DETECTION OF BISTABLE PHASE SPACE OF A REAL GALAXY, USING A NEW NON-PARAMETRIC BAYESIAN TEST OF HYPOTHESIS

BY DALIA CHAKRABARTY[*,†],

University of Warwick[†]

In lieu of direct detection of dark matter, inverse learning of the distribution of the gravitational mass in distant galaxies is of crucial importance in Cosmology. Such inverse learning is typically invoked within under-abundant information domains, which in turn motivates the need for undertaking simplifying model assumptions. Assuming the topology of the phase space structure of a galaxy allows for the learning of the density of the gravitational mass in galaxies, using partial and noisy data on velocity and location vectors of individual particles that reside in such a galactic system. The simplest assumption about the phase space topology is that of an isotropic phase space. This is equivalent to the assumption that the phase space pdf that the particle data are drawn from, is an isotropic function of the particle phase space coordinates.

We present a new distribution-free test of hypothesis that tests for relative support in two or more measured particle data sets, for the assumption that the respective data sets are drawn from such isotropic phase space pdfs. This test is designed to work in parameter space, for disparate sample sizes, in situations characterised by little and/or differential information about the prior for the null, i.e. the prior probability that a given particle data set is drawn from an isotropic phase space. The problem of differential sensitivity to a choice of the prior, given disparate sample sizes, is circumvented by this test. It works by computing the fraction of the inferred state space vectors, the posterior probability for which exceeds the maximal posterior achieved under the null. We illustrate applications of this test to two independent particle data sets in simulated as well as a real galactic system. The dynamical implications of the results of application to the real galaxy is indicated to be the residence of the observed particle samples in disjoint volumes of the galactic phase space. This result is used to suggest the serious risk borne in attempts at learning of gravitational mass density of galaxies, using particle data.

1. Introduction. One of the burning questions in science today is the understanding of dark matter. The quantification of the distribution of dark matter in our Universe, at different scales, is of major interest in Cosmology (Roberts and Whitehurst, 1975; Sofue and Rubin, 2001; Salucci and Burkert, 2000; de Blok, Bosma and McGaugh, 2003; Hayashi, Navarro and Springel, 2007). At scales of individual galaxies, the estimation of the density of the gravitational mass of luminous as well as dark matter content of these systems, $\rho(\mathbf{x}) \geq 0$, $\mathbf{X} \in \mathbb{R}^3$, is the relevant version of this exercise. Readily available data on galactic images, i.e. photo-

[*]Research fellow at Department of Statistics, University of Warwick; supported by a Warwick Centre for Analytical Sciences Research Fellowship

Keywords and phrases: Distribution free tests, Inverse problems, Bayesian Learning, Gravitational mass estimation in galaxies

metric observations from galaxies, can in principle, be astronomically modelled to quantify the gravitational mass density of the luminous matter in the galaxy, $\rho_L(\mathbf{x})$, (Gallazzi and F., 2009; Bell and de Jong, 2001); such luminous matter is however, only a minor fraction of the total that is responsible for the gravitational field of the galaxy since the major fraction of the galactic gravitational mass is contributed to by dark matter. Thus, the gravitational mass density of luminous matter, along with constraints on the gravitational mass density of the dark matter content, if available, can help the learning of the total gravitational mass density $\rho(\mathbf{x})$ in galaxies.

However, herein lies the difficulty - while photometric measurements are more readily available, direct detection of dark matter has hitherto been impossible, implying that measurements that can allow for quantification of the gravitational mass density of dark matter, are not achievable. Instead, there are effects of the total (dark+luminous) gravitational mass that can be measured, though astronomical data that bear signature of such effects are hard to achieve in "early-type" galaxies, the observed image of which is typically elliptical in shape[1]. Of some such measurables, noisy and "partially sampled" velocities of individual galactic particles have been implemented to learn $\rho(\mathbf{x})$ (Côté et al., 2001; Genzel et al., 2003; Chakrabarty and Raychaudhury, 2008). From this, when the astronomically modelled $\rho_L(\mathbf{x})$ is subtracted, the density of the gravitational mass of the dark matter content of early-type galaxies can be learnt. In this paradigm of learning $\rho(\mathbf{x})$, the data is referred to as "partially sampled" since the noisy measurements of only one component - namely, the component along the line-of-sight that joins the observer to the particle - of the three-dimensional velocity vector of galactic particles, are typically available. We view these measurables as sampled from the phase space $\mathcal{W}$ of the system, where $\mathcal{W}$ is the space of all the states that the system can achieve. In a mechanical system such as a galaxy, $\mathcal{W}$ is the space of the spatial vectors and velocity vectors of all galactic particles. On other occasions, the dispersion of the line-of-sight component of the velocity vector comprises the measurement. In either case, such kinematic data is expected to track the gravitational field of the galaxy in which the sampled particles play. .

That motion tracks gravitational field is not entirely alien to our experience - after all, the revolution of the Earth around the Sun, at a speed of $V_\oplus$ and radius $R_\oplus$, allows for an estimate of the solar gravitational mass $M_\odot$ via Newton's law, as $V_\oplus^2 = GM_\odot/R_\oplus$, where G is a known constant and the assumption of a circular orbit is made. In contrast to this, when the topology of the orbits in phase space is unknown, a closed-form relation between the orbital speed and gravitational mass enclosed within the orbit cannot be written. Then, in the absence of information about the topology of the orbits of galactic particles, a parametric approximation for the probability density function of $\mathcal{W}$ maybe too naive a model if the system manifests even moderate complexity.

We ask the question that if the available data include kinematic measurements of particles in distinct samples that live in mutually disjoint volumes of $\mathcal{W}$, what are the implications for the inferred $\rho(\mathbf{x})$ that is learnt using such data sets? One major implication

[1]The intrinsic global morphology of such "early-type" galaxies is approximated as a triaxial ellipsoid; in this paper we discuss gravitational mass density determination of such galaxies only

is inconsistent estimates of the gravitational mass density, obtained from any methodology that aims at such learning, using the aforementioned data. In this work, we encounter such inconsistency that stems from the implementation of distinct data sets in an inverse Bayesian methodology that exists in the literature (D.Chakrabarty, 2011, 2009; Chakrabarty and Portegies Zwart, 2004) and is discussed below. Such an inconsistency however cannot be accepted as physically possible - the same galactic system cannot be ascribed two distinct gravitational mass density functions. In other words, inconsistent mass density estimates are viewed as spuriously caused by one or more factors.

The possible factors are scrutinised, including the suggestion that the observed samples are drawn from insular volumes in $\mathcal{W}$. This paper introduces a new nonparametric test of hypothesis that is used to test for the support in two or more distinct kinematic data sets in a given system, for an undertaken model assumption. In the considered application, this is the assumption about the existence of global symmetries in the topological structure of the phase space that the data set is sampled from. The test is designed to work in parameter space, in the context of non-parametric Bayesian inference when for the two or more cases of observation, no or differential prior information are available for the null. At the same time, the data sets being of disparate sizes, imply differential sensitivity to any *ad hoc* choice of priors (Kass and Raftery, 1995). The new test discussed herein works by estimating the probability that the posterior probability of learnt model parameter vectors given a data set, exceeds the maximal posterior of such parameters given that data, achieved in an isotropy-assuming model, i.e. under the null (de B. Pereira and Stern, 1999; de B. Pereira, Stern and Wechsler, 2008). The test is designed to work without explicitly declaring priors for the null that the relevant model assumption is undertaken. However, we could also view such a model assumption to reduce to a constraint on all or some of the model parameters, in general. Therefore, priors on the null, when available, imply priors on a function of (all or some of) the model parameters, and vice versa. If priors on the model parameters are identified, these can be incorporated in the computation of the posterior of the learnt model parameters, given the relevant data, under the null or otherwise. However, in the absence of such prior knowledge, weak - or even improper priors - may be used in the learning of the model parameters, which are then implemented in the testing procedure. In this sense, the test is an improvement over tests that ignore available prior information on the null.

We configure the question of inferring inconsistent gravitational mass estimates using distinct particle samples, caused by residence of the samples in non-overlapping volumes of the galactic phase space, to the context of the real galaxy NGC 3379. For this system, multiple kinematic data sets are measured for two distinct types of galactic particles (Douglas et al., 2007; Bergond et al., 2006).

While some of the model assumptions involved in the learning of the total gravitational mass density of the galaxy, given such data, will be tested as suggested above, others that are implicit in the learning, but cannot be tested within the scope of this paper, will be identified, and the characterisation of this real galaxy will be qualified accordingly.

The paper is organised as follows. In the following section, the relevance of the assump-

tion of an isotropic phase space pdf, to the inverse learning of $\rho(\mathbf{x})$ is motivated. The resulting formulation of the phase space pdf is discussed in Section 2.1. The inverse learning methodology is itself discussed in Section 2.2. Thereafter, we discuss the shortcomings of the Bayes factor computations in regard to this application. In Section 4, the details of the new test are delineated. The test is illustrated on simulated data in Section 5 and on real data in Section 7. The paper is concluded with Section 8.

2. Inverse learning of gravitational mass density. We reinvoke the definition of the total galactic gravitational mass density of luminous and dark matter as $\rho(\mathbf{x})$, where $\rho(\mathbf{x}) \geq 0$, and the 3-D spatial location vector is $\mathbf{X} \in \mathbb{R}^3$, with $\mathbf{X} = (X_1, X_2, X_3)^T$. Also, the 3-D velocity vector is $\mathbf{V} = (V_1, V_2, V_3)^T$, with only the component V_3 being a measurable. Then the galactic phase space $\mathcal{W}$ is the space of $\mathbf{V}$ and $\mathbf{X}$ of all galactic particles. Thus, $\mathcal{W} \subseteq \mathbb{R}^6$.

2.1. *Assuming an isotropic phase space pdf.* We formulate the learning of $\rho(\mathbf{x})$ by declaring the data to be unknown functional of $\rho(\mathbf{x})$. Thus the inverse learning of $\rho(\mathbf{x})$ is motivated, but this inverse problem is ill-posed given that only one component of the particle velocity vector is a measurable, namely V_3. In order to render the problem tractable, we invoke simplifying model assumption about the topology of the system state space. In fact, the imposed assumption involves assuming the phase space $\mathcal{W}$ to be isotropic, i.e. for the phase space probability density function to be $f(\mathbf{x}, \mathbf{v}) \geq 0$. In general, the pdf of $\mathcal{W}$ has a time dependence, but here we assume the system to have attained stationarity. Under this model assumption of stationarity, we interpret the phase space pdf as as a scalar-valued, function of 2 vectors, $\mathbf{X}$ and $\mathbf{V}$. We will then undertake testing of this assumption in the available data. Such testing will address the often-expressed anxiety featured in the astrophysical literature, over the lack of identifiability among solutions for $\rho(\mathbf{x})$, caused by lack of knowledge of isotropy in the structure of the phase space pdf.

For $\mathbf{X} = \mathbf{x}$ and $\mathbf{V} = \mathbf{v}$, $f(\mathbf{x}, \mathbf{v})$ is an isotropic function if $f(\mathbf{x}, \mathbf{v}) = f(\mathbf{Q}\mathbf{x}, \mathbf{Q}\mathbf{v})$, for any orthogonal transformation matrix $\mathbf{Q}$ (Truesdell, Noll and Antman, 2004; Wang, 1969). We recall from the theory of scalar valued functions of two vectors, that if $f(\cdot, \cdot)$ is isotropic, the set of invariants (with respect to $\mathbf{Q}$) of this isotropic scalar function is $\Upsilon_Q = \{\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v}, \mathbf{x}\cdot\mathbf{v}\}$. Then, under this assumption that the phase space pdf is isotropic, it can be proved that the phase space pdf admits the representation $f(\Upsilon_Q) \equiv f(\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v}, \mathbf{x}\cdot\mathbf{v})$ (Truesdell, Noll and Antman, 2004; Liu, 2002).

This is achieved, if for example, the phase space is represented as $f(E)$, where $E := \Phi(\sqrt{\mathbf{x}\cdot\mathbf{x}}) + h(\mathbf{v}\cdot\mathbf{v})$. The physical interpretation of this representation is that a phase space pdf chosen as a function of the particle energy E, is isotropic. In order to allow for this physical interpretation of E as the particle energy, we choose $\Phi(\sqrt{\mathbf{x}\cdot\mathbf{x}})$ as the gravitational potential energy, written as a function of $\mathbf{x}\cdot\mathbf{x} = x_1^2 + x_2^2 + x_3^2$ and $h(\mathbf{v}\cdot\mathbf{v})$ to be $\mathbf{v}\cdot\mathbf{v}/2$ (kinetic energy)[2].

[2]The above identification of the dependence of f on particle energy E is not motivated merely to invoke familiar ideas of physics into the modelling but is reinforced by the fact that f depends on $\mathbf{x}$ and $\mathbf{v}$ only via con-

It merits mention that in general, $\rho(\mathbf{x})$ is a known function of $\Phi(\mathbf{x})$ as $\nabla^2\Phi(\mathbf{x}) = -4\pi G\rho(\mathbf{x})$, (Poisson Equation), where ∇^2 is the Laplacian operator and G is a known constant.

However, in lieu of information about the structure of the phase space density, and given at least the moderate level of complexity that is envisaged to manifest in phase spaces of galaxies, there is no justification for opting for an isotropic, over an anisotropic prescription for galactic phase spaces. There is a logistical motivation for doing so however - the assumption of an isotropic $f(\mathbf{x},\mathbf{v})$ renders the calculations involved in the learning of $\rho(\mathbf{x})$ relatively easier. The details of this learning, given the data and the assumed prescription of isotropy for $f(\mathbf{x},\mathbf{v})$, are discussed in Section 2.2. Since the learnt $\rho(\mathbf{x})$ is sensitive to the prescription chosen for the phase space - isotropic or not - the problem is referred to as the mass-anisotropy degeneracy in the astrophysical literature. Limited phase space information - in particular, information about only one out of the 3 components of $\mathbf{v}$ - triggers this mass-anisotropy degeneracy.

In the astrophysical literature, the aforementioned anxiety has been a pressing concern of Dekel et al. (2005) when they advance the possibility that anisotropy might be introduced into $f(\mathbf{x},\mathbf{v})$ of NGC 3379 due to evolutionary reasons, which Romanowsky et al. (2003) had ignored. They state this in an effort to explain the "spurious" results that the latter had reported for the aforementioned galactic systems, among 4 others. Douglas et al. (2007) disagree that this could explain the identification of the low dark matter content in NGC 3379 by Romanowsky et al. (2003) given that the existence of anisotropy was included in the data analysis employed in the earlier work. They stress that "anisotropy parameter" recovered from their kinematic data and that obtained from the simulations presented in Dekel et al. (2005), are similar in nature Here, the anisotropy parameter is a parametrisation of the deviation from an isotropic phase space density, defined as $1-\sigma_k^2(\mathbf{x})/\sigma_3^2(\mathbf{x})$ where $\sigma_j^2(\mathbf{x})$ is the variance in the j-th component of $\mathbf{V}$, $j=1,2,3$, $k=1,2$. $\sigma_j^2(\mathbf{x})$ is related to the second order moment of $f(\mathbf{x},\mathbf{v})$ as $\sigma_j^2(\mathbf{x}) = \frac{\int_V v_j^2 f(\mathbf{x},\mathbf{v})d\mathbf{v}}{\int_V f(\mathbf{x},\mathbf{v})d\mathbf{v}} - \left(\frac{\int_V v_j f(\mathbf{x},\mathbf{v})d\mathbf{v}}{\int_V f(\mathbf{x},\mathbf{v})d\mathbf{v}}\right)^2$ (Binney and Tremaine, 1987).

However, the suggestion advanced in this paper is more general. Firstly, the learning of $\rho(\mathbf{x})$ is nonparametric, so an explicit parametrisation of anisotropy is not of relevance. Importantly, if the test of hypothesis that we develop here, suggests less support in one data set over another, to the model assumption of an isotropic phase space pdf, it would imply that the two data sets have been sampled from distinct phase space density functions that will then be interpreted as fundamentally different in that these cannot necessarily be reconciled with each other, merely by adjusting one parameter, the physical interpretation of which is the anisotropy parameter as given in astrophysical literature.

Thus, differential support in the two sets for an isotropic pdf of $\mathcal{W}$, will imply that

---

stants of motion since $\frac{df}{dt} = 0$, i.e. $f(\mathbf{x},\mathbf{v})$ obeys the Collisionless Boltzmann Equation(Binney and Tremaine, 1987). The only constant of motion I that depends on $\mathbf{x}$ and $\mathbf{v}$ as in $I(\mathbf{x}\cdot\mathbf{x},\mathbf{v}\cdot\mathbf{v})$, is the particle energy E. Thus, an isotropic phase space pdf depends on E or any regular function of E. Without loss of generality, for our purpose, we suggest that the isotropic phase space pdf is $f(E)$.

any methodology that attempts the learning of $\rho(\mathbf{x})$ while using measurements of galactic phase space coordinates or moments of $f(\mathbf{x},\mathbf{v})$, will be pre-disposed to offer distinct $\rho(\mathbf{x})$, given the distinct data sets. Of course, for the same system, distinct $\rho(\mathbf{x})$ are impossible - in fact, such a fallacious result, if achieved, can be explained as due to the fact that the used data sets have been drawn from mutually insulated volumes of the galactic phase space, the pdfs of which do not concur. Such is possible, if the galactic phase space is characterised by disjoint volumes, motions in which do not communicate with each other - this is a viable scenario in non-linear dynamics even for systems with moderate complexity when the data are sampled from (at least two) distinct basins of attraction that characterise the multistable galactic phase space (Thompson and Stewart, 2001).

2.2. *Non-parametric Bayesian learning of* $\rho(\mathbf{x})$. If we are convinced that motion tracks the gravitational field due to a given gravitational mass density function, then we can write down $\mathbf{v}=\xi[\rho(\mathbf{x})]$, where $\xi[\cdot]$ is an unknown functional of $\rho(\mathbf{x})$. In Chakrabarty and Portegies Zwart (2004); D.Chakrabarty (2009), the effort has been to learn $\rho(\mathbf{x})=\xi^{-1}(\mathbf{v})$. This inverse learning is achieved by attempting the forward problem iteratively. D.Chakrabarty (2009) adopted a fully discretised model such that $\mathbf{x}$ is binned, with $\mathbf{x}\in[\mathbf{x}_i-\delta\mathbf{x},\mathbf{x}_i)$ defining the i-th $\mathbf{x}$-bin of width $\delta\mathbf{x}$. The forward problem consists of writing the posterior probability of $\{\rho_h\}_{h=1}^{N_x}$ given the kinematic data $\mathbf{D}$, where the gravitational mass density in the h-th $\mathbf{X}$-bin is ρ_h. $[\{\rho_h\}_{h=1}^{N_x}|\mathbf{D}]$ is written in terms of the likelihood of the data given the model for the gravitational mass density and the prior. Now, in addition to the variable V_3, the measurables include the spatial locations X_1,X_2 of the particle on the image of the galaxy, i.e $\mathbf{D}:=\{x_1^{(k)},x_2^{(k)},v_3^{(k)}\}_{k=1}^{N_{tot}}$. Let $(X_1,X_2,V_3)^T\in\mathcal{M}\subset\mathcal{W}$. The data is then sampled from the pdf $\nu(x_1,x_2,v_3)$ of the subspace $\mathcal{M}\subset\mathcal{W}$, so that likelihood is defined in terms of the projection of $f(\mathbf{x},\mathbf{v})$ into this subspace $\mathcal{M}$ of these observables, i.e. $\nu(x_1,x_2,v_3)$. Assuming the observed vectors $(x_1^{(k)},x_2^{(k)},v_3^{(k)})^T$, $k=1,\ldots,N_{tot}$, to be $i.i.d$, the likelihood is defined as $\mathcal{L}=\prod_{k=1}^{N_{tot}}\nu(x_1^{(k)},x_2^{(k)},v_3^{(k)})$. Chakrabarty and Portegies Zwart (2004); Chakrabarty (2006); Chakrabarty and Raychaudhury (2008) assumed an isotropic phase space pdf, so that

$$\begin{aligned}\nu(x_1^{(k)},x_2^{(k)},v_3^{(k)}) &= \int_{X_3}\int_{V_1}\int_{V_2} f[E(x_1^{(k)},x_2^{(k)},x_3,v_1,v_2,v_3^{(k)})]dx_1dx_2dv_3 \quad \text{where,}\\ E(x_1^{(k)},x_2^{(k)},x_3,v_1,v_2,v_3^{(k)}) &= \Phi(\sqrt{\{x_1^{(k)}\}^2+\{x_2^{(k)}\}^2+\{x_3\}^2})+\frac{(\{v_1\}^2+\{v_2\}^2+\{v_3^{(k)}\}^2)}{2}. \qquad (2.1)\end{aligned}$$

Identifying $\Phi(\cdot)$ with the gravitational potential energy and $r^{(k)}\equiv\sqrt{\{x_1^{(k)}\}^2+\{x_2^{(k)}\}^2+\{x_3\}^2}$ as the radial location of the k-th particle from the system centre, Poison equation was recalled as $\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\Phi(r)}{dr}\right)=-4\pi G\rho(r)$ in this geometry. This identification of $\sqrt{\mathbf{xx}}$, with the spherical radius r is a manifestation of the undertaken model assumption of phase space isotropy. The definition of the phase space pdf in Equation 2.1 borrows from the admissible representation of an isotropic phase space pdf as a function of energy,

as discussed in Section 2.1. The likelihood is defined as above and its computation was achieved in the forward problem by discretising the relevant range of energy values such that $E \in [E_j - \delta E, E_j)$ defines the j-th E-bin of width δ_E, inside which, $f = f_j$, $j = 1, \ldots, N_E$. The unknown functions abide by the physically motivated constraints that $\rho(r) \geq 0$, $f(E) \geq 0$ and $\frac{d\rho(r)}{dr} \leq 0$. In the adopted discrete model, a simultaneous learning of discretised versions of two univariate functions $f(E)$ and $\rho(r)$ is attempted, i.e. the target is to learn the vector $\boldsymbol{\rho} := (\rho_1, \ldots, \rho_{N_x})^T$ and the vector $\mathbf{f} := (f_1, \ldots, f_{N_E})^T$ as proxies for the unknown functions. Given that the gravitational mass density is non-negative and monotonically non-increasing function of r, $\boldsymbol{\rho} \in \mathcal{R} \subset \mathbb{R}^{N_x}$. Again, given that $f(E) \geq 0$, $\mathbf{f} \in \mathcal{F} \subset \mathbb{R}^{N_E}$. Then $\mathbf{f}$ and $\boldsymbol{\rho}$ are learnt by sampling from the posterior $[\mathbf{f}, \boldsymbol{\rho} | \{x_1^{(k)}, x_2^{(k)}, v_3^{(k)}\}_{k=1}^{N_{tot}}]$, using random-walk Metropolis-Hastings.

To ensure monotonic decline in the gravitational mass density function, first the difference $\tilde{\Delta}_\rho^{(h)}$ between the gravitational mass densities in the h-th and $h+1$-th radial bins is proposed to be $\tilde{\Delta}_\rho^{(h)} \sim \mathcal{N}_F(\Delta_\rho^{(h)}, s^2)$, where $\mathcal{N}_F(c, d)$ is the folded normal distribution (Leone, Nottingham and Nelson, 1961) with mean c and variance d, $c \geq 0$, $d > 0$. In our implementation, the current difference between the gravitational mass density values in the h and $h+1$-th radial bins is $\Delta_\rho^{(h)} := \rho_h - \rho_{h+1}$. The variance s^2 is the empirical variance of $\Delta_\rho^{(h)}$, computed using values of this difference variable from step number t_0 to $t-1$, where t is the current step number and t_0 is chosen experimentally to be post-burnin (Haario et al., 2006), $h = 1, \ldots, N_{rad}$. Here ρ_{N_x+1} and $\tilde{\rho}_{N_x+1}$ are defined as 0. Then as h is varied from N_x to 1, the proposed h-th component of the unknown gravitational mass density vector is $\tilde{\rho}_h = \tilde{\rho}_{h+1} + \tilde{\Delta}_\rho^{(h)}$. f_j is updated similarly, while following an imposed constraint that $f(E)$ is monotonically non-decreasing with energy E, $j = 1, \ldots, N_{eng}$. However, unlike in the case of the updating of the gravitational mass density value, a monotonically non-decreasing phase space pdf with energy is not motivated by any physically justifiable constraints, though astronomical literature often suggests phase space density functions that typically increase with E (Binney and Tremaine, 1987).

Uniform priors for ρ_h and f_j were used, $h = 1, \ldots, N_x$, $j = 1, \ldots, N_{eng}$, such that $\rho \sim \mathcal{U}[\rho_{lo}, \rho_{hi}]$ and $f \sim \mathcal{U}[0, 1]$. Here $\rho_{lo}^{(h)}, \rho_{hi}^{(h)}$ are experimentally chosen constants and the prior on the phase space density is uniform in [0,1] since $\mathbf{f}$ is normalised to be 1 for the most bound orbit, i.e. for the maximum E value.

3. Testing for isotropy. The apparently inconsistent $M(r)$ learnt from two different kinematic data sets, (Figure 3), translates physically to inconsistent gravitational potential estimates for the given galaxy - this is physically unacceptable. Here we have identified the serious risk in the approach that such a result entails - the inverse Bayesian nonparametric learning of $\boldsymbol{\rho}$, as above, using an available sample of velocity measurements, might mislead the astronomer in her quantification of the total gravitational mass in a galaxy, leading to erroneous estimation of the fraction of the galactic gravitational mass that is contributed to by dark matter. Such mis-quantification, if pertinent to a sample of galaxies can in turn bias cosmological ideas about the distribution of dark matter on galactic scales.

In appreciation of this serious risk, we need to check if the anxiety is inherent in the very notion of attempting the learning of gravitational mass density using the available (incomplete) phase space coordinates of sampled galactic particles, or if it is a problem in our Bayesian inverse methodology discussed above, or if the data implemented in this example are biased to produce the recovered inconsistency in $\boldsymbol{\rho}^{(1)}$ and $\boldsymbol{\rho}^{(2)}$ learnt using the 2 data sets.

We begin the check by examining if the assumption of an isotropic phase space that was used in the above Bayesian methodology can result in the recovered inconsistency in the $\boldsymbol{\rho}$ learnt using the 2 data sets. This question is addressed by testing for the support in each of the used data sets, towards the assumption of an isotropic phase space.

We test for the support towards the assumption of isotropy in the measured data, using a new test of hypothesis that we present here. As there are two data sets - the relative support in which towards isotropy is sought - we define two null hypotheses, such that the i-th null, H_i, is that the i-th data set is drawn from an isotropic phase space pdf, $i = 1, 2$. We first test for H_i as supported by $\mathbf{D}_i$; conventionally, this is suggested as computing $\Pr(H_i|\mathbf{D}_i)$ for each $i = 1, 2$. The relative support in the 2 data sets to isotropy is gauged by then comparing $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$. For our purpose,

$$H_i : f_i(\mathbf{x}, \mathbf{v}) = \Psi_i[E(v^2/2 + \Phi(r))], \quad , i = 1, 2. \tag{3.1}$$

where $f_i(\mathbf{x}, \mathbf{v})$ is the phase space pdf that $\mathbf{D}_i$ are drawn from and $\Psi_i(E)$ is some real-valued function of energy E, such that $\Psi_i(E) \geq 0$ if $E \leq 0$ and $\Psi_i(E) = 0$ otherwise, for $i = 1, 2$.

Here we are not making any pre-emptive assumptions about the concurrence of the phase space densities that the two data sets are drawn from. If of course $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ are isotropic and coincide with each other, then $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will be similar and both $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will be high. If $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ coincide but are not isotropic, $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will be similar but neither $\Pr(H_1|\mathbf{D}_1)$ nor $\Pr(H_2|\mathbf{D}_2)$ will be high. If $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ do not coincide but are both isotropic, $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will both be high. If $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ do not coincide and only one is an isotropic function, it is likely that either of $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will be high and the other low. Given this framework, we now motivate the need for a new test developed to work in this context of Bayesian non-parametric learning of the unknown functions.

3.1. *Handling priors.* We prefer to work in parameter space rather than sample space. At the same time, given the radically different sample sizes of the GC and PNe data, it is imperative that the test be independent of sample-size effects. We recall that the sample size of the PNe data is about 5.6 times the size of the GC data. Another relevant question is that of the availability of the prior for the null, i.e. $\Pr(H_i)$, or at least of the ratio of the prior for the null to the alternative hypothesis, i.e. the prior odds $\Pr(H_i)/(1 - \Pr(H_i))$, $i = 1, 2$. There are real-life situations when this information might be unavailable, the application to the testing of isotropy in galactic phase spaces being an example. When the

aim is to test for the support in the data for a model assumption, the null is in general, a statement of the form $H_0 : \xi(\boldsymbol{\theta}) = 0$ where $\boldsymbol{\theta}$ is the model parameter vector, $\boldsymbol{\theta} \in \mathbb{R}^n$ and the function $\xi : \mathbb{R}^n \longrightarrow \mathbb{R}$ expresses the undertaken assumption. Indeed, the null could also be express as an inequality, but we do not consider such a case here; instead our focus is on sharp hypotheses of the above form. Now, a prior probability density on the null implies a prior on a certain functional of $\xi(\boldsymbol{\theta})$, given $H_0 : \xi(\boldsymbol{\theta}) = 0$. Similarly, a prior on all or some components of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ can be expressed as a prior on the null. Thus, in this application, identifying priors on null is equivalent to selecting priors on the model parameters and vice versa. Such equivalence follows from the premise of the problem, i.e. is due to the aim of quantifying support in the data for an undertaken model assumption.

Indeed, the test that we present here can work when no prior information is available in the observations for the null H_i, but if such information becomes available, full use of the same should be made. In this sense, the test discussed here is different from a classical distribution-free test. With this in mind, the test is designed to employ all available prior information on the null, i.e. available prior information on $\{\theta_i\}_{i=i_{min}}^{i_{max}}$, where $i_{min} \geq 1$ and $i_{max} \leq n$, which is used in the computation of the posterior $\Pr(\boldsymbol{\theta}|\mathbf{D}_i)$ of the model parameter vector, given data $\mathbf{D}_i$ under the null or otherwise. Such posterior computation is at the heart of the test, as we shall shortly see. On the other hand, when priors on the null are unavailable, weak priors - or even improper priors - on θ_i can be implemented, $i = 1, \ldots, n$. Similarly, differential availability of priors for the null in the two (or more cases of observation) can be handled within this test.

Conventional wisdom may then be invoked to suggest that these demands will be met by comparing the posterior odds $\Pr(H_1|\mathbf{D}_1)/[1-\Pr(H_1|\mathbf{D}_1)]$ and $\Pr(H_2|\mathbf{D}_2)/[1-\Pr(H_2|\mathbf{D}_2)]$ (Kass and Raftery, 1995; Goodman, 1999) of the isotropy-assuming model and the model that does not assume isotropy, respectively, for the data sets $\mathbf{D}_1$ and $\mathbf{D}_2$. This would entail computation of the Bayes factors $\Pr(\mathbf{D}_1|H_1)/[1 - \Pr(\mathbf{D}_1|H_1)]$ and $\Pr(\mathbf{D}_2|H_2)/[1 - \Pr(\mathbf{D}_2|H_2)]$ and the prior odds $\Pr(H_1)/[1 - \Pr(H_1)]$ and $\Pr(H_2)/[1 - \Pr(H_2)]$. There is no apriori physical reason to constrain the ratio of the prior odds of models that can describe data $\mathbf{D}_1$ as equal to that of models that describe data $\mathbf{D}_2$. In allowing for such flexibility, we are allowing for the system phase space structure to be characterised by at least two volumes, the state of isotropy in which are not necessarily the same. Since in complex mechanical systems phase spaces characterised by such insulated pockets are possible, as arising out of non-linear dynamical mechanisms, we do not ignore such a possibility.

We consult the astronomical literature, with the aim of eliciting estimation of $\Pr(H_i)/[1-\Pr(H_i)]$, $i = 1, 2$. However, there is no underlying physical theory to describe the status of isotropy in any part of galactic phase spaces; most work to this effect are in the form of numerical simulations of models of generic galaxies. In fact, it is precisely because consensus over the status of isotropy in individual real galaxies has not been achieved by such generic numerical modelling, that the aforementioned mass-anisotropy degeneracy continues to be an anxiety in the astronomy community.

In contradiction to such numerical simulations, Côté et al. (2001) discuss isotropy in phase spaces that observed samples of GCs are drawn from, using data from multiple

sources that are available for these 2 galactic systems. However, extrapolating these suggestions to the particular instance of the galaxy of interest in our work (NGC 3379) is risky. Douglas et al. (2007) do indeed discuss the issues regarding isotropy in the phase space that the observed sample of PNe in the galaxy NGC 3379 are drawn from, though they do not provide estimates of the probability of this phase space being isotropic. In other words, the work by Douglas et al. (2007) cannot be employed to constrain $\Pr(H_1)$. The generic simulations of Dekel et al. (2005), will have to be considered to be applicable to the case of this particular galaxy, to suggest $\Pr(H_1)/1 - \Pr(H_1)$. Secondly, they do not consider the possibility that distinct samples of different types of galactic particles could bear independent information about the respective volumes of the galactic phase space that the data are drawn from. Thus, using these simulation-based studies as suggestions for priors is risky. The risk is accentuated in light of the fact that the two sample sizes are very disparate in our application.

Given the situation described above - which is not atypical of real-life complex physical systems - we could opt for weak priors for H_1 and H_2, such as $\Pr(H_i)/1 - \Pr(H_i) = \alpha_i$, $\alpha_i \in (0, 1]$, $i = 1, 2$, except that the sensitivity of the posterior odds to such ad-hoc choices of the priors will be different given the drastically different sample sizes of $\mathbf{D}_1$ and $\mathbf{D}_2$ (Kass and Raftery, 1995). In other words, the relative support in the two data sets for the null, will be rendered crucially sensitive to the choice of the priors $\Pr(H_i)$, $i = 1, 2$. Thus, a method that relies on the explicit declaration of the prior odds in each distinct case of observation, can be misguiding, given the application at hand. Thus, Bayes factor computation is not a viable way to get the the sought relative support towards the assumption of isotropy in the two data sets.

It is to be noted that even if the priors on the unknown $\boldsymbol{\rho}$ and $\mathbf{f}$ vectors are chosen as improper, subject to a multiplicative constant, (Berger and Pericchi, 1996; O'Hagan, 1995), such a constant would not contribute to the comparison of the posterior under the null and otherwise, as long as the same improper priors are used in the posterior computation under the null and otherwise.

In view of this, we present a new distribution-free test of hypothesis that works by taking the full posterior structure into account and does not rely heavily on the prior for the null. This test is inspired by the Fully Bayesian Significance Test or FBST advanced by de B. Pereira and Stern (1999); de B. Pereira, Stern and Wechsler (2008).

4. Method. In our test, a measure of the support in the data, for the null is defined; we interpret this measure of support as the evidence in favour of the null. This evidence in favour of H_i is given by the complement of the posterior probability of $\boldsymbol{\rho}$ given data $\mathbf{D}_i$ exceeds the posterior of $\boldsymbol{\rho}^\star$, given $\mathbf{D}_i$, where $\boldsymbol{\rho}^\star$ is the gravitational mass density that maximises the posterior $\Pr(\boldsymbol{\rho}|\mathbf{D}_i)$ under the null.

In this section we first describe FBST and then proceed to describe its adaptation for our application.

4.1. *FBST.* FBST tests the sharp null hypothesis that the relevant parameter, $\theta \in \Theta$ has a value θ_0, i.e. $H_0 : \theta = \theta_0$. Here θ is assumed to be distributed continuously in the

parameter space Θ. In FBST, the evidence in favour of H_0 is measured by first identifying the parameter $\theta^\star$ for which posterior probability under the null, given the data, is a maximum. This is followed by the identification of the θ, the posterior of which, given the data, is higher than the posterior of $\theta^\star$. Such θ is then considered to comprise the tangential set T, the volume of which is synonymous to the probability of identifying θ that are more likely than $\theta^\star$. Thus, the evidence in favour of H_0 is:

$$\begin{aligned} ev &= 1 - \Pr(\theta \in T|\mathbf{D}_i), \quad \text{where} \\ T &= \{\theta : \Pr(\theta|\mathbf{D}_i) > \Pr(\theta^\star|H_i)\}, \end{aligned} \tag{4.1}$$

$i = 1, 2$. Intuitively, the higher the probability of identifying $\theta \in T$, the less is the support for H_i. $\theta \in T$ are more consistent with the observed data than $\theta^\star$. In other words, FBST quantifies support in the data against the null hypothesis as the volume under the posterior probability distribution of θ, for $\theta \geq \theta^\star$. Thus, FBST involves identification of $\theta^\star$ via optimisation, followed by integration over T. A highly satisfying aspect of this test, is that the measure of evidence in favour of the null, the evidence value ev, obeys the Likelihood Principle (Basu, 1975; Birnbaum, 1962).

One shortcoming of this formulation is that the way the evidence value is defined above, it is not invariant to re-parametrisation of the parameter space. This problem has been addressed by Madruga, Pereira and Stern (2003) and de B. Pereira, Stern and Wechsler (2008) via the suggestion that the formulation of the ev be in terms of the surprise function relative to a reference density, i.e. $s_i(\theta) := \dfrac{\Pr(\theta|\mathbf{D}_i)}{r(\theta)}$, where the reference density on θ is $r : \Theta \longrightarrow \mathbb{R}$. Then, the FBST procedure involves identifying $\theta^\star = \arg\max_{H_0} s_i(\theta)$ and $s_i^\star = \max_{H_0} s_i(\theta^\star)$. The evidence in favour of the null, in this approach is

$$\begin{aligned} \bar{ev}_i &= 1 - \Pr(\theta \in \bar{T}_i|\mathbf{D}_i), \quad \text{where} \\ \bar{T}_i &= \{\theta : s_i(\theta) > s_i(\theta^\star)\}, \end{aligned} \tag{4.2}$$

The purpose of the normalisation of the posterior of θ, given the data, by the reference density function $r(\theta)$ is to ensure that for any transformation of θ, such as $\omega = \xi(\theta)$, results in a measure of evidence $\bar{ev}_i = ev_i$, against the null, $i = 1, 2$.

de B. Pereira, Stern and Wechsler (2008); Madruga, Pereira and Stern (2003) suggests possible interpretations of $r(\theta)$. If $r(\theta)$ is treated as the uniform density $U(\theta)$, $\bar{ev}_i$ and ev_i are equivalent.

4.2. *Implementation of FBST.* In contrast to the parametric FBST discussed by de B. Pereira, Stern and Wecl (2008), in our work, the the posterior of the system variable, given the data, does not conform to any parametric form. Given this, we present our test below.

The null hypothesis H_i that we aim to test for, is given above in Section 3, $i = 1, 2$. Both H_1 and H_2 represent sharp hypotheses. In our work, the system variable that we define the i-th null in terms of, is the phase space *pdf* $f_i(\mathbf{x}, \mathbf{v})$ that the data $\mathbf{D}_i$ is sampled from, and the total gravitational mass density $\rho_i(\mathbf{x})$ that is given uniquely by $\Phi_i(\mathbf{x})$ that in turn

defines the energy of the particles in the sample $\mathbf{D}_i$, for given $|\mathbf{v}|$. In the discretised model that is implemented to learn the unknown functions, the normalised phase space density is rendered the vector $\mathbf{f}$ and the total gravitational mass density is rendered the vector $\boldsymbol{\rho}$. As discussed in Section 2.2, $\mathbf{f}$ and $\boldsymbol{\rho}$ are continuous vector variables in $\mathcal{F}$ and $\mathcal{R}$ respectively. The system variable is then considered to be the vector $(\mathbf{f}, \boldsymbol{\rho})^T \in \mathcal{F} \times \mathcal{R}$.

We compute the measure of support in data $\mathbf{D}_i$ for the null hypothesis H_i, by working with the state space vector $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$, for $i = 1, 2$, in our adaptation of FBST. In our test, we consider the reference density $r(\theta)$ to be the uniform prior that is implemented in the inverse Bayesian learning of $f(\mathbf{x}, \mathbf{v})$ and $\boldsymbol{\rho}(\mathbf{x})$. This effectively equates the measure $\bar{ev}_i$ of evidence against the null, defined in terms of the surprise function and ev_i, defined in terms of the posterior of $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$, given $\mathbf{D}_i$, for $i = 1, 2$. Hence, we discuss our implementation of the test in terms of this posterior probability and compute the measure ev_i against H_i. In this non-parametric implementation of the test, we replace the integration over the tangential set T - which is difficult in this no-parametric setting - by a case-counting scheme. Our implementation of the test is as follows.

4.2.1. *Identification of posterior-maximising state space vector, under the null.* For the i-th case, $i = 1, 2$, we identify the $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$ that maximises the posterior $[\mathbf{f}_i, \boldsymbol{\rho}_i|\mathbf{D}_i]$ under the null, given the data. This posterior maximising, null abiding vector is referred to as $(\mathbf{f}_i^\star, \boldsymbol{\rho}_i^\star)^T$, motivated by the notation used to describe FBST above. In order to identify this $\boldsymbol{\rho}^*, \mathbf{f}^*$ pair, the following construct is used.

- During the posterior inference on $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$, performed with random-walk Metropolis, let the current state vector be $(\mathbf{f}_i^{(j)}, \boldsymbol{\rho}_i^{(j)})^T$, in the j-th iterative step of the chain, $j = 1, \ldots, N_0$, where the chain is N_0 steps long. Upon convergence, the unknown $\mathbf{f}_i$ and $\boldsymbol{\rho}_i$ are learnt within 95% credible regions. From a given chain, the state space vector $(\mathbf{f}_i^M, \boldsymbol{\rho}_i^M)^T$ corresponding to the mode of the posterior $[\mathbf{f}_i, \boldsymbol{\rho}_i|\mathbf{D}_i]$ is identified.
- Then from the identified $\mathbf{f}_i^M$, at the identified $\boldsymbol{\rho}_i^M$, we simulate n number of N_i-sized data sets of the observables, such that the ℓ-th of these data sets is $\mathbf{D}_i^\ell := \{(X_1^{(k\ell)}, X_2^{(k\ell)}, V_3^{(k\ell)})\}_{k=1}^{N_i}$. We understand that by construction, these n data sets are sampled from an isotropic phase space density since $\mathbf{f}_i^M$ was learnt under this model assumption. These n data sets are simulated from $\mathbf{f}_i^M$ at $\boldsymbol{\rho}_i^M$ using rejection sampling, according to the following algorithm.

  1. We recall that $\Phi(r) = \dfrac{-GM(r)}{r}$, where $M(r) = \displaystyle\int_{s=0}^{r} 4\pi\rho(s)s^2 ds$ and G is a known constant. Then discretising the integration in our discrete model used

for learning $\rho(r)$, we define

$$\begin{aligned} M(r) &= \sum_{q=1}^{p} \frac{4\pi}{3}[q^3\delta^3 - (q-1)^3\delta^3]\rho_q + \frac{4\pi}{3}[r^3 - p^3\delta^3]\rho_{p+1} \\ \text{for} \quad & r \in [p\delta, (p+1)\delta), \quad 1 \le p \le N_x - 1, \\ M(r) &= \sum_{q=1}^{N_x} \frac{4\pi}{3}[q^3\delta^3 - (q-1)^3\delta^3]\rho_q \quad \text{for} \quad r \ge N_x\delta, \\ M(r) &= \frac{4\pi}{3}[r^3\delta^3]\rho_1 \quad \text{for} \quad 0 \le r \le \delta. \end{aligned} \tag{4.3}$$

Here $N_x\delta$ is the maximum radius to which data are available and ρ_q is the gravitational mass density in the q-th radial bin, i.e. $\rho(r) = \rho_q$ if $r \in [(q-1)\delta, q\delta)$, $q = 1, \ldots, N_x$ (introduced in Section 2.2). This defines $\Phi(r)$ for any $r \ge 0$, for the identified $\boldsymbol{\rho}_i^M$.

2. The 3 components of the velocity vector, V_1, V_2, V_3 are continuous in $[-\sqrt{-2\Phi(0)}, \sqrt{-2\Phi(0)}]$, while the components of the spatial vector are continuous in $[-N_x\delta, N_x\delta]$. We set, $X_1, X_2, X_3 \sim \mathcal{U}[-N_x\delta, N_x\delta]$ and $V_1, V_2, V_3 \sim \mathcal{U}[-\sqrt{-2\Phi(0)}, \sqrt{-2\Phi(0)}]$. Here $\mathcal{U}[a, b]$ is the uniform distribution over the range $[a, b]$, for any $a, b \in \mathbb{R}$.
3. Draw $X_1, X_2, X_3, V_1, V_2, V_3$ from the respective uniform distributions defined above. Using the drawn phase space coordinates, define the parameter $\epsilon = \dfrac{\Phi(x_1^2 + x_2^2 + x_3^2) + (v_1^2 + v_2^2 + v_3^2)/2}{\Phi_0}$. Thus, the physical interpretation of ϵ is that it is the particle energy, normalised by its maximum possible value $\Phi_0 = \Phi(0)$ that is determined by the choice of the form of $\Phi(r)$. Thus, $\epsilon \in [-1, 0]$.
4. Scan over the integer s, from $s = 1$ to $s = N_E$, where f_s is the s-th component of $\mathbf{f}_i^M$. The s-th component of $\mathbf{f}_i^M$ is the phase space density over the energy-bin $[(s-1)\delta_E, s\delta_E]$, where δ_E is the constant width of the energy-bins. As we scan across the range of s, let the computed value of ϵ (defined above) be such that ϵ lies in the t-th energy bin, i.e. $\epsilon \in [(t-1)\delta_E, t\delta_E]$.
5. Over this identified t-th energy bin, let the component of $\mathbf{f}_i^M$ be $f_t^{(iM)}$. Then, if $f_t^{(iM)}/g(\epsilon) > \mathcal{S}$, where $\mathcal{S} \sim \mathcal{U}[0, 1]$, for the drawn phase space coordinates, we allow an integer-valued flag, γ, an increment of 1 and accept the drawn X_1, X_2, V_3 as the γ-th line in $\mathbf{D}_i^\ell$. We iterate over points 2 to 5, until γ equals N_i. Here $g(E)$ is the proposal density function that is chosen to envelope over $f(E)$, $\forall E$, invoked in the context of our implementation of rejection sampling.
6. When $\gamma = N_i$, the ℓ-th data set $\mathbf{D}_i^\ell$ has been fully generated. Then we reset γ to 0, ℓ incremented by 1 and the iterating over points 2 to 5 begins again.
7. When $\ell = n$, the n generated data sets, $\{\mathbf{D}_i^\ell\}_{\ell=1}^n$, have been constructed. The rejection sampling is stopped at this point.

- We input each $\mathbf{D}_i^\ell$, $\ell = 1, \ldots, n$, one at a time, to CHASSIS, to start n new chains. In each such chain, post burn-in, samples of $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$ vectors are generated in each iterative step. It is to be noted that all such $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$ samples are generated under the null hypothesis H_i since the data, $\mathbf{D}_i^\ell$, $\ell = 1, \ldots, n$, implemented for recovering these vectors, are drawn from phase space densities that are isotropic by construct. Of these recovered $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$, the vector which maximises the posterior $[\mathbf{f}_i, \boldsymbol{\rho}_i|\mathbf{D}_i^\ell]$, $\ell = 1, \ldots, n$, is recognised as $(\mathbf{f}_i^\star, \boldsymbol{\rho}_i^\star)$. Let this vector be sampled during the chain obtained with the generated data $\mathbf{D}_i^{\ell^\star}$.

4.2.2. *Quantifying membership of tangential set.* Once this null-abiding, posterior-maximising state space vector $(\mathbf{f}_i^\star, \boldsymbol{\rho}_i^\star)$ is identified, the next step, as in FBST, is to identify other state space vectors, the posterior of which exceeds the posterior probability $[\mathbf{f}_i^\star, \boldsymbol{\rho}_i^\star|\mathbf{D}_i^{\ell^\star}]$. We adopt a case counting scheme in which we quantify the number of state space vectors sampled at posterior probability in excess of $[\mathbf{f}_i^\star, \boldsymbol{\rho}_i^\star|\mathbf{D}_i^{\ell^\star}]$, in the equilibrium stage of chains run with the real data $\mathbf{D}_i$, expressed as a fraction of the total number of state space vectors sampled in these chains. Post-burnin, at each iterative step, there is a state space vector sample recovered.

Thus, let there be a total of $A_i \in \mathbb{Z}_+$ number of samples recovered in the post burn-in stage in these chains, i.e. there are A_i $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$ vectors recovered in chains run with data $\mathbf{D}_i$. Out of these, let B_i samples of $(\mathbf{f}_i, \boldsymbol{\rho}_i)^T$ vectors be such that the posterior probability $[\mathbf{f}_i, \boldsymbol{\rho}_i|\mathbf{D}_i] > [\mathbf{f}_i^\star, \boldsymbol{\rho}_i^\star|\mathbf{D}_i^{\ell^\star}]$. Here, $B \in \mathbb{Z}_+$, $B_i \leq A_i$. Then the fraction B_i/A_i is the equivalent of computing the volume of the tangential set T. Thus, we state that the evidence in favour of null H_i is

$$ev_{H_i} = 1 - \frac{A}{B} \tag{4.4}$$

In line with FBST, the null is rejected for "small" values of the ev, where this "small"ness can be objectively qualified in terms of minimisation of the loss function (de B. Pereira, Stern and Wechsler, 2008; Madruga, Pereira and Stern, 2003). Here however, we compute ev_{H_1} and ev_{H_2} and compare the values of the two.

5. Support for the assumption of isotropic phase spaces that synthetic data sets are simulated from. In this section, we implement the new test to quantify support for the assumption of isotropy, in data that have been sampled from synthetic phase spaces. We choose an isotropic phase space density function $f_{iso}(bfx, \mathbf{v})$ and an anisotropic phase space pdf $f_{aniso}(\mathbf{x}, \mathbf{v})$, and sample 2 data sets $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$ are drawn from these phase space density functions respectively. We choose the sample size of $\mathbf{D}_{aniso}$ to be 54 while the sample size of $\mathbf{D}_{iso}$ is chosen to be 270, i.e. the 5 times that of $\mathbf{D}_{aniso}$. The chosen

parametric forms of the phase space density functions are

$$f_{iso}(\mathbf{x},\mathbf{v}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\epsilon}{2\sigma^2}\right), \tag{5.1}$$

$$f_{aniso} = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\epsilon}{2\sigma^2}\right)\exp\left(-\frac{|\mathbf{L}|^2}{r_a^2\sigma^2}\right)$$

$$\text{where} \quad \epsilon = \frac{\mathbf{v}^2/2+\Phi(r)}{\Phi_0}, \qquad \mathbf{L} = \mathbf{x}\wedge\mathbf{v}.$$

Here, we choose the anisotropy scale $r_a = 4$ in units of kpc (the astronomical unit of length at galactic scales) and $\sigma = 220$ in units of km s^{-1}. The chosen form of $\Phi(r)$ is maintained as same in both definitions of the synthesis phase space pdfs, as that of the Plummer model (Binney and Tremaine, 1987),

$$\Phi(r) = -\frac{GM_0}{\sqrt{r_c^2+r^2}}, \tag{5.2}$$

where we chose the total gravitational mass of the synthetic system to be $M_0 = 4.06\times10^{11}$ times the mass of the Sun and the scale length of the Plummer model to be r_c=8 kpc. G is a known physical constant, the universal gravitational constant.

Having defined the true model, we then simulated $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$ from respective phase space densities, where each data set contained information on X_1, X_2 and V_3. The sampled V_3 data was chosen to be characterised by Gaussian noise $\sim \mathcal{N}(0, 20^2)$ which is typical of real-life galaxies that are nearby (Douglas et al., 2007). The model parameter vectors $\boldsymbol{\rho}$ and $\mathbf{f}$ learnt by porting $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$ to the algorithm CHASSIS is shown in black in the right and left panels respectively, of Figure 1. The modal values of the learnt components of these model parameter vectors are displayed with open circles while the error-bars in the plots depict the 95% HPDs. Given the lack of information that typifies such learning in real-life galaxies, here, the learning was performed in both cases, under the uniform priors $\pi_0(\rho_i) = \frac{1}{10^2-10^{-5}}$ and $\pi_0(f_j) = 1$, where for the implementation of $\mathbf{D}_{iso}$, we used $i = 1,\ldots,19$, $j = 1,\ldots,9$ and for the implementation of $\mathbf{D}_{aniso}$, we used $i = 1,\ldots,14$, $j = 1,\ldots,9$.

Thus, these vectors are learnt under the assumption of phase space isotropy, using data - one of which is, and one of which is not, sampled from an isotropic phase space density. We refer to the vectors learnt using $\mathbf{D}_{iso}$ as $\boldsymbol{\rho}_{iso}$ and $\mathbf{f}_{iso}$ and those learnt using $\mathbf{D}_{aniso}$ as $\boldsymbol{\rho}_{aniso}$ and $\mathbf{f}_{aniso}$. In order to check for the support in these 2 data sets for the assumption of phase space isotropy, we now check for the model parameter vector pair $(\mathbf{f}_i^\star, \boldsymbol{\rho}_i^\star)$ that maximises the posterior under the null, by simulating 10 data sets $\{\mathbf{D}_i^{(\ell)}\}_{\ell=1}^{10}$ from $\mathbf{f}_i$ that is defined using $\boldsymbol{\rho}_i$, where i stands for the subscript iso or $aniso$ (see Section 4.2) for details of this sampling). Here the null $H_i : f_i(\mathbf{x},\mathbf{v}) = \Psi_i(E(v^2/2+\Phi(r))]$.

For the implementation of $\mathbf{D}_{iso}$, we used $g(\epsilon) = 1.5\exp(\epsilon)$ and for the other case, we used $g(\epsilon) = \epsilon^{1.5}+0.66$. The choices of $g(\epsilon)$ are marked in broken lines in the relevant panels of Figure 1.

TABLE 1

Table displaying support in synthetic data $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$ simulated respectively from known isotropic and anisotropic parametric phase space density functions for the null that the data is sampled from an isotropic phase space density. Column 2 shows the maximal posterior probability under the null, computed by generating 10 distinct null-abiding data sets that have each been sampled from the modal $(\mathbf{f}^M, \boldsymbol{\rho}^M)$ pairs learnt using $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$. Column 3 delineates the support in the synthetic data for the null, computed using after identification of the null-abiding data set out of the 10 such data sets, the posterior of $(\mathbf{f}, \boldsymbol{\rho})$ given which, is the maximum, under the null.

Null-abiding generated data $\mathbf{D}_i^{(\star)}$	$\max[\pi(\mathbf{f}, \boldsymbol{\rho}\|\mathbf{D}_i^{(\star)})]$	ev_{H_i}
$\mathbf{D}_{iso}^{(\star)}$	-3006.8	0.9614
$\mathbf{D}_{aniso}^{(\star)}$	-844.58	0.0301

In Figure 2, the log posterior probability density $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}_{iso})$ and $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}_{aniso})$ are shown in black in the right and left panels respectively. The log posterior probability density under the null, i.e. $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}_{iso}^{(\star)})$ and $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}_{aniso}^{(\star)})$, are shown in red, in the right and left panels respectively, where $\mathbf{D}_i^{(\star)}$ is the null-abiding generated data set, the chain run with which yields the posterior maximising sample $(\mathbf{f}_i^\star, \boldsymbol{\rho}_i^\star)$ under the null. Here i stands for "iso" or "$aniso$". The horizontal lines in these panels represent the value of the log of the maximum of the posterior achieved under the null, with the data $\mathbf{D}_i^{(\star)}$ that are simulated from the modal $\mathbf{f}_i^M$, at $\boldsymbol{\rho}_i^M$, where $(\mathbf{f}_i, \boldsymbol{\rho}_i)$ are learnt from $\mathbf{D}_i$. The number of MCMC samples in the chain run with $\mathbf{D}_i$ that correspond to posterior probability density values that exceed the maximal posterior achieved under the null, defines the probability for a sample $\mathbf{f}, \boldsymbol{\rho}$ to live in the tangential set. The complement of this probability is a measure of support in the data in favour of the null.

The computed support in $\mathbf{D}_i$ for the null H_i is tabulated in Table 1. As is expected, for the data set $\mathbf{D}_{aniso}$ that is synthetically generated via rejection sampling from an isotropy-defying phase space density, the support in the data for the null that the data is sampled from an isotropy-assuming phase space pdf is $\leq$0.04. On the other hand, when the implemented data $\mathbf{D}_{iso}$ is truly sampled from an isotropic distribution, the method recovers a support in the data for the null that is ≥ 0.96. We attribute the small deviations from the values of 0 and 1 for the support in the two cases, respectively, to uncertainties triggered both by measurement noise (the data sets are all chosen to manifest noise) as well as learning the model parameters within 95% HPDs.

6. Support for the assumption of isotropic phase space of a real galaxy. To present the quantification of the support in the real data sets $\mathbf{D}_1$ and $\mathbf{D}_2$ each, towards the assumption of an isotropic phase space pdf, we first recall the results of the Bayesian learning of the model parameters of using these data sets

7. Bayesian inverse learning of model parameter vectors in real galaxy NGC 3379. NGC 3379, which is one of the few elliptical galaxies, for which kinematic information is available for individual members of two different populations of galactic particles - referred to as planetary nebulae (PNe) and globular clusters (GCs) - over an extensive radial range

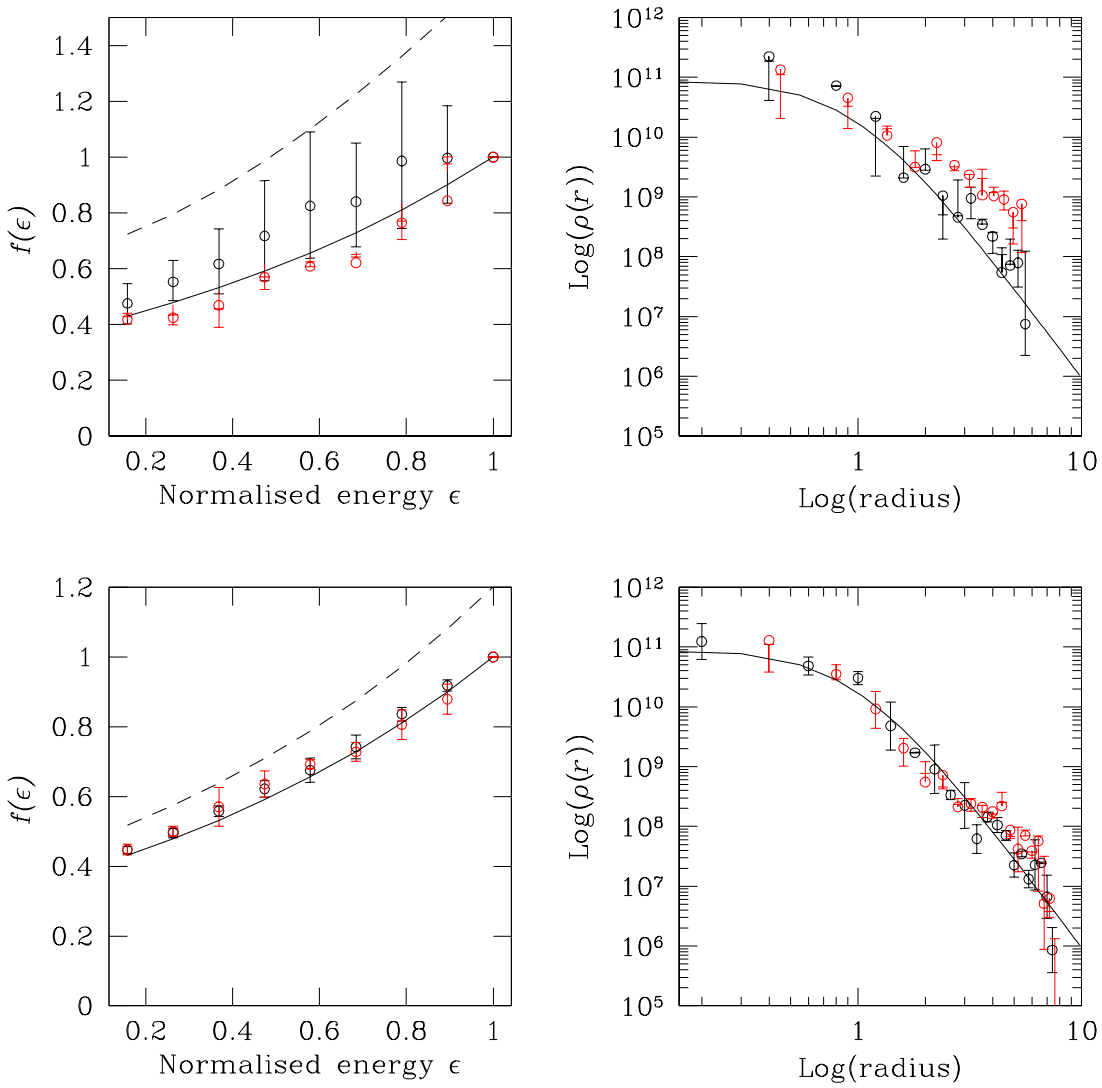

FIG 1. *Figure showing the phase space density vector* $\mathbf{f}$ *(left panels) and gravitational mass density vector* $\boldsymbol{\rho}$ *(right panels) learnt using the synthetic data* $\mathbf{D}_{iso}$*) (lower panels) that is sampled from a known isotropic phase space density and* $\mathbf{D}_{aniso}$*) (upper panel) that is learnt from a chosen anisotropic phase space density. The true gravitational mass density is presented in the solid black lime in the left panels. The true isotropic phase space density that* $\mathbf{D}_{iso}$ *is sampled from, is shown in the lower left panel in the black solid line. The true anisotropic phase space density for the value of the variable* $\mathbf{L} := \mathbf{x} \wedge \mathbf{v} = 0$ *is depicted in black solid line in the upper left panel. The figure also shows in red, the posterior maximising system parameter vectors* $\mathbf{f}^\star$ *and* $\boldsymbol{\rho}^\star$*, learnt using null-abiding data that are (rejection) sampled from the corresponding modal* $\mathbf{f}^M$ *that uses* $\boldsymbol{\rho}^M$ *for its definition. The instrumental distributions* $g(\epsilon)$ *employed in the rejection sampling are shown in broken black lines in the left panels. Modal values of all learnt parameters (components of learnt vectors) are represented by open circles while the 95% HPD credible region is shown by the error bar.*

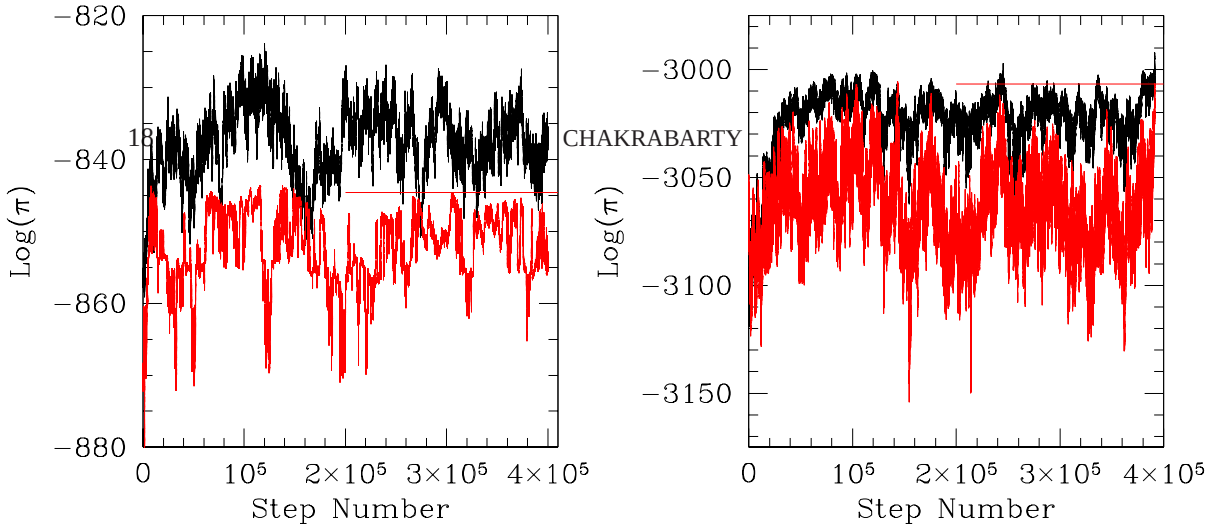

FIG 2. *Figure showing log posterior density $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}_{iso})$ (right) and $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}_{aniso})$ (left), in black, for chains that were run for 400,000 steps. Convergence is achieved by the 200,000-th step. The log of the posterior $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}^{(\star)}_{iso})$ and $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}^{(\star)}_{aniso})$ are shown in red in the right and left panels respectively. Here, the posterior under the null is maximised during chains run with the null-abiding data $\mathbf{D}^{\star}_{iso}$ and $\mathbf{D}^{\star}_{aniso}$, generated from modal model parameter vectors $(\mathbf{f}^M, \boldsymbol{\rho}^M)$ that are learnt using the data $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$, respectively. The red horizontal lines depict the values $\max[\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}^{(\star)}_{iso})]$ and $\max[\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}^{(\star)}_{aniso})]$. Here $\mathbf{D}_{iso}$ has a sample size of 270 and $\mathbf{D}_{aniso}$ bears information about 54 particles.*

spanning the outer parts of the galaxy. The input data used in the work include X_1, X_2, V_3 measurements of 164 PNe reported by Douglas et al. (2007) and of 29 GCs that were reported by (Bergond et al., 2006)[3]. We refer to the PNe data set as $\mathbf{D}_1$ and the GC data set as $\mathbf{D}_2$, with respective sample sizes of N_1=164 and N_2=29.

In Figure 3, we present the vectors $\boldsymbol{\rho}^{(i)} := (\rho^{(i)}_1, \ldots, \rho^{(i)}_{N_x})^T$ and $\mathbf{f}^{(i)} := (f^{(i)}_1, \ldots, f^{(i)}_{N_E})^T$, learnt from using the i-th data $\mathbf{D}_i$, $i = 1, 2$. This learning is performed using the aforementioned Bayesian nonparametric methodology, under the model assumption of an isotropic phase space pdf, i.e. a phase space density that is expressed as $f(E)$ and approximated in the discretised model of CHASSIS as the vector $\mathbf{f}$, (the j-th component of which is the value of the phase space density in the j-th energy-bin) and the vector $\boldsymbol{\rho}$, (the h-th component of which is the value of the phase space density in the h-th radial-bin). The learnt 95% HPDs are represented as error bars and the modal values are shown as open circles.

7.1. *Relative support in two data sets for the assumption of isotropic phase space of real galaxy.* The implementation of FBST, as described above, is invoked to estimate if the assumption of isotropy should be rejected, in the two different particle samples that we deal with.

The computed measures of evidence in the two data sets $\mathbf{D}_1$ and $\mathbf{D}_2$ for phase space densities $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ being isotropic, respectively, are ev_1 and ev_2 respectively. These are represented in Table 3.

The details of the computational and inferential procedure are as follows. Convergence

[3]We use the velocities of only 29 of the GCs that Bergond et al. (2006) advance as unambiguously bound to the gravitational field of NGC 3379, as distinguished by others that might be in the shared by the fields of this galaxy and its neighbours.

TABLE 2

Table displaying seeds that individual chains are started with. The initial choice of the gravitational mass density function is one that is sometimes used in astrophysical literature, $\frac{\rho_0}{\left(\frac{r}{r_c}\right)^{\alpha_1}\left(1+\frac{r}{r_c}\right)^{\alpha_2}}$. *The starting phase space density is chosen to be either of the form* $\exp^{-E}$ *or* $(-E)^{\beta}$. *Here the parameters* $\rho_0, \alpha_1, \alpha_2, r_c, \beta \in \mathbb{R}_{\geq 0}$. *The chains run with PNe data* $\mathbf{D}_1$ *are assigned names characterised by the pre-fix "PNe-RUN", while runs performed with the GC data* $\mathbf{D}_2$ *are labelled with pre-fix "GC-RUN".*

Name	ρ_o	r_c (kpc)	α_1	α_2	f_{seed}
$PNe-RUN\ I$	10^3	30	2.8	1	$=\exp(-E)$
$PNe-RUN\ II$	10^{10}	20	3.8	2	$=(-E)^3$
$PNe-RUN\ III$	10^{14}	10	1.8	3	$=(-E)^5$
$GC-RUN\ I$	10^5	30	2.8	1	$=\exp(-E)$
$GC-RUN\ II$	10^8	5	3.6	2	$=(-E)^5$
$GC-RUN\ III$	10^{10}	10	3.2	3	$=(-E)^2$

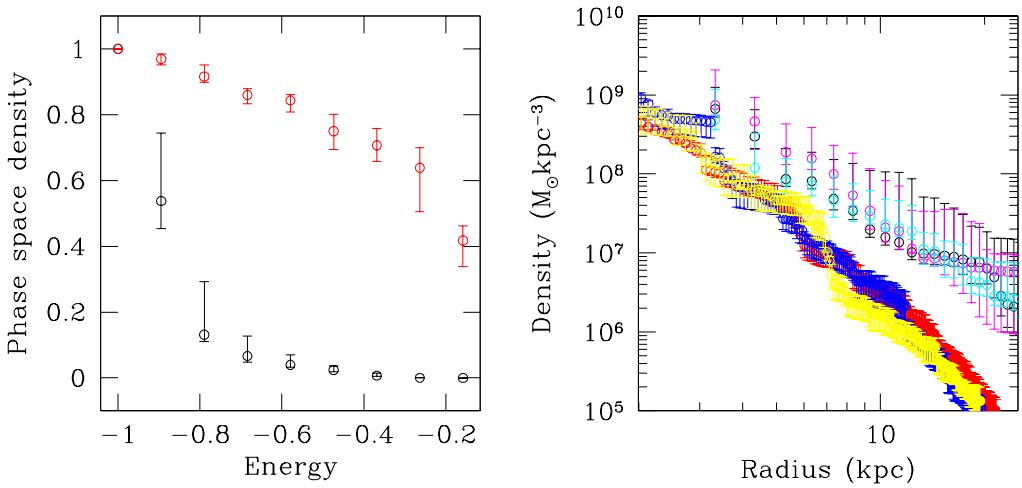

FIG 3. *Right panel: density profiles recovered from the chains named in Table 1. The* $\boldsymbol{\rho}_1$ *learnt from the chains run with PNe data* $\mathbf{D}_1$ *are shown in red, yellow and blue while* $\boldsymbol{\rho}_2$ *learnt with the GC data* $\mathbf{D}_2$ *kinematics are in black, magenta and cyan. As is apparent from the figure, starting from distinct functional forms does not affect results. Importantly,* $\boldsymbol{\rho}$ *vectors learnt using distinct data sets are not consistent with each other. Left panel: isotropic (normalised), phase space density vector* $\mathbf{f}_1$ *recovered from* $PNe-RUN\ I$ *(in red) and* $\mathbf{f}_2$ *from* $GC-RUN\ I$ *(in black), plotted in corresponding energy-bins.*

TABLE 3
Table showing measure of support, ev, in data $\mathbf{D}_i$ for null H_i, $i = 1, 2$, computed using different chains.

Chain name	Data set used	ev
$PNe - RUN\ I$	$\mathbf{D}_1$	0.61
$PNe - RUN\ II$	$\mathbf{D}_1$	0.58
$PNe - RUN\ III$	$\mathbf{D}_1$	0.62
$GC - RUN\ I$	$\mathbf{D}_2$	0.96
$GC - RUN\ II$	$\mathbf{D}_2$	0.96
$GC - RUN\ III$	$\mathbf{D}_2$	0.93

is achieved with chains that are 400,000 steps long, when data $\mathbf{D}_i$ is used to inversely learn $\boldsymbol{\rho}_i$ and $\mathbf{f}_i$, $i = 1, 2$, under the assumption of phase space isotropy, in the Bayesian non-parametric method CHASSIS. Uniform priors on $f(E)$ and $\rho(r)$ are used in these chains. 10 samples $\mathbf{D}_i^\ell$, $\ell = 1, \ldots, 10$ were drawn from the learnt $\mathbf{f}_i^M$ at the learnt $\boldsymbol{\rho}_i^M$, for each i, i.e. n=10. The chains run with each of these generated data $\mathbf{D}_i^\ell$, are each 400,000 steps long. For the rejection sampling, the form of the normalised phase space density that we learn using the real data $\mathbf{D}_1$ and $\mathbf{D}_2$, motivate us to choose $g(E) = 1$ when the GC data $\mathbf{D}_2$ is used and $g(E) = 1.3(-E)^5$ when $\mathbf{D}_1$ is used and E is normalised to live in $[-1, 0]$.

Comparing the computed ev_1 and ev_2 across the chains implies that the assumption of isotropy is more likely to be invalid for the phase space from which the PNe data $\mathbf{D}_1$ are drawn than from which the GC data $\mathbf{D}_2$ are drawn. Then $\mathbf{D}_1$ and $\mathbf{D}_2$ must be drawn from distinct phase space density functions since support in the two data sets to phase space isotropy is different. In terms of the phase space structure of this system, the phase space of the real galaxy NGC 3379 is marked by at least two distinct volumes, motions in which do not communicate with each other and the PNe and GC samples are drawn from such disjoint volumes of the galactic phase space. Of course, such an interpretation would hold up if we can rule out extraneous reasons that might be invoked to explain the differential support in $\mathbf{D}_1$ and $\mathbf{D}_2$ to the assumption of phase space isotropy. Such extraneous factors are systematically dealt with in Section 8.

Our result that $ev_2 \geq 0$ also suggest that the phase space density that the observed GCs in this galaxy live in is isotropic.

8. Discussions. In the above test, the high support in the GC sample towards an isotropic phase space pdf, along with the moderate support in the sampled PNe, indicate that the two samples are drawn from two distinct phase space density functions.

The expectation that the implementation of the PNe and GC data sets will lead to concurring gravitational mass density estimates is foreshadowed by the assumption that both data sets are sampled from the same - galactic - phase space density $f(\mathbf{x}, \mathbf{v})$. The apparent motivation behind this assumption is that since both samples live in the galactic phase space $\mathcal{W}$, they are expected to be sampled from the same galactic phase space density, at the galactic gravitational potential. However, consider the case that $f(\mathbf{x}, \mathbf{v})$ is a non-analytic function, such that

$$f(\mathbf{w}) = f_p(\mathbf{w}), \quad \forall \mathbf{w} \in (\mathbf{w}_1^{(p)}, \mathbf{w}_2^{(p)}), \quad p = 1, \ldots, p_{max} \tag{8.1}$$

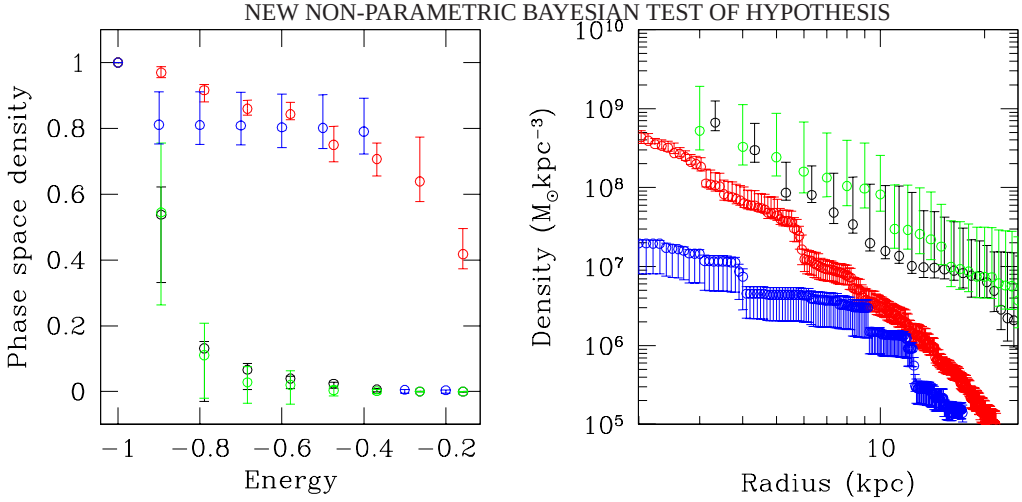

FIG 4. *Right panel: gravitational mass density vector $\boldsymbol{\rho}_2$ (in black) recovered from chain $GC - RUN\ I$ run using data $\mathbf{D}_2$ and $\boldsymbol{\rho}_1$ from chain $PNe - RUN\ I$ run using $\mathbf{D}_1$ (in red). These gravitational mass density results were obtained under the assumption of an isotropic phase space, the support for which in either data is indicated in Table 2. Overlaid on these are the identified vectors $\boldsymbol{\rho}_1^\star$ (in blue) and $\boldsymbol{\rho}_2^\star$ (in green), which are respectively, the posterior-maximising, gravitational mass density vectors identified in chains run with data $\mathbf{D}_1^\star$ and $\mathbf{D}_2^\star$, that are part of the sets of data sets simulated from the respective isotropic phase space density functions. The concurrence of $\boldsymbol{\rho}_2$ and $\boldsymbol{\rho}_2^\star$ is noted, along with the lack of consistency between $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_1^\star$. In the left panel, the phase space density vectors $\mathbf{f}_1$ (in red) and $\mathbf{f}_2$ (in black), learnt from the chains $PNe - RUN\ I$ and $G - RUN\ I$, are shown, compared respectively to $\mathbf{f}_1^\star$ (in blue) and $\mathbf{f}_2^\star$ (in green). Again, the overlap of $\mathbf{f}_2$ and $\mathbf{f}_2^\star$ is noted, as is the discord between $\mathbf{f}_1$ and $\mathbf{f}_1^\star$, especially at higher energies. The $\mathbf{f}$ vectors are normalised to unity at $E = -1$.*

where $\mathbf{w} := (\mathbf{x}, \mathbf{v})^T$. Then, if the GC data $\mathbf{D}_2$ are sampled from the density f_ζ and PNe data $\mathbf{D}_1$ are sampled from f_η, where $\eta, \zeta \in \{1, \ldots, p_{max}\}$, $\eta \neq \zeta$, then the assumption that both the observed samples are drawn from equal phase space densities is erroneous. If this assumption is erroneous, firstly it implies that the phase portrait of the galaxy NGC 3379 is split into at least two sub-spaces, the phase space densities of which are unequal, and secondly that the observed GCs live in a sub-space of $\mathcal{W}$ defined by $\mathbf{w} \in (\mathbf{w}_1^{(\zeta)}, \mathbf{w}_2^{(\zeta)})$ and the observed PNe live in a distinct sub-space defined by $\mathbf{w} \in (\mathbf{w}_1^{(\eta)}, \mathbf{w}_2^{(\eta)})$, $\eta \neq \zeta$. Indeed, if the galactic phase space $\mathcal{W}$ is split into at least two distinct and isolated volumes, such that the motions in these volumes do not mix and are therefore distinctly distributed in general, leading to unequal phase space densities. This is synonymous to saying that $\mathcal{W}$ is marked by two distinct basins of attraction and the two observed samples reside in the such distinct basins.

Such a situation will readily explain lack of consistency in the estimate of support in the two data, towards the assumption of isotropy in the phase space density function that these two data sets are drawn from. However, the fundamental question is really about the inverse of this statement. Does differential adherence of $\mathbf{D}_1$ and $\mathbf{D}_2$ to isotropic pase space *pdf* necessarily imply a split $\mathcal{W}$? Indeed it does, since differential support in $\mathbf{D}_1$ and $\mathbf{D}_2$

towards isotropy of pdf of the native phase space space implies that the distribution of the phase space vector $\mathbf{w}$ in these native sub-spaces of $\mathcal{W}$ are distinct, which implies that the phase portrait of this galaxy manifests at least two volumes, the motions in which are isolated from each other. Such separation of the motions is possible if these distinct sub-spaces that the two data sets reside in, are separated by separitrices (Thompson and Stewart, 2001).

8.1. *What could cause a split galactic phase space?.* Galactic phase spaces can be split given that a galaxy is expectedly a complex system, built of multiple components with independent evolutionary histories and distinct dynamical timescales. As an example, at least in the neighbourhood of the Sun, the phase space structure of the Milky Way is highly multi-modal and the ensuing dynamics is highly non-linear, marked by significant chaoticity. The standard causes for the splitting of $\mathcal{W}$ include the development of basins of attraction leading to attractors, generated in a multistable galactic gravitational potential. Basins of attraction could also be triggered around chaotic attractors, which in turn could be due to resonance interaction with external perturbers or due to merging events in the evolutionary history of the galaxy.

8.2. *Not stationary?.* A model assumption that we undertake is that of stationarity; we assumed that the phase space density is independent of time as is the gravitational mass density function. Deviation from stationarity cannot be checked in the present data since the data are snapshots in the evolution of the system. There are reports of observational signatures of non-stationarity; in NGC 3379 these could be considered to include the X-ray image of NGC 3379 (Pellegrini and Ciotti, 2006)[4].

All in all, at the level of this paper, we cannot say that deviation from equilibrium is ruled out.

8.3. *Assuming spherical geometry?.* One worry that astronomers have expressed in the literature about the galaxy NGC 3379 is that the spatial geometry of the system is triaxial and not spherical. For us, the relevant question to ask is if there is support in the data for the consideration of the gravitational mass density to depend on the spherical radius r alone. The fact that the methodology assumes the gravitational mass density to be dependent on $\mathbf{x} \cdot \mathbf{x} (= \sqrt{x_1^2 + x_2^2 + x_3^2})$ is a manifestation of the broader assumption of phase space isotropy (see Section 2.1). Thus, our test for phase space isotropy includes testing for the assumption that the gravitational mass density of the galaxy NGC 3379 bears a dependence on r. In other words, we have already tested for the assumption that the gravitational mass density depends on the components of the spatial vector, via the spherical radius r.

[4] Pellegrini and Ciotti (2006) report the gravitational mass of the galaxy modelled using X-ray measurements, as overestimated by a factor of about 2, compared to hydrodynamical modelling that assume hydrostatic equilibrium. Pellegrini and Ciotti (2006) invoke the imaged out-flowing nature of the hot gas from this galaxy to support deviation from hydrostatic equilibrium. However, the contribution of such deviations from hydrostatic equilibrium, towards the estimate of gravitational mass has not been satisfyingly addressed.

8.4. *True Mass Distribution.* In contrast to the PNe and GC samples observed in this galaxy, if two sets of tracer kinematics can be inferred to have been drawn from the same phase space density, we will expect consistency in the gravitational matter density that is recovered by using such data sets in a mass determination formalism. At the end of the discussion presented above, we will naturally want to know what the true gravitational mass density function of NGC 3379. However, if distinct $\rho(r)$ estimates are available at a given radius, from n independent tracer samples, and these estimates are sorted as $\{\rho_1(r), \rho_2(r), \ldots,\ \rho_n(r)\}$, then the lower limit on the galactic gravitational mass density at r is $\sup\{\rho_1(r), \rho_2(r), \ldots,\ \rho_n(r)\}$. Hence we can at least put a lower bound on the total gravitational mass content of the galaxy, as we now present for NGC 3379. NGC 3379 is advanced as a dark matter rich galaxy, with the gravitational mass inside a radius of about 20 kpc to be at least as high as about 4 to $10\times10^{11}\mathrm{M}_\odot$, where $\mathrm{M}_\odot$ denotes the mass of the Sun and the astronomical unit of length, kiloparsec, is abbreviated as kpc.

8.5. *Risk of using particle kinematics.* The above results and arguments suggest that it is inherently risky to refer to the gravitational mass density recovered using an observed particle sample - and the gravitational potential computed therefrom - as the gravitational potential of the galaxy. We have demonstrated this with the example of NGC 3379 and shown that inconsistencies in $\rho(r)$ learnt using distinct observed samples, cannot be attributed to any other factor except that these observed samples are drawn from distinct and insular sub-spaces within the galactic phase space, and/or the lack of time-independence in the gravitational mass density or phase space density function.

APPENDIX A: COMPARISON WITH THE CONVENTIONAL POINT OF VIEW

To sum up the results obtained above, we state that in general, kinematic data drawn from distinct phase space densities will yield distinct gravitational mass density functions and this is not solely because of the lack of sufficient information to help constrain the state of anisotropy in the phase space of the system. Here we illustrate the idea that even when phase space anisotropy parameters - as defined in astronomical literature - are equal, learnt gravitational mass estimates will be in general be unequal.

The method that is conventionally employed to learn gravitational mass enclosed within radius r, $M(r) = \int_0^r \rho(s) 4\pi s^2 ds$, is the Jeans equation that implements the (typically smoothed) empirical dispersion σ_3 of the values of V_3 of the galactic particles in an observed sample, and the number density function of the sampled particles, in the assumed spherical spatial geometry, i.e. the number of particles $\nu(r)$ that lie in the interval $(r, r+\delta r]$ where δr is the width of the radial bin chosen by the astronomer. Jeans equation gives

$$M(r) = -\frac{r\sigma_3^2}{G}\left[\frac{d\ln\nu(r)}{d\ln r} + \frac{d\ln\sigma_3^2(r)}{d\ln r} + \beta(r)\right] \tag{A.1}$$

where the anisotropy parameter is defined as $\beta(r) = 1 - \sigma_2^2/\sigma^3$. Then we see that it is possible to obtain distinct $M(r)$ using two data sets that are drawn from two distinct phase space distributions that are equally anisotropic (as parameterised by the anisotropy

parameter $\beta(r)$) - then $d\ln\sigma_3(r)/d\ln r$ and $d\ln\nu(r)/d\ln r$ terms would in general be different in the two cases, even if $\beta(r)$ is the same. Since $\beta(\cdot)$ is a function of r, we refer to it as the anisotropy function rather than the anisotropy parameter from now. This follows from the fact that the second moments (leading to the dispersion $\sigma_3(r)$) and the zeroth moments (leading to $\nu(r)$) of unequal phase space density functions are unequal in general, even if the anisotropy functions are equal. Thus, for example, even if two observed samples are inferred to be drawn from phase space densities, each of which is isotropic, but are unequal functions, the anisotropy functions will be identically zero in each case, but the second and zeroth order moments would not necessarily enjoy equal rate of change with $\ln r$, i.e. the $M(r)$ learnt using the two samples will be inconsistent even in this case. Thus, unequal gravitational mass density estimates necessarily imply distinct phase space density functions that the observed data are sampled from but not necessarily distinct anisotropy functions.

ACKNOWLEDGMENTS

I am thankful to Dr. John Aston for his kind comments. I acknowledge the support of a Warwick Centre for Analytical Sciences Fellowship.

REFERENCES

BASU, D. (1975). Statistical Information and Likelihood. *Sankhya A* **37** 1-71.

BELL, E. F. and DE JONG, R. S. (2001). Stellar Mass-to-Light Ratios and the Tully-Fisher Relation. *Astrophysical Journal* **550** 212.

BERGER, J. and PERICCHI, L. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association* **57** 109-122.

BERGOND, G., E., Z. S., J., R. A., M., S. R. and L., R. K. (2006). Wide-field kinematics of globular clusters in the Leo I group. *Astronomy & Astrophysics* **448** 155-164.

BINNEY, J. and TREMAINE, S. (1987). *Galactic Dynamics*. Princeton University Press, Princeton.

BIRNBAUM, A. (1962). On the foundations of statistical inference. *Jl. of the American Statistical Association* **57** 269-326.

CHAKRABARTY, D. (2006). An Inverse Look at the Center of M15. *Astronomical Jl.* **131** 2561-2570.

CHAKRABARTY, D. and PORTEGIES ZWART, S. (2004). An Inverse-Problem Approach to Cluster Dynamics. *Astronomical Jl.* **128** 1046-1057.

CHAKRABARTY, D. and RAYCHAUDHURY, S. (2008). The Distribution of Dark Matter in the Halo of the Early-Type Galaxy NGC 4636. *Astronomical Journal* **135** 2350-2357.

CÔTÉ, P., MCLAUGHLIN, D. E., HANES, D. A., BRIDGES, T. J., GEISLERAND D. MERRITT, D., HESSER, J. E., HARRIS, G. L. H. and LEE, M. G. (2001). Dynamics of the Globular Cluster System Associated with M87 (NGC 4486). II. Analysis. *Astrophysical Jl.* **559** 828-850.

CHAKRABARTY, D. (2009). CHASSIS - Inverse Modelling of Relaxed Dynamical Systems. In *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation* (R. D. B. ANDERSSEN R. S. and L. T. H. N. (ED.), eds.) 362-368. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation.

CHAKRABARTY, D. (2011). Galactic Phase Spaces. In *Science: Image In Action - Proceedings Of The 7Th International Workshop On Data Analysis In Astronomy, "Livio Scarsi And Vito Digesu"* (B. ZAVIDOVIQUE and G. L. B. (ED.), eds.) 327. World Scientific Publishing Company, Incorporated.

DE B. PEREIRA, C. A. and STERN, J. M. (1999). Evidence and credibility: full Bayesian significance test for precise hypotheses. *Entropy* **1** 69-80.

DE B. PEREIRA, C. A., STERN, J. M. and WECHSLER, S. (2008). Can a Significance Tet be Genuinely Bayesian? *Bayesian Analysis* **3** 79-100.

DE BLOK, W. J. G., BOSMA, A. and MCGAUGH, S. (2003). Simulating observations of dark matter dominated galaxies: towards the optimal halo profile. *Monthly Notices of the Royal Astronomical Soc* **340** 657-678.

DEKEL, A., F., S., A., M. G., J., C. T., S., N. G. and R., P. J. (2005). Lost and found dark matter in elliptical galaxies. *Nature* **437** 707-710.

DOUGLAS, N. G., R., N. N., J., R. A., L., C., K., K., R., M. M., M., A., O., G., C., F. K., R., M. H., E., N. and M., C. (2007). The PN.S Elliptical Galaxy Survey: Data Reduction, Planetary Nebula Catalog, and Basic Dynamics for NGC 3379. *Astrophysical Jl.* **664** 257-276.

GALLAZZI, A. and F., B. E. (2009). Stellar Mass-to-Light Ratios from Galaxy Spectra: How Accurate Can They Be? *Astrophysical Jl. Supplement* **185** 253-272.

GENZEL, R., SCHÖDEL, R., OTT, T., EISENHAUER, F., HOFMANN, R., LEHNERT, M., ECKART, A., ALEXANDER, T., STERNBERG, A., LENZEN, R., CLÉNET, Y., LACOMBE, F., ROUAN, D., RENZINI, A. and TACCONI-GARMAN, L. E. (2003). The Stellar Cusp around the Supermassive Black Hole in the Galactic Center. *Astrophysical Jl.* **594** 812-832.

GOODMAN, S. (1999). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine* **130** 1005-1013.

HAARIO, H., LAINE, M., MIRA, A. and SAKSMAN, E. (2006). DRAM: Efficient adaptive MCMC. *Statistics and Computing* **16** 339.

HAYASHI, E., NAVARRO, J. F. and SPRINGEL, V. (2007). The shape of the gravitational potential in cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society* **377** 50-62.

KASS, R. E. and RAFTERY, A. E. (1995). Bayes Factors. *JASA* **90** 773-795.

LEONE, F. C., NOTTINGHAM, R. B. and NELSON, L. S. (1961). The Folded Normal Distribution. *Technometrics* **3** 543.

LIU, I.-S. (2002). *Continuum Mechanics*. Springer-Verlag, New York.

MADRUGA, M. R., PEREIRA, C. A. B. and STERN, J. M. (2003). Bayesian evidence test for precise hypotheses. *Journal of Statistical Planning and Inference* **117** 185-198.

O'HAGAN, A. (1995). Fractional Bayes Factors for Model Comparisons. *Journal of Royal Statistical Society, Ser. B* **57** 99-138.

CÔTÉ, P., MCLAUGHLIN, D. E., HANES, D. A., BRIDGES, T. J., GEISLER, D., MERRITT, D., HESSER, J. E., HARRIS, G. L. . H. and LEE, M. G. (2001). Dynamics of the Globular Cluster System Associated with M87 (NGC 4486). II. Analysis. *Astrophysical Jl* **559** 828.

PELLEGRINI, S. and CIOTTI, L. (2006). Reconciling optical and X-ray mass estimates: the case of the elliptical galaxy NGC3379. *Monthly Notices of the Royal Astronomical Soc* **370** 1797.

ROBERTS, M. S. and WHITEHURST, R. N. (1975). The rotation curve and geometry of M31 at large galactocentric distances. *Astrophysical Jl.* **201** 327-346.

ROMANOWSKY, A. J., G., D. N., M., A., K., K., R., M. M., R., N. N., M., C. and C., F. K. (2003). A Dearth of Dark Matter in Ordinary Elliptical Galaxies. *Science* **301** 1696-1698.

SALUCCI, P. and BURKERT, A. (2000). Dark Matter Scaling Relations. *Astrophysical Jl. Letters* **537** L9-L12.

SOFUE, Y. and RUBIN, V. (2001). Rotation Curves of Spiral Galaxies. *Annual Review of Astronomy & Astrophysics* **39** 137-174.

THOMPSON, J. M. T. and STEWART, H. B. (2001). *Non-linear dynamics and chaos.* John Wiley & Sons, Chichester.

TRUESDELL, C., NOLL, W. and ANTMAN, S. S. (2004). *The non-linear field theories of mechanics* Volume 3. Springer-Verlag, New York.

WANG, C. C. (1969). On representations for isotropic functions. *Archive for Rational Mechanics and Analysis* **33** 249-267. 10.1007/BF00281278.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL, U.K.
d.chakrabarty@warwick.ac.uk