCMC. The problem is rather complex, however, because every likelihood computation involves an N-dimensional numerical integral.

With further assumptions we can simplify the likelihood considerably. First, let us assume that the N parallax measurements are independent. In this case the measurement model can be written as a product of N 1-dimensional integrals (equation 5), and so the likelihood – equation 10 – can be written (also using equation 15) also as a product of N 1-dimensional integrals

$$P(\{\varpi\}|C, \{\boldsymbol{\alpha}\}, r_c, \boldsymbol{s}) = \int \left[\prod_{i=1}^{N} P(\varpi_i|r_i, \sigma_{\varpi_i})\, P(x_i|\boldsymbol{\alpha}_i, \boldsymbol{s}) \left(\frac{r_i - r_c\cos\theta_i}{x_i}\right)\right] d\{r\} \tag{16}$$

$$= \prod_{i=1}^{N} \int P(\varpi_i|r_i, \sigma_{\varpi_i})\, P(x_i|\boldsymbol{\alpha}_i, \boldsymbol{s}) \left(\frac{r_i - r_c\cos\theta_i}{x_i}\right) dr_i \,. \tag{17}$$

The order of integral and product could be swapped because the integrals are independent. While the integrals must still be computed numerically (in general), they are one dimensional, so can be done by methods such as Gaussian quadrature which are often much faster and more precise than MCMC.

We can simplify even further. All known stellar clusters (except the Hyades) have angular extents of at most a few degrees. In this case $\cos\theta \simeq 1$ to a high degree of precision (e.g. $\cos 3^\circ = 0.9986$), and $x \simeq r - r_c$, in which case the geometric factor in the above likelihood is about 1, so can be dropped (as can any dependence on θ or $\boldsymbol{\alpha}$). It turns out that although the integrals still have to be computed numerically in general, they are now faster and more robust to compute.

So far I have not made any assumptions about the cluster model. The posterior has dimensionality $p + 1$, where p is the number of parameters in the cluster model. How well these parameters can be constrained depends on the data. The most extreme simplification here is to assume an isotropic model for the cluster, i.e. it is described by a single length scale parameter, s_c. In this case, and using also the small θ approximation, the likelihood is

$$P(\{\varpi\}|C, r_c, s_c) \simeq \prod_{i=1}^{N} \int P(\varpi_i|r_i, \sigma_{\varpi_i})\, P(x_i|s_c)\, dr_i \,. \tag{18}$$

3 Summary of key steps in many inference problems

The previous section involves a number of key steps which are used in many inference problems. These are summarized below.

Identify the following:

1. the data and the parameters: what do we measure, and what do we want to infer?
2. the generative model (the thing which predicts noise-free data given some values of the parameters);
3. the measurement model, which is the likelihood;
4. our background knowledge, which is the prior.

We proceed by writing down the posterior, and then manipulating it in the following ways:

1. use Bayes' theorem to write the posterior in terms of the prior and likelihood;
2. use the marginalization law to write the likelihood as a marginalization over variables which connect the generative model and the measurement model (here that was the set of true but unknown distances, $\{r\}$);
3. use the product rule to express this likelihood as a product of known quantities (related to the measurement model and generative model);
4. use conditional independence to remove irrelevant terms from these quantities;
5. use a Jacobian to convert any PDFs to be over more convenient variables;
6. apply reasonable assumptions which simplify the computations (here assuming independent parallax measurements and small clusters).

4 Tutorial example

In the notebook "Distance inference – multiple sources" on `https://github.com/ehalley/parallax-tutorial-2018`, case 2, I use the likelihood in equation 18 with an isotropic Gaussian model for the cluster

$$P(x_i \,|\, s_c) \sim \mathcal{N}(x_i; \mu = 0, \sigma = s_c) \; . \tag{19}$$

As we adopt $\cos\theta \simeq 1$, $x_i \simeq r_i - r_c$. Thus the likelihood (equation 18) for this case becomes

$$P(\{\varpi\} |\, \{\sigma_{\varpi_i}\}, r_c, s_c) \simeq \prod_{i=1}^{N} \int \frac{1}{\sqrt{2\pi}\sigma_{\varpi_i}} \exp\left[-\frac{(\varpi_i - 1/r_i)^2}{2\sigma_\varpi^2}\right] \frac{1}{\sqrt{2\pi}s_c} \exp\left[-\frac{(r_i - r_c)^2}{2s_c^2}\right] dr_i \; . \tag{20}$$