

A new maximum likelihood method for luminosity calibrations

X. Luri¹, M.O. Mennessier², J. Torra¹ and F. Figueras¹

¹ Departament d'Astronomia i Meteorologia, Universitat de Barcelona, Avda. Diagonal 647, E-08028, Barcelona, Spain

² Université Montpellier II, Groupe de Recherche en Astronomie et Astrophysique du Languedoc, Unité Associée au CNRS 1368, F-34095 Montpellier Cedex 5, France

Received July 31; accepted November 3, 1995

Abstract. — A new statistical parallax method using the Maximum Likelihood principle is presented, allowing the simultaneous determination of a luminosity calibration, kinematic characteristics and spatial distribution of a given sample. This method has been developed for the exploitation of the Hipparcos data and presents several improvements with respect to the previous ones: the effects of the selection of the sample, the observational errors, the galactic rotation and the interstellar absorption are taken into account as an intrinsic part of the formulation (as opposed to external corrections). Furthermore, the method is able to identify and characterize physically distinct groups in inhomogeneous samples, thus avoiding biases due to unidentified components. Moreover, the implementation used by the authors is based on the extensive use of numerical methods, so avoiding the need for simplification of the equations and thus the bias they could introduce. Several examples of application using simulated samples are presented, to be followed by applications to real samples in forthcoming articles .

Key words: methods: statistical — stars: fundamental parameters — stars: kinematics — Galaxy: stellar content

1. Introduction

Several implementations of the Maximum Likelihood (hereafter ML) principle to obtain luminosity calibrations from kinematical data have been developed – see Rigal (1958), Jung (1970), Clube & Jones (1971) and Heck (1975) –, each one more complete than the previous ones. However, several difficulties remain:

1. Given the complexity of the equations involved, no analytical solution can be found for the ML estimator in the general case. Some crude approximations have to be introduced to solve the problem analytically. This can be found even in Heck's implementation, where numerical methods are more widely used. For instance, the dispersion in absolute magnitudes σ_M is never directly determined. Although the effects of the simplifications are tested by the authors – using samples with known trigonometric parallaxes or simulated samples – they reduce the precision of the method and may introduce undetected bias when applied to a real problem.
2. None of the implementations take directly into account the effect of the sample's selection process or observational constraints. To apply the ML principle to a given sample correctly, *its own* density law has to be used, which is determined not only by the density law of the group of stars from which the sample was extracted but also by the selection process by which the stars were chosen. A classical example of this is the case of a magnitude-limited sample. Its mean absolute magnitude is brighter than that of the original group of stars, and the difference is known as the *Malmquist bias* – see Malmquist (1936) and Luri et al. (1993). As the selection effects are not taken into account directly, the results still have to be corrected for some biases (such as the Malmquist bias cited above). These a posteriori corrections reduce the precision of the results and, in some cases, are not enough to correct all the biases completely (Luri et al. 1993).
3. The effects of the observational errors are taken into account only in Heck's implementation but, again, not directly. They are estimated using simulated samples and then subtracted from the results, thus reducing the precision of the method.
4. The interstellar absorption is never completely taken into account. The apparent magnitudes are usually corrected beforehand or iteratively but, as we will see in the next section, this is not enough to take into

account all the effects in play.

5. The characteristics of the spatial distribution of the stars around the Sun are not taken into account, usually assuming it as being homogeneous. This is only valid if the sample is very close to the Sun and for samples extending to greater distances the presence of the Galactic Disk should be taken into account.
6. All the implementations implicitly assume that the sample being treated is homogeneous. In a real case, however, the samples will usually be inhomogeneous, containing more than one type of star. Only in Heck's implementation is this problem taken into account by iteratively eliminating stars with a velocity too far from the mean. However, the procedure can be dangerous in some cases. The distance to a star, needed to calculate its velocity, is estimated using the mean absolute magnitude obtained for the sample. For samples with large absolute magnitude dispersions there will be stars much fainter than the mean, for which the distance estimation will be too large and lead to a very high velocity. Faint stars will then be systematically eliminated, biasing the sample towards bright magnitudes.

In the next sections our implementation of the ML principle is presented, showing how these problems have been overcome by taking advantage of the power of modern computers to minimize the approximations. For a more detailed description see Luri (1995).

2. Mathematical formalism

The principle of ML can be briefly described as follows: let \mathbf{x} be a random variable following the density law given by $\mathcal{D}(\mathbf{x} | \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0 = (\theta_1, \theta_2, \dots, \theta_{n_\theta})$ is the set of unknown parameters on which it depends. Given a sample of realizations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_x})$ of the random variable, its *Likelihood Function* is defined as

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n_x} \mathcal{D}(\mathbf{x}_i | \boldsymbol{\theta}). \quad (1)$$

The value of $\boldsymbol{\theta}$ which maximizes this function is the ML estimator, $\boldsymbol{\theta}_{\text{ML}}$, of the parameters $\boldsymbol{\theta}_0$ characterizing the density law of the sample. It can be shown that $\boldsymbol{\theta}_{\text{ML}}$ is asymptotically non-biased, asymptotically gaussian and that for large samples it is the most efficient estimator – see Kendall & Stuart (1979).

To apply the ML principle correctly the density law *describing the sample* is needed. In this section we will study, step by step, how the different effects – selection of the sample, observational errors, galactic rotation, interstellar absorption, and inhomogeneities of the sample

– can be introduced in the density law, so it can describe the sample being studied as realistically as possible.

Even though a much more compact deduction of this density law is possible – see Luri et al. (1992) – this step-by-step presentation has been chosen for the sake of clarity and also to study how the different effects are treated individually.

2.1. Density law of the Base Group

We use the term *Base Group* to refer to the group of stars from which the sample is extracted, that is to say, the base population in statistical terms. For the moment we will assume that this Base Group is homogeneous, i.e. that it is composed of a single type of stars. In this case its physical characteristics can be described by a single set of distribution functions, which in this work we will assume as:

Absolute magnitude distribution

$$\varphi_M(M) = e^{-\frac{1}{2} \left(\frac{M-M_0}{\sigma_M} \right)^2} \quad (2)$$

Velocity distribution

$$\varphi_v(U, V, W) = e^{-\frac{1}{2} \left(\frac{U-U_0}{\sigma_U} \right)^2 - \frac{1}{2} \left(\frac{V-V_0}{\sigma_V} \right)^2 - \frac{1}{2} \left(\frac{W-W_0}{\sigma_W} \right)^2} \quad (3)$$

Spatial distribution

$$\varphi_e(r, l, b) = e^{-\frac{|r \sin(b)|}{Z_0}} r^2 \cos(b) \quad (4)$$

Notice, however, that the mathematical formalism presented below does not depend on this choice and, as we are going to use numerical methods extensively, the ML estimation can also be performed with a different set of hypotheses. For instance, a different absolute magnitude distribution can be used, assuming a different shape – see Jaschek & Gómez (1985) – or even parametrizing the mean absolute magnitude as a function of a color index – as is performed in forthcoming articles.

A realistic choice of the distribution functions, correctly describing the physical characteristics of the base group, is very important, as they are the basis for the development.

The physical quantities on which these functions depend can be joined in a single random variable, $\mathbf{x} = (M, r, l, b, U, V, W)$ and the distribution functions themselves define the *density law of \mathbf{x} for the Base Group*:

$$\mathcal{D}(\mathbf{x} | \boldsymbol{\theta}) = \varphi_M(M) \varphi_v(U, V, W) \varphi_e(r, l, b) \quad (5)$$

(not normalized) .

It is more convenient, however, to work with observable quantities and so we use a different random variable

$\mathbf{y} = (m, r, l, b, \mu_l, \mu_b, v_r)$. The density law for \mathbf{y} can be obtained from that for \mathbf{x} using the law of transformation of probability distributions:

$$\mathcal{D}'(\mathbf{y} | \boldsymbol{\theta}) = \mathcal{J} \mathcal{D}(\mathbf{x} | \boldsymbol{\theta}), \quad (6)$$

where the parameter vector of the density law is $\boldsymbol{\theta} = (M_0, \sigma_M, U_0, V_0, W_0, \sigma_U, \sigma_V, \sigma_W, Z_0)$ and \mathcal{J} is the jacobian of the $\mathbf{x} \rightarrow \mathbf{y}$ transformation:

$$\mathcal{J} = k^2 r^2 \cos(b), \quad (7)$$

with $k = 4.74 \frac{\text{Km yr}}{\text{s pc}}$.

2.2. Observational selection

The density law $\mathcal{D}'(\mathbf{y} | \boldsymbol{\theta})$ gives the distribution of the random variable \mathbf{y} for the base group but, as stated above, its distribution for the sample of stars being used may be different due to the selection process applied to the choice of its stars.

The selection of a star for inclusion in a sample is based on its observational characteristics. The selection can be due to observational limitations – such as the magnitude limit of a telescope – or to criteria fixed by the observer – e.g. only to take stars with high proper motions. In any case the selection process, being based on the values of \mathbf{y} , can be modelled by a function of this variable. We will call this function the *Selection Function of the Sample*, defined as follows: given a star with values of its observational quantities between \mathbf{y} and $\mathbf{y} + d\mathbf{y}$, its probability of being selected for the sample is

$$dP = \mathcal{S}(\mathbf{y}) d\mathbf{y}. \quad (8)$$

Using this function, the *Density Law of \mathbf{y} for the Sample* can be built simply taking into account that its stars verify two conditions:

1. they belong to the base group
2. they have been selected for the sample

The joint density law for these two independent events can be written as:

$$\mathcal{M}(\mathbf{y} | \boldsymbol{\theta}) = \mathcal{C}_M^{-1} \mathcal{D}'(\mathbf{y} | \boldsymbol{\theta}) \mathcal{S}(\mathbf{y}). \quad (9)$$

Notice that a normalization constant \mathcal{C}_M has been explicitly introduced, whereas it is missing in the definition of $\mathcal{D}(\mathbf{x} | \boldsymbol{\theta})$. This is not normalizable because no limits were introduced in the spatial distribution, which is mathematically correct but has not physical meaning. On the other hand, $\mathcal{M}(\mathbf{y} | \boldsymbol{\theta})$ is normalizable if a realistic selection function is chosen – as for example a limit in apparent magnitude – even if $\mathcal{D}(\mathbf{x} | \boldsymbol{\theta})$ is not. The normalization constant plays an important role in the application of the ML principle. It depends on the same parameters $\boldsymbol{\theta}$ as the functional part of the density law, so it has to be taken into account when maximizing the likelihood with respect

to them.

An example of a selection function is that for the case of a sample complete up to a certain limiting apparent magnitude m_{lim} . In this case all the stars of the base group with $m \leq m_{\text{lim}}$ have a probability one of being selected for the sample, while those with $m > m_{\text{lim}}$ have a zero probability. This kind of selection function can be described with a Heaviside's function, so that

$$\mathcal{S}(\mathbf{y}) = \Theta(m - m_{\text{lim}}). \quad (10)$$

Only one more remark is required to our discussion about the selection function. In many cases the selection process leading to a sample will not be known or will be very difficult to model in a function. In these cases a *Parametric Selection Function* can be used: some parameters are left in the selection function to be estimated by ML, together with those in $\boldsymbol{\theta}$. In this way the selection function adjusts itself to the sample in the process of the ML estimation.

2.3. Observational errors and elimination of the distance

The distribution of the variable \mathbf{y} given by the density law $\mathcal{M}(\mathbf{y} | \boldsymbol{\theta})$ corresponds to the ideal case without observational errors. However, in a real sample these are present and have to be taken into account to describe the distribution of the observational quantities correctly.

Furthermore, in the development presented in this article, the use of trigonometric parallaxes is avoided since it decreases the size of the available samples, severely restricting the application of the method – see Gómez & Luri (1992). For this reason, the distance r has been kept as an integration parameter (see below) among the components of the variable \mathbf{y} instead of introducing the trigonometric parallax, which is directly observable. However, the development can easily be modified to include it, if available, and even adapted to a mixed case, where some stars have measurements of this quantity and others have not. These further developments are included in our plans for future work.

To meet both goals we have to redefine our variables:

- $\mathbf{y} = (m_0, r, l_0, b_0, \mu_{l0}, \mu_{b0}, v_{r0})$ containing the “real” values
- $\mathbf{z} = (m, l, b, \mu_l, \mu_b, v_r)$ containing the measured values

The measured values in \mathbf{z} are distributed around the “real” values in \mathbf{y} according to a certain *Error Distribution* $\mathcal{E}(\mathbf{z} | \mathbf{y})$. The joint distribution of “real” values and measurements will then be determined by the joint density law:

$$\mathcal{M}_{\mathcal{E}}(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}) = \mathcal{C}_{M\mathcal{E}}^{-1} \mathcal{D}'(\mathbf{y} | \boldsymbol{\theta}) \mathcal{S}(\mathbf{z}) \mathcal{E}(\mathbf{z} | \mathbf{y}). \quad (11)$$

Notice that the selection function obviously depends on \mathbf{z} because the selection criteria are applied to the measured values and not to the “real” ones.

These “real” values (including the distance r) are never known, only the measured ones are available. We do not want the joint density law but the one for only \mathbf{z} . This law can be obtained from $\mathcal{M}_{\mathcal{E}}$ by integration (convolution) with respect to \mathbf{y} :

$$\begin{aligned} \mathcal{O}(\mathbf{z} | \boldsymbol{\theta}) &= C_O^{-1} \int_{\forall \mathbf{y}} \mathcal{M}_{\mathcal{E}}(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}) d\mathbf{y} = \\ &= C_O^{-1} \mathcal{S}(\mathbf{z}) \int_{\forall \mathbf{y}} \mathcal{D}'(\mathbf{y} | \boldsymbol{\theta}) \mathcal{E}(\mathbf{z} | \mathbf{y}) d\mathbf{y}, \end{aligned} \quad (12)$$

which constitutes the *Density Law for the Measurements of the Sample*. Notice, as stated above, that the distance r only plays the role of an integration parameter, as it does not appear in \mathbf{z} .

At the present state of our implementation of the method the error distribution has been chosen as follows:

- negligible errors for (m, l, b)
- gaussian errors for (μ_l, μ_b, v_r) with dispersions $(\epsilon_{\mu_l}, \epsilon_{\mu_b}, \epsilon_{v_r})$

Neglecting the errors in m will be discussed more thoroughly in Sect. 2.5. The ones in (l, b) are negligible because their effects (given the precision of position measurements) are well below the statistical errors of the estimation – as can be seen using simulated samples.

As the dispersions $(\epsilon_{\mu_l}, \epsilon_{\mu_b}, \epsilon_{v_r})$ of the measured values around the “real” ones – which are sometimes called the *measurement errors* – vary from star to star, the error distribution does also and so the measurements follow a different density law for each star:

$$\mathcal{O}^{(i)}(\mathbf{z} | \boldsymbol{\theta}) = C_{O_i}^{-1} \mathcal{S}(\mathbf{z}) \int_{\forall \mathbf{y}} \mathcal{D}'(\mathbf{y} | \boldsymbol{\theta}) \mathcal{E}^{(i)}(\mathbf{z} | \mathbf{y}) d\mathbf{y}. \quad (13)$$

Using these individual density laws to define the likelihood function, the different magnitude of the observational errors for each star is naturally taken into account.

2.4. Galactic rotation

The velocity distribution assumed for the base group does not take into account the presence of the galactic differential rotation. While this is accurate enough for samples of stars in the solar neighbourhood, for samples at greater distances its effects have to be included.

The velocities of the stars can be decomposed as the sum of a peculiar velocity, the reflection of the solar motion and the differential rotation. The first is well modelised by an ellipsoidal distribution with zero mean. The second is the same for all the stars, (U_0, V_0, W_0) . The

third one depends on the position of the star. The resulting velocity distribution is then an ellipsoid with means (U'_0, V'_0, W'_0) depending on the position.

$$\varphi_v(U, V, W | r, l, b) = e^{-\frac{1}{2} \left(\frac{U-U'_0}{\sigma_U} \right)^2 - \frac{1}{2} \left(\frac{V-V'_0}{\sigma_V} \right)^2 - \frac{1}{2} \left(\frac{W-W'_0}{\sigma_W} \right)^2}. \quad (14)$$

Using the Oort-Lindblad rotation model at first order, these mean components can be written as:

$$U'_0 = U_0 + \left[\left(3 - \frac{2R}{R_{\odot}} \right) A - B \right] R \sin(\phi) \quad (15)$$

$$V'_0 = V_0 + \left[\left(3 - \frac{2R}{R_{\odot}} \right) A - B \right] R \cos(\phi) - (A - B) R_{\odot} \quad (16)$$

$$W'_0 = W_0, \quad (17)$$

where (R, ϕ) are the galactocentric coordinates of the star, (A, B) the Oort’s constants and R_{\odot} the distance of the Sun to the galactic center.

Thus, using this velocity distribution to construct $\mathcal{O}(\mathbf{z} | \boldsymbol{\theta})$, the presence of the galactic differential rotation is integrated in the ML estimation.

Notice that in this case the spatial and velocity distributions are not independent. While this does not affect the mathematical formalism presented here, it can make its practical implementation more difficult. Notice also that (A, B) and R_{\odot} can either be estimated by ML with the rest of the parameters or fixed in advance.

The velocity distribution could still be completed to take into account the fact that the characteristics and the shapes of ellipsoids of peculiar velocities may depend on the position.

2.5. Interstellar absorption

The usual approach to take into account the presence of interstellar absorption is to correct the apparent magnitudes of the stars using some procedure that is external to the estimation. For instance, the absorption can be estimated a priori, using Strömgren photometry, or iteratively, using the estimated distances to apply Parenago’s (1940) formula and then iterating the estimation – as in Heck (1975).

However, such a correction is not enough to give a correct ML estimation. The spatial distribution of the sample will be changed due to the presence of the absorption. For instance, for a given absolute magnitude the distance limit in a magnitude limited sample is higher in the direction of the galactic pole than in the direction of the galactic equator. This change has to be taken into account in the

density law, otherwise the estimations may be biased.

In our formalism the interstellar absorption has already been included implicitly. In the transformation $\mathbf{x} \rightarrow \mathbf{y}$ the absolute magnitude M is replaced by the “true” apparent magnitude m_0 , its relationship being given by

$$M = m_0 + 5 - 5 \log_{10}(r) - A_v(r, l, b), \quad (18)$$

so that the interstellar absorption $A_v(r, l, b)$ is naturally taken into account in the formulae. Notice again that an external correction for the apparent magnitudes is not necessary, because the fact that they are affected by the interstellar absorption is already contained in the equations.

Even if the formalism is the same, its implementation becomes more complicated when taking into account the interstellar absorption. A model, giving its value as a function of the position (r, l, b) , has to be introduced. In our case the Arenou et al. (1992) model has been chosen. It divides the sky in 199 regions giving for each one the absorption as a function of the distance only, so making the implementation easier.

However, even such a detailed model is only a representation of the *mean* absorption in each region. The real values of the absorption for the stars in a sample will present “fluctuations” around this mean value. The model of Arenou et al. (1992) itself gives the expected magnitude for each region. These fluctuations are superposed to the observational errors in apparent magnitude and give the measured value of this quantity, m , as opposed to m_0 .

These effects can be taken into account in the density law by including both the distribution of the fluctuations and the distribution of the observational errors (in a similar way that the error distributions of other quantities have already been introduced in the previous section). However, when doing so, the implementation becomes complicated and unpractical. Luckily, the simulations show that the combined effect of both phenomena does not introduce remarkable biases in the ML estimation when not taken into account, so they can be neglected (see Sect. 6). In spite of this, we strongly recommend testing their influence for each particular case using simulated samples to confirm the absence of biases.

In Appendix A the detailed expression of $\mathcal{O}(\mathbf{z} | \boldsymbol{\theta})$ for the case of a magnitude-limited sample is presented as an example.

3. Group separation

Until now we have kept the hypothesis of homogeneity of the sample, i.e. we have assumed that the sample has been extracted from a single base group with homogeneous characteristics. Now we are going to relax this assumption.

We will assume that the sample is a mixture of stars coming from several base groups, their physical characteristics being described by the distributions introduced

in previous sections but each one with different values of the corresponding parameters in $\boldsymbol{\theta}$. Given this, the *Density Law for the Measurements of the Part of the Sample Belonging to the j th Group* is

$$\mathcal{O}_j(\mathbf{z} | \boldsymbol{\theta}_j) = c_{\mathcal{O}_j}^{-1} \mathcal{S}(\mathbf{z}) \int_{\forall \mathbf{y}} \mathcal{D}'(\mathbf{y} | \boldsymbol{\theta}_j) \mathcal{E}(\mathbf{z} | \mathbf{y}) d\mathbf{y}. \quad (19)$$

Notice that a common error distribution has been assumed, but individual error distributions could easily be introduced.

If we assume that the group to which each star belongs is not known (which is the usual case), the density law of \mathbf{y} for the sample has to reflect this uncertainty. If $(w_1, w_2, \dots, w_{n_g})$ are the relative abundances of the groups in the sample, the *Density Law for Multiple Groups* can be written as

$$\mathcal{O}(\mathbf{z} | \boldsymbol{\Theta}) = \sum_{j=1}^{n_g} w_j \mathcal{O}_j(\mathbf{z} | \boldsymbol{\theta}_j), \quad (20)$$

where $\boldsymbol{\Theta}$ contains all the parameters to be determined by ML: the parameters of the groups $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{n_g})$ and the relative abundances $(w_1, w_2, \dots, w_{n_g})$.

This density law must be used when the sample being treated is inhomogeneous. However, the number of groups composing it, n_g , is not usually known. In these cases a likelihood test – like Wilk’s test – is used to determine it: ML estimations are performed with $n_g = 1, 2, 3, \dots$ and the maximum likelihoods obtained for each case are compared using the test, to decide on the correct value of n_g .

In many applications the possibility of separating groups is very important, as in the fundamental problem of recognising different galactic populations in a sample.

A further improvement of the *density law for Multiple Groups* can be obtained by introducing different physical distributions – not only different values of the parameters – for each one.

3.1. Assignment

With the density law introduced in the previous section – Eq. (20) – the different groups composing a sample can be identified and characterized. However, if we want to apply the different luminosity calibrations so obtained – one for each group – to the stars of the sample, they have to be assigned to a group.

The problem can then be stated as follows: given a star of the sample, with values of the measurements $\mathbf{z} = (m, l, b, \mu_l, \mu_b, v_r)$, what is the probability that it belongs to a given group? Notice that the problem has been posed in probabilistic terms, because in general there will be no means to classify the stars with complete certainty.

Our mathematical formalism provides all the tools needed to answer the question. The a priori probability of

a star of the sample belonging to the j th group is w_j , and the distribution of \mathbf{z} in this group is given by $\mathcal{O}_j(\mathbf{z} | \boldsymbol{\theta}_j)$. Then, using Bayes formula, the a posteriori probability of a star of the sample belonging to the j th group given its measured values \mathbf{z}_* is:

$$P(* \in G_j | \mathbf{z}_*) = \frac{w_{j \text{ ML}} \mathcal{O}(\mathbf{z}_* | \boldsymbol{\theta}_{j \text{ ML}})}{\sum_{k=1}^{n_g} w_{k \text{ ML}} \mathcal{O}_k(\mathbf{z}_* | \boldsymbol{\theta}_{k \text{ ML}})}. \quad (21)$$

Using this formula the probabilities of the star belonging to each group can be compared, and the star can be assigned to the most likely one.

Notice that this procedure, like any method of statistical classification, will have a certain percentage of misclassification. However, the reliability of each assignation is clearly indicated by the probability given. Notice also that in the given formula the parameters of the groups $\boldsymbol{\theta}_{j \text{ ML}}$ and the relative abundances $w_{j \text{ ML}}$ used are the result of the ML estimation.

4. Estimation of individual distances

The main use of a luminosity calibration is the estimation of distances. However, the existing distance estimators only take into account the characteristics of the sample in a very limited way, and use only a part of the information available for the star. In this section we present a new tool to estimate individual distances of stars which overcomes these limitations. All the effects in play (selection effects, observational errors, interstellar absorption, galactic rotation, etc.) are naturally taken into account, and the estimation error is provided.

Once a star has been assigned to a group, the corresponding ML estimator of the parameters, $\boldsymbol{\theta}_{j \text{ ML}}$ can be used to estimate its distance. Here we present a method to obtain this estimation using all the information available.

When the ML estimation has been performed and the star has been classified we can use

1. The joint density law of (\mathbf{y}, \mathbf{z}) for the group of the star $\mathcal{M}_{\mathcal{E}}(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}_j)$
2. The ML estimation of the parameters of the group $\boldsymbol{\theta}_{j \text{ ML}}$
3. The measurements of the star \mathbf{z}_*

Taking $\boldsymbol{\theta}_j = \boldsymbol{\theta}_{j \text{ ML}}$ in the joint density law and fixing \mathbf{z} to the measured values of the star \mathbf{z}_* , a function depending only on $\mathbf{y} = (m_0, r, l_0, b_0, \mu_{10}, \mu_{b0}, v_{r0})$ is obtained, $\mathcal{M}_{\mathcal{E}}(\mathbf{y} | \mathbf{z}_*, \boldsymbol{\theta}_{j \text{ ML}})$. By convolving this function with respect to $\mathbf{z}_0 = (m_0, l_0, b_0, \mu_{10}, \mu_{b0}, v_{r0})$ a new function depending only on the distance is obtained:

$$\mathcal{R}(r) = \int_{\forall \mathbf{z}_0} \mathcal{M}_{\mathcal{E}}(\mathbf{y} | \mathbf{z}_*, \boldsymbol{\theta}_{j \text{ ML}}) d\mathbf{z}_0. \quad (22)$$

This is the marginal density law of the distance for the star. In other words, it is the distribution of probability of

r for the star. It can then be used to obtain the expected value of the distance

$$\bar{r} = \int_0^{\infty} r \mathcal{R}(r) dr \quad (23)$$

and its dispersion

$$\varepsilon_r^2 = \int_0^{\infty} (r - \bar{r})^2 \mathcal{R}(r) dr. \quad (24)$$

The first can be used as a distance estimator and the second as its error, with the following advantages:

1. It uses all the information available for the star
2. It is not biased, as all the effects in play in the sample are automatically taken into account
3. An individual error estimation is given for each star

The estimation of distances can also be used to improve the classification process presented in the previous section. For a given star, distance estimations can be obtained for each of the groups detected in the sample. Comparison of these distances with external estimations or with trigonometric parallaxes can reveal an assignation error (too great a discrepancy between the distance estimation obtained with the assigned group and the external estimation) and suggest a better classification (the group for which the discrepancy is smaller).

5. Numerical methods

An important part of our implementation is the use of numerical methods to maximize the likelihood. In this way the density laws developed in the previous section can be used without any approximation. A detailed account of all the methods used would be too long so only a brief account is presented. More details can be found in Luri (1995).

5.1. Maximization

The classical approach to maximize the likelihood is based on the *likelihood equations*:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = 0 \quad \forall i. \quad (25)$$

This nonlinear system has to be solved to obtain the ML estimation of the parameters of the density law $\boldsymbol{\theta}_{\text{ML}}$.

To deduce these equations the likelihood function has to be derived with respect to each of the parameters. This is a long and tedious process that makes it difficult to introduce even small changes in the density law. It is much more convenient to use a different approach: if \mathcal{L} is an N-dimensional function, algorithms for N-dimensional maximization can be used directly. In this way it is only necessary to program the function itself and then use the appropriate maximization routine, since modern computers are powerful enough to give the solution in a reasonable

amount of time. Among the N-dimensional maximization algorithms, the ones not using the gradient should be used. Otherwise the derivatives should be calculated and the advantages are lost. In our case the *Powell Method Discarding the Direction of Largest Decrease* has been selected. Details about this method can be found in Press et al. (1986).

In both cases – use of the likelihood equations and direct maximization – a starting point for the numerical process is needed. It is highly recommended to perform several maximizations, with different starting points, in case the function presents multiple local maxima, and as a check against numerical problems.

5.2. Integration

The density law developed in the previous section contains several integrations that in some cases can not be solved analytically. For these cases numerical integration has been used, with the added advantage of making it easier to introduce modifications in the density law. In our case a *modified Gauss method* has been selected. See also Press et al. (1986).

5.3. Simulated samples

Simulated samples play a key role in our implementation of the ML principle. On the one hand they are used to test the equations and programs developed for the estimation. On the other they allow a good estimation of the errors of the results and the eventual detection of biases: once an ML estimation Θ_{ML} has been obtained, several samples are simulated using the values given and the method is applied to them. The comparison of the results (which could be called “the estimation of the estimation”) with the introduced values allows the detection of biases. The dispersion of these results $\sigma(\Theta_{\text{ML}})$ can be taken also as the error of the estimation.

The base of the simulation of samples is the generation of random numbers following a given distribution. More details about this subject can be found in Press et al. (1986) and Luri (1995).

6. Test using simulated samples

In this section some tests of the method using simulated samples are presented. The case of a sample with the mean characteristics of K0 III stars and complete up to apparent magnitude $m_{\text{lim}} = 7.2^m$ (limit of the Hipparcos survey) has been chosen as a realistic example. Its size has been fixed to 800 stars, which is approximately the number of K0 III stars in the Hipparcos survey.

6.1. Correction of all the effects

The values of the parameters assumed for K0 III stars have been taken from Egret et al. (1982) – M_0 and σ_M – and Mihalas & Binney (1981) – (U_0, V_0, W_0) , $(\sigma_U, \sigma_V, \sigma_W)$ and Z_0 . They are listed in Table 1.

Table 1. Assumed parameters for K0 III stars

M_0 (m)	-0.5
σ_M (m)	1.2
U_0 (km s $^{-1}$)	-7
σ_U (km s $^{-1}$)	30
V_0 (km s $^{-1}$)	-16
σ_V (km s $^{-1}$)	21
W_0 (km s $^{-1}$)	-6
σ_W (km s $^{-1}$)	16
Z_0 (pc)	270

Using these values 10 sets of 20 samples have been generated. Interstellar absorption (with fluctuations around the mean value) and galactic rotation have been included in all cases and observational errors have been increased from set to set. The *base value* of the observational errors has been taken as representative of the ones expected when Hipparcos results are available, that is to say $\epsilon_m = 0.05^m$ – Tycho photometry – , $\epsilon_{\mu\alpha} = \epsilon_{\mu\delta} = 0.002'' \text{ yr}^{-1}$ – Hipparcos measurements – and $\epsilon_{v_r} = 0.5 \text{ km s}^{-1}$ – CORAVEL radial velocities. For the following sets these errors have been multiplied by an increasing *error factor*. Two kinds of ML fit have been applied:

1. Using a likelihood function not taking into account the presence of observational errors, interstellar absorption and galactic rotation (*partial likelihood function*), so taking only into account the limit in apparent magnitude.
2. Using a likelihood function taking into account all the effects, except the observational errors in apparent magnitude and the variations of interstellar absorption around the mean law of Arenou et al. (1992) (*complete likelihood function*).

The mean results for each set are compared in Figs. 1 to 4, where the values used to generate the samples θ_0 (Table 1) are indicated with horizontal lines. The error bars shown in the figures correspond to the dispersion of the individual results (not to be mistaken with the error of the mean) for the complete likelihood function case.

It is obvious from these figures that the results in the partial likelihood function case are strongly biased. Even in the best case of error (factor one, the Hipparcos case), a bias is present – mainly due to the effects of interstellar absorption and galactic rotation – and the biases become worse as the errors increase. Although the extreme case (error factor of ten) is not realistic, the medium cases (error

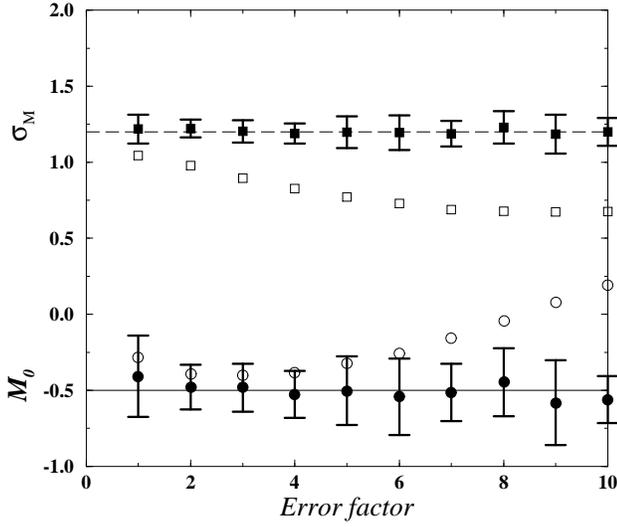


Fig. 1. Results for M_0 and σ_M for ten sets of samples with different *error factors*. Empty symbols: partial likelihood function. Filled symbols: complete likelihood function

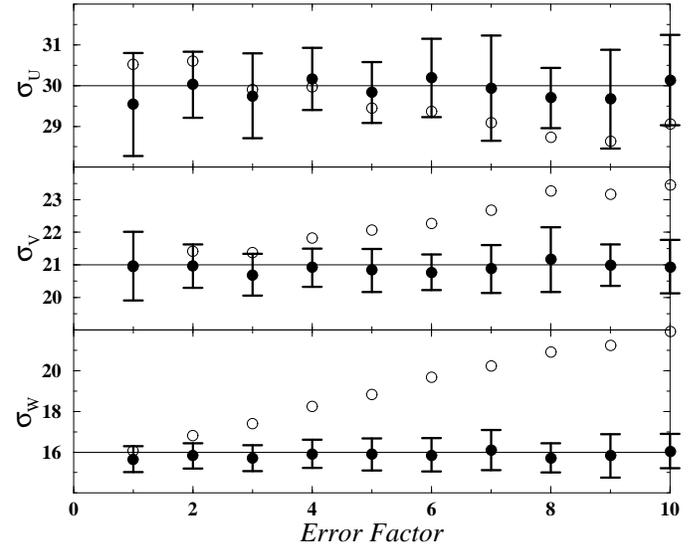


Fig. 3. Results for $(\sigma_U, \sigma_V, \sigma_W)$ for ten sets of samples with different *error factors*. Empty symbols: partial likelihood function. Filled symbols: complete likelihood function

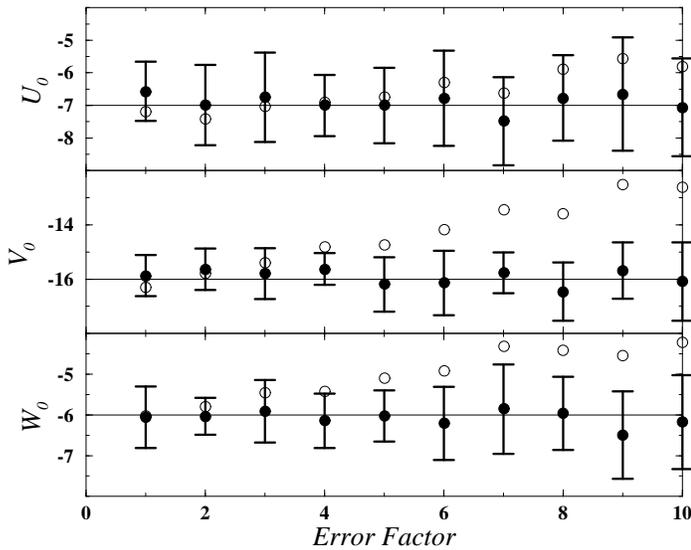


Fig. 2. Results for (U_0, V_0, W_0) for ten sets of samples with different *error factors*. Empty symbols: partial likelihood function. Filled symbols: complete likelihood function

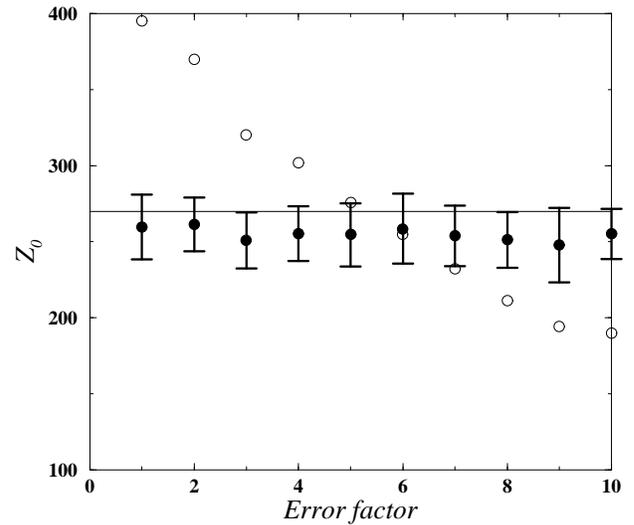


Fig. 4. Results for Z_0 for ten sets of samples with different *error factors*. Empty symbols: partial likelihood function. Filled symbols: complete likelihood function

factor around four) correspond approximately to the quality of the present catalogues, showing how important it is to take into account the effects in play correctly.

On the other hand, in the complete likelihood function case, the estimations are non biased even with an error factor of ten. The only exception are the estimations of Z_0 , which have a slight bias of about 10 pc. This bias is due to the fact that neither the errors in apparent magnitude nor the fluctuations in interstellar absorption have been taken into account. If samples are generated without them, the bias disappears. As this bias is small enough to be neglected (always smaller than the error bars) the use of a likelihood function not including the errors in apparent magnitude and the interstellar absorption fluctuations can be considered acceptable.

6.2. Group separation

In the previous section the samples have been generated as composed by a single group. Now we are going to study the case of two-group samples. As suggested by the results of Ratnatunga (1989) the K0 III stars could be a mixture of two different groups with characteristics similar to the ones listed in Table 2.

Table 2. Assumed parameters for K0 III stars in the case of a mixture of two groups

	group 1	group 2
M_0 (m)	-0.8	1
σ_M (m)	0.4	1
U_0 (km s^{-1})	-7	-10
σ_U (km s^{-1})	22	52
V_0 (km s^{-1})	-12	-25
σ_V (km s^{-1})	15	36
W_0 (km s^{-1})	-6	-7
σ_W (km s^{-1})	12	28
Z_0 (pc)	200	600
%	84	16

To test our method in this case, 20 samples of 800 stars were generated using the above mentioned parameters and the base values of the observational errors given in the previous section. A first fit to these samples was performed including only one group in the likelihood function. The results are listed in Table 3. A second fit was then performed including two groups in the likelihood function, and the results are listed in Table 4.

The results for the one-group fit present a striking resemblance to the parameters used in the previous section, thus suggesting that the K0 III stars could be a mixture of at least two different groups. On the other hand, the results for the two-group fit show that the parameters introduced are correctly estimated.

Table 3. Results for the one-group fitting

	mean	σ
M_0	0.4	0.25
σ_M	1.2	0.11
U_0	-8.0	1.3
σ_U	28.5	1.8
V_0	-14.8	1.3
σ_V	20.1	0.9
W_0	-6.9	0.6
σ_W	15.8	0.6
Z_0	260.	26

Table 4. Results for the two-group fitting

	group 1		group 2	
	mean	σ	mean	σ
M_0	-0.85	0.11	0.89	0.28
σ_M	0.31	0.17	0.95	0.12
U_0	-7.0	0.98	-11.3	5.0
σ_U	22.2	0.92	52.5	4.5
V_0	-12.2	0.90	-27.3	4.8
σ_V	14.9	0.60	37.3	2.9
W_0	-6.3	0.58	-6.8	2.4
σ_W	12.3	0.42	27.7	1.3
Z_0	207.	21.	689.	701.
%	84.3	2.2	15.7	2.2

The classification algorithm described in Sect. 3.1 has been tested with two of the previous samples (1600 stars). The results of the classification are presented in Table 5.

Table 5. Results of the assignment algorithm

real	assigned group	
	1	2
group		
1	82.7%	1.3%
2	4.4%	11.6%

The overall classification error is around 5%, which can be considered satisfactory. However, the percentage of missclassifications for the second group is around 30%, due to its low abundance in the samples.

6.3. Estimation of individual distances

To test the distance estimator defined in Sect. 4, two samples of 800 stars have been generated using the parameters of group 1 given in the previous section and the *base values* of the observational errors. In Fig. 5 the real distances

of the simulated stars (generated during the process of the sample's simulation) are compared with their estimations.

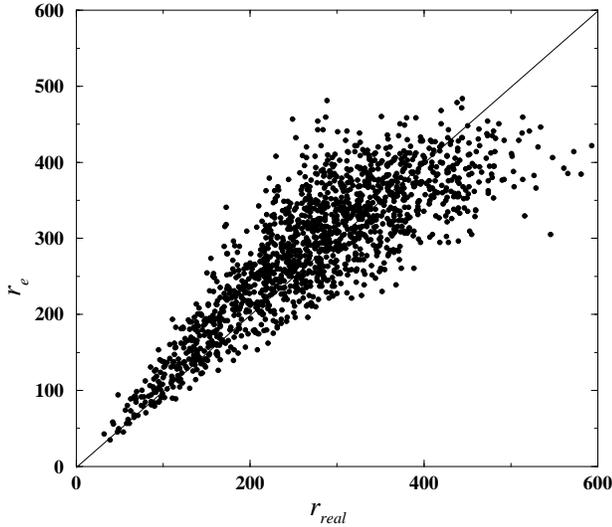


Fig. 5. Real distance (r) vs. estimated distance (r_e) for two samples of one-group K0 III stars

The mean value of the difference ($r - r_e$) is -11 pc, showing that the estimation has no significant bias. Notice, however, that the error distribution (or, equivalently, the distribution around the line $r = r_e$) is not symmetrical. This asymmetry comes from the properties of the density law function of the samples, and in this case is due to the limitation in apparent magnitude. For more details see Luri (1995). Finally, the mean error of the estimations is 16.7%, which can be considered satisfactory.

7. Conclusions

As shown by the tests presented in previous section, the ML method presented in this paper correctly takes into account the effects of sample selection, observational errors, interstellar absorption and galactic rotation, providing bias-free estimations of the parameters of the sample, in particular its luminosity calibration. In addition, it is able to identify and separate physically distinct groups of stars mixed in a sample, to classify its stars and to provide non-biased estimations of individual distances (including an individual estimation of its error). Several preliminary applications to real samples of stars extracted from the *Hipparcos Input Catalogue*, – Turon et al. (1992) – can be found in Luri (1995).

The resulting method can take full profit of the high quality data of the Hipparcos mission, which will be soon available. In particular, our team is engaged in a program

to recalibrate the HR diagram using the preliminary release of these data.

Acknowledgements. We wish to thank Ana Gómez and Frederic Arenou of the *Observatoire de Meudon* for their great help and useful suggestions. This work was supported by the CICYT under contracts ESP94-1311-E and ESP95-0180, and by the PICASSO program-AI 93135.

A. Appendix: example of detailed equations

The detailed expression of the density law $\mathcal{O}(z | \theta)$ – as defined in Sect. 2 – for the case of a magnitude-limited sample is given in Eq. (A1),

$$\mathcal{O}(z | \theta) = \mathcal{C}_{\mathcal{O}}^{-1} \mathcal{S}(z) \int_0^{\infty} \mathcal{F}_M \mathcal{F}_r \mathcal{F}_V dr, \quad (\text{A1})$$

where

$$\mathcal{S}(z) = \Theta(m - m_{\text{lim}}) \quad (\text{A2})$$

$$\mathcal{F}_M = e^{-\frac{1}{2} \left(\frac{M - M_0}{\sigma_M} \right)^2} \quad (\text{A3})$$

$$M = m + 5 - 5 \log_{10}(r) - A_v(r, l, b) \quad (\text{A4})$$

$$\mathcal{F}_r = e^{-\frac{|r \sin(b)|}{z_0}} k^2 r^4 \cos^2(b) \quad (\text{A5})$$

$$\mathcal{F}_V = \sqrt{-\frac{\pi^3}{J_2 E_1 X}} e^{\left(\frac{Y^2}{4X} - Z\right)} \quad (\text{A6})$$

$$X = \frac{D_1^2}{4E_1} - F_1 \quad (\text{A7})$$

$$Y = \frac{B_1 D_1}{2E_1} - C_1 \quad (\text{A8})$$

$$Z = \frac{B_1^2}{4E_1} - A_1 \quad (\text{A9})$$

$$A_1 = \frac{A_2^2}{4J_2} - D_2 \quad (\text{A10})$$

$$B_1 = \frac{A_2 B_2}{2J_2} - E_2 \quad (\text{A11})$$

$$C_1 = \frac{A_2 C_2}{2J_2} - F_2 \quad (\text{A12})$$

$$D_1 = \frac{B_2 C_2}{2J_2} - G_2 \quad (\text{A13})$$

$$E_1 = \frac{B_2^2}{4J_2} - H_2 \quad (\text{A14})$$

$$F_1 = \frac{C_2^2}{4J_2} - I_2 \quad (\text{A15})$$

$$A_2 = -\frac{\mu_l}{\epsilon_{\mu l}^2} - \left(\frac{a_1 U'_0}{\sigma_U^2} + \frac{a_2 V'_0}{\sigma_V^2} + \frac{a_3 W_0}{\sigma_W^2} \right) r \quad (\text{A16})$$

$$B_2 = \left(\frac{a_1 b_1}{\sigma_U^2} + \frac{a_2 b_2}{\sigma_V^2} + \frac{a_3 b_3}{\sigma_W^2} \right) r^2 \quad (\text{A17})$$

$$C_2 = \left(\frac{a_1 c_1}{\sigma_U^2} + \frac{a_2 c_2}{\sigma_V^2} + \frac{a_3 c_3}{\sigma_W^2} \right) r \quad (\text{A18})$$

$$D_2 = \frac{1}{2} \left(\frac{\mu_l^2}{\epsilon_{\mu l}^2} + \frac{\mu_b^2}{\epsilon_{\mu b}^2} + \frac{v_r^2}{\epsilon_{v_r}^2} + \frac{U_0'^2}{\sigma_U^2} + \frac{V_0'^2}{\sigma_V^2} + \frac{W_0^2}{\sigma_W^2} \right) \quad (\text{A19})$$

$$E_2 = -\frac{\mu_b}{\epsilon_{\mu b}^2} - \left(\frac{b_1 U'_0}{\sigma_U^2} + \frac{b_2 V'_0}{\sigma_V^2} + \frac{b_3 W_0}{\sigma_W^2} \right) r \quad (\text{A20})$$

$$F_2 = -\frac{v_r}{\epsilon_{v_r}^2} - \left(\frac{c_1 U'_0}{\sigma_U^2} + \frac{c_2 V'_0}{\sigma_V^2} + \frac{c_3 W_0}{\sigma_W^2} \right) \quad (\text{A21})$$

$$G_2 = \left(\frac{b_1 c_1}{\sigma_U^2} + \frac{b_2 c_2}{\sigma_V^2} + \frac{b_3 c_3}{\sigma_W^2} \right) r \quad (\text{A22})$$

$$H_2 = \frac{1}{2} \frac{1}{\epsilon_{\mu b}^2} + \frac{1}{2} \left(\frac{b_1^2}{\sigma_U^2} + \frac{b_2^2}{\sigma_V^2} + \frac{b_3^2}{\sigma_W^2} \right) r^2 \quad (\text{A23})$$

$$I_2 = \frac{1}{2} \frac{1}{\epsilon_{v_r}^2} + \frac{1}{2} \left(\frac{c_1^2}{\sigma_U^2} + \frac{c_2^2}{\sigma_V^2} + \frac{c_3^2}{\sigma_W^2} \right) \quad (\text{A24})$$

$$J_2 = \frac{1}{2} \frac{1}{\epsilon_{\mu l}^2} + \frac{1}{2} \left(\frac{a_1^2}{\sigma_U^2} + \frac{a_2^2}{\sigma_V^2} + \frac{a_3^2}{\sigma_W^2} \right) r^2 \quad (\text{A25})$$

$$a_1 = -k \cos(b) \sin(l) \quad (\text{A26})$$

$$b_1 = -k \sin(b) \cos(l) \quad (\text{A27})$$

$$c_1 = \cos(l) \cos(b) \quad (\text{A28})$$

$$a_2 = k \cos(b) \cos(l) \quad (\text{A29})$$

$$b_2 = -k \sin(b) \sin(l) \quad (\text{A30})$$

$$c_2 = \sin(l) \cos(b) \quad (\text{A31})$$

$$a_3 = 0 \quad (\text{A32})$$

$$b_3 = k \cos(b) \quad (\text{A33})$$

$$c_3 = \sin(b) \quad (\text{A34})$$

And the detailed expression of the normalisation constant \mathcal{C}_O is given in Eq. (A35),

$$\mathcal{C}_O = \mathcal{K} \sum_{i=1}^{199} \left[\Delta l^{(i)} \int_{\chi_{\text{lim}}} \text{cferr}(-\chi_{\text{lim}}^{(i)}) I_b^{(i)}(r) r dr \right], \quad (\text{A35})$$

where

$$\mathcal{K} = \frac{1}{2} (2\pi)^{7/2} \epsilon_{\mu l} \epsilon_{\mu b} \epsilon_{v_r} \sigma_U \sigma_V \sigma_W \sigma_M Z_0 \quad (\text{A36})$$

$$\Delta l^{(i)} = l_{\text{max}}^{(i)} - l_{\text{min}}^{(i)} \quad (\text{A37})$$

$$\chi_{\text{lim}}^{(i)} = \frac{m_{\text{lim}} + 5 - 5 \log_{10}(r) - A_v^{(i)}(r) - M_0}{\sqrt{2} \sigma_M} \quad (\text{A38})$$

$$I_b^{(i)}(r) = \begin{cases} e^{-\frac{|r \sin(b_{\text{min}}^{(i)})|}{Z_0}} - e^{-\frac{|r \sin(b_{\text{max}}^{(i)})|}{Z_0}} & \text{if } b_{\text{min}}^{(i)}, b_{\text{max}}^{(i)} > 0 \\ 2 - e^{-\frac{|r \sin(b_{\text{min}}^{(i)})|}{Z_0}} - e^{-\frac{|r \sin(b_{\text{max}}^{(i)})|}{Z_0}} & \text{if } b_{\text{min}}^{(i)} < 0, b_{\text{max}}^{(i)} > 0 \\ e^{-\frac{|r \sin(b_{\text{max}}^{(i)})|}{Z_0}} - e^{-\frac{|r \sin(b_{\text{min}}^{(i)})|}{Z_0}} & \text{if } b_{\text{min}}^{(i)}, b_{\text{max}}^{(i)} < 0 \end{cases}, \quad (\text{A39})$$

being $(b_{\text{min}}^{(i)}, b_{\text{max}}^{(i)})$ and $(l_{\text{min}}^{(i)}, l_{\text{max}}^{(i)})$ the limits of the sky sectors in the model of Arenou et al. (1992).

References

- Arenou F., Grenon M., Gómez A.E., 1992, *A&A* 258, 104
 Clube S.V.M., Jones A.H.P., 1971, *MNRAS* 151, 231
 Egret D., Keenan P.C., Heck A., 1982, *A&A* 106, 115
 Gómez A., Luri X., 1992, "Hipparcos: une nouvelle donnée pour l'Astrométrie", (Comptes rendus de l'école de Goutelas, 1992), 369. In: Benest D., Froeschlé F. (eds). Éditions de la S.F.S.A.
 Heck A., 1975, Ph. D. Thesis, Univ. of Liege
 Jung J., 1970, *A&A* 4, 53
 Jaschek C., Gómez A., 1985, *A&A* 146, 387
 Kendall Sir M., Stuart A., 1979, "The advanced theory of statistics", Charles Griffin & Co. Ltd, Vol. II, 38
 Luri X., 1995, Ph. D. Thesis, Univ. of Barcelona
 Luri X., Mennessier M.O., Torra J., Figueras F., 1992, in "Distancia'92. Congrès International Sur l'Analyse en Distance", 123. In: Joly S. & Lecalve G. (eds.)
 Luri X., Mennessier M.O., Torra J., Figueras F., 1993, *A&A* 267, 305
 Malmquist K.G., 1936, *Stockholms Ob. Medd.*, No. 26
 Mihalas D., Binney J., 1981, "Galactic Astronomy", W.H. Freeman & Co., 2nd Ed.
 Parenago P.P., 1940, *Astron. Zn.* 17, 3
 Press W.H., Flannery B.P., Teukolsky S.A., Vetterling W.T., 1986, "Numerical Recipes (The Art of Scientific Computing)". Cambridge University Press
 Ratnatunga K.U., 1989, in "The Gravitational Force Perpendicular to the Galactic Plane". In: Philip A.G.D. & Lu P.K. (eds.). L. Davis Press New York
 Rigal J.L., 1958, *Bull. Astr. Paris* 22, 171
 Turon C., Crézé M., Egret D., et al., 1992, "The Hipparcos Input Catalogue", ESA SP-1136