Astronomy & Astrophysics

Large and small-scale structures of the local Galactic disc

A maximum entropy approach to the stellar velocity distribution

R. Cubarsi

Dept. Matemàtica Aplicada IV, Universitat Politècnica de Catalunya, 08034 Barcelona, Catalonia, Spain
e-mail: rcubarsi@mat.upc.edu

ABSTRACT

An analytical model based on the maximum entropy approach is proposed to describe the eventual asymmetries of the velocity distribution, which are collected through its sample moments. If an extended set of moments is available, the current method provides a linear algorithm, associated with a Gramian system of equations, that leads to a fast and suitable estimation of the velocity distribution. In particular, it could be used to model multimodal distributions that cannot be described through Gaussian mixtures. The method is used with several samples from the HIPPARCOS and Geneva-Copenhagen survey catalogues. For the large-scale distribution, the phase density function may be obtained by fitting moments up to sixth order as a product of two exponential functions, one giving a background ellipsoidal shape of the distribution and the other accounting for the skewness and for the slight shift in the ellipsoidal isocontours in terms of the rotation velocity. The small-scale distribution can be deduced from truncated distributions, such as velocity-bounded samples with $|V| \leq 51$ km s^{-1}, which contain a complex mixture of early-type and young disc stars. By fitting up to ten-order moments, the maximum entropy approach gives a realistic portrait of actual asymmetries, showing a clear bimodal pattern: (i) around the Hyades-Pleiades stream, with negative radial mean velocity and (ii) around the Sirius-UMa stream, with slightly positive radial mean velocity. Among metallicity, colour, and other star properties, the eccentricity of the star's orbit behaves as a very good sampling parameter to find a more detailed structure for the disc velocity distribution, allowing distinctions between different eccentricity layers. For subsamples with eccentricities $e < 0.15$, star velocities are approximately symmetrically distributed around the LSR in the radial direction, with a dearth of stars at the LSR. For $e = 0.15$, the core distribution of the thin disc is supported by two major stellar groups with opposite radial velocities. Several simulations confirm that such a double-peaked distribution comes from the lognormal distribution of the velocity amplitudes. For maximum eccentricity 0.3 and maximum distance to the Galactic plane 0.5 kpc a representative thin disc sample is obtained. The "U-anomaly" along the radial direction is estimated straightforwardly 30−35 km s^{-1} from the contour plots. An explanation of the apparent vertex deviation of the disc from the swinging of those major kinematic groups around the LSR is then possible, which predicts a continuously changing orientation of the disc's pseudo ellipsoid.

Key words. stars: kinematics and dynamics – Galaxy: kinematics and dynamics – galaxies: statistics – methods: statistical

1. Introduction

The asymmetry of the local velocity distribution was first studied in 1905 by Kapteyn in his theory of two star streams and further developed by Kapteyn (1922), Strömberg (1925), and Charlier (1926), which considered up to fourth moments of the velocity distribution. However, those moments were not determined with a sufficient degree of accuracy up to Erickson (1975). During the past decade, higher order velocity moments with better precision could be obtained from large and representative stellar samples of the solar neighbourhood, like the HIPPARCOS catalogue (ESA 1997) or, more recently, the Geneva-Copenhagen Survey (GCS) (Nordtröm et al. 2004), accounting for velocity discontinuities and kinematic populations in the solar neighbourhood (Cubarsi & Alcobé 2004; Alcobé & Cubarsi 2005). Several approaches have been tried to describe the asymmetry of the velocity distribution. In the beginning, an anisotropic velocity distribution was obtained by superposition of isotropic phase-density functions with different means. Later, the Schwarzschild distribution, based on a single trivariate Gaussian distribution, could easily handle the basic anisotropic features, and more parameters could be controlled by assuming non-Gaussian ellipsoidal distributions. However, to account for non-null, odd-order

central moments, it was once again necessary to return to mixture models.

In addition to the works describing the actual velocity distribution from a mixture of stellar populations (e.g. Soubiran & Girard 2005; Vallerani et al. 2006), there is a wide variety of approaches that generally do not make use of the velocity moments, such as the two- or three-integral models based in Fricke (1952) components (Evans et al. 1997; Famaey et al. 2002; Jiang & Ossipkov 2007) or even a combination of a Gaussian part of the density function with a perturbation factor expressed in a polynomial form in terms of the integrals of motion (van der Marel & Franx 1993; Gerhard 1993; Kormendy et al. 1998). The velocity distribution is sometimes numerically estimated (Dehnen 1998; Skuljan et al. 1999; Bovy et al. 2009), although it is also frequently the analytical modelling (Famaey et al. 2005; Veltz et al. 2008). However, in the latter case, according to today's observational data, some intricate trivariate distribution functions (or with a very high number of components) may be obtained. In most of these works, there is the job of describing the detailed structure of the velocity distribution, or of associating specific moving groups with the density function components, although in most cases the small groups do not have a clear visual impact on the overall density function.

There is also a desire for a simple, qualitative description of the distribution in terms of basic measures of spread or asymmetry like the skew or for a comparison to Gaussian distributions, like the curtosis. For trivariate distributions with strong asymmetries, e.g. the structure that lies under the groups of young and early-type stars, the statistical moments are the natural tool for such a description of the basic geometric trends. To this purpose, the method of moments is revisited here.

An alternative analytical model based on the maximum entropy approach is proposed to describe the eventual asymmetries of the velocity distribution, which are collected through its sample moments. Even though such an approach has been widely used to solve many univariate technical and scientific problems, to my knowledge there has been no general application to stellar kinematics. There are several numeric algorithms for estimating the maximum entropy density function, which are not computationally trivial for the trivariate case. However, if an extended set of moments is available, the method described in this work allows a parameter estimation by solving a linear system of equations. Its simplicity makes it worthwhile using it to construct any ad hoc velocity distribution function.

The maximum entropy approach will be used to describe the main kinematical features of solar neighbourhood stars by working from the two formerly mentioned large and kinematically representative local stellar samples. In the first case, the large-scale distribution of the local disc is inferred from Sample I (Cubarsi & Alcobé 2004), obtained by crossing the HIPPARCOS Catalogue with radial velocities from the HIPPARCOS Input Catalogue (ESA 1992). In the second case, the method is applied to Sample II from the GCS catalogue. It has new and more accurate radial velocity data than the HIPPARCOS sample, and contains the total velocity space of F and G dwarf stars, which are considered the favourite tracer populations of the history of the disc. In both cases, the largest samples providing stable velocity moments are used. The preceding applications provide and confirm some general and well known trends in the background velocity distribution, such as the overall vertex deviation, the skewness, or the symmetry plane of the distribution. These stellar samples, which mainly contain thin and thick disc stars, can be sufficiently described from an exponential density function with a four-degree polynomial, although a six-degree polynomial provides a more accurate portrait of the local velocity distribution. According to the maximum entropy modelling, it is possible to interpret the velocity distribution as a product of two exponential functions, the one giving a background ellipsoidal shape of the distribution and the other, which is even and at least quadratic in the rotation velocity alone, acting as a perturbation factor that breaks the distribution symmetry.

On the other hand, the small-scale velocity distribution of the local disc can be deduced from truncated distributions. According to Alcobé & Cubarsi (2005), hereafter Paper I, a selection of stars with an absolute value of the total space motion $|V| \leq 51$ km s^{-1} leaves the older disc stars aside. Such a selection is analysed in more depth and their properties described better. It contains a complex mixture of early-type and young disc stars for which a Gaussian mixture approach is not feasible. Thus, Sample III is built as a subsample of Sample I with $|V| \leq 51$ km s^{-1}. Finally, Sample IV is drawn from Sample II under the same condition on the absolute velocity.

It is also possible to obtain a more detailed structure of the velocity distribution for specific subsamples, allowing the results of our approach to be compared with the small-scale structure sustained by moving groups. Among metallicity, colour, and other star properties, the eccentricity of the star's orbit is found

to behave as a very good sampling parameter that allows distinguishing between different *eccentricity layers* within the thin disc, and allowing visualisation of the underlying structure of the distribution. In particular, for maximum eccentricity 0.3 and maximum distance to the Galactic plane 0.5 kpc, we get a representative thin disc sample.

For these truncated distributions, the density function needs a six-degree polynomial to describe their strong asymmetries and their main kinematic features. The improvement in the GCS catalogue over the HIPPARCOS catalogue provides a higher resolution contour plot for the inner thin disc, which in addition to describing a velocity distribution far from the ellipsoidal hypothesis, explains a clear bimodal structure. Therefore, the maximum entropy modelling can be presented as an alternative way instead of mixture models.

The paper is organised as follows. In Sect. 2 the notation is introduced while reviewing some basic concepts of stellar statistics. Maximum entropy density functions are introduced and the meaning they have in mathematical statistics and statistical mechanics discussed. The mathematical formulation of the current functional approach is developed in Appendix A. In Sect. 3 the method is applied to local stellar samples from the HIPPARCOS and the GCS catalogues, for either complete (large-scale) or truncated (small-scale) distributions. Some aspects of the results and of the subsamples are analysed. In Sect. 4 a link to a dynamical model allows interpretation of the major kinematical groups sustaining the disc structure and of its possible swinging vertex deviation. Finally, in Sect. 5, the conclusions are presented.

2. Method

We study the necessary complexity of the velocity distribution for satisfying a set of moment constraints. The current approach simplifies both analytical dependence and parameter estimation of the distribution function under the following circumstances. We choose a density function maximising Shannon's information entropy. The maximum entropy approach to the solution of inverse problems was introduced long ago by Jaynes (1957a,b), so that it provides a unique solution that is the best one for not having to deal with missing information. It agrees with what is known, but expresses maximum uncertainty with respect to all other matters. It is a flexible and powerful tool for density approximation, which collects a complete family of generalised exponential distributions, including the exponential, normal, lognormal, gamma, and beta as special cases. Other properties of maximum entropy distributions are outlined in Appendix A.

An interesting application of the maximum entropy approach is the problem of moments (Mead & Papanicolaou 1984), which is described along with introducing the notation accordingly to the astronomical formulation.

2.1. Stellar statistics

For fixed values of time t and position $\boldsymbol{r}$, the macroscopic properties of a stellar system can be described from the moments of the distribution, which provide indirect information on the phase-space density function $f(t, \boldsymbol{r}, \boldsymbol{V})$, which is normalised with regard to the velocities. It is well known that the first moments, accounting for the mean, give the more elementary property of the distribution; the second central moments describe how much the distribution is spread around the mean; the third moments describe distribution asymmetries like the skewness; the fourth moments are used to quantify how peaked the distribution is;

and so forth (e.g. Stuart & Ord 1987). In general, the symmetric tensor of the nth-order, non-centred trivariate moments is obtained from the expected value

$$\boldsymbol{m}_n(t, \boldsymbol{r}) = \langle(\boldsymbol{V})^n\rangle \equiv \int_{\Gamma_V} (\boldsymbol{V})^n f(t, \boldsymbol{r}, \boldsymbol{V})\,\mathrm{d}\boldsymbol{V}, \quad n \geq 0 \qquad (1)$$

where $(\cdot)^n$ stands for the nth-tensor power, and Γ_V the velocity domain. The tensor $\boldsymbol{m}_n$ then has $\binom{n+2}{2}$ different elements according to the expression

$$m_{i_1 i_2 \ldots i_n} = \langle V_{i_1} V_{i_2} \ldots V_{i_n}\rangle, \qquad (2)$$

so that the indices belong to the set $\{1, 2, 3\}$, depending on the velocity component. Sometimes, instead of the component notation, namely Latin indices, it is also used a notation to make the velocity powers explicit, namely Greek indices, according to

$$m_{\alpha\beta\gamma} = \langle V_1^\alpha V_2^\beta V_3^\gamma\rangle. \qquad (3)$$

Obviously, $m_0 = 1$ and $\boldsymbol{m}_1 = \boldsymbol{v}(t, \boldsymbol{r})$, which is the mean velocity, or velocity of the centroid.

In a similar way, the symmetric tensor of the nth-order centred moments is obtained by working from the peculiar velocity

$$\boldsymbol{u} = \boldsymbol{V} - \boldsymbol{v}(t, \boldsymbol{r}), \qquad (4)$$

defined as $\boldsymbol{\mu}_n = \langle(\boldsymbol{u})^n\rangle$, with elements $\mu_{i_1 i_2 \ldots i_n} = \langle u_{i_1} u_{i_2} \ldots u_{i_n}\rangle$. In this case, $\mu_0 = 1$ and $\boldsymbol{\mu}_1 = 0$.

Hereafter, when studying the velocity dependence of the distribution function from a statistical viewpoint, the variables of time and position are omitted, although they might be used in the framework of a dynamical model for the whole phase-space distribution function.

Ellipsoidal distributions, such as the Schwarzschild distribution, can be described in terms of their central second moments μ_{ij}, which sometimes are written with Latin indices, such as $\sigma_{ij}^2 = \langle V_i V_j\rangle - \langle V_i\rangle\langle V_j\rangle$ (e.g. Binney & Tremaine 1987, pp. 194–211). However, in other standard astronomy reference books, the Greek index notation is used (e.g. Gilmore et al. 1989, p. 135–138), in particular when the velocity variables are expressed in the (U, V, W) coordinate system (without subindices), where the nth moments $m_{\alpha\beta\gamma}$ satisfy $\alpha + \beta + \gamma = n$. The second central moments account for the shape and orientation of the velocity ellipsoid and for the variance σ_l^2 of the velocity distribution function in an arbitrary direction l of the peculiar velocity space. According to the coordinate system, if c_1, c_2, and c_3 are the corresponding direction cosines, we have

$$\sigma_l^2 = \langle(c_1 u_1 + c_2 u_2 + c_3 u_3)^2\rangle = \sum_{i,j} c_i c_j \mu_{ij}; \; i, j \in \{1, 2, 3\}. \qquad (5)$$

The symmetric tensor $\boldsymbol{\mu}_2^{-1}$ (inverse of the second central moments $\boldsymbol{\mu}_2$) is then associated with the peculiar velocity ellipsoid

$$\boldsymbol{u}^{\mathrm{T}} \cdot \boldsymbol{\mu}_2^{-1} \cdot \boldsymbol{u} = 1, \qquad (6)$$

so that the velocity dispersions σ_1, σ_2, and σ_3 are the semiaxes of the ellipsoid that refers to the same coordinate axes. For these distributions, higher order central moments can be computed depending on the second ones; in other words, they cannot take arbitrary values, as shown in Appendix B. However, for arbitrary distributions, the variances and the velocity ellipsoid are meaningless, unless they could be used as a Gaussian approximation. Similarly, the skewness and the curtosis are also meaningless for multivariate distributions far from Gaussians, which have to be qualitatively described from moments of order higher than two, up to a sufficient degree of approximation of the basic distribution trends.

2.2. Maximum entropy

Therefore, more general and anisotropic distributions have a wider set of independent moments, and, in the more general case, the exact distribution may be univocally determined by the infinite hierarchy of independent moments. Provided an order for a set of moments (for example according to the Latin indices notation 0, 1, 2, 3, 11, 12, 13, 22, and so on) if the first m moments are known, it is possible to find an infinite variety of functions whose first m moments coincide with the above set. Various approximation procedures exist to find a sequence of functions f_m, which fulfils the foregoing moment constraints and converges to the true distribution as m approaches infinity. Fortunately, between those sequences of functions, a uniquely maximum entropy sequence exists that maximises the entropy functional

$$W(f_m) = -\int_{\Gamma_V} f_m(\boldsymbol{V}) \ln f_m(\boldsymbol{V})\,\mathrm{d}\boldsymbol{V}. \qquad (7)$$

Then, the maxima $f = f_m$ is usually called the least biased sequence of approximations, and, by using Lagrangian multipliers, it can be shown (e.g. Kagan et al. 1973) that it has the form

$$f(\boldsymbol{V}) = \mathrm{e}^{\mathcal{P}(\boldsymbol{V})}, \qquad (8)$$

where $\mathcal{P}(\boldsymbol{V})$ is a power series of the velocity components containing m terms, as many terms as the number of moment constraints, so that each coefficient is related to a single moment constraint. Then, the sequence f_m of maximum entropy functions, such as increasing m, is able to fit more complex and informative systems, so that the higher the number of moment constraints, the lower the entropy, the symmetry, and the algebraic simplicity of f_m.

The solution of the maximum entropy problem usually consists in solving a set of m nonlinear equations in the form

$$\boldsymbol{m}_k = \int_{\Gamma_V} (\boldsymbol{V})^k\,\mathrm{e}^{\mathcal{P}(\boldsymbol{V})}\,\mathrm{d}\boldsymbol{V}. \qquad (9)$$

However, these solution techniques are typically not easy to generalise to the multidimensional problem. On the other hand, even for the unidimensional problem, an analytical solution generally does not exist for higher than second moments. Generally, the numerical techniques for solving the coefficients of the polynomial $\mathcal{P}$ are based on nonlinear optimisation, Legendre transformation, etc. (e.g. de Bruin et al. 1999; Kouskoulas et al. 2004), and either way, they are not easy to implement. However, if an extended set of moments is known, then the parameters can be estimated by solving a linear system of equations, as explained in Appendix A.2. In the case of trivariate distributions, for a polynomial of degree n in three variables, it is necessary to compute moments up to order $2(n - 1)$.

Thus, the current purpose is to infer the trivariate velocity distribution from a finite set of moment constraints. To simplify estimation of the polynomial coefficients of $\mathcal{P}(\boldsymbol{V})$, an alternative method has been developed, based on a unique assumption that the velocity distribution satisfies the boundary conditions associated with the stellar hydrodynamic equations, also known as moment equations.

If the phase-space distribution function f satisfies the collisionless Boltzmann equation, $\frac{\mathrm{D}f}{\mathrm{D}t} = 0$[1], then by multiplying

---

[1] The Stokes operator $\frac{\mathrm{D}(\cdot)}{\mathrm{D}t}$ is generally used to simplify the notation of the Lagrangian derivative $\frac{\partial(\cdot)}{\partial t} + \boldsymbol{V} \cdot \nabla_{\boldsymbol{r}}(\cdot) + \dot{\boldsymbol{V}} \cdot \nabla_{\boldsymbol{V}}(\cdot)$.

it by the nth-tensor power of the star velocity and by integrating over the whole velocity space, the family of stellar hydrodynamic equations is obtained:

$$\int_{\Gamma_V} (V)^n \frac{\mathrm{D}f}{\mathrm{D}t} \, \mathrm{d}^3V = (0)^n, \quad n \geq 0. \tag{10}$$

In Cubarsi (2007), the above equations were derived in terms of the central or comoving moments, in a completely analytical way, for any order n and without any additional hypotheses. Then, if the above integrals exist and since there are no stars with velocity beyond Γ_V, the following boundary conditions were, as usual, assumed in the integration process:

$$V \to \partial\Gamma_V \implies (V)^n f(t, r, V) \to (0)^n, \ n \geq 0. \tag{11}$$

These boundary conditions are satisfied by a wide family of distributions that are bell-shaped in any direction of the velocity domain. One of the integral properties that was derived in Cubarsi (2007) will allow, in Appendix A.1, establishment of a Gramian system of equations for solving our estimation problem. From a purely statistical inference viewpoint, the requirement of estimating the distribution parameters is not that the phase density function is the solution to the collisionless Boltzmann equation, but it is enough that it satisfies, or approximately satisfies, the above boundary conditions.

The entropy functional $W(f)$, as defined in Eq. (7), is far from containing all the information about the Boltzmann equation (with or without collisions) since $W(f)$ only depends on the velocity space, similar to the collision operator of the complete Boltzmann equation. In the following section, we discuss how such a maximum entropy density function may or may not be a solution to the collisionless Boltzmann equation.

In review, two typical cases of maximum entropy distribution function are solutions to the whole set of moment equations. The simplest case is an isothermal velocity distribution of Maxwell type in the peculiar velocities, which according to the Maxwell-Boltzmann law, represents a system with the more basic thermal equilibrium

$$\psi(t, r, u) = \mathrm{e}^{-\frac{1}{2}\mu^{-1}|u|^2} \tag{12}$$

where $\mu(t, r) > 0$ is a continuous and differentiable function in both arguments accounting for the variance in the distribution, which is the only constraint. Therefore, the distribution is totally isotropic. It has equal diagonal second central moments, vanishing off-diagonal second moments, and zero odd-order moments as well.

Another well known example is the Schwarzschild distribution, that is, an exponential density function that depends on the peculiar velocities in a quadratic way (Chandrasekhar 1942),

$$\psi(t, r, u) = \mathrm{e}^{-\frac{1}{2}(Q+\sigma)}, \quad Q = u^{\mathrm{T}} \cdot A_2 \cdot u \tag{13}$$

where Q is a quadratic, positive-definite form, with $A_2(t, r)$ a second-rank symmetric tensor and $\sigma(t, r)$ a scalar function, which are continuous and differentiable in both arguments. As a result, the distribution is Gaussian in the peculiar velocities, although it is multiplied by an arbitrary function of time and position. In such a way, the quadratic form Q can explain three isolating integrals of star motions, and the distribution may have some different diagonal second central moments and non-vanishing off-diagonal moments, although the odd-order moments still vanish. Therefore it is a maximum entropy function constrained by the whole set of moments up to second order.

The above examples, which are integrable functions in an infinite velocity domain, satisfy the boundary conditions, Eq. (11), and can be generalised according to an exponential function, Eq. (8), with as many polynomial terms as available moment constraints, under the necessary conditions over the polynomial coefficients to obtain an integrable distribution function. For higher-degree polynomials, the distribution function is integrable if the polynomial is upper bounded, and therefore the polynomial must be even. On the other hand, truncated distributions, which are associated with velocity-bounded stellar samples, $|V - V_0| \leq$ const., have a finite velocity domain. Then the boundary conditions are still a good approximation if the truncated distribution vanishes enough when approaching the contour of the velocity domain, so that the density function may be assumed null out of this boundary. Thus, for a domain that is either bounded or unbounded, we assume that the velocity distribution is continuous, differentiable, and positive in the interior of the velocity domain Γ_V and that the boundary conditions are fulfiled in its contour $\partial\Gamma_V$.

2.3. Information entropy

Let us briefly explain how to interpret a maximum entropy density function, or better, what is the appropriate context for its use. Up to a change of sign, Shannon's information entropy is defined as the Boltzmann H-functional, which first appeared in statistical mechanics in works by Boltzmann and Gibbs in the 19th century. However, it is not exactly the same concept.

Boltzmann's functional is used for non-equilibrium systems and is related to the irreversibility of dynamical processes in a uniform gas. For elastic collisions involving short-range forces and in the absence of boundaries, mass, momentum, and energy are conserved in binary encounters (e.g. Cercignani 1988). They are usually referred to as collisional invariants. There is only one distribution function, the Maxwellian distribution, fulfiling all of the following properties: it depends on a linear combination of the collisional invariants, the collision term of the Boltzmann equation is exactly zero, and it minimises Boltzmann's entropy. This solution represents a local equilibrium state, in the sense that other solutions to the Boltzmann equation will become closer to it as the time goes by. Depending on the potential, boundary conditions, and dissipative or collision effects (e.g. Villani 2002), maximum entropy solutions can be non-Maxwellian.

Shannon's information entropy[2] was introduced in communication theory to measure the redundancy of a language and the maximal compression rate, which is applicable to a message without any loss of information. It is defined for complex systems and is related to Boltzmann's entropy as a measure of the number of microstates associated with a given macroscopic configuration. On the other hand, the Fisher information was introduced as part of his theory of coefficient statistics as a measure of the uncertainty. It is also related to Shannon's entropy, so that the entropy quantifies the variation of information. If we maximise the entropy subject to some constraints (e.g. statistics

---

[2] The quotation by Shannon, extracted from Martin & England (1981), is amusing: my greatest concern was how to call it. I thought of calling it "information". But the word was overly used, so I decided to call it "uncertainty". When I discussed it with John Von Neumann, he had a better idea. He told me: "you should call it entropy, for two reasons. In the first place your uncertainty has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage".

describing macroscopic properties) we get distributions containing maximum uncertainty that is compatible with these constraints.

For given mass and energy, the Fisher information takes its minimum value and Shannon's entropy its maximum value in the form of Maxwellian distributions. For a given covariance matrix, they take extreme values for Gaussian distributions. The number of constraints involved in the Lagrange multipliers may reach higher order moments, by reflecting more complex situations in which the stars interact with the potential and with themselves, as well as having different masses.

We quote Jaworsky ([1987](#)) to point out that these two typical viewpoints for interpreting the entropy as uncertainty. In mathematical statistics and information theory, the entropy functional is maximised by attending to some constraints that express any available information of a complex physical system, which depend on the actual experimental situation. In statistical mechanics the entropy is used to study the thermodynamic equilibrium or non-equilibrium of a physical system, generally a uniform gas, in terms of the mean values of some physical quantities, which describe the macroscopic state of a physical system as a whole, like energy or number of particles. Thus, statistical mechanics based on this principle can be interpreted as a special type of statistical inference. The use of higher order statistical moments in addition to the mean values represents a generalisation of the thermodynamic concept of entropy, which is used to approximate the exact probability distributions for a few specified random variables when a finite number of their moments is known.

The maximum entropy principle implies that the the resulting distribution belongs to the exponential family. The actual moment constraints are a direct consequence of the isolating integrals of the stellar motion, or more precisely, they reflect particular combinations of the isolating integrals that are conserved. More complex distributions exist than the Maxwellian, which are maximum entropy distributions and are solution of the collisionless Boltzmann equation. These solutions are generally obtained by assuming that Liouville's theorem is satisfied, so that the essential information about the density function is provided by the isolating integrals of the motion of the stars. Thus, if we assume that the polynomial form $\mathcal{P}$ of Eq. ([7](#)) depends on the integrals of motion and is itself an integral of motion, Liouville's theorem is equivalent to the collisionless Boltzmann approximation. Then, the collisionless Boltzmann equation obviously takes the form

$$\frac{\mathrm{d}f(\mathcal{P})}{\mathrm{d}\mathcal{P}} \frac{\mathrm{D}\mathcal{P}}{\mathrm{D}t} = 0 \tag{14}$$

so that the factor $\frac{\mathrm{d}f}{\mathrm{d}\mathcal{P}}$ accounts for the maximum entropy condition, and the factor $\frac{\mathrm{D}\mathcal{P}}{\mathrm{D}t}$ is, in fact, the collisionless Boltzmann condition. Thus, both conditions are independent and compatible. If the maximum entropy criterion is fulfiled, then the function $f(\mathcal{P}) = \mathrm{e}^{\mathcal{P}}$ takes the smoothest possible form, while the dependence of $\mathcal{P}$ in terms of the powers of the velocity, as well as in terms of time and position through its polynomial coefficients, is, in this approach, independent of the maximum entropy condition. Thus, we may affirm that the maximum entropy procedure is non-essential to the solution of the collisionless Boltzmann equation.

The physical mechanism providing such a maximum entropy function is irrelevant to the statistical approach. In contrast, what is important is the set of statistical moments accounting for the macroscopic state, which, of course, have a dynamical significance in terms of viscosity, conductivity, or diffusion effects. The present statistical approach adopts the opposite viewpoint of studying possible warming mechanisms that modify a Schwarzschild distribution, and to then test how the distribution fits the actual velocity moments (e.g. Dehnen [1999](#)). In the current method, the available information is condensed within the polynomial $\mathcal{P}(V)$. The maximum entropy approach then gives a very good mathematical estimation of the density function and of its velocity derivatives involved in $\frac{D\mathcal{P}}{Dt} = 0$, although it may or may not match any physical model. On the other hand, the maximum uncertainty in the light of the missing information is guaranteed by the function $f(\mathcal{P}) = \mathrm{e}^{\mathcal{P}}$.

The maximum entropy density function may be explicitly written as

$$f = \psi_n \equiv \mathrm{e}^{\mathcal{P}_n}$$
$$\mathcal{P}_n = \sum_{k=0}^{n} \sum_{\alpha+\beta+\gamma=k} \lambda_{\alpha\beta\gamma}(t, \boldsymbol{r}) \, V_1^\alpha V_2^\beta V_3^\gamma \tag{15}$$

where the subindex n does not represent the number of polynomial terms, but rather the maximum polynomial power.

If the velocity domain Γ_V is all the space R^3, the polynomial $\mathcal{P}_n$ must be upper bounded to satisfy the integrability conditions. As a result, the power series of the velocities reaches a natural value n, which must be *even*, and, for the highest degree $k = n$, the n-adic form $\sum_{\alpha+\beta+\gamma=n} \lambda_{\alpha\beta\gamma}(t, \boldsymbol{r}) V_1^\alpha V_2^\beta V_3^\gamma$ must be negative definite.

Equation ([15](#)), in addition to including Eqs. ([12](#)) and ([13](#)) as particular cases, also contains, in general, any desired type of two- or three-integral functions (e.g. Hénon [1973](#); Dejonghe [1983](#); White [1985](#)). It represents a general functional approach, in a similar way to Fricke ([1952](#)), with the difference that, while the distribution function in the Fricke-based models is either a linear combination or product of the powers of integrals of motion, in Eq. ([15](#)) the linear combination of powers of integrals of motion appears as the argument in the exponential function.

The mathematical formulation of the maximum entropy functional approach is detailed in Appendix A. The smoothest density function that is consistent with an extended set of moment constraints is provided by the Gramian system of equations in Eq. ([42](#)). The resulting system allows computation of the elements of tensors λ_k; $1 \leq k \leq n$ in terms of the velocity moments up to order $2(n - 1)$, which is the highest order involved in Eq. ([40](#)), as discussed in Appendix A.2. For the case $n = 2$, it is easy to solve the Gramian system in an analytical way and to find out how moments of order higher than two depend on the second ones (Appendix B). For higher values of n, however, it must be done by using the numerical procedure outlined in Appendix A.3. Also, for $n = 2$, the integrability of the distribution function in an infinite velocity domain is easily derived from the tensor λ_2, since $\lambda_{ij} = -\frac{1}{2}\mu_{ij}^{-1}$, where the tensor of second central moments μ_2 is positive-definite. For $n \geq 4$, it is impossible to guarantee the definiteness of the tensor λ_n in a general way. This is a problem related to Hilbert's 17th problem, which is obviously beyond the scope of the present work. However, by using a finite velocity domain, one as wide as needed, according to the working stellar sample, such a problem may be easily avoided for truncated distributions, as described in Appendix A.1.

3. Application

Several illustrations of the current functional approach are used to describe the main kinematical features of the solar neighbourhood. The first two cases give the whole velocity distribution
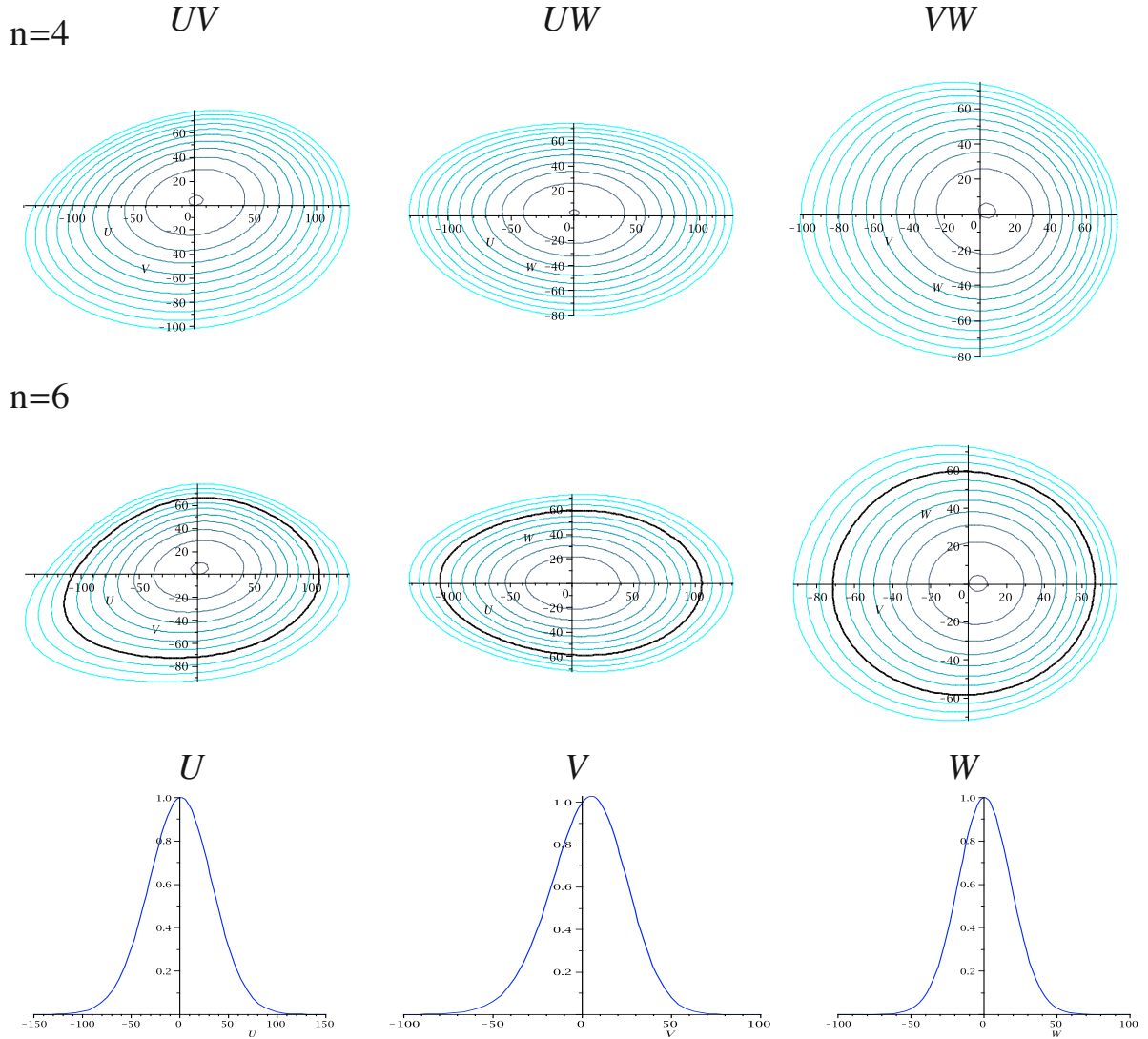
Fig. 1. Contour plots of the local velocity distribution in terms of the peculiar velocities for HIPPARCOS' Sample I and Sample I'. The plots are centred on the mean heliocentric velocity $(-10.85, -19.93, -7.49)$ km s^{-1} of Sample I', with radial velocity errors up to 2.5 km s^{-1}. The case $n = 4$, by fitting up to sixth moments, leads to more realistic contour plots than a pure ellipsoidal distribution ($n = 2$), although $n = 6$ provides a slightly improvement, by fitting up to tenth moments. The contours indicate levels $(\frac{1}{2})^k$; $k = 0, ..., 10$, and the black contour line corresponds to an approximate level 10^{-2} surrounding nearly the whole distribution, as confirmed in the last row, also for $n = 6$, from the sections of the velocity density function (not normalised) in each velocity component.

of the local disc, which is usually fitted by a mixture of trivariate Gaussian distributions. In the first application, a nearly complete and kinematically representative local sample, Sample I (Cubarsi & Alcobé 2004) with 13 678 stars, is used. It was obtained by crossing the HIPPARCOS Catalogue (ESA 1997) with radial velocities from the HIPPARCOS Input Catalogue (ESA 1992). To get a representative sample of the solar neighbourhood, it was limited to a trigonometric distance of 300 pc, where the only input data points were the velocity components (U, V, W) in a cartesian heliocentric coordinate system, with U toward the Galactic centre, V in the rotational direction, and W perpendicular to Galactic plane, positive in the direction of the North Galactic pole. In Paper I it was found that the optimal subsample containing thin and thick disc stars could be obtained by selecting stars with absolute heliocentric velocity $|V| \leq 210$ km s^{-1}. The sample has undergone a deeper statistical analysis in Cubarsi et al. (2010), hereafter Paper II, where the subsamples selected from $|V| > 400$ km s^{-1} contain, in addition to above disc populations, a fraction up to 1% of halo

stars, with very stable computed moments. To compare results with the next sample, the velocity domain is limited to the absolute space velocity of 500 km s^{-1}, which only excludes five stars from the whole sample. The resulting Sample I is then composed of 13 673 stars. The distribution smoothly vanishes in reaching the velocity boundary, as shown in the last row of Fig. 1. For all practical purposes, the sample may be considered as unbounded. To compare between different fittings, Eq. (42) up to $n = 6$ is used, by taking up to tenth moments into account. According to Appendix A.3, by normalising to the number N of equations, the squared error of the fit is then given by the expression

$$\chi^2 = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\varepsilon_i^2} \left| y_i - g_{ij} x_j \right|^2. \tag{16}$$

The maximum entropy procedure with $n = 2$ tries to represent the whole distribution from an unique ellipsoidal distribution. Thus, odd-order moments and even-order moments higher than four are not fitted. The resulting fitting error $\chi^2 = 83.93$ is not acceptable. Let us point out, however, that more than the value χ^2

itself, which is more significant is the increase or decrease in this quantity, since it may become distorted because over or under-estimating observational errors, or undesired error distribution. The approach with $n = 4$ by using up to sixth moments gives a clearly improved result $\chi^2 = 1.05$. A symmetric distribution around the plane $W = 0$ is also quite admissible, with $\chi^2 = 2.85$, which is the same order as the previous error. The approach with $n = 6$ is also computed by fitting up to the tenth moments, which gives a very accurate fit, $\chi^2 = 0.02$, even with symmetry around the plane $W = 0$, $\chi^2 = 0.62$. The contour plots of the velocity distribution on each velocity plane are displayed in Fig. 1, as well as the bell-shaped sections of the density function in each velocity component. The coordinate system is centred in the mean velocity of the sample $(-10.83, -20.47, -7.32)$ km s^{-1}, referring to the Sun.

In a second example, the method is applied to Sample II, drawn from the GCS catalogue (Nordtröm et al. 2004; Holmberg et al. 2007). It has new and more accurate radial velocity data than the HIPPARCOS sample and contains the total velocity space of 13 240 F and G dwarf stars, which are considered the favourite tracer populations of the history of the disc. According to the authors, the main essential features of the sample are the lack of kinematic selection bias and the radial velocity data, which allowed to reject stars that have not taken part in the evolution of the local disc. The same cartesian heliocentric coordinate system is used. For this sample, according to the analysis in Paper II, the moments are computationally stable for all the velocity components in the range $400 \leq |V| \leq 500$ km s^{-1}. The limitation up to an absolute velocity space of 500 km s^{-1} excludes five stars. The halo component is present in the total sample in a fraction less than 0.5%. Therefore, for practical purposes, this sample may also be considered as unbounded. The results and the graphs are similar to those of Sample I.

In the next section we confirm that the GCS sample provides more accurate moments of the disc velocity distribution than HIPPARCOS' due to its more precise radial velocities. In Table 1, centred and non-centred velocity moments up to order four are listed for the GCS Sample II, along with their standard errors. These are moments for a mixture of thin disc (94%), tick disc (5.5%), and halo (0.5%), as discussed in Paper II. They allow some measures of spread and asymmetry of the distribution in the desired variables to be computed, as the non null skewness in the rotation velocity V, $\gamma_V = \mu_{030} \mu_{020}^{-\frac{3}{2}} = -3 \pm 0.5$ (in the Greek indices notation), which is zero in the other components. The moments also lead to a non-significant curtosis in the vertical velocity W, $c_W = \mu_{004} \mu_{002}^{-2} - 3 = 34 \pm 40$, or the non-vanishing vertex deviation on the UV plane $\delta = \frac{1}{2} \arctan[2\mu_{110} (\mu_{200} - \mu_{020})^{-1}] = 11.3° \pm 2.3$.

Therefore, the main features of the maximum entropy distribution for Samples I and II, which show a reasonable deviation from an ellipsoidal distribution, may be easily deduced from Fig. 1, either for $n = 4$ or $n = 6$. The main features are:

(i) the velocity distribution is not symmetric around the mean, mainly in the rotation direction;
(ii) the whole distribution has a clear vertex deviation on the plane UV and no deviation on other planes;
(iii) there is some skewness in the variable V. As a consequence of both previous situations there is a wider distribution wing towards lower U and V velocities, which is likely caused by thick disc stars;
(iv) the curtosis in the W variable vanishes and is zero or very small in the U velocity (see Table 3);
(v) the plane $W = 0$ is basically a symmetry plane.

Table 1. Centred moments $\mu_{\alpha\beta\gamma}$ and non-centred moments $m_{\alpha\beta\gamma}$ with their standard errors up to fourth order for the GCS Sample II.

$\alpha\beta\gamma$	$\mu_{\alpha\beta\gamma}$	$\Delta\mu_{\alpha\beta\gamma}$	$m_{\alpha\beta\gamma}$	$\Delta m_{\alpha\beta\gamma}$
100	0.00	±0.30	−9.97	±0.30
010	0.00	0.22	−18.59	0.22
001	0.00	0.16	−7.16	0.16
200	1205.49	±30.08	1304.84	±30.94
110	114.33	20.02	299.63	22.96
020	657.33	28.67	1002.95	34.28
101	−28.26	11.81	43.09	12.32
011	14.09	10.97	147.17	12.65
002	332.93	17.33	384.17	17.82
300	−5410.36	±6929.41	−42 446.74	±7693.93
210	−29 820.51	4514.01	−56 357.61	5178.37
120	−1883.35	3918.62	−16 130.85	4722.18
030	−50 726.61	6922.64	−93 813.02	8810.45
201	−1529.39	2007.04	−10 306.27	2143.30
111	1567.29	1482.29	−192.55	1670.70
021	−1486.53	2027.51	−9189.67	2515.32
102	−2581.18	2450.85	−6005.75	2637.40
012	−9440.44	1724.77	−16 784.25	2090.32
003	−5661.87	5733.57	−13 178.21	6024.02
400	13 430 320.65	±2 125 166.40	14 374 454.53	±2 291 035.07
310	1 460 621.61	1 270 588.16	3 175 498.30	1 469 075.82
220	5 317 165.76	948 783.69	7 064 506.56	1 153 512.90
130	322 235.30	930 673.62	1 480 869.49	1 171 917.48
040	11 308 141.98	1 962 964.12	16 562 919.85	2 562 819.68
301	−1 192 280.95	478081.09	−851 132.23	485 712.04
211	62 741.79	322 480.47	454 274.09	354 935.98
121	−315 560.92	289 850.40	−248 097.44	344 247.93
031	234 834.16	517 841.17	1 003 877.53	647 019.02
202	1 845 208.39	421 200.10	2 010 427.12	459 870.47
112	315 298.09	262 885.04	506 475.65	323 652.27
022	1 593 831.34	251 847.69	2 140 083.38	327 148.87
103	139 487.68	845 660.69	321 923.19	926 678.48
013	492 111.63	552 721.46	941 999.67	688 288.39
004	4 084 082.40	2 190 306.74	4 351 176.92	2 325 987.63

The resulting density function, according to the most significant polynomial coefficients, can be expressed as a product of two exponential functions in the form

$$f = \varphi_1(Q) \, \varphi_2(h) \qquad (17)$$

where Q is a quadratic negative-definite form, which gives the background ellipsoidal shape of the distribution, with axis ratios 1:0.7:0.5, symmetry plane $W = 0$, as expected for disc stellar samples, and overall vertex deviation in the UV velocity components of about 12°. The function $\varphi_2(h)$ can be expressed in terms of the angular momentum integral h, and may be interpreted as a perturbation factor. It is even and is at least quadratic in the V velocity alone, which accounts for the skewness and the shift in the ellipsoidal isocontours in terms of the rotation velocity.

The bulk of the local velocity distribution does not show any substructure reflecting the existing moving groups, even by associating these moving groups with different proxy Gaussian components (Bovy et al. 2009). Then, it results in a smooth background distribution. However, by selecting specific subsamples by colour, or by using different analysis techniques where the resolution scale may vary, the substructures of the velocity distribution arise. We discuss it in the next section.
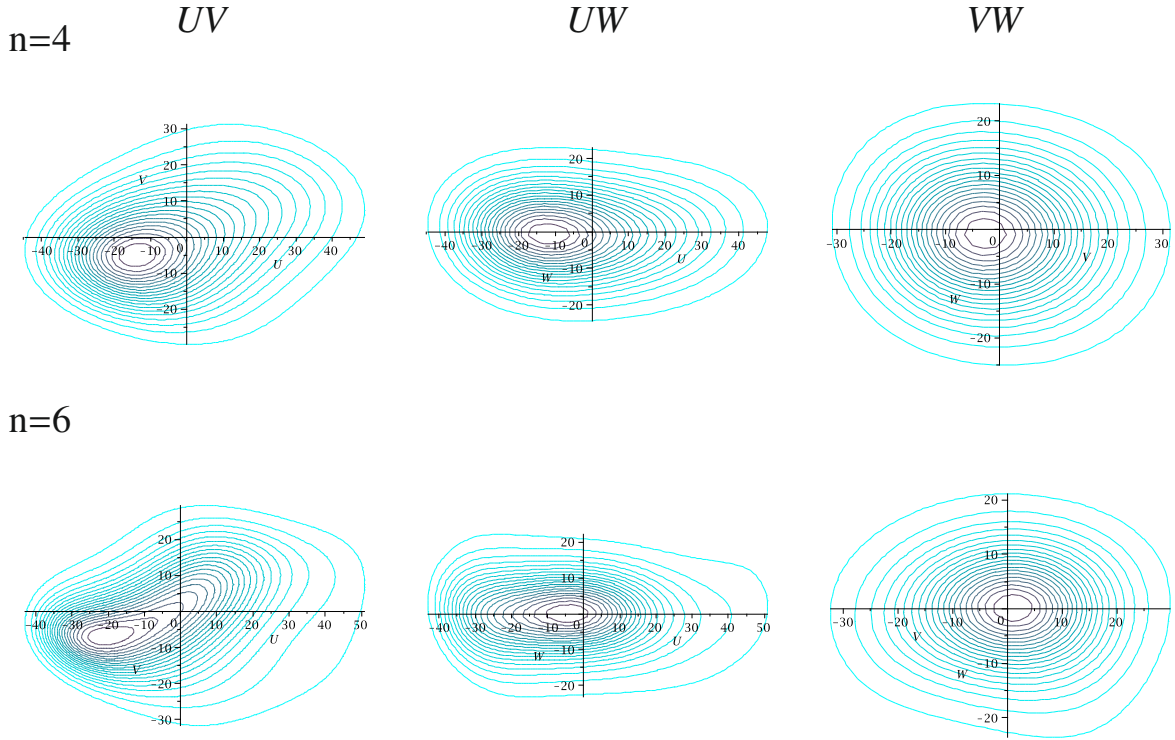
Fig. 2. Contour plots of the velocity distribution for Sample III, from the HIPPARCOS catalogue, for stars with $|V| \leq 51$ km s^{-1}. The peculiar velocities are centred on the heliocentric mean velocity $(-7.49, -11.25, -6.41)$ km s^{-1}. The approach $n = 4$ uses moments up to sixth order, and $n = 6$ up to tenth order. The asymmetry of the velocity distribution, mainly on the UV plane, may be sufficiently described in the case $n = 6$.

The next two examples are used for two new purposes: first, to test the ability of the maximum entropy method in reconstructing a truncated velocity distribution associated with a velocity bounded sample; second, to try a magnifying glass effect over the distribution and to focus on a specific velocity domain. According to Paper I, the selection of local stars with an absolute value of the total space motion $|V| \leq 51$ km s^{-1} had left the older disc stars aside, which are the originators of an important softening of the distribution. Such a selected group of stars contained a complex mixture of early-type and young disc stars for which a Gaussian mixture approach was unreliable because of the large fitting errors. This small-scale structure of the velocity distribution was strongly asymmetric in comparison to the background distribution. It was also observed in other analyses of the solar neighbourhood (e.g. Famaey et al. 2005; Soubiran & Girard 2005).

Sample III is then composed of 10 195 stars from the HIPPARCOS Sample I, with $|V| \leq 51$ km s^{-1}. The maximum entropy approach for $n = 2$ gives a fitting error $\chi^2 = 0.66$, according to Eq. (16). Although it could seem a very low value compared to previous samples, we might bear in mind that Samples I and II contain stars with higher velocity than Sample III, which increases the uncertainty of the computed moments. Because of this, the fitting errors for Sample III are expected to be much smaller. Once again, we must pay attention to the variation in χ^2.

For $n = 4$, the approach is able to provide a more realistic, non-ellipsoidal map of the truncated distribution by fitting moments up to sixth order. In this case the fitting error is $\chi^2 = 0.12$. Nevertheless, for $n = 6$, the maximum entropy approach gives a much improved portrait by fitting up to tenth moments. The fitting error $\chi^2 = 0.008$ is about 10^2 times lower than the ellipsoidal approach. The contour plots of the velocity distribution on each velocity plane are displayed in Fig. 2. The

coordinate system is centred in the heliocentric mean velocity $(-7.49, -11.25, -6.41)$ km s^{-1}.

Finally, Sample IV is drawn from the GCS catalogue by selecting 9733 stars with absolute velocity lower than 51 km s^{-1}. As in the above example, the approaches with $n = 2$ ($\chi^2 = 1.09$) and $n = 4$ ($\chi^2 = 0.29$) are not able to provide a realistic map of the truncated distribution. However, for $n = 6$, with a fitting error $\chi^2 = 0.002$, more than 10^2 lower than the case $n = 4$, the maximum entropy approach gives a detailed portrait of actual asymmetries, in particular on the UV plane. The contour plots of the velocity distribution on each velocity plane are displayed in Fig. 3. The coordinate system is centred in the mean heliocentric velocity $(-6.12, -11.23, -6.18)$ km s^{-1}. In Table 2, centred and non-centred velocity moments up to fourth order are listed as well with their standard errors. The skewness in the rotation velocity V is small, $\gamma_V = 0.23 \pm 0.03$, but non-zero, being similar in the U direction. The curtosis in the vertical velocity W, $c_W = 0.7 \pm 0.3$, is also very low. The vertex deviation on the UV plane is $\delta = 8.7° \pm 0.6$. Although it is caused by both subjacent structures, it is nearly the same as the one obtained in Paper II for the thin disc component. The results are summarised in Table 3.

The improvement of the GCS catalogue over the HIPPARCOS catalogue, mainly for the bounded sample, provides a higher resolution contour plot of the velocity distribution, which in addition to describing a velocity distribution far from the ellipsoidal hypothesis, shows a clear bimodal structure, as displayed in Fig. 4.

The results are consistent with the contour plots obtained by Dehnen (1998) when inferring the velocity distribution of his total sample (AL), in particular for the innermost dark contour. Also, the shape of the velocity distribution for early-type stars (Skuljan et al. 1999) is similar to ours, which is now
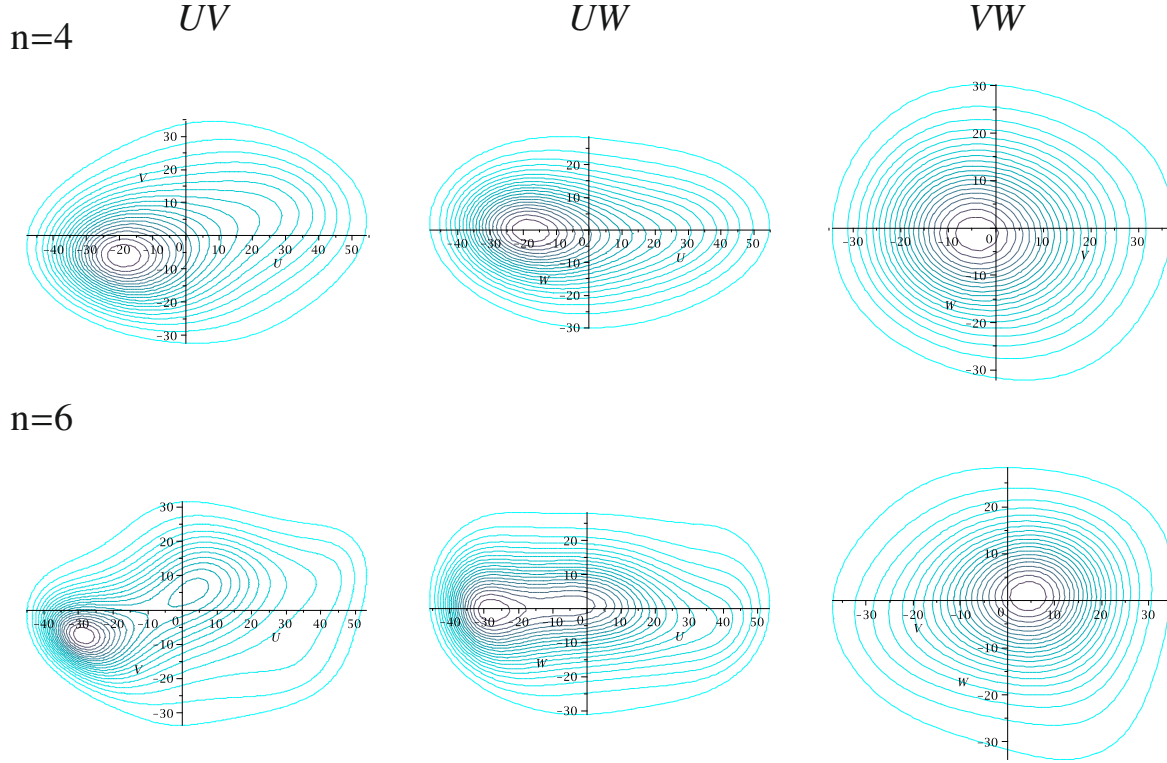
Fig. 3. Contour plots of the velocity distribution for Sample IV, from the GCS catalogue, with $|V| \leq 51$ km s^{-1}. The peculiar velocities are centred on the heliocentric mean velocity $(-6.12, -11.23, -6.18)$ km s^{-1}. The strong asymmetry of the velocity distribution, mainly on the UV plane, may only be described in the case $n = 6$, by fitting moments up to tenth order.

derived only from velocity moments. By using the GCS catalogue, Famaey et al. (2007) describe a similar small-scale structure of local stars; however, the entropy approach provides the smoothest density function that is also consistent with the data. In Figs. 3 and 4, two regions with higher probability densities are clearly identified, even using a large sample containing most of the thin disc with 9733 stars. The highest peak is placed around the Hyades-Pleiades moving groups, and the lower peak around the Sirius-UMa stream. However, our method works in the opposite direction of methods based on an arbitrary number of mixture components, or on wavelet transforms on arbitrary smaller scales. As Bovy et al. (2009) point out, adding a new component could substantially increase the goodness of the fit over the model with less complexity, while still being far from the truth. Similarly, Dehnen (1998) points out that structures on scales of a few km s^{-1} are likely to be spurious. On the contrary, the maximum entropy approach is a technique for computing the simplest and smoothest approach to the distribution function that fulfils the provided set of moment constraints. For a good estimation, the only requirement is that the sample is bell-shaped enough and the moments have enough accuracy. The method tends to smoothing all the statistical fluctuations of the sample, since the moments are obviously means. However, as shown in the above examples and in the next sections, if more complexity or resolution is desirable, either a larger set of constraints must be taken into account or specific subsamples must be selected.

3.1. Analysis of samples

In the preceding sections, the method has been applied to four case examples to show how the fitting of the distribution function is getting more informative depending on the degree of the polynomial $\mathcal{P}_n$ and on the complexity of the sample. We

now discuss some aspects of the results and samples. Samples I and II were chosen because they contain the maximum number of available stars with known velocity space, so that the velocity moments have minimum sampling variances. The main goal was to build the largest samples with stable velocity moments. However, these samples contain data with great uncertainty that could hamper the fitting of the distribution function.

The main source of data error, a matter of consequence for the HIPPARCOS sample, is the radial velocity, which is mostly measured from high proper motion stars. This may introduce a kinematical bias into the sample (Binney et al. 1997), although Skuljan et al. (1999) proved that the kinematic bias does not significantly affect the core of the disc distribution. Therefore, the description of disc kinematics from star velocities lower than $|V| \approx 75$ km s^{-1} should not reflect such a bias. To see how the error in the radial velocity could change the shape of the distribution function, and in particular the computed velocity moments, we select some new samples (Sample I' and Sample II') with radial velocity errors up to 2.5 km s^{-1}, a similar value to the mean observational error (Figueras et al. 1997). Sample I' from the HIPPARCOS catalogue now contains 9534 stars (70% of Sample I) and Sample II' from the GCS catalogue contains 11 514 stars (87% of Sample II). Clearly the GCS catalogue has stars with more accurate radial velocities. The velocity moments of Sample I' correspond now to a colder sample, with similar standard errors despite the small size of the sample. The diagonal second central moments are $(1310.09 \pm 45.20, 951.40 \pm 60.85, 345.39 \pm 16.85)$ instead of $(1431.46 \pm 45.23, 1073.89 \pm 54.95, 372.73 \pm 16.25)$ of Sample I. The velocity moments of Sample II' are not significantly changed and also have similar standard errors. Now, the diagonal second central moments are $(1236.14 \pm 31.60, 681.60 \pm 31.59, 344.70 \pm 19.60)$ compared to $(1205.49 \pm 30.08, 657.33 \pm 28.67, 332.93 \pm 17.33)$ of
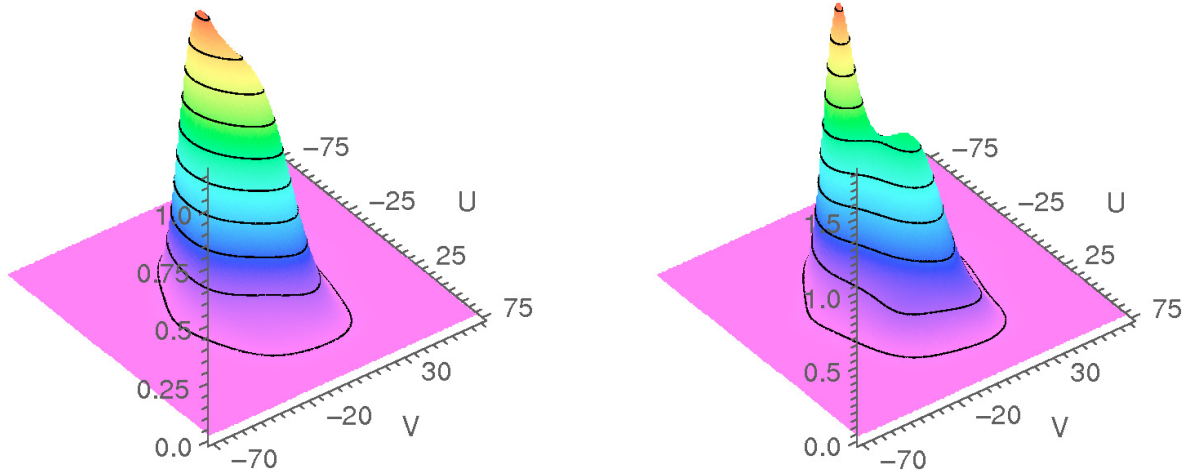
Fig. 4. Density functions on the plane UV for the HIPPARCOS Sample III (*left*) and the GCS Sample IV (*right*). The plots show a bimodal structure around the Hyades stream (highest peak) and the Sirius-UMa stream (lowest peak) for a distribution far from the ellipsoidal hypothesis.

Table 2. Centred and non-centred moments with their standard errors up to fourth order for the GCS Sample IV.

$\alpha\beta\gamma$	$\mu_{\alpha\beta\gamma}$	$\Delta\mu_{\alpha\beta\gamma}$	$m_{\alpha\beta\gamma}$	$\Delta m_{\alpha\beta\gamma}$
100	0.00	±0.22	−6.12	±0.22
010	0.00	0.14	−11.23	0.14
001	0.00	0.13	−6.18	0.13
200	462.82	±5.41	500.27	5.56
110	41.10	2.75	109.85	3.82
020	199.40	2.87	325.63	3.95
101	−5.00	2.53	32.82	2.81
011	−1.72	1.90	67.71	2.35
002	153.41	2.55	191.61	2.92
300	2173.59	±157.78	−6551.81	±276.03
210	−145.28	70.79	−6268.73	109.93
120	530.94	53.23	−2385.18	97.37
030	638.37	79.37	−7500.56	158.75
201	330.59	20.62	−2700.19	82.60
111	131.81	34.76	−480.42	47.48
021	96.46	40.55	−1877.37	64.93
102	337.15	46.93	−773.54	62.06
012	308.20	37.83	−1823.19	53.07
003	142.89	65.05	−2937.69	110.67
400	499 489.15	±10 136.14	551 669.32	±9932.67
310	34 551.28	3316.06	115 444.19	4479.28
220	75 113.83	1635.32	153 794.57	2498.65
130	6595.67	1901.12	50 161.86	3041.40
040	119 660.84	4596.79	257 920.40	6223.25
301	−1064.18	2965.11	32 799.70	3070.77
211	−299.90	1022.52	32 365.54	1356.19
121	−102.22	838.49	10 219.59	1199.62
031	3803.90	1762.48	46 257.99	2182.27
202	62 272.91	1472.05	78 158.49	1552.90
112	2098.65	788.56	8713.72	991.04
022	35 099.21	1149.57	58 305.32	1340.04
103	1025.32	1640.64	12 177.32	1815.65
013	−1390.25	1394.34	25702.11	1649.39
004	86 642.60	3447.49	119 730.52	4100.98

Sample II. The same criterion is applied to the bounded samples. Sample III' is 71% of Sample III, while Sample IV' is 86% of Sample IV. In these cases the velocity moments and their standard errors do not change at all. Although the respective fractions are similar to those in the previous samples, the moments remain stable, which confirms that the kinematic bias is

associated with higher velocity stars. In all the cases a similar shape of the velocity distribution is obtained as well as a slightly improvement of the χ^2 fitting error (in particular for the complete Samples I' and II' with $n = 4$).

Another issue to clarify is the cut $|V| \leq 51$ km s^{-1} for the bounded Samples III and IV. The main reason to choose them is the discontinuity noticed in Paper I, which has also been borne out in Paper II for the GCS sample. The recurrent segregation method used in these works (MEMPHIS algorithm) analysed the variations of two parameters accounting for a mixture approach: the entropy of the partition and the fitting error. By increasing a sampling parameter, in that case the absolute star velocity, the first significative discontinuity of those parameters took place at 51 km s^{-1}. After this value the method was able to segregate thin and thick discs with a decreasing fitting error. Therefore, this is not an astronomical reason but a statistical fact. We can now investigate the astronomical facts. Since the GCS samples have more accurate velocities, the analysis is centred in this catalogue. In Fig. 5 the graphs show how the stars are distributed into populations in terms of absolute velocity, eccentricity, metallicity, and colour. According to Paper II, the three bands of the vertical axis represent the expected value of a star to belong to any Galactic component (thin disc in the bottom, thick disc in the middle and halo at the top). Except for the eccentricity plot, the blue dots correspond to stars with $|V| \leq 51$ km s^{-1}, which clearly belong to the thin disc. This is also true for eccentricities, but now the blue dots correspond to stars with $|z_{max}| \leq 0.5$ kpc. In the $|V|$ plot we see that a large fraction of thin disc stars are still beyond 51 km s^{-1}. They are mixed with thick disc stars, especially from 65 km s^{-1} onward. From the eccentricity plot we deduce that thin disc stars are below 0.3, as discussed in Paper II, but beyond 0.1 they are increasingly mixed with the thick disc. However, when the condition $|z_{max}| \leq 0.5$ is applied, no thick disc or halo stars are included for eccentricities below ≈ 0.3. On the other hand, it is well known that the metallicity is appropriate for distinguishing the halo from the disc, [Fe/H] ≤ -0.5, but not between thin and thick discs. For the Strömgren photometry, the $b − y$ colour is spread along the three main populations. Most of the thin disc stars of the sample may be found at any index between 0.2 and 0.6, with mode 0.3, while the thick disc has mode 0.4 and the halo 0.5, with slightly narrower distributions. Similarly, for the maximum height over the Galactic plane, $|z_{max}|$ (not shown), thin and thick disc stars are also mixed

Table 3. Distribution parameters for HIPPARCOS and GCS samples with radial velocity errors up to 2.5 km s^{-1} (Samples I', II', III', and IV')[†].

Sample	σ_U	σ_V	σ_W	$\delta\,[°]$	c_U	c_V	c_W	γ_U	γ_V	γ_W
HIP total	36.2 ± 0.6	30.8 ± 1.0	18.6 ± 0.5	7.6 ± 6.6	9.3 ± 4.0	37.0 ± 13.9	20.7 ± 14.8	0.2 ± 0.2	-4.5 ± 0.8	-0.9 ± 0.5
$\|V\| \leq 51$	19.9 ± 0.1	13.2 ± 0.1	10.7 ± 0.1	12.3 ± 0.8	-0.4 ± 0.1	0.1 ± 0.3	1.5 ± 0.5	0.3 ± 0.0	0.2 ± 0.0	0.2 ± 0.1
GCS total	35.2 ± 0.4	26.1 ± 0.6	18.6 ± 0.5	10.7 ± 2.4	5.5 ± 2.8	22.7 ± 9.3	35.2 ± 42.4	-0.1 ± 0.2	-3.1 ± 0.5	-0.9 ± 1.0
$\|V\| \leq 51$	21.5 ± 0.1	14.1 ± 0.1	12.4 ± 0.1	8.7 ± 0.6	-0.7 ± 0.1	0.0 ± 0.2	0.7 ± 0.3	0.2 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
$e \leq 0.3$	29.1 ± 0.2	18.1 ± 0.1	11.6 ± 0.1	9.6 ± 0.6	0.2 ± 0.2	0.4 ± 0.2	-0.4 ± 0.1	0.1 ± 0.0	-0.3 ± 0.0	0.0 ± 0.0

Notes. [†] The last sample contains representative thin disc stars, obtained from eccentricities $e \leq 0.3$ and $|z_{max}| \leq 0.5$ kpc. The displayed parameters are dispersions σ_U, σ_V, σ_W, vertex deviation δ on the UV-plane, curtosis c_U, c_V, c_W, and skewness γ_U, γ_V, γ_W.
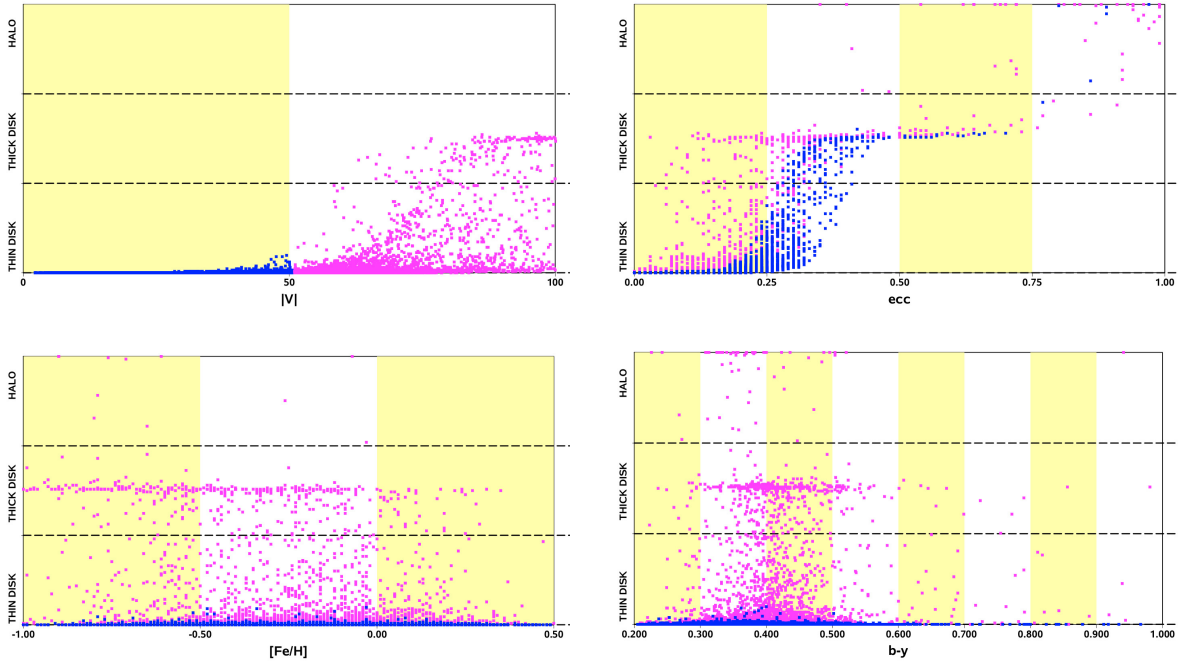
Fig. 5. Distribution of GCS sample stars into populations in terms of absolute velocity, eccentricity, metallicity, and colour. The blue dots indicate stars with $|V| \leq 51$ km s^{-1}, except for the eccentricity plot, which are for stars with $|z_{max}| \leq 0.5$ kpc.

in the interval $|z_{max}| < 1$ kpc, but, like metallicity, the halo can be segregated.

In Fig. 6 the distributions obtained from the current method with $n = 6$ are plotted in terms of metallicity and colour on the three velocity planes for subsamples with $|V| \leq 51$ km s^{-1}. All of them reproduce the bimodal structure of Fig. 3 with $n = 6$, except metalicities in the range $0 \leq$ [Fe/H] < 0.5 and colours with $0.205 \leq b - y < 0.300$, which correspond to the earliest F dwarfs of the thin disc, with negative radial mean velocity. However, such a bimodal shape comes from the velocity cut. For the whole CGS sample, the distributions for different metalicities and colours are similar to the deformed velocity ellipsoid of the whole sample, as shown in Fig. 7, though with a slightly different mean, depending on the colour and metallicity range.

3.2. Smaller scale

It is however possible to obtain a more detailed shape for the velocity distribution for specific subsamples, allowing comparison of the results of our approach with the small-scale structure sustained by moving groups as described by other authors. By selecting samples with bounded peculiar velocity, such as $|u| \leq 7.5$ km s^{-1} (256 stars), 10 km s^{-1} (498 stars), or 20 km s^{-1} (2817 stars), a more complex structure is manifest

on the UV plane, but also in the vertical direction. The shape of the distribution becomes softer while increasing the size of the sample. Because of this, the substructure of thin disc subsamples with less stars become statistical fluctuations within larger subsamples, up to describing a sufficiently complete distribution of the thin disc. Thus the clue is to find a clean and representative thin disc sample. The cut $|V| \leq 51$ km s^{-1} therefore seems to be a good value that includes most of thin disc stars and excludes thick disc stars, but it is still far from being a complete thin disc sample. Samples selected from small peculiar velocities have some limitations. On one hand, they contain few stars, so that their distribution may not be bell-shaped enough. Furthermore, their moments have greater uncertainties. On the other hand, the boundary of the distribution is fixed by the velocity limit of the sample, which may cut down some well-defined structures. Fortunately, there is a way to avoid this problem. In Papers I and II, consecutive stellar populations were merged to nested subsamples in terms of several sampling parameters: maximum absolute velocity, peculiar velocity, vertical velocity, etc. Optimal values of these sampling parameters allowed the segregation of these populations. For the complete GCS sample, once the stars are classified according to the probability of belonging to any of the local Galactic components (Paper II), a highly significant correlation is obtained between the expected population of a star and its absolute velocity $|V|$.
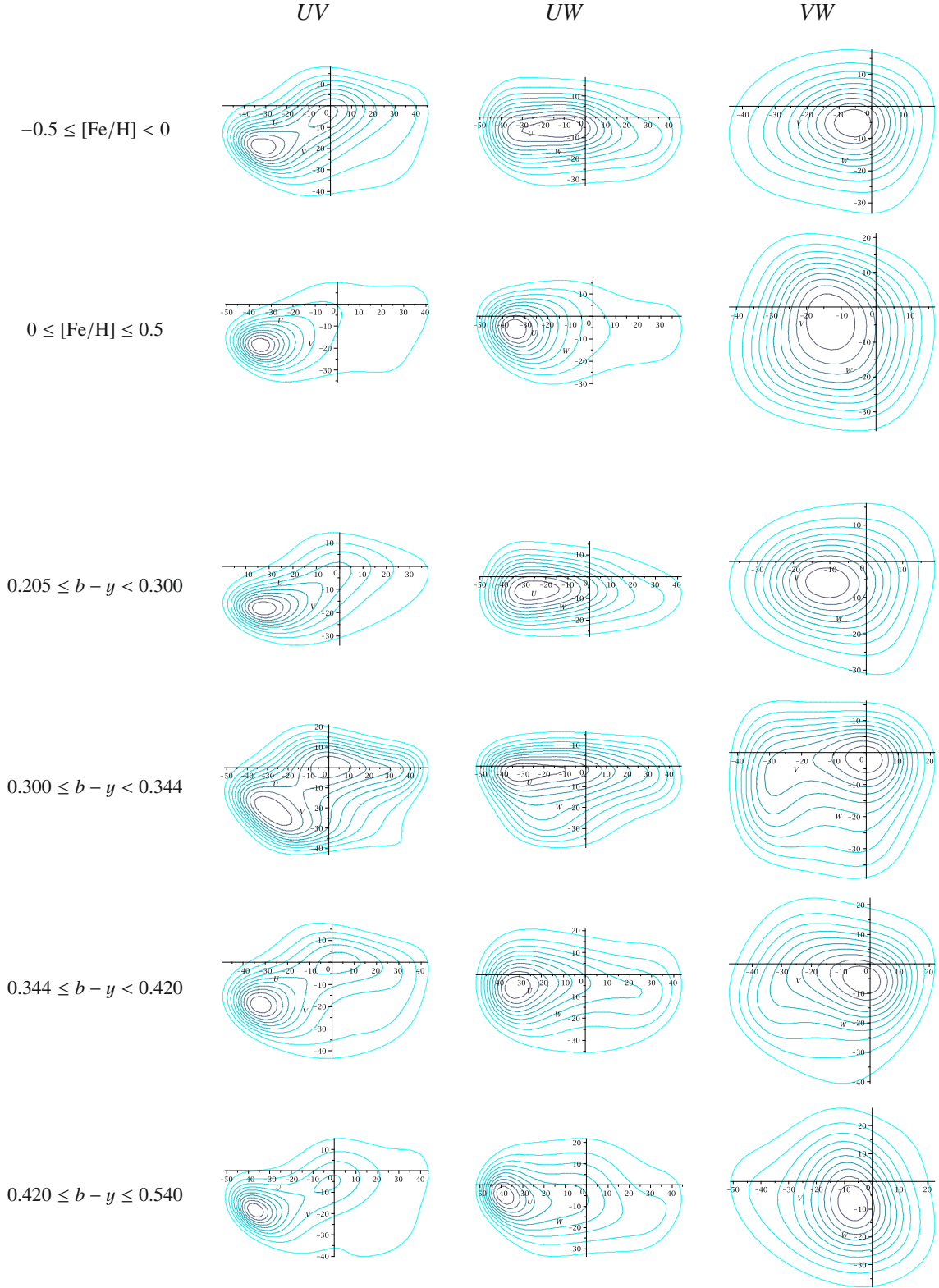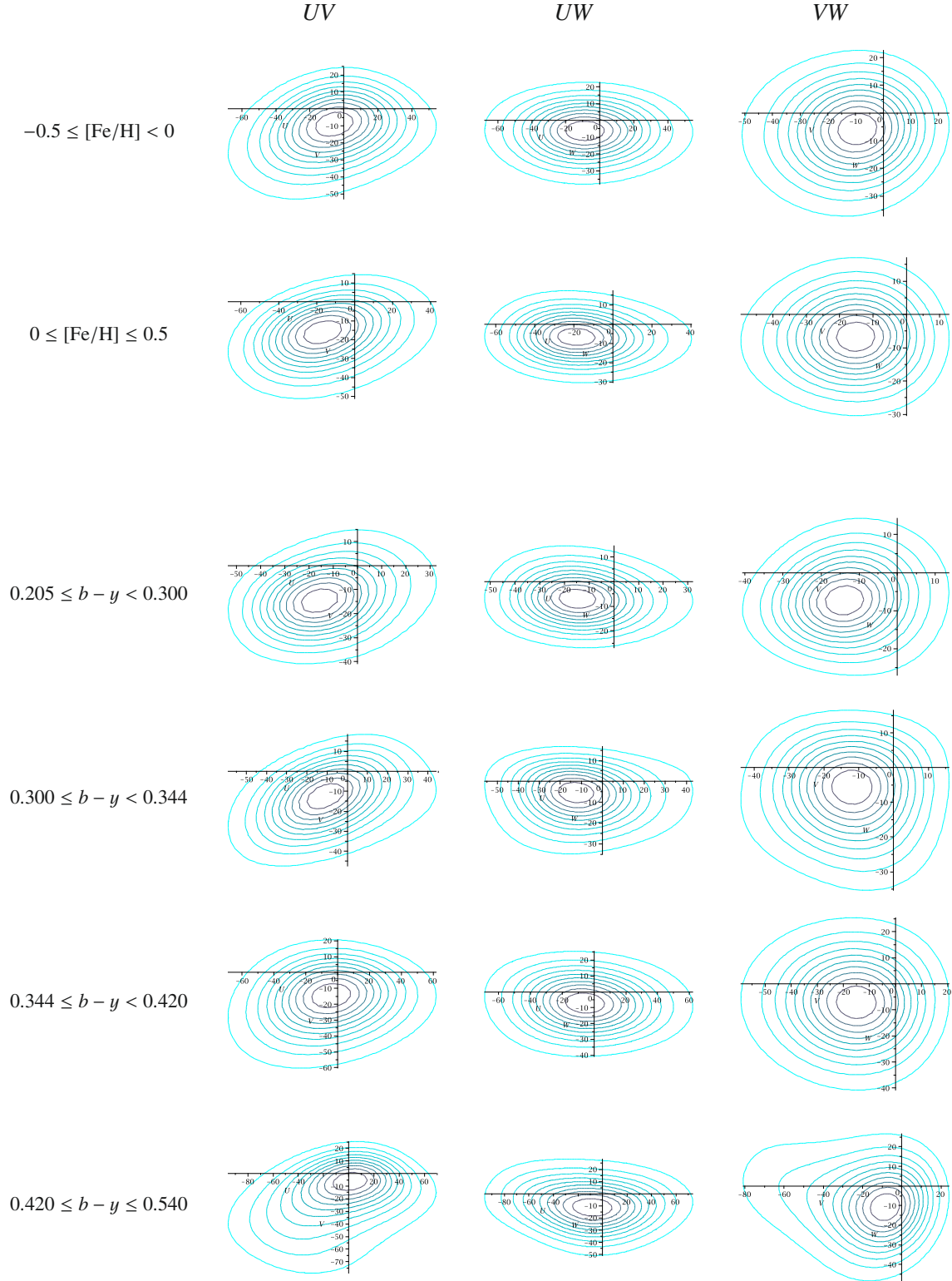
Fig. 6. Contour plots of the velocity distribution for stars from the GCS catalogue with $|V| \leq 51$ km s^{-1}, obtained for $n = 6$ in terms of metallicity and colour. The axes are labelled according to heliocentric velocities.

The expected value is similarly highly correlated with the planar eccentricity, and also correlated with rotation velocity, $|z_{\max}|$ and metallicity. The colour is few correlated with the expected population and the other preceding properties. Therefore significant partial correlations between couples of the former star properties exist. However, when the sample is bounded to $|V| \leq 51$ km s^{-1},

by leaving aside thick disc and halo stars, the only significant partial correlation that is maintained is the absolute velocity and the eccentricity, as well as the expected population with them. That means that the other properties are only relevant for segregating thick-disc or halo stars, but are not useful within the very thin disc. As discussed in Paper II, the sampling parameter

Fig. 7. Contour plots of the velocity distribution for stars from the total GCS catalogue, obtained from the entropy approach with $n = 6$, in terms of metallicity and colour.

is related to the isolating integrals of the star motion. Both the absolute velocity and the eccentricity satisfy this requirement. The former is less discriminant, but is a direct measure from the star. The latter is more discriminant, but requires computing the orbital parameters, with the need of additional hypothesis on the potential, symmetries, stationarity, mean motion, solar position, etc. Therefore, it is possible to use the eccentricity not for segregating populations, as e.g., Pauli et al. (2005) or Vidojević & Ninković (2009), but as an improved sampling parameter to select subsamples.

For samples with maximum eccentricities 0.01 (220 stars), 0.02 (591 stars), 0.03 (1058 stars), 0.05 (2465 stars), 0.1
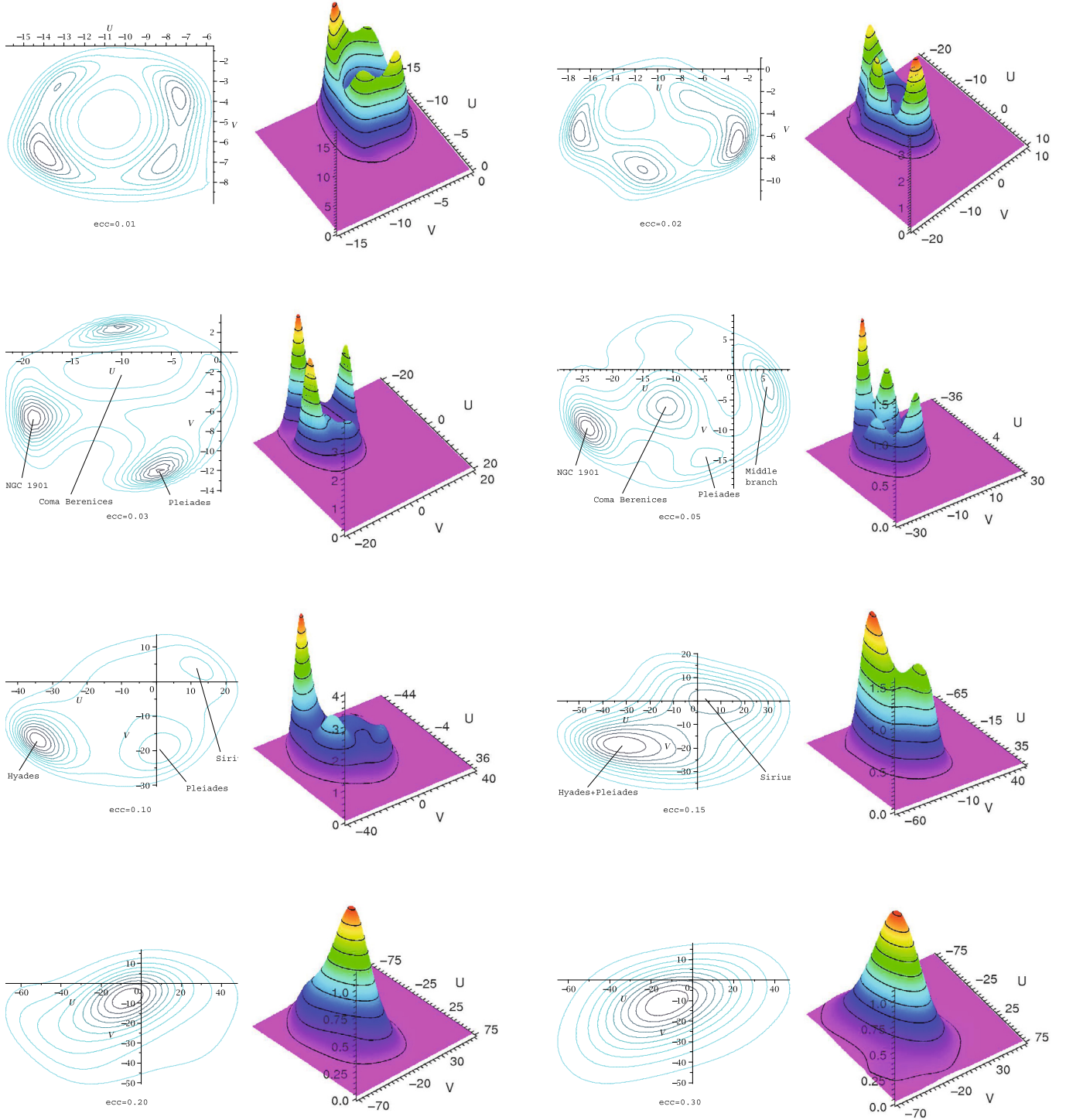
Fig. 8. Series of contour plots and distributions on the *UV* plane for GCS subsamples selected from $|z_{max}| < 0.5$ kpc and eccentricities up to 0.01, 0.02, 0.03, 0.05, 0.1, 0.15, 0.2 and 0.3. The origin is at the Solar velocity.

(7095 stars), 0.15 (9545 stars), 0.2 (10 903 stars), and 0.3 (11 826 stars), with the additional condition $|z_{max}| < 0.5$ kpc to avoid contamination from stars not belonging to the thin disc, the maximum entropy approach provides the series of plots in Fig. 8. Both previous limitations introduced by the peculiar velocity boundary have disappeared. For example, the structure described by the plot $|\boldsymbol{u}| \leq 10$ km s^{-1} with 498 stars is now more completely described from the plot with maximum eccentricity 0.05 with 2465 stars. Similarly, the shape of the distribution is no longer forced by the sampling parameter. The eccentricity

then behaves as a very good sampling parameter that allows us to distinguish between different *eccentricity layers* within the thin disc and enables us to visualise the structure below each layer. In the lower layers, with maximum eccentricities 0.01 and 0.02, the velocity distribution shows a hole around the local standard of rest (LSR), taken as $(-10., -5.23, 7.17)$ km s^{-1} (Dehnen & Binney 1998), which is the mean of the distribution. Those lowest eccentricity stars are moving around the LSR and have velocities distributed on a ring with some peaks around the LSR. The radial velocities are symmetrically grouped into two main bulks

at each side of the LSR. This behaviour is maintained up to eccentricity $e = 0.03$, where the LSR hole begins to be filled by the group of stars corresponding to the Coma Berenices moving group, nearly at the same LSR velocity. In addition, three stellar groups around the LSR conform the basic structure: NGC 1901, a group that can be part of the middle branch (Skuljan et al. 1999), and a part of the Pleiades group. The structure is the same as described by Bovy et al. (2009) and by previous works of Dehnen (1998), Skuljan et al. (1999), Famaey et al. (2005, 2008), with the greatest peak in NGC 1901.

For $e = 0.05$ the structure is maintained and enlarged. It incorporates a new group of stars also associated with the middle branch, which is not referred to as a moving group by Bovy et al. (2009), but is the centre of their Gaussian component with the largest weight. In the range of eccentricities from 0.05 to 0.1, the small previous structures are diluted in a background distribution, and only the Pleiades group remains. The main weight of the distribution is now in the stars around the Hyades group. A stellar group around the Sirius/UMa stream arises at positive radial velocities. For $e = 0.15$, that is, approximately the higher eccentricity before appearing thick disc stars, the distribution is divided into about half: one bulk with negative radial and rotation mean velocities with respect to the LSR, which contain the main groups Hyades and Pleiades; and another one with positive values around Sirius and UMa stream. For higher eccentricities, the distribution becomes similar to the one corresponding to the thin disc. In particular, for $e = 0.3$, with a fraction of 90% of the whole sample, we get a distribution similar to the thin disc of Paper II (obtained by two different methods: MEMPHIS algorithm and the method of Galactic orbits), with dispersions and vertex deviation $(\sigma_U, \sigma_V, \sigma_W; \delta) = (29.1 \pm 0.2, 18.1 \pm 0.1, 11.6 \pm 0.1; 10° \pm 1)$.

Thus, for eccentricities below 0.15, there is a general trend in the radial direction: the main weight of the distribution is symmetrically placed around the LSR. Thus, the velocity distribution of the thin disc is supported by two major stellar groups with opposite radial velocities around the LSR, with a dearth of stars at the LSR. One bulk, with positive radial velocity, has a mean velocity similar to the Sun or slightly higher, with a lower peak but a wider distribution. The other one, with radial velocity ≈ -30 km s^{-1}, has a mean rotation ≈ -20 km s^{-1} and a higher peak. This behaviour is definitively broken for eccentricities $e \geq 0.2$.

4. Discussion

The situation described in the preceding section, where the star velocities are distributed for low eccentricities along the U direction with a local minimum at the LSR velocity, may have a simple explanation: a mixture of stars with a discrete number of radial oscillation periods. In the special case of a nearly planar orbit, where the planar eccentricity is low enough that the amplitude of the vertical motion becomes independent of the radial motion, the motion of a star can be studied as a case of epicyclic orbit. Let us remember that, if a nearly planar orbit is projected onto the Galactic plane, its distance to the Galactic centre oscillates between two limiting values R_p and R_a. The planar eccentricity e is then defined as

$$e = \frac{R_a - R_p}{R_a + R_p}, \tag{18}$$

which is a dimensionless measure of the deviation from the circular motion in the plane of symmetry.

The orbits in the Galactic disc are nearly planar, and their planar eccentricities may be significantly different from zero (Vidojević & Ninković 2009). Thus, the epicyclic approximation can be used for the current disc sample (Paper II). It is commonly assumed that moving groups of young stars are born in nearly circular orbits (e.g. Dehnen 1998) and, with age, they are transformed into more eccentric orbits, which oscillate locally around the LSR (assumed to be in circular motion, $U_{LSR} = 0$, for steady state and axisymmetric systems). Thus, the radial velocity of a star oscillating around the LSR with a period T may be written as

$$U = a \sin \frac{2\pi t}{T}, \tag{19}$$

where the amplitude is proportional to the eccentricity, $a = \frac{2\pi}{T}(R_a + R_p) e$. If the time t is measured in oscillation periods, we may assume $T = 1$.

For a sufficiently great t, e.g., taking several periods, we may also assume that t is uniformly distributed within the interval $[0, 1]$, so that its probability density function is $f_t(t) = 1$, $t \in [0, 1]$, and zero otherwise. We may ask for the distribution of U around the LSR velocity, that is, for the probability of finding a star velocity at any given value within $[-a, a]$. Since $t(U) = \frac{1}{2\pi} \arcsin(\frac{U}{a})$, and this is a two-valued function, the probability density function $f_U(U; a)$, for a given amplitude a, is easily obtained as $2 f_t(t) |t'(U)|$. Thus,

$$f_U(U; a) = \frac{1}{\pi} \frac{1}{\sqrt{a^2 - U^2}}, \quad U \in [-a, a], \tag{20}$$

and zero out of this interval. As seen in Fig. 21b, for an arbitrary value $a = 1$, it is less probable to find the star with nearly zero velocity, which means near the extreme positions R_p or R_a, than around the mean position $(R_p + R_a)/2$ with maximum absolute velocity, either negative by going toward the Galactic centre or positive toward the anticentre.

We may think of a mixture of stars with the same oscillation period and different eccentricities, from zero eccentricity and amplitude, $a = 0$, up to greater amplitudes, say $a = A$. Let us assume a normalised density function $\rho(a)$ of stars in terms of the amplitude a. Then the cumulative density function $h_U(U; A)$ is obtained by integration over a as

$$h_U(U; A) = \int_0^A f_U(U; a) \rho(a) \, da. \tag{21}$$

Depending on how the stars are distributed in terms of the amplitude, we may get different symmetric distributions around the LSR. For a fixed period T, this depends on the distribution of eccentricity. The distribution of eccentricity is approximately lognormal, similar to the distribution of wealth in a country, as shown in Fig. 9 from the histogram for the GCS sample, where the interval $[0, 1]$ of eccentricities is divided into 50 bins on the x-axis ($x = 50 e$ is approximately lognormal with $m = 1.75$ and $\sigma = 0.5$).

To find out the shape of $h_U(U; A)$ for a group of stars with the same period, several simulations are carried out for arbitrary values of the amplitude, by assuming $\rho(a)$ lognormal. Bimodal distributions around the origin (LSR) are always obtained, like the plots (c), (d), and (e) of Fig. 10. The mathematical reason is that the lognormal distribution vanishes in a neighbourhood of zero. (It is tangent to zero at the origin.) However, the wider the distribution wing, the less significant the bimodality, since the behaviour for low amplitudes becomes a smaller structure when larger amplitudes are considered. The distribution tends to be
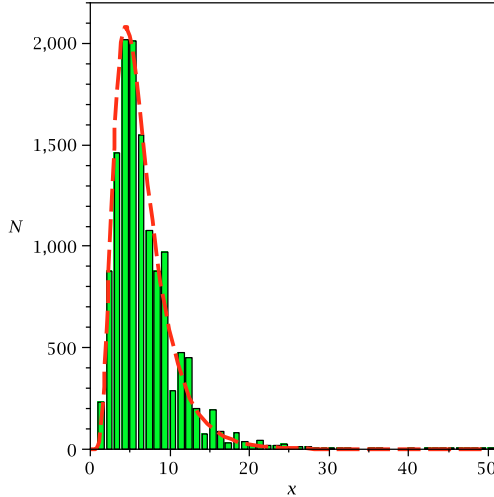
Fig. 9. Distribution of eccentricities for the GCS sample. The probability density function obtained from the histogram is approximately lognormal. In the x-axis, the interval $[0, 1]$ of eccentricities is divided into 50 bins. The variable $x = 50\,e$ is lognormal with $m = 1.75$ and $\sigma = 0.5$ (red dashed line).

Gaussian. In the case of a discrete mixture of stars with different periods, we should get a mixture of densities $h_U(U; A)$ with similar properties, which could explain the radial velocity shape obtained for the inner thin disc. However, for a continuous mixture of populations in terms of oscillation periods, the bimodal structure may be irrelevant. For example, apart from statistical fluctuations, as increasing T, the density function $\rho(a)$ may become more populated around zero, so that it is no longer tangent to zero at the origin of amplitudes. Then, a high-peaked $h_U(U; A)$ may be obtained at the origin, for low values of A. In general, if in a neighbourhood of the origin, $\rho(a)$ behaves as a^p, with $0 < p \leq 1$ we then get smooth unimodal distributions centred at the origin. On the other hand, it is easy to prove that, if the radial velocity does not oscillate symmetrically around the LSR meaning a slight deviation from the epicyclic approximation, the peaks of $h_U(U; A)$ become non-symmetric around the LSR. Therefore, these simple simulations reproduce the actual situation for low eccentricities approximately.

As a result, the thin disc contain two major streams moving with opposite radial directions around the LSR, one with small positive radial mean velocity and rotation similar to the Sun, and the other with negative radial mean velocity and lower rotation. For each subsystem, we could assume some less restrictive hypotheses, such as point-axial symmetry (opposite points through an axis, allowing, in particular, spiral structures) or a time-dependent model, in order to describe the non-vanishing radial velocity of their centroids and the vertex deviation of their approximate velocity ellipsoids. Thus, a general Chandrasekhar point-axial model (Sanz-Subirana & Català-Poch 1987; Juan-Zornoza et al. 1990, 1995) should be the simplest approximation, where, despite the non-cylindrical symmetry of the system, the solution of Chandrasekhar's equations system yields an axisymmetric potential. In this case, it is interesting to recall the relationship between the vertex deviation and the radial mean velocity.

It is well known that the vertex deviation δ of a velocity ellipsoid depends on the second central moments in the form

$$\tan(2\delta) = \frac{2\mu_{UV}}{\mu_{UU} - \mu_{VV}}. \qquad (22)$$

Thus, if $\mu_{UU} - \mu_{VV} > 0$, the angle δ and the moment μ_{UV} have the same sign. As explained in Cubarsí & Alcobé (2006), the radial mean velocity referred to an axisymmetric (cylindrical) system can be related to the increment of rotation mean velocity with respect to the same axisymmetric system and, in particular, to the vertex deviation of the velocity ellipsoid, through the expression

$$U_0 - U_0^{(\text{cyl})} = \frac{\mu_{UV}}{\mu_{VV}}\,\Theta_0 \qquad (23)$$

where U_0 is the radial mean velocity, $U_0^{(\text{cyl})}$ the radial mean velocity of an ideal axisymmetric system (which is now associated with the LSR), and Θ_0 the galactocentric rotation mean velocity. This is not a special situation, and recent studies of the shape of velocity ellipsoids in spiral galaxies (Vorobyov & Theis 2008) confirm that the magnitude of the vertex deviation is not correlated with the gravitational potential (even though it is assumed non-axisymmetric), but is strongly correlated with the spatial gradients of the mean stellar velocities, in particular with the radial gradient of the mean radial velocity.

Thus, according to Eq. (23), a velocity ellipsoid with a positive [negative] vertex deviation might be associated with a loss of axisymmetry and with a radial motion towards [against] the Galactic centre.

In Fig. 11 (left) the two major stellar groups with opposite radial velocities around the LSR, which support the largest structure within the thin disc, are associated with two velocity ellipsoids with similar total dispersions. The one moving toward the Galactic centre, with radial and rotation galactocentric mean velocities $U_0 = 15, \Theta_0 = 220$, and the other one, toward the anticentre, with mean motion $U_0 = -15, \Theta_0 = 200$. The LSR is placed in the middle of both ellipsoids in a similar situation to the sample with maximum eccentricity 0.15. By assuming the same mixture proportions $n' = n''$, the partial diagonal central moments $\mu'_{ii} = \mu''_{ii}$ are obtained from totals and from their deferential mean velocity $u'_i - u''_i$, from the usual relationship

$$\mu_{ii} = n'\mu'_{ii} + n''\mu''_{ii} + n'n''(u'_i - u''_i)^2; \quad i = 1, 2, 3. \qquad (24)$$

The vertex deviation of each ellipsoid is obtained from Eq. (23). The total moments are taken from the sample with maximum eccentricity $e = 0.3$. The graph shows the partial ellipsoids $\boldsymbol{u}^{\text{T}} \cdot \boldsymbol{\mu}_2^{-1} \cdot \boldsymbol{u} = 1$, in blue, and the thin disc velocity ellipsoids $\boldsymbol{u}^{\text{T}} \cdot \boldsymbol{\mu}_2^{-1} \cdot \boldsymbol{u} = 2, 3$, in red. The ellipsoid with positive radial velocity has a positive vertex deviation (although small), and the one with negative radial peculiar velocity has negative vertex deviation (also small). The shape of the thin disc, in particular its apparent positive vertex deviation, is generated from the inner structure. The green dashed partial ellipsoids represent a situation with the opposite radial motions, so that, in such a case, the apparent total vertex deviation should be negative. On the right, the contour plots in the UV plane (in heliocentric velocities) for the samples with maximum eccentricity $e = 0.15$ (blue) and $e = 0.3$ are superposed. Simulated and actual plots are totally consistent.

5. Conclusions

Although successfully applied to a wide disparity of actual problems, the maximum entropy approach has rarely been used to solve the classical moment problem of stellar kinematics. Instead, a number of statistical techniques, maximum likelihood-based multivariate sampling algorithms, wavelet-based algorithms, among others, up to the present date had been proved to be more appropriate than the moment method for accurately

(a) $\rho(a) = \text{lognormal}(0.1, 0.5)$ **(b)** $f_U(U; 1)$ **(c)** $h_U(U; 0.25)$ **(d)** $h_U(U; 1.25)$ **(e)** $h_U(U; 6)$
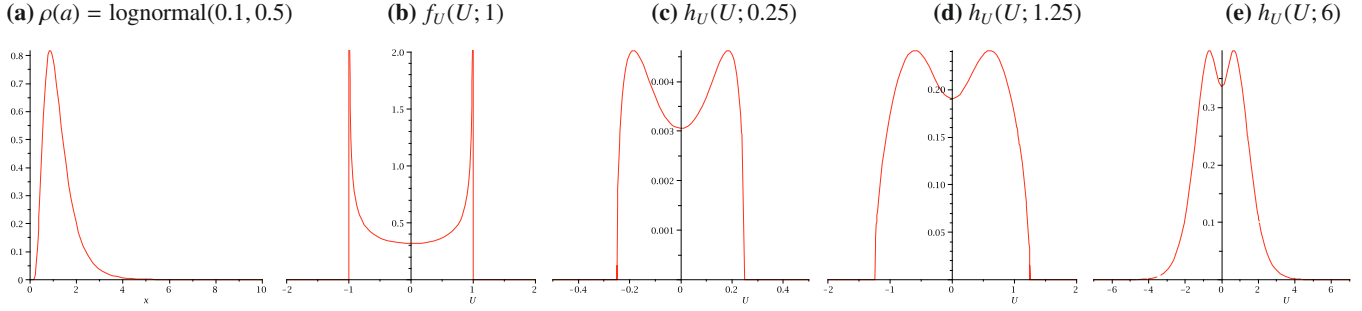
Fig. 10. Simulated distribution of radial velocities for arbitrary values of the amplitude, by assuming the epicyclic approximation and $\rho(a)$ lognormal. Bimodal distributions around the origin (LSR) are obtained. **a)** Probability density function $\rho(a)$, assumed to be lognormal (0.1, 0.5). **b)** Probability density function $f_U(U; a)$ for amplitude $a = 1$. **c)** Cumulative density function $h_U(U; A)$ integrated up to $A = 0.25$. **d)** Cumulative density function $h_U(U; A)$ integrated up to $A = 1.25$. **e)** Cumulative density function $h_U(U; A)$ integrated up to $A = 6$.
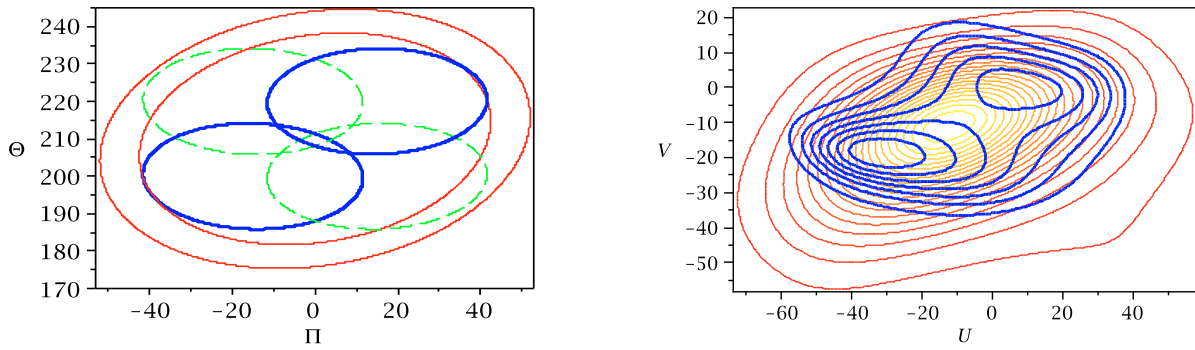
Fig. 11. (*Left*) Velocity ellipsoids, in blue, depicted according to Eqs. (23) and (24), from total moments corresponding to the sample with eccentricities $e \leq 0.3$. They are centred in galactocentric velocities $\Pi_0 = 15$, $\Theta_0 = 220$, and $\Pi_0 = -15$, $\Theta_0 = 200$, with the LSR placed in the middle of them. Thin disc isocontours, in red, with positive vertex deviation, are generated from the inner structure. The green dashed partial ellipsoids represent a situation with the opposite radial motions. (*Right*) Contour plots in the UV plane (heliocentric velocities) for the samples with maximum eccentricity $e = 0.15$ (blue) and $e = 0.3$ (red).

describing the local stellar velocity distribution. The moment approach had two basic difficulties compared to other methods: the low accuracy of available data and the complexity of the trivariate model for estimating the parameters. Nowadays, with larger and more precise stellar catalogues, it is possible to compute a reasonable number of moments with sufficient accuracy. On the other hand, the parameter estimation for a maximum entropy function requires some complex computational procedures when working from the minimum set of moment constraints. In contrast, if an extended set of moments is used, the current method provides a linear estimation algorithm, so that a Gramian system of equations leads to a fast and suitable estimation of the smoothest velocity distribution that is consistent with the available constraints. It therefore tends to smooth all the statistical fluctuations of the sample out, so that it is more appropriate to study the velocity distribution of large stellar samples, instead of particular moving groups. The estimation method is, however, specific to bell-shaped distributions.

The resulting features of the local velocity distribution are similar for the complete HIPPARCOS Sample I and GCS Sample II, although, as expected, the latter provides more accurate results due to the lower uncertainties of radial velocity measurements. The whole distribution shows a nearly constant vertex deviation in the pseudo ellipsoidal level curves, as well as a nearly constant axis ratio. According to the most significant polynomial coefficients, the resulting density function can be expressed as the product of two exponential functions in the form of Eq. (17). The background ellipsoidal shape has axis ratios 1:0.7:0.5 and a symmetry plane $W = 0$. The

overall vertex deviation in the UV velocity components is approximately 11°. Some characteristic parameters of the distribution are summarised in Table 3. The function $\varphi_2(h)$ may be interpreted as a perturbation factor, and is even and at least quadratic in the V velocity alone. It accounts for the skewness and the shift in the velocity ellipsoids in terms of the rotation velocity. This is clearly visible on the UV and VW planes of Fig. 1, and may be interpreted with regard to the heating of disc stars, which is also correlated with a decreasing galactocentric rotation velocity, as expected from Strömberg's law. The resulting overall distribution has zero curtosis in the W velocity, and, within 2σ level, in the U component for the more accurate GCS sample.

On the other hand, the entropy method offers an excellent estimation of the truncated velocity distributions of Samples III and IV, which only contain thin disc stars. For these subsamples, a Gaussian mixture approach was impossible in Paper I. This method can therefore be used as an alternative way to study multimodal distributions. For star velocities $|V| \leq 51$ km s^{-1}, the tentative mixture (with a very large chi-squared fitting error) obtained in Paper I suggested a superposition of two enlarged pseudo-ellipsoidal distributions, mainly along the radial direction, with very overlapped wings, and a separation of 28 ± 9 km s^{-1} between means. For Sample IV, the separation of the two peaks (the "U-anomaly") along the radial direction may now be straightforwardly estimated in 30−35 km s^{-1} from the contour plots obtained from the moment constraints. Figures 2 and 3 show a strongly asymmetric velocity dispersion on the plane UV, and nearly laminar isocontours of the W velocity component along the radial direction. There is a large radial

velocity dispersion on the UV plane, in the direction of the gravitational gradient, and a very small dispersion in the direction perpendicular to the galactic plane. On the VW plane, the isotropy is slightly recovered. However, the higher resolution contour plots obtained from Sample IV allow more detailed analysis. A first look at the UW plane of Fig. 3 ($n = 6$) shows a core distribution with negative radial peculiar velocity, while there is a clear unimodal behaviour with positive peculiar rotation velocity on the VW plane. However, those high-density regions of the distribution are not simultaneous on the UV plane, but are associated with different large stellar groups.

In addition to limited velocity distributions, other truncated distributions were analysed in terms of metallicity, colour, maximum distance to the Galactic plane, and eccentricity. Those star properties were correlated with the star's expected population obtained in Paper II. The only significant correlations that are maintained within thin disc stars are the absolute velocity and the eccentricity, since metallicity and $|z_{max}|$ are more appropriate for segregating thick disc and halo stars from the thin disc. In particular, the eccentricity, which is directly related to the isolating integrals of the star motion, is more discriminating than the absolute velocity for selecting subsamples. However, a part of thin and thick disc stars may have similar eccentricities. Then, to isolate thin disc stars, it is necessary to combine two sampling parameters: eccentricity and $|z_{max}|$. A representative thin disc, containing 90% of the whole sample, is selected from maximum eccentricity 0.3 and $|z_{max}| \leq 0.5$ kpc. Its central moments are similar to the ones obtained for the thin disc in Paper II. Furthermore, within the thin disc sample, the eccentricity behaves as an excellent sampling parameter that distinguishes between different eccentricity layers allowing subjacent structures to be visualised.

For subsamples obtained from eccentricities $e < 0.15$, the maximum entropy method is able to plot the classical moving groups composing the small-scale structure of the velocity distribution, as described from other algorithms based on an arbitrary number of mixture components, wavelet transforms, or maximum likelihood (e.g., Famaey et al. 2008; Veltz et al. 2008; Skuljan et al. 1999; Dehnen 1998), especially those providing a modest amount of complexity (Bovy et al. 2009). For these subsamples with small eccentricities, there is a general trend: the star velocities are approximately symmetrically distributed around the LSR in the radial direction. In most cases the distribution is bimodal, with a dearth of stars at the LSR. At the end, for $e = 0.15$, the core distribution of the thin disc is supported by two major stellar groups with opposite radial velocities, referred to the LSR. One bulk, with zero or small positive heliocentric radial mean velocity, has a lower peak but a wider distribution around Sirius/UMa stream. The other one, with radial velocity about -30 km s^{-1}, has mean rotation ≈ -20 km s^{-1} and a higher peak, containing the main groups Hyades and Pleiades.

For stars with a similar period of oscillation around the LSR in the radial direction (under the epicyclic approximation), several simulations allow us to confirm that such a two-peaked distribution of radial velocities is due to a lognormal distribution of the eccentricities. For a mixture of stars with different periods and a lognormal distribution of the velocity amplitude of the stellar orbits, the bimodal shape is maintained. However, if the number of stars with nearly vanishing amplitude increases, then the radial velocity distribution becomes unimodal, similar to the total thin disc sample with $e \leq 0.3$.

The bimodal behaviour of the central disc associated with the previous major subsystems may then be explained from two different phenomena. On one hand, it may be a perturbation similar to a pressure wave acting in part along the radial direction that induces an oscillation of the radial velocity around the LSR. Let us remember that the oscillation of each subsystem centroid along the U direction is also the expected motion of axisymmetric systems under steady state potentials (e.g. Cubarsi et al. 1990). On the other hand, both kinematical major groups, which actually are placed at the solar position, are in opposite oscillation states. In addition both groups have a difference of about 20 km s^{-1} in rotation mean velocity, so that one group of stars actually surpasses the other group. Therefore, the apparent vertex deviation of the thin disc may stem from the swinging of those major kinematic groups. A scenario of a continuously changing orientation of the disc pseudo ellipsoid is then possible.

Appendix A

We need to write $\mathcal{P}_n$ in Eq. (15) by using a slightly different notation, with Latin indices instead of Greek indices, so that each term accounts for products of the same degree in the velocities. Thus, by using hereafter Einstein's summation criterion for repeated indices, we may write

$$\mathcal{P}_n = \lambda_0 + \lambda_i V_i + \lambda_{ij} V_i V_j + \cdots + \lambda_{i_1 \ldots i_n} V_{i_1} \cdots V_{i_n}. \qquad (25)$$

In the summation term corresponding to the kth-power of the velocities we have $\binom{k+2}{2}$ different coefficients, since the coefficients $\lambda_{i_1 \ldots i_k}$ are symmetric; hence, up to the nth-power there are $\sum_{k=0}^{n} \binom{k+2}{2} = \binom{n+3}{3}$ different coefficients. In addition, we use the relationship

$$\lambda_{\alpha\beta\gamma} = \frac{k!}{\alpha!\beta!\gamma!} \lambda_{\underbrace{1 \ldots 1}_{\alpha} \underbrace{2 \ldots 2}_{\beta} \underbrace{3 \ldots 3}_{\gamma}}, \quad k = \alpha + \beta + \gamma, \qquad (26)$$

which establishes the correspondence between the Greek and Latin index notations for the coefficients of $\mathcal{P}_n$. Some practical aspects of the maximum entropy distribution function may still be pointed out. Under maximum entropy distributions, the sample moments are maximum likelihood estimators of the population moments.

When $n \to \infty$, Eq. (15) converges to the true distribution. Then, if the velocity distribution is expressed as a power series of the velocities, we have

$$
\begin{aligned}
f(t, \boldsymbol{r}, \boldsymbol{V}) &= c_0 \left(1 + \sum_{k=1}^{\infty} \sum_{\alpha+\beta+\gamma=k} c_{\alpha\beta\gamma}(t, \boldsymbol{r}) V_1^\alpha V_2^\beta V_3^\gamma \right) \\
&= \exp\left(\sum_{k=0}^{\infty} \sum_{\alpha+\beta+\gamma=k} \lambda_{\alpha\beta\gamma}(t, \boldsymbol{r}) V_1^\alpha V_2^\beta V_3^\gamma \right) \qquad (27)
\end{aligned}
$$

which is a similar relationship between generalised moments $c_{\alpha\beta\gamma}$ and cumulants $\lambda_{\alpha\beta\gamma}$ (Stuart & Ord 1987, p. 437), where the coefficient $c_0 = \exp(\lambda_0)$ provides the normalisation of the distribution[3].

A maximum entropy distribution function can exhibit several modes. In the trivariate case, if Eq. (15) is a polynomial of even degree n, the distribution can exhibit $(n/2)^3$ modes, since an univariate exponential with a polynomial of degree n may have up to $n/2$ modes. In general, it is necessary to estimate less number of parameters for Eq. (15) than for a mixture of trivariate Gaussian distributions accounting for the same number of modes.

---

[3] When a similar relation holds for the characteristic function Φ, which is the Fourier transform of density function f, then the coefficients $c_{\alpha\beta\gamma}$ become proportional to the population moments $m_{\alpha\beta\gamma}$, and $\lambda_{\alpha\beta\gamma}$ become proportional to the cumulants of the distribution $\kappa_{\alpha\beta\gamma}$, by a factor $\frac{1}{\alpha!\beta!\gamma!}$.

A.1. Boundary conditions

We now study a quite general case for fitting a defined set of velocity moments, up to order $2(n - 1)$, with a maximum entropy velocity distribution containing a polynomial of degree n, which allows a simple and linear estimation of the polynomial coefficients. By using Latin indices notation for $\mathcal{P}_n$, according to Eq. (25), we assume that all the moments in Eq. (2) exist, which is equivalent to considering the distribution function to be a square-integrable function in the velocity domain Γ_V. The scalar λ_0 is the normalisation factor, and, in general, all of the above coefficients are symmetric elements of k-rank tensors λ_k; $k = 0, \ldots n$ involved in Eq. (25).

The other coefficients than λ_0 can be obtained by using the property

$$\int_{\Gamma_V} \nabla_V \left[(V)^m e^{\mathcal{P}_n} \right] d^3 V = (0)^{n+1}, \tag{28}$$

which is a direct consequence of Eq. (11) and, in particular, is fulfilled by any solution of the moment equations.

The above integral is an $(n + 1)$-rank tensor, which is symmetric with respect to the indices of the tensor power $(V)^m$. Thus, when integrating Eq. (28) by components and the conditions of Eq. (11) are applied over the domain of any variable $V_{i_{m+1}}$, we get

$$\int_{\Gamma_{V_{i_{m+1}}}} \frac{\partial (V_{i_1} \ldots V_{i_m} e^{\mathcal{P}_n})}{\partial V_{i_{m+1}}} dV_{i_{m+1}} = V_{i_1} \ldots V_{i_m} e^{\mathcal{P}_n} |_{\Gamma_{V_{i_{m+1}}}} = 0. \tag{29}$$

In the case of a finite velocity domain, if the density function is bell-shaped, the null value of the righthand side might be substituted by a tolerance error, namely

$$\varepsilon_{i_{m+1}} = V_{i_1} \ldots V_{i_m} e^{\mathcal{P}_n} |_{\Gamma_{V_{i_{m+1}}}}, \tag{30}$$

such that this value can be neglected on condition of being significantly small[4].

In particular, for $m = 0$, since

$$\frac{\partial e^{\mathcal{P}_n}}{\partial V_k} = e^{\mathcal{P}_n} \frac{\partial \mathcal{P}_n}{\partial V_k} \tag{31}$$

we have

$$\int_{\Gamma_V} \frac{\partial \mathcal{P}_n}{\partial V_k} e^{\mathcal{P}_n} d^3 V = 0. \tag{32}$$

Similarly, for $m = 1$,

$$\int_{\Gamma_V} \frac{\partial}{\partial V_k} (V_i e^{\mathcal{P}_n}) d^3 V = \int_{\Gamma_V} \delta_{ik} e^{\mathcal{P}_n} d^3 V$$
$$+ \int_{\Gamma_V} V_i \frac{\partial \mathcal{P}_n}{\partial V_k} e^{\mathcal{P}_n} d^3 V = 0 \tag{33}$$

where δ_{ik} is the Kronecker delta.

---

[4] For example, the velocity density function f of local stars is approximately Gaussian in the component W perpendicular to the galactic plane, with dispersion $\sigma_W \simeq 19$ km s^{-1}. A finite velocity domain $I_W = [-W_{\max}, W_{\max}]$ could be then assumed, with $W_{\max} = 220$ km s^{-1}, where the integral $\int_{I_W} \frac{\partial (W^n f)}{\partial W} dW$ is exactly null for even values of n, and remains less than $\sim 10^{-5}$ for odd values $n < 13$. Obviously, the integral is still lower for wider intervals. Similarly, the local young-disc stars, with absolute heliocentric velocity up to 51 km s^{-1}, have a velocity dispersion $\sigma_W \simeq 11$ km s^{-1}. In the similar situation above, we may then assume a finite velocity domain for the truncated velocity distribution with $W_{\max} = 120$ km s^{-1}, where such an integral can be neglected up to powers $n = 12$.

And, in general, for $m \geq 2$, we get

$$\frac{\partial (V_{i_1} \ldots V_{i_m} e^{\mathcal{P}_n})}{\partial V_{i_{m+1}}} = (\delta_{i_1 i_{m+1}} V_{i_2} \ldots V_{i_m} + \ldots$$
$$+ \delta_{i_j i_{m+1}} V_{i_1} \ldots \widehat{V_{i_j}} \ldots V_{i_m} + \ldots$$
$$+ \delta_{i_m i_{m+1}} V_{i_1} \ldots V_{i_{m-1}}) e^{\mathcal{P}_n} + V_{i_1} \ldots V_{i_m} \frac{\partial e^{\mathcal{P}_n}}{\partial V_{i_{m+1}}} \tag{34}$$

where the hat indicates the omitted factors. Once more, bearing Eq. (31) in mind, the identity Eq. (28) yields

$$\int_{\Gamma_V} \left(\delta_{i_1 i_{m+1}} V_{i_2} \ldots V_{i_m} + \ldots + \delta_{i_j i_{m+1}} V_{i_1} \ldots \widehat{V_{i_j}} \ldots V_{i_m} + \ldots \right.$$
$$\left. + \delta_{i_m i_{m+1}} V_{i_1} \ldots V_{i_{m-1}} \right) e^{\mathcal{P}_n} d^3 V$$
$$+ \int_{\Gamma_V} V_{i_1} \ldots V_{i_m} \frac{\partial \mathcal{P}_n}{\partial V_{i_{m+1}}} e^{\mathcal{P}_n} d^3 V = 0. \tag{35}$$

Since the first integral is symmetric with respect to permutation of indices, and in general it is not null, then the second integral

$$q_{m+1} = \int_{\Gamma_V} (V)^m \otimes (\nabla_V \mathcal{P}_n) e^{\mathcal{P}_n} d^3 V \tag{36}$$

must be symmetric, too. Indeed, Eqs. (35) and (36) are equivalent to those obtained in Cubarsi (2007) as Eqs. (22) and (29), which were derived for expressing the conservation of pressures.

In this new context, the above identities will provide a linear method of fitting any desired set of moments. In contrast to the usual maximum entropy methods for the moments problem, which are nonlinear and not well conditioned enough, the present method allows all of the coefficients to be determined with accuracy.

First we evaluate $\nabla_V \mathcal{P}_n$ starting from Eq. (25),

$$\frac{\partial \mathcal{P}_n}{\partial V_k} = \lambda_k + 2\lambda_{j_1 k} V_{j_1} + 3\lambda_{j_1 j_2 k} V_{j_1} V_{j_2} + \cdots$$
$$+ n\lambda_{j_1 j_2 \ldots j_{n-1} k} V_{j_1} \cdots V_{j_{n-1}}. \tag{37}$$

To obtain all of the elements of tensors λ_k; $k = 1, \ldots, n$, we compute the integrals of Eq. (28) for m from 0 to $n - 1$. For $m = 0$, by taking Eqs. (32) and (37) into account, and by using the moments definition Eq. (1), since $m_0 = 1$, we have

$$\lambda_k + 2\lambda_{j_1 k} m_{j_1} + 3\lambda_{j_1 j_2 k} m_{j_1 j_2} + \cdots + n\lambda_{j_1 j_2 \ldots j_{n-1} k} m_{j_1 \ldots j_{n-1}} = 0, \tag{38}$$

which stands for a set of 3 scalar equations, $k = 1, 2, 3$. For $m = 1$, also by taking Eqs. (33) and (37) into account, we get

$$\delta_{ik} + \lambda_k m_i + 2\lambda_{j_1 k} m_{j_1 i} + 3\lambda_{j_1 j_2 k} m_{j_1 j_2 i} + \cdots$$
$$+ n\lambda_{j_1 j_2 \ldots j_{n-1} k} m_{j_1 j_2 \ldots j_{n-1} i} = 0. \tag{39}$$

Hence, this set of relations, for $i, k = 1, 2, 3$, thanks to the symmetry of Eq. (36), provides 6 independent scalar equations.

And, in general, for $m = n - 1$, from Eq. (35) we likewise get

$$\delta_{i_1 i_n} m_{i_2 \ldots i_{n-1}} + \ldots + \delta_{i_j i_n} m_{i_1 \ldots \widehat{i_j} \ldots i_{n-1}} + \ldots + \delta_{i_{n-1} i_n} m_{i_1 \ldots i_{n-2}} + \lambda_{i_n} m_{i_1 \ldots i_{n-1}}$$
$$+ 2\lambda_{j_1 i_n} m_{j_1 i_1 \ldots i_{n-1}} + 3\lambda_{j_1 j_2 i_n} m_{j_1 j_2 i_1 \ldots i_{n-1}} + \cdots$$
$$+ n\lambda_{j_1 j_2 \ldots j_{n-1} i_n} m_{j_1 \cdots j_{n-1} i_1 \ldots i_{n-1}} = 0, \tag{40}$$

which consists in a set of $\binom{n+2}{2}$ independent scalar equations, $i_1, \ldots, i_n = 1, 2, 3$, owing to the symmetry of Eq. (36). Therefore, we have as many independent linear equations as unknowns composing the elements of the symmetric tensors λ_k, for $k = 1, \ldots n$, whose elements are the coefficients of $\mathcal{P}_n$. Such

Table A.1. Matrix $\mathbf{G}_2$ is a symmetric matrix of inner products of the velocity components $\langle V_0 V_i V_j ..., \ V_0 V_p V_q ...\rangle$, according to Latin indices, with $V_0 \equiv 1$ and the other indices sorted as $1 \le i \le j \le ... \le 3$ and $1 \le p \le q \le ... \le 3^{\dagger}$.

	0	1	2	3	11	12	13	22	23	33	111	112	$\cdots$	333	$\cdots$
0	1	m_1	m_2	m_3	m_{11}	m_{12}	m_{13}	m_{22}	m_{23}	m_{33}	m_{111}	m_{112}	$\cdots$	m_{333}	$\cdots$
1		m_{11}	m_{12}	m_{13}	m_{111}	m_{112}	m_{113}	m_{122}	m_{123}	m_{133}	m_{1111}	m_{1112}	$\cdots$	m_{1333}	$\cdots$
2			m_{22}	m_{23}	m_{112}	m_{122}	m_{123}	m_{222}	m_{223}	m_{233}	m_{1112}	m_{1122}	$\cdots$	m_{2333}	$\cdots$
3				m_{33}	m_{113}	m_{123}	m_{133}	m_{223}	m_{233}	m_{333}	m_{1113}	m_{1123}	$\cdots$	m_{3333}	$\cdots$
11					m_{1111}	m_{1112}	m_{1113}	m_{1122}	m_{1123}	m_{1133}	m_{11111}	m_{11112}	$\cdots$	m_{11333}	$\cdots$
12						m_{1122}	m_{1123}	m_{1222}	m_{1223}	m_{1233}	m_{11112}	m_{11122}	$\cdots$	m_{12333}	$\cdots$
13							m_{1133}	m_{1223}	m_{1233}	m_{1333}	m_{11113}	m_{11123}	$\cdots$	m_{13333}	$\cdots$
22								m_{2222}	m_{2223}	m_{2233}	m_{11122}	m_{11222}	$\cdots$	m_{22333}	$\cdots$
23									m_{2233}	m_{2333}	m_{11123}	m_{11223}	$\cdots$	m_{23333}	$\cdots$
33										m_{3333}	m_{11133}	m_{11233}	$\cdots$	m_{33333}	$\cdots$
111											m_{111111}	m_{111112}	$\cdots$	m_{111333}	$\cdots$
112												m_{111122}	$\cdots$	m_{112333}	$\cdots$
$\vdots$													$\vdots$	$\vdots$	$\vdots$
333														m_{333333}	$\cdots$
$\vdots$															$\vdots$

Notes. [†] The external first row and first column refer to the velocity indices. Since the matrix is symmetric, only the diagonal and upper triangular part are written.

a non-homogeneous system can be associated with a Gramian matrix, as we see in the next section.

Finally, the scalar λ_0 left to be evaluated may be obtained as the normalisation factor to satisfy

$$\int_{\Gamma_V} e^{\mathcal{P}_n} d^3 V = 1$$

$$e^{-\lambda_0} = \int_{\Gamma_V} e^{\lambda_i V_i + \lambda_{ij} V_i V_j + \cdots + \lambda_{i_1...i_n} V_{i_1} \cdots V_{i_n}} d^3 V. \quad (41)$$

A.2. Gramian system

The three scalar equations involved in Eq. (38), corresponding to $m = 0$, for $k = 1, 2, 3$, are homogeneous in the elements of tensors λ_k. In Eq. (39), for $m = 1$, we group the terms containing the elements of λ_k, by writing the other ones on the righthand side, and likewise for the general equation with $m = n - 1$, Eq. (40). Thus we obtain the following linear system of equations for the elements of tensors λ_k,

$$\lambda_k + 2\lambda_{j_1 k} m_{j_1} + 3\lambda_{j_1 j_2 k} m_{j_1 j_2} + \cdots + n\lambda_{j_1 j_2 ... j_{n-1} k} m_{j_1...j_{n-1}} = 0$$

$$\lambda_k m_i + 2\lambda_{j_1 k} m_{j_1 i} + 3\lambda_{j_1 j_2 k} m_{j_1 j_2 i} + \cdots$$

$$+ n\lambda_{j_1 j_2 ... j_{n-1} k} m_{j_1 j_2 ... j_{n-1} i} = -\delta_{ik}$$

$$\vdots$$

$$\lambda_{i_n} m_{i_1...i_{n-1}} + 2\lambda_{j_1 i_n} m_{j_1 i_1...i_{n-1}} + 3\lambda_{j_1 j_2 i_n} m_{j_1 j_2 i_1...i_{n-1}} + \cdots$$

$$+ n\lambda_{j_1 j_2 ... j_{n-1} i_n} m_{j_1 \cdots j_{n-1} i_1 ... i_{n-1}}$$

$$= -\left(\delta_{i_1 i_n} m_{i_2...i_{n-1}} + ... + \delta_{i_j i_n} m_{i_1...\widehat{i_j}...i_{n-1}} + \cdots \right.$$

$$\left. + \delta_{i_{n-1} i_n} m_{i_1...i_{n-2}}\right). \quad (42)$$

Such a system of equations can be grouped according to three different vectors on its righthand side, for $k = 1, 2, 3$ in the

first two equations and for $i_n = 1, 2, 3$ in the general expression. A similar procedure can be applied to the λ_k coefficients.

The system matrix, $\mathbf{G}_2$ shown in Table A.1, can be interpreted as a symmetric matrix of inner products with respect to the weight $e^{\mathcal{P}_n}$ of the velocity components $\langle V_0 V_i V_j ..., V_0 V_p V_q ...\rangle$ according to the Latin indices notation, with $V_0 \equiv 1$ and the other indices sorted as $1 \le i \le j \le ... \le 3$ and $1 \le p \le q \le ... \le 3$. Therefore, $\mathbf{G}_2$ is a Gram matrix, symmetric, positive-definite and, among other well-known properties, it is invertible, meaning the system has a unique solution. Thus, we may define, according to Table A.2, the following three column matrices $\mathbf{X} = [a, b, c]$ and $\mathbf{Y} = [A, B, C]$ so that the following equality is satisfied:

$$\mathbf{Y} = \mathbf{G}_2 \mathbf{X}. \quad (43)$$

This is the numerical form of the system of equations Eq. (42), ready to be solved.

The coefficients to compute are elements of the symmetric tensors λ_k, for orders $k = 1, ..., n$, since the zero order coefficient is the normalisation factor. In total there are $\binom{n+3}{3} - 1$ independent coefficients. Each column of the matrix $\mathbf{X}$ is composed of:

– one element of the symmetric tensor λ_1, which multiplies the moments of orders $0, 1, ..., n - 1$ in the first column of matrix $\mathbf{G}_2$;
– three elements of λ_2, which multiply the moments of orders $1, 2, ..., n$ in the next three columns of matrix $\mathbf{G}_2$;
– and, in general, $\binom{k+2}{2}$ elements of the symmetric tensor λ_{k+1}, which multiply the moments of orders $k, ..., k + n - 1$, up to the value $k = n - 1$.

Therefore the matrix $\mathbf{G}_2$ has $\binom{n+2}{3}$ rows and columns, where the moments up to order $2(n-1)$ are involved. For example, for $n = 2$ we use the matrix $\mathbf{G}_2$ with the first row containing moments up to first order ($1 + 3 = 4$ columns in total) and the last row containing moments up to order 2, a 4×4 matrix. For $n = 4$ we use the matrix $\mathbf{G}_2$ with the first row containing moments up to order 3

Table A.2. The system of Eqs. (42) grouped as a three column matrix, on their righthand side.

	A	B	C	a	b	c
$m=0$	0	0	0	λ_1	λ_2	λ_3
$m=1$	-1	0	0	$2\lambda_{11}$	$2\lambda_{21}$	$2\lambda_{31}$
	0	-1	0	$2\lambda_{12}$	$2\lambda_{22}$	$2\lambda_{32}$
	0	0	-1	$2\lambda_{13}$	$2\lambda_{23}$	$2\lambda_{33}$
$m=2$	$-2\,m_1$	0	0	$3\,\lambda_{111}$	$3\,\lambda_{211}$	$3\,\lambda_{311}$
	$-m_2$	$-m_1$	0	$2\cdot3\,\lambda_{112}$	$2\cdot3\,\lambda_{212}$	$2\cdot3\,\lambda_{312}$
	$-m_3$	0	$-m_1$	$2\cdot3\,\lambda_{113}$	$2\cdot3\,\lambda_{213}$	$2\cdot3\,\lambda_{313}$
	0	$-2\,m_2$	0	$3\,\lambda_{122}$	$3\,\lambda_{222}$	$3\,\lambda_{322}$
	0	$-m_3$	$-m_2$	$2\cdot3\,\lambda_{123}$	$2\cdot3\,\lambda_{223}$	$2\cdot3\,\lambda_{323}$
	0	0	$-2\,m_3$	$3\,\lambda_{133}$	$3\,\lambda_{233}$	$3\,\lambda_{333}$
$m=3$	$-3\,m_{11}$	0	0	$4\,\lambda_{1111}$	$4\,\lambda_{2111}$	$4\,\lambda_{3111}$
	$-2\,m_{12}$	$-m_{11}$	0	$3\cdot4\,\lambda_{1112}$	$3\cdot4\,\lambda_{2112}$	$3\cdot4\,\lambda_{3112}$
	$-2\,m_{13}$	0	$-m_{11}$	$3\cdot4\,\lambda_{1113}$	$3\cdot4\,\lambda_{2113}$	$3\cdot4\,\lambda_{3113}$
	$-m_{22}$	$-2\,m_{12}$	0	$3\cdot4\,\lambda_{1122}$	$3\cdot4\,\lambda_{2122}$	$3\cdot4\,\lambda_{3122}$
	$-m_{23}$	$-m_{13}$	$-m_{12}$	$6\cdot4\,\lambda_{1123}$	$6\cdot4\,\lambda_{2123}$	$6\cdot4\,\lambda_{3123}$
	$-m_{33}$	0	$-2\,m_{13}$	$3\cdot4\,\lambda_{1133}$	$3\cdot4\,\lambda_{2133}$	$3\cdot4\,\lambda_{3133}$
	0	$-3\,m_{22}$	0	$4\,\lambda_{1222}$	$4\,\lambda_{2222}$	$4\,\lambda_{3222}$
	0	$-2\,m_{23}$	$-m_{22}$	$3\cdot4\,\lambda_{1223}$	$3\cdot4\,\lambda_{2223}$	$3\cdot4\,\lambda_{3223}$
	0	$-m_{33}$	$-2\,m_{23}$	$3\cdot4\,\lambda_{1233}$	$3\cdot4\,\lambda_{2233}$	$3\cdot4\,\lambda_{3233}$
	0	0	$-3\,m_{33}$	$4\,\lambda_{1333}$	$4\,\lambda_{2333}$	$4\,\lambda_{3333}$
$m=4$	$-4\,m_{111}$	0	0	$5\,\lambda_{11111}$	$5\,\lambda_{21111}$	$5\,\lambda_{31111}$
	$-3\,m_{112}$	$-m_{111}$	0	$4\cdot5\,\lambda_{11112}$	$4\cdot5\,\lambda_{21112}$	$4\cdot5\,\lambda_{31112}$
	$-3\,m_{113}$	0	$-m_{111}$	$4\cdot5\,\lambda_{11113}$	$4\cdot5\,\lambda_{21113}$	$4\cdot5\,\lambda_{31113}$
	$-2\,m_{122}$	$-2\,m_{112}$	0	$6\cdot5\,\lambda_{11122}$	$6\cdot5\,\lambda_{21122}$	$6\cdot5\,\lambda_{31122}$
	$-2\,m_{123}$	$-m_{113}$	$-m_{112}$	$12\cdot5\,\lambda_{11123}$	$12\cdot5\,\lambda_{21123}$	$12\cdot5\,\lambda_{31123}$
	$-2\,m_{133}$	0	$-2\,m_{113}$	$6\cdot5\,\lambda_{11133}$	$6\cdot5\,\lambda_{21133}$	$6\cdot5\,\lambda_{31133}$
	$-m_{222}$	$-3\,m_{122}$	0	$4\cdot5\,\lambda_{11222}$	$4\cdot5\,\lambda_{21222}$	$4\cdot5\,\lambda_{31222}$
	$-m_{223}$	$-2\,m_{123}$	$-m_{122}$	$12\cdot5\,\lambda_{11223}$	$12\cdot5\,\lambda_{21223}$	$12\cdot5\,\lambda_{31223}$
	$-m_{233}$	$-m_{133}$	$-2\,m_{123}$	$12\cdot5\,\lambda_{11233}$	$12\cdot5\,\lambda_{21233}$	$12\cdot5\,\lambda_{31233}$
	$-m_{333}$	0	$-3\,m_{133}$	$4\cdot5\,\lambda_{11333}$	$4\cdot5\,\lambda_{21333}$	$4\cdot5\,\lambda_{31333}$
	0	$-4\,m_{222}$	0	$5\,\lambda_{12222}$	$5\,\lambda_{22222}$	$5\,\lambda_{32222}$
	0	$-3\,m_{223}$	$-m_{222}$	$4\cdot5\,\lambda_{12223}$	$4\cdot5\,\lambda_{22223}$	$4\cdot5\,\lambda_{32223}$
	0	$-2\,m_{233}$	$-2\,m_{223}$	$6\cdot5\,\lambda_{12233}$	$6\cdot5\,\lambda_{22233}$	$6\cdot5\,\lambda_{32233}$
	0	$-m_{333}$	$-3\,m_{233}$	$4\cdot5\,\lambda_{12333}$	$4\cdot5\,\lambda_{22333}$	$4\cdot5\,\lambda_{32333}$
	0	0	$-4\,m_{333}$	$5\,\lambda_{13333}$	$5\,\lambda_{23333}$	$5\,\lambda_{33333}$
$m=5$	$-5\,m_{1111}$	0	0	$6\,\lambda_{111111}$	$6\,\lambda_{211111}$	$6\,\lambda_{311111}$
	$-4\,m_{1112}$	$-m_{1111}$	0	$5\cdot6\,\lambda_{111112}$	$5\cdot6\,\lambda_{211112}$	$5\cdot6\,\lambda_{311112}$
	$-4\,m_{1113}$	0	$-m_{1111}$	$5\cdot6\,\lambda_{111113}$	$5\cdot6\,\lambda_{211113}$	$5\cdot6\,\lambda_{311113}$
	$-3\,m_{1122}$	$-2\,m_{1112}$	0	$10\cdot6\,\lambda_{111122}$	$10\cdot6\,\lambda_{211122}$	$10\cdot6\,\lambda_{311122}$
	$-3\,m_{1123}$	$-m_{1113}$	$-m_{1112}$	$20\cdot6\,\lambda_{111123}$	$20\cdot6\,\lambda_{211123}$	$20\cdot6\,\lambda_{311123}$
	$-3\,m_{1133}$	0	$-2\,m_{1113}$	$10\cdot6\,\lambda_{111133}$	$10\cdot6\,\lambda_{211133}$	$10\cdot6\,\lambda_{311133}$
	$-2\,m_{1222}$	$-3\,m_{1122}$	0	$10\cdot6\,\lambda_{111222}$	$10\cdot6\,\lambda_{211222}$	$10\cdot6\,\lambda_{311222}$
	$-2\,m_{1223}$	$-2\,m_{1123}$	$-m_{1122}$	$30\cdot6\,\lambda_{111223}$	$30\cdot6\,\lambda_{211223}$	$30\cdot6\,\lambda_{311223}$
	$-2\,m_{1233}$	$-m_{1133}$	$-2\,m_{1123}$	$30\cdot6\,\lambda_{111233}$	$30\cdot6\,\lambda_{211233}$	$30\cdot6\,\lambda_{311233}$
	$-2\,m_{1333}$	0	$-3\,m_{1133}$	$10\cdot6\,\lambda_{111333}$	$10\cdot6\,\lambda_{211333}$	$10\cdot6\,\lambda_{311333}$
	$-m_{2222}$	$-4\,m_{1222}$	0	$5\cdot6\,\lambda_{112222}$	$5\cdot6\,\lambda_{212222}$	$5\cdot6\,\lambda_{312222}$
	$-m_{2223}$	$-3\,m_{1223}$	$-m_{1222}$	$20\cdot6\,\lambda_{112223}$	$20\cdot6\,\lambda_{212223}$	$20\cdot6\,\lambda_{312223}$
	$-m_{2233}$	$-2\,m_{1233}$	$-2\,m_{1223}$	$30\cdot6\,\lambda_{112233}$	$30\cdot6\,\lambda_{212233}$	$30\cdot6\,\lambda_{312233}$
	$-m_{2333}$	$-m_{1333}$	$-3\,m_{1233}$	$20\cdot6\,\lambda_{112333}$	$20\cdot6\,\lambda_{212333}$	$20\cdot6\,\lambda_{312333}$
	$-m_{3333}$	0	$-4\,m_{1333}$	$5\cdot6\,\lambda_{113333}$	$5\cdot6\,\lambda_{213333}$	$5\cdot6\,\lambda_{313333}$
	0	$-5\,m_{2222}$	0	$6\,\lambda_{122222}$	$6\,\lambda_{222222}$	$6\,\lambda_{322222}$
	0	$-4\,m_{2223}$	$-m_{2222}$	$5\cdot6\,\lambda_{122223}$	$5\cdot6\,\lambda_{222223}$	$5\cdot6\,\lambda_{322223}$
	0	$-3\,m_{2233}$	$-2\,m_{2223}$	$10\cdot6\,\lambda_{122233}$	$10\cdot6\,\lambda_{222233}$	$10\cdot6\,\lambda_{322233}$
	0	$-2\,m_{2333}$	$-3\,m_{2233}$	$10\cdot6\,\lambda_{122333}$	$10\cdot6\,\lambda_{222333}$	$10\cdot6\,\lambda_{322333}$
	0	$-m_{3333}$	$-4\,m_{2333}$	$5\cdot6\,\lambda_{123333}$	$5\cdot6\,\lambda_{223333}$	$5\cdot6\,\lambda_{323333}$
	0	0	$-5\,m_{3333}$	$6\,\lambda_{133333}$	$6\,\lambda_{233333}$	$6\,\lambda_{333333}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$(1+3+6+10=20$ columns in total) and the last row containing moments up to order 6, as a 20×20 matrix. Similarly, for $n=6$ we use the matrix $\mathbf{G}_2$ with the first row containing moments up to order 5 $(1+3+6+10+15+21=56$ columns in total) and the last row containing moments up to order 10, as a 56×56 matrix.

On the other hand, since the matrices $\mathbf{X}$ and $\mathbf{Y}$ consist of three column vectors, we dispose of a number of $3\binom{n+2}{3}$ equations. This number, for $n>1$, is always greater than the number of independent unknowns (leaving out the normalisation factor). For example, in the case $n=2$, we have 12 equations and

Table A.3. Coefficient submatrix relating the first column A of matrix $\mathbf{Y}$ and the first column a of matrix $\mathbf{X}$.

$$
\begin{bmatrix}
1 & 0 & 0 & m_1 & m_2 & m_3 & 0 & 0 & 0 & m_{11} & 2m_{12} & 2m_{13} & m_{22} & 2m_{23} & m_{33} & 0 & 0 & 0 & 0 & \cdots \\
m_1 & 0 & 0 & m_{11} & m_{12} & m_{13} & 0 & 0 & 0 & m_{111} & 2m_{112} & 2m_{113} & m_{122} & 2m_{123} & m_{133} & 0 & 0 & 0 & 0 & \cdots \\
m_2 & 0 & 0 & m_{12} & m_{22} & m_{23} & 0 & 0 & 0 & m_{112} & 2m_{122} & 2m_{123} & m_{222} & 2m_{223} & m_{233} & 0 & 0 & 0 & 0 & \cdots \\
m_3 & 0 & 0 & m_{13} & m_{23} & m_{33} & 0 & 0 & 0 & m_{113} & 2m_{123} & 2m_{133} & m_{223} & 2m_{233} & m_{333} & 0 & 0 & 0 & 0 & \cdots \\
m_{11} & 0 & 0 & m_{111} & m_{112} & m_{113} & 0 & 0 & 0 & m_{1111} & 2m_{1112} & 2m_{1113} & m_{1122} & 2m_{1123} & m_{1133} & 0 & 0 & 0 & 0 & \cdots \\
m_{12} & 0 & 0 & m_{112} & m_{122} & m_{123} & 0 & 0 & 0 & m_{1112} & 2m_{1122} & 2m_{1123} & m_{1222} & 2m_{1223} & m_{1233} & 0 & 0 & 0 & 0 & \cdots \\
m_{13} & 0 & 0 & m_{113} & m_{123} & m_{133} & 0 & 0 & 0 & m_{1113} & 2m_{1123} & 2m_{1133} & m_{1223} & 2m_{1233} & m_{1333} & 0 & 0 & 0 & 0 & \cdots \\
m_{22} & 0 & 0 & m_{122} & m_{222} & m_{223} & 0 & 0 & 0 & m_{1122} & 2m_{1222} & 2m_{1223} & m_{2222} & 2m_{2223} & m_{2233} & 0 & 0 & 0 & 0 & \cdots \\
m_{23} & 0 & 0 & m_{123} & m_{223} & m_{233} & 0 & 0 & 0 & m_{1123} & 2m_{1223} & 2m_{1233} & m_{2223} & 2m_{2233} & m_{2333} & 0 & 0 & 0 & 0 & \cdots \\
m_{33} & 0 & 0 & m_{133} & m_{233} & m_{333} & 0 & 0 & 0 & m_{1133} & 2m_{1233} & 2m_{1333} & m_{2233} & 2m_{2333} & m_{3333} & 0 & 0 & 0 & 0 & \cdots \\
m_{111} & 0 & 0 & m_{1111} & m_{1112} & m_{1113} & 0 & 0 & 0 & m_{11111} & 2m_{11112} & 2m_{11113} & m_{11122} & 2m_{11123} & m_{11133} & 0 & 0 & 0 & 0 & \cdots \\
m_{112} & 0 & 0 & m_{1112} & m_{1122} & m_{1123} & 0 & 0 & 0 & m_{11112} & 2m_{11122} & 2m_{11123} & m_{11222} & 2m_{11223} & m_{11233} & 0 & 0 & 0 & 0 & \cdots \\
m_{113} & 0 & 0 & m_{1113} & m_{1123} & m_{1133} & 0 & 0 & 0 & m_{11113} & 2m_{11123} & 2m_{11133} & m_{11223} & 2m_{11233} & m_{11333} & 0 & 0 & 0 & 0 & \cdots \\
m_{122} & 0 & 0 & m_{1122} & m_{1222} & m_{1223} & 0 & 0 & 0 & m_{11122} & 2m_{11222} & 2m_{11223} & m_{12222} & 2m_{12223} & m_{12233} & 0 & 0 & 0 & 0 & \cdots \\
m_{123} & 0 & 0 & m_{1123} & m_{1223} & m_{1233} & 0 & 0 & 0 & m_{11123} & 2m_{11223} & 2m_{11233} & m_{12223} & 2m_{12233} & m_{12333} & 0 & 0 & 0 & 0 & \cdots \\
m_{133} & 0 & 0 & m_{1133} & m_{1233} & m_{1333} & 0 & 0 & 0 & m_{11133} & 2m_{11233} & 2m_{11333} & m_{12233} & 2m_{12333} & m_{13333} & 0 & 0 & 0 & 0 & \cdots \\
m_{222} & 0 & 0 & m_{1222} & m_{2222} & m_{2223} & 0 & 0 & 0 & m_{11222} & 2m_{12222} & 2m_{12223} & m_{22222} & 2m_{22223} & m_{22233} & 0 & 0 & 0 & 0 & \cdots \\
m_{223} & 0 & 0 & m_{1223} & m_{2223} & m_{2233} & 0 & 0 & 0 & m_{11223} & 2m_{12223} & 2m_{12233} & m_{22223} & 2m_{22233} & m_{22333} & 0 & 0 & 0 & 0 & \cdots \\
m_{233} & 0 & 0 & m_{1233} & m_{2233} & m_{2333} & 0 & 0 & 0 & m_{11233} & 2m_{12233} & 2m_{12333} & m_{22233} & 2m_{22333} & m_{23333} & 0 & 0 & 0 & 0 & \cdots \\
m_{333} & 0 & 0 & m_{1333} & m_{2333} & m_{3333} & 0 & 0 & 0 & m_{11333} & 2m_{12333} & 2m_{13333} & m_{22333} & 2m_{23333} & m_{33333} & 0 & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{bmatrix}
$$

9 independent unknowns, because the symmetric coefficients λ_{12}, λ_{13}, and λ_{23} are equivalent to λ_{21}, λ_{31}, and λ_{32}, respectively, and similarly for higher values of n. In general, if the true distribution is indeed a maximum entropy distribution, the actual moments will be consistent with the symmetry of the coefficients, but a significant deviation from the maximum entropy property will produce some non-symmetric coefficients[5]. To avoid this situation, an equivalent overdeterminate system of equations is built up, as explained in the next section, where the symmetric coefficients of tensors λ_n will not be repeated in the vector of unknowns. The system is solved by applying a least squares method, so that to get the minimum squared error of the fit, it is weighted in terms of the inverse sampling variances σ^2 of the moments up to order $n - 2$, in the righthand side of Eq. (42). In addition, a predictor-corrector method is applied to evaluate the variance matrix of the unknowns, as detailed below.

A.3. Parameter estimation

The overdeterminate system of equations, which is equivalent to Eq. (43), is written as

$$y = \mathbf{g}_2 x, \tag{44}$$

where the only unknowns are the non-identical elements of the symmetric tensors λ_n. It takes the following form:

The first column A (Table A.2) of matrix $\mathbf{Y}$ and the first column a of matrix $\mathbf{X}$ are related by the coefficient submatrix of Table A.3. The second column B of matrix $\mathbf{Y}$ and the first column b of matrix $\mathbf{X}$ are related by the coefficient submatrix of Table A.4. The third column C of matrix $\mathbf{Y}$ and the third column c of matrix $\mathbf{X}$ are related by the coefficient submatrix of Table A.5.

The resulting $\mathbf{g}_2$ matrix is obtained by stacking the three foregoing submatrices. Vector x now takes the form

$$x = (\lambda_1, \lambda_2, \lambda_3, 2\lambda_{11}, 2\lambda_{12}, 2\lambda_{13}, 2\lambda_{22}, 2\lambda_{23}, 2\lambda_{33}, 3\lambda_{111}, 3\lambda_{112},$$

$$3\lambda_{113}, 3\lambda_{122}, 3\lambda_{123}, 3\lambda_{133}, 3\lambda_{222}, 3\lambda_{223}, 3\lambda_{233}, 3\lambda_{333}, ...)^{\mathrm{T}}. \tag{45}$$

The factors multiplying the elements of tensors λ_n in Table 2, other than those appearing in vector x, have been carried over the elements of matrix $\mathbf{g}_2$.

It is well known that, if V_y is the variance matrix for vector y, which is taken as the diagonal matrix of its sampling variances σ_y^2, then the least squares system weighted by V_y^{-1} provides minimum variance estimates for x according to (e.g. Stuart & Ord 1987)

$$x = (\mathbf{g}_2^{\mathrm{T}} V_y^{-1} \mathbf{g}_2)^{-1} \mathbf{g}_2^{\mathrm{T}} V_y^{-1} y. \tag{46}$$

---

[5] This is true for $n > 2$, but for $n = 2$ Appendix B shows that the coefficients λ_{ij} are related to the second central moments μ_{ij}^{-1}, which are necessarily symmetric.

Table A.4. Coefficient submatrix relating the second column B of matrix $\mathbf{Y}$ and the first column b of matrix $\mathbf{X}$.

$$
\begin{bmatrix}
0 & 1 & 0 & 0 & m_1 & 0 & m_2 & m_3 & 0 & 0 & m_{11} & 0 & 2\,m_{12} & 2\,m_{13} & 0 & m_{22} & 2\,m_{23} & m_{33} & 0 & \cdots \\
0 & m_1 & 0 & 0 & m_{11} & 0 & m_{12} & m_{13} & 0 & 0 & m_{111} & 0 & 2\,m_{112} & 2\,m_{113} & 0 & m_{122} & 2\,m_{123} & m_{133} & 0 & \cdots \\
0 & m_2 & 0 & 0 & m_{12} & 0 & m_{22} & m_{23} & 0 & 0 & m_{112} & 0 & 2\,m_{122} & 2\,m_{123} & 0 & m_{222} & 2\,m_{223} & m_{233} & 0 & \cdots \\
0 & m_3 & 0 & 0 & m_{13} & 0 & m_{23} & m_{33} & 0 & 0 & m_{113} & 0 & 2\,m_{123} & 2\,m_{133} & 0 & m_{223} & 2\,m_{233} & m_{333} & 0 & \cdots \\
0 & m_{11} & 0 & 0 & m_{111} & 0 & m_{112} & m_{113} & 0 & 0 & m_{1111} & 0 & 2\,m_{1112} & 2\,m_{1113} & 0 & m_{1122} & 2\,m_{1123} & m_{1133} & 0 & \cdots \\
0 & m_{12} & 0 & 0 & m_{112} & 0 & m_{122} & m_{123} & 0 & 0 & m_{1112} & 0 & 2\,m_{1122} & 2\,m_{1123} & 0 & m_{1222} & 2\,m_{1223} & m_{1233} & 0 & \cdots \\
0 & m_{13} & 0 & 0 & m_{113} & 0 & m_{123} & m_{133} & 0 & 0 & m_{1113} & 0 & 2\,m_{1123} & 2\,m_{1133} & 0 & m_{1223} & 2\,m_{1233} & m_{1333} & 0 & \cdots \\
0 & m_{22} & 0 & 0 & m_{122} & 0 & m_{222} & m_{223} & 0 & 0 & m_{1122} & 0 & 2\,m_{1222} & 2\,m_{1223} & 0 & m_{2222} & 2\,m_{2223} & m_{2233} & 0 & \cdots \\
0 & m_{23} & 0 & 0 & m_{123} & 0 & m_{223} & m_{233} & 0 & 0 & m_{1123} & 0 & 2\,m_{1223} & 2\,m_{1233} & 0 & m_{2223} & 2\,m_{2233} & m_{2333} & 0 & \cdots \\
0 & m_{33} & 0 & 0 & m_{133} & 0 & m_{233} & m_{333} & 0 & 0 & m_{1133} & 0 & 2\,m_{1233} & 2\,m_{1333} & 0 & m_{2233} & 2\,m_{2333} & m_{3333} & 0 & \cdots \\
0 & m_{111} & 0 & 0 & m_{1111} & 0 & m_{1112} & m_{1113} & 0 & 0 & m_{11111} & 0 & 2\,m_{11112} & 2\,m_{11113} & 0 & m_{11122} & 2\,m_{11123} & m_{11133} & 0 & \cdots \\
0 & m_{112} & 0 & 0 & m_{1112} & 0 & m_{1122} & m_{1123} & 0 & 0 & m_{11112} & 0 & 2\,m_{11122} & 2\,m_{11123} & 0 & m_{11222} & 2\,m_{11223} & m_{11233} & 0 & \cdots \\
0 & m_{113} & 0 & 0 & m_{1113} & 0 & m_{1123} & m_{1133} & 0 & 0 & m_{11113} & 0 & 2\,m_{11123} & 2\,m_{11133} & 0 & m_{11223} & 2\,m_{11233} & m_{11333} & 0 & \cdots \\
0 & m_{122} & 0 & 0 & m_{1122} & 0 & m_{1222} & m_{1223} & 0 & 0 & m_{11122} & 0 & 2\,m_{11222} & 2\,m_{11223} & 0 & m_{12222} & 2\,m_{12223} & m_{12233} & 0 & \cdots \\
0 & m_{123} & 0 & 0 & m_{1123} & 0 & m_{1223} & m_{1233} & 0 & 0 & m_{11123} & 0 & 2\,m_{11223} & 2\,m_{11233} & 0 & m_{12223} & 2\,m_{12233} & m_{12333} & 0 & \cdots \\
0 & m_{133} & 0 & 0 & m_{1133} & 0 & m_{1233} & m_{1333} & 0 & 0 & m_{11133} & 0 & 2\,m_{11233} & 2\,m_{11333} & 0 & m_{12233} & 2\,m_{12333} & m_{13333} & 0 & \cdots \\
0 & m_{222} & 0 & 0 & m_{1222} & 0 & m_{2222} & m_{2223} & 0 & 0 & m_{11222} & 0 & 2\,m_{12222} & 2\,m_{12223} & 0 & m_{22222} & 2\,m_{22223} & m_{22233} & 0 & \cdots \\
0 & m_{223} & 0 & 0 & m_{1223} & 0 & m_{2223} & m_{2233} & 0 & 0 & m_{11223} & 0 & 2\,m_{12223} & 2\,m_{12233} & 0 & m_{22223} & 2\,m_{22233} & m_{22333} & 0 & \cdots \\
0 & m_{233} & 0 & 0 & m_{1233} & 0 & m_{2233} & m_{2333} & 0 & 0 & m_{11233} & 0 & 2\,m_{12233} & 2\,m_{12333} & 0 & m_{22233} & 2\,m_{22333} & m_{23333} & 0 & \cdots \\
0 & m_{333} & 0 & 0 & m_{1333} & 0 & m_{2333} & m_{3333} & 0 & 0 & m_{11333} & 0 & 2\,m_{12333} & 2\,m_{13333} & 0 & m_{22333} & 2\,m_{23333} & m_{33333} & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{bmatrix}
$$

The minimum fitting error is then obtained from the weighted norm of the difference between the observed values and their theoretical predictions,

$$\chi^2 = (\boldsymbol{y}^{\mathrm{T}} - \boldsymbol{x}^{\mathrm{T}}\mathbf{g}_2^{\mathrm{T}})V_y^{-1}(\boldsymbol{y} - \mathbf{g}_2\boldsymbol{x}). \tag{47}$$

The error on the results of the least squares fit, that is the variance σ_x^2 of vector $\boldsymbol{x}$, is obtained from the diagonal matrix of

$$V_x = (\mathbf{g}_2^{\mathrm{T}} V_y^{-1} \mathbf{g}_2)^{-1}. \tag{48}$$

Some aspects of the fitting procedure must be pointed out:

(1) Some elements of vector $\boldsymbol{y}$ are exact values, -1 or 0 (left part of Table A.2), thus they have no associated error. However, an initial tolerance error may be assumed for vector $\boldsymbol{y}$, which can be associated with the finite domain beyond which the density function is negligible, as pointed out in Eq. (30). This tolerance error is assumed to be constant and significantly small for all the components (e.g. 10^{-6}), and it is added to the sampling error σ_y^2 of the data. The final norm of the quadratic error is computed as $f_0 = \|\sigma_y^2\|$.

(2) In the least squares method, it is generally assumed that the elements of matrix $\mathbf{g}_2$ are evaluated from exact values, and does not contribute to the error of the estimates, although it is not true in the current case. To evaluate the part of the total fitting error due to the elements of matrix $\mathbf{g}_2$, an iterative procedure is started by assuming equal uncertainties for all equations, which are normalised to constant norm f_0. Let us call its predicted quadratic error ϵ_I^2.

(3) By starting from the predicted quadratic error, successive evaluations of the variance matrix V_y from the error propagation formula

$$V_y = \mathbf{g}_2 V_x\, \mathbf{g}_2^{\mathrm{T}} \tag{49}$$

are carried out, so that we obtain a corrected quadratic error ϵ_F^2, which is used as a new predicted error, by normalising it to a constant norm f_0.

(4) The algorithm is stopped when a fixed point ϵ_0^2 is reached, that is, when ϵ_I^2 and ϵ_F^2 have the same direction. It is found that the final quadratic error ϵ_0^2 of the iterative process does not depend on the initial predicted error, so that the final variance matrix V_y is the result of a redistribution of weights provided by the matrix $\mathbf{g}_2$ of the least squares system.

(5) To compute the final fitting error χ^2, a total sampling variance $\varepsilon^2 = \epsilon_0^2 + \sigma_y^2$ is assumed as the sum of both independent quadratic errors. The fitting error χ^2 is only partially significant, since it is related to the initial value f_0, which depends on the initial errors of the data $\boldsymbol{y}$, although this way it is possible to compare the goodness of different fittings.

Appendix B

The Gramian system and the moment recurrence can be solved straightforwardly for the case $n = 2$, which corresponds to a Schwarzschild distribution. For the sake of simplicity, and

Table A.5. Coefficient submatrix relating the third column C of matrix $\mathbf{Y}$ and the third column c of matrix $\mathbf{X}$.

$$
\begin{bmatrix}
0 & 0 & 1 & 0 & 0 & m_1 & 0 & m_2 & m_3 & 0 & 0 & m_{11} & 0 & 2m_{12} & 2m_{13} & 0 & m_{22} & 2m_{23} & m_{33} & \cdots \\
0 & 0 & m_1 & 0 & 0 & m_{11} & 0 & m_{12} & m_{13} & 0 & 0 & m_{111} & 0 & 2m_{112} & 2m_{113} & 0 & m_{122} & 2m_{123} & m_{133} & \cdots \\
0 & 0 & m_2 & 0 & 0 & m_{12} & 0 & m_{22} & m_{23} & 0 & 0 & m_{112} & 0 & 2m_{122} & 2m_{123} & 0 & m_{222} & 2m_{223} & m_{233} & \cdots \\
0 & 0 & m_3 & 0 & 0 & m_{13} & 0 & m_{23} & m_{33} & 0 & 0 & m_{113} & 0 & 2m_{123} & 2m_{133} & 0 & m_{223} & 2m_{233} & m_{333} & \cdots \\
0 & 0 & m_{11} & 0 & 0 & m_{111} & 0 & m_{112} & m_{113} & 0 & 0 & m_{1111} & 0 & 2m_{1112} & 2m_{1113} & 0 & m_{1122} & 2m_{1123} & m_{1133} & \cdots \\
0 & 0 & m_{12} & 0 & 0 & m_{112} & 0 & m_{122} & m_{123} & 0 & 0 & m_{1112} & 0 & 2m_{1122} & 2m_{1123} & 0 & m_{1222} & 2m_{1223} & m_{1233} & \cdots \\
0 & 0 & m_{13} & 0 & 0 & m_{113} & 0 & m_{123} & m_{133} & 0 & 0 & m_{1113} & 0 & 2m_{1123} & 2m_{1133} & 0 & m_{1223} & 2m_{1233} & m_{1333} & \cdots \\
0 & 0 & m_{22} & 0 & 0 & m_{122} & 0 & m_{222} & m_{223} & 0 & 0 & m_{1122} & 0 & 2m_{1222} & 2m_{1223} & 0 & m_{2222} & 2m_{2223} & m_{2233} & \cdots \\
0 & 0 & m_{23} & 0 & 0 & m_{123} & 0 & m_{223} & m_{233} & 0 & 0 & m_{1123} & 0 & 2m_{1223} & 2m_{1233} & 0 & m_{2223} & 2m_{2233} & m_{2333} & \cdots \\
0 & 0 & m_{33} & 0 & 0 & m_{133} & 0 & m_{233} & m_{333} & 0 & 0 & m_{1133} & 0 & 2m_{1233} & 2m_{1333} & 0 & m_{2233} & 2m_{2333} & m_{3333} & \cdots \\
0 & 0 & m_{111} & 0 & 0 & m_{1111} & 0 & m_{1112} & m_{1113} & 0 & 0 & m_{11111} & 0 & 2m_{11112} & 2m_{11113} & 0 & m_{11122} & 2m_{11123} & m_{11133} & \cdots \\
0 & 0 & m_{112} & 0 & 0 & m_{1112} & 0 & m_{1122} & m_{1123} & 0 & 0 & m_{11112} & 0 & 2m_{11122} & 2m_{11123} & 0 & m_{11222} & 2m_{11223} & m_{11233} & \cdots \\
0 & 0 & m_{113} & 0 & 0 & m_{1113} & 0 & m_{1123} & m_{1133} & 0 & 0 & m_{11113} & 0 & 2m_{11123} & 2m_{11133} & 0 & m_{11223} & 2m_{11233} & m_{11333} & \cdots \\
0 & 0 & m_{122} & 0 & 0 & m_{1122} & 0 & m_{1222} & m_{1223} & 0 & 0 & m_{11122} & 0 & 2m_{11222} & 2m_{11223} & 0 & m_{12222} & 2m_{12223} & m_{12233} & \cdots \\
0 & 0 & m_{123} & 0 & 0 & m_{1123} & 0 & m_{1223} & m_{1233} & 0 & 0 & m_{11123} & 0 & 2m_{11223} & 2m_{11233} & 0 & m_{12223} & 2m_{12233} & m_{12333} & \cdots \\
0 & 0 & m_{133} & 0 & 0 & m_{1133} & 0 & m_{1233} & m_{1333} & 0 & 0 & m_{11133} & 0 & 2m_{11233} & 2m_{11333} & 0 & m_{12233} & 2m_{12333} & m_{13333} & \cdots \\
0 & 0 & m_{222} & 0 & 0 & m_{1222} & 0 & m_{2222} & m_{2223} & 0 & 0 & m_{11222} & 0 & 2m_{12222} & 2m_{12223} & 0 & m_{22222} & 2m_{22223} & m_{22233} & \cdots \\
0 & 0 & m_{223} & 0 & 0 & m_{1223} & 0 & m_{2223} & m_{2233} & 0 & 0 & m_{11223} & 0 & 2m_{12223} & 2m_{12233} & 0 & m_{22223} & 2m_{22233} & m_{22333} & \cdots \\
0 & 0 & m_{233} & 0 & 0 & m_{1233} & 0 & m_{2233} & m_{2333} & 0 & 0 & m_{11233} & 0 & 2m_{12233} & 2m_{12333} & 0 & m_{22233} & 2m_{22333} & m_{23333} & \cdots \\
0 & 0 & m_{333} & 0 & 0 & m_{1333} & 0 & m_{2333} & m_{3333} & 0 & 0 & m_{11333} & 0 & 2m_{12333} & 2m_{13333} & 0 & m_{22333} & 2m_{23333} & m_{33333} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{bmatrix}
$$

without losing generality, we use the central moments $\boldsymbol{\mu}_n$, so that $\boldsymbol{\mu}_1 = 0$. Then, the Eqs. (38) and (39), for $m = 0, 1$, become

$$\lambda_k = 0, \quad k = 1, 2, 3 \tag{50}$$

and

$$\delta_{ik} + 2\lambda_{jk}\mu_{ji} = 0 \implies \lambda_{ik} = -\frac{1}{2}\mu_{ik}^{-1}, \quad i, k = 1, 2, 3 \tag{51}$$

where μ_{ik}^{-1} are the elements of the inverse of the covariance matrix. Therefore, the above relation shows that the tensor of elements λ_{ik} is a definite negative form, and it leads to an integrable distribution function.

Now we can apply the same procedure for $m \geq 2$, to obtain higher order moments in terms of the second moments. Thus, for $m = 2$, according to Eq. (40), with $m = n - 1$, and bearing in mind Eq. (50), we have

$$\lambda_k\mu_{ij} + 2\lambda_{lk}\mu_{ijl} = 0 \implies \mu_{ijk} = 0, \quad i, j, k = 1, 2, 3. \tag{52}$$

The result accounts for the obvious symmetry of the distribution, with vanishing odd-order central moments.

Similarly, for $m = 3$, we get

$$\delta_{i_1i_4}\,\mu_{i_2i_3} + \delta_{i_2i_4}\,\mu_{i_1i_3} + \delta_{i_3i_4}\,\mu_{i_1i_2} + \lambda_{i_4}\mu_{i_1i_2i_3}$$
$$+2\lambda_{ji_4}\mu_{ji_1i_2i_3} = 0, \quad i_1, i_2, i_3, i_4 = 1, 2, 3. \tag{53}$$

Then, taking into account Eq. (51), we multiply by μ_{ki_4}. Since the third moments are null, by reordering indices we obtain the following moment recurrence relation:

$$\mu_{i_1i_2i_3i_4} = \mu_{i_1i_4}\,\mu_{i_2i_3} + \mu_{i_2i_4}\,\mu_{i_1i_3} + \mu_{i_3i_4}\,\mu_{i_1i_2}. \tag{54}$$

The above relationship is the well known property of a Gaussian distribution, which characterises it from having vanishing fourth cumulants.

And, in general, according to Eq. (40), for even m, we obtain a vanishing set of odd-order central moments, and, for odd m, we obtain the relation

$$\delta_{i_1i_{m+1}}\,\mu_{i_2\dots i_m} + \dots + \delta_{i_ji_{m+1}}\,\mu_{i_1\dots\widehat{i_j}\dots i_m} + \dots$$
$$+\delta_{i_mi_{m+1}}\,\mu_{i_1\dots i_{m-1}} + 2\lambda_{ji_{m+1}}\mu_{ji_1\dots i_m} = 0. \tag{55}$$

Once again, by multiplying by $\mu_{ki_{m+1}}$ and by reordering indices, we get

$$\mu_{i_1\dots i_mi_{m+1}} = \mu_{i_1i_{m+1}}\,\mu_{i_2\dots i_m} + \dots$$
$$+\mu_{i_ji_{m+1}}\,\mu_{i_1\dots\widehat{i_j}\dots i_m} + \dots + \mu_{i_mi_{m+1}}\,\mu_{i_1\dots i_{m-1}}. \tag{56}$$

This is the general relationship of moment recurrence for trivariate normal distributions, which leads to a vanishing set higher-order cumulants.

References

Alcobé, S., & Cubarsi, R. 2005, A&A, 442, 929

Binney, J., & Tremaine, S. 1987, in Galactic Dynamics (Princeton: Princeton University Press)

Binney, J., Dehnen, W., Houk, N., Murray, C. A., & Penston, M. J. 1997, in Hipparcos- Venice'97, ESA SP-402, 473

Bovy, J., Hogg, D. W., & Roweis, S. T. 2009, ApJ, 700, 1794

Cercignani, C. 1988, in The Boltzmann Equation and its Applications (Berlin: Springer-Verlag)

Chandrasekhar, S. 1942, in Principles of Stellar Dynamics (New York: Dover Publications Inc.)

Charlier, C. W. L. 1926, in The Motion and the Distribution of the Stars, Berkeley, University of California

Cubarsi, R. 2007, MNRAS, 207, 380

Cubarsi, R., & Alcobé, S. 2004, A&A, 427, 131

Cubarsi, R., & Alcobé, S. 2006, A&A, 457, 537

Cubarsi, R., Sanz, J., & Juan, J. M. 1990, Ap&SS, 170, 197

Cubarsi, R., Alcobé, S., Vidojević, S., & Ninković, S. 2010, A&A, 510, A102

de Bruin, R., Salomé, D., & Schaafsma, W. 1999, Comput. Statist. Data Anal., 30, 19

Dehnen, W. 1998, AJ, 115, 2384

Dehnen, W. 1999, AJ, 118, 1201

Dehnen, W., & Binney, J. J. 1998, MNRAS, 298, 387

Dejonghe, H. 1983, A&A, 133, 225

Erickson, R. R. 1975, ApJ, 195, 343

ESA 1992, The Hipparcos Input Catalogue, ESA SP-1136

ESA 1997, The Hipparcos Catalogue, ESA SP-1200

Evans, N. W., Haefner, R. M., & de Zeeuw, P. T. 1997, MNRAS, 286, 315

Famaey, B., Van Caelenberg, K., & Dejonghe, H. 2002, MNRAS, 335, 201

Famaey, B., Jorissen, A., Luri, X., et al. 2005, A&A, 430, 165

Famaey, B., Pont, F., Luri, X., et al. 2007, A&A, 461, 957

Figueras, F., Gomez, A. E., Asiain, R., et al. 1997, in Hipparcos-Venice'97, ed. B. Battrick (Noordwijk: ESA), ESA SP-402, 519

Fricke, W. 1952, AN, 280, 193

Gerhard, O. E. 1993, MNRAS, 265, 213

Gilmore, G., King, I., & Kruit, P.v.d. 1989, in The Milky Way as a Galaxy, ed. R. Buser, & I. King (Geneva: Publ. Geneva Observatory)

Hénon, M. 1973, A&A, 24, 229

Holmberg, J., Nordström, B., & Andersen, J. 2007, A&A, 475, 519

Jaynes, E. T. 1957a, Phys. Rev., 106, 620

Jaynes, E. T. 1957b, Phys. Rev., 108, 171

Jaworsky, W. 1987, J. Phys. A: Math. Gen., 20, 915

Jiang, Z., & Ossipkov, L. 2007, MNRAS, 379, 1133

Juan-Zornoza, J. M. 1995, Ph.D. Thesis, Barcelona: Universitat de Barcelona

Juan-Zornoza, J. M., Sanz-Subirana, J., & Cubarsi, R. 1990, Ap&SS, 170, 343

Kagan, A. M., Linnik, Y. V., & Rao, C. R. 1973, in Characterization Problems in Mathematical Statistics (New York: Wiley)

Kapteyn, J. C. 1922, ApJ, 55, 302

Kormendy, J., Bender, R., Evans, A. S., & Richstone, D. 1998, AJ, 115, 1823

Kouskoulas, Y., Pierce, L. E., & Ulaby, F. T. 2004, IEEE Trans. Geosci. Rem. Sens., 42, 457

Martin, N. F. G., & England, J. W. 1981, in Encyclopaedia of Mathematics and its applications, Mathematical theory of entropy, Reading, Mass (Addison-Wesley), 12

Mead, L. R., & Papanicolaou, N. 1984, J. Math. Phys., 25, 2405

Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989

Pauli, E.-M., Napiwotzki, R., Heber, U., Altmann, M., & Odenkirchen, M. 2006, A&A, 384, 879

Sanz-Subirana, J., & Català-Poch, M. A. 1987, in 10th ERAM of the IAU, 4, 267

Skuljan, J., Hearnshaw, J. B., & Cottrell, P. L. 1999, MNRAS, 308, 731

Soubiran, C., & Girard, P. 2005, A&A, 438, 139

Strömberg, G. 1925, ApJ, 61, 363

Stuart, A., & Ord, J. K. 1987, in Kendall's Advanced Theory of Statistics, Distribution Theory (London: Ch. Griffin & Co.), 1

Vallenari, A., Pasetto, S., Bertelli, G., et al. 2006, A&A, 451, 125

van der Marel, R. P., & Franx, M. 1993, ApJ, 407, 525

Veltz, L., Bienaymé, O., Freeman, K. C., et al. 2008, A&A, 480, 753

Villani, C. 2002, A review of mathematical topics in collisional kinetic theory, in Handbook of Mathematical Fluid Dynamics, ed. S. Friedlander, & D. Serre (Amsterdam: North-Holland), 1, 71

Vidojević, S., & Ninković, S. 2009, AN, 330, 46

Vorobyov, E. I., & Theis, Ch. 2008, MNRAS, 383, 817

White, S. D. M. 1985, ApJ, 294, L99