

Near-Gaussian distributions for modelling discrete stellar velocity data with heteroskedastic uncertainties

Jason L. Sanders ^{1,2}★ and N. Wyn Evans¹

¹*Institute of Astronomy, University of Cambridge, Madingley Rise, Cambridge CB3 0HA, UK*

²*Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*

Accepted 2020 September 16. Received 2020 September 9; in original form 2020 July 17

ABSTRACT

The velocity distributions of stellar tracers in general exhibit weak non-Gaussianity encoding information on the orbital composition of a galaxy and the underlying potential. The standard solution for measuring non-Gaussianity involves constructing a series expansion (e.g. the Gauss–Hermite series) that can produce regions of negative probability density. This is a significant issue for the modelling of discrete data with heteroskedastic uncertainties. Here, we introduce a method to construct positive-definite probability distributions by the convolution of a given kernel with a Gaussian distribution. Further convolutions by observational uncertainties are trivial. The statistics (moments and cumulants) of the resulting distributions are governed by the kernel distribution. Two kernels (uniform and Laplace) offer simple drop-in replacements for a Gauss–Hermite series for negative and positive excess kurtosis distributions with the option of skewness. We demonstrate the power of our method by an application to real and mock line-of-sight velocity data sets on dwarf spheroidal galaxies, where kurtosis is indicative of orbital anisotropy and hence a route to breaking the mass–anisotropy degeneracy for the identification of cusped versus cored dark matter profiles. Data on the Fornax dwarf spheroidal galaxy indicate positive excess kurtosis and hence favour a cored dark matter profile. Although designed for discrete data, the analytic Fourier transforms of the new models also make them appropriate for spectral fitting, which could improve the fits of high-quality data by avoiding unphysical negative wings in the line-of-sight velocity distribution.

Key words: methods: statistical – techniques: radial velocities – techniques: spectroscopic – galaxies: kinematics and dynamics.

1 INTRODUCTION

The velocity distributions of stars within a galaxy exhibit a range of morphologies. Non-Gaussianity, particularly kurtosis, in the velocity distributions is linked to anisotropy of the orbits of the stars: populations of stars on more radial orbits typically produce ‘peakier’ or long-tailed velocity distributions, whilst populations of stars on circular orbits produce more flat-topped, short-tailed distributions (Gerhard 1993; van der Marel & Franx 1993). Furthermore, rotating populations generically produce skewed distributions, as demonstrated by the azimuthal velocities of the Milky Way disc populations (e.g. Binney & Merrifield 1998, chap. 10) or the h_3 – V/σ relation observed in elliptical galaxies (Bender, Saglia & Gerhard 1994). However, in the main, these deviations from Gaussianity are weak and so can be captured via a Gauss–Hermite series where the zeroth-order model is a Gaussian and orthogonal polynomial corrections produce non-Gaussianity. Typically, only two correction terms are used with coefficients h_3 quantifying the skewness and h_4 quantifying the (excess) kurtosis (Gerhard 1993; van der Marel & Franx 1993).

Gauss–Hermite coefficients have seen widespread use in the summary of integral field unit spectroscopy data (Emsellem et al. 2004; Cappellari et al. 2011; van de Sande et al. 2017) where typically one simultaneously extracts information on the stellar

populations and their kinematics via template fitting (Cappellari & Emsellem 2004; Cappellari 2017). The coefficients provide useful intermediate data products for comparison with dynamical models (e.g. Rix et al. 1997), whilst also giving physical insight into the orbital distributions and hence potential. They were originally developed to characterize the velocity profiles near the centres of elliptical galaxies and hence to measure the masses of central black holes (e.g. van der Marel et al. 1994, 1998). Applications to barred galaxies have also been suggested. Debattista et al. (2005) identified negative h_4 (negative excess kurtosis) as an indicator of viewing a face-on peanut-shaped bar/bulge (e.g. Méndez-Abreu et al. 2019). Recently, Sellwood & Gerhard (2020) have suggested the sign of h_4 of the vertical velocity distribution distinguishes different formation channels for a peanut-shaped bar/bulge with positive h_4 (longer tails) indicative of a peanut formed through resonances, whilst negative h_4 (flat-topped distribution) indicative of a more violent buckling event that depletes the bar of low vertical velocity stars.

Higher order moments have also been used in the analysis of dwarf spheroidal (dSph) galaxy line-of-sight velocity data. Typically, they have been used to address the question of the nature of the dark matter density at the centre – whether it is cusped like the famous Navarro, Frenk & White (1997) model or cored like a harmonic potential. This information is summarized in the excess kurtosis (Łokas 2002) and can be extracted from data using Gauss–Hermite coefficients (Amorisco & Evans 2012b), the virial shape parameters (Merrifield & Kent 1990; Read, Walker & Steger 2019), higher order

* E-mail: jason.sanders@ucl.ac.uk

Jeans equations (Richardson & Fairbairn 2013), or through the use of a dynamical distribution functions (Łokas 2002; Łokas, Mamon & Prada 2005; Amorisco & Evans 2011; Breddels & Helmi 2013; Pascale et al. 2018).

In the application to discrete velocity measurements with heteroskedastic uncertainties, as for dSph galaxies, we hit an immediate stumbling block when using the Gauss–Hermite series: in general a Gauss–Hermite series is not positive definite and so cannot be interpreted as a probability density function. This means it cannot be used in a probabilistic framework that incorporates observational uncertainties or membership probabilities (in a mixture model). Although unphysical, typically the regions of negativity are small, so for spectral fitting this is not a significant issue. However, negative probabilities in any probabilistic framework produce awkward numerical problems. For the case of dSph galaxies, there are data sets of at most a few thousand stars (e.g. Walker, Mateo & Olszewski 2009a) and the varying stellar types produce a range of differing uncertainties with which velocities can be measured. In order to extract the maximal information from such a data set, we wish to avoid either binning the data or assuming the uncertainties are small. Kuijken & Merrifield (1993) explored the possibility of ensuring positive definiteness by using a Gaussian mixture, which also allows for simple convolution with measurement uncertainties. Amorisco & Evans (2012b) provided an alternative solution to the issue by constructing probability density functions inspired by dynamical galaxy models. This approach produced expressions that required numerical integration, although the authors demonstrate how to make their procedure numerically efficient. Here, we provide an alternative and simpler method for the construction of appropriate analytic probability density functions. Although slightly more restrictive, they offer a drop-in replacement for the Gauss–Hermite series that guarantees a positive-definite probability density function. Here, we focus on data from dSph galaxies, but our methods can be applied to any sets of discrete tracers such as globular clusters and planetary nebulae (Agnello et al. 2014; Oldham & Auger 2016) or clusters of galaxies (Wojtak & Łokas 2010).

We begin in Section 2 by describing how the Gauss–Hermite series can be used to model discrete tracers with heteroskedastic uncertainties and highlight its limitations, before presenting an alternative method for constructing appropriate non-Gaussian probability distribution functions (PDFs) in Section 3. We provide specific examples of the new families in Section 4 and apply the new probability density functions to the problem of dSph galaxies, specifically Fornax, in Section 5. We close with a brief discussion of the use of the new models in the context of spectral fitting in Section 6 where the guaranteed non-negativity could potentially improve fits to high-quality data.

2 THE GAUSS–HERMITE SERIES

An approximation to non-Gaussian velocity distributions is given by the Gauss–Hermite series (Gerhard 1993; van der Marel & Franx 1993),

$$f(x) = \frac{\lambda\alpha(w)}{\sigma} \left(1 + \sum_{i \geq 3} h_i H_i(w) \right); \quad w = \frac{x - V}{\sigma}. \quad (1)$$

Here, $H_i(w)$ are Hermite polynomials and

$$\alpha(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} \quad (2)$$

is the weight function. The Hermite polynomials satisfy the orthogonality condition

$$\int_{-\infty}^{\infty} dw H_m(w) H_n(w) \alpha^2(w) = \frac{1}{\sqrt{4\pi}} \delta_{mn}. \quad (3)$$

In Appendix A, we give the expression for the n th Hermite polynomial and the explicit formulae for the first five. The Gauss–Hermite series is similar to the Gram–Charlier and Edgeworth series familiar from the theory of statistics (Kendall & Stuart 1977) and discussed in Appendix D.

Typically, only two terms in the series (1) are used giving five fitting parameters $p = (\lambda, V, \sigma, h_3, h_4)$. The first three terms give the amplitude, mean, and dispersion of the distribution, whilst h_3 describes the skewness and h_4 the excess kurtosis (positive h_4 produces broader tails, whilst negative h_4 produces more truncated tails). Although V and σ are not strictly equal to the mean and dispersion for non-zero h_i , by the orthogonality properties of the series, they are the mean and dispersion of the best Gaussian fit. As the series is constructed from orthogonal polynomials, for some parameter choices the Gauss–Hermite series can have regions of negativity (e.g. see fig. 1 of van der Marel & Franx 1993). Assuming we can interpret $f(x)$ as a PDF, the choice of λ that normalizes $f(x)$ such that $\int dx f(x) = 1$ is $\lambda = (1 + \sqrt{3/8}h_4)^{-1}$.

2.1 Fitting a Gauss–Hermite series to data

The Gauss–Hermite series was introduced as an approximation to a galaxy’s line-of-sight velocity distribution (LOSVD) as measured from spectra (Gerhard 1993; van der Marel & Franx 1993). Given a profile $\mathcal{L}(x)$ measured with uncertainty $\sigma_{\mathcal{L}}(x)$, the best-fitting parameters p are measured by minimizing

$$\int dx \frac{(\mathcal{L}(x) - f(x))^2}{\sigma_{\mathcal{L}}^2(x)}. \quad (4)$$

In practice, this integral is performed as a sum over spectral pixels. For a model galaxy template spectrum, the kinematics is incorporated through convolution with a Gauss–Hermite series (Cappellari & Emsellem 2004). As discussed by Cappellari (2017), when the dispersion is smaller than half the pixel spacing, this convolution is most efficiently performed by first taking the analytic Fourier transform of the Gauss–Hermite series, before taking the discrete inverse transform of the product of the Fourier template and the Fourier Gauss–Hermite series. Further Gaussian broadening is trivially incorporated in Fourier space. In these methods, the requirement of a well-defined probability density function ($f(x) > 0$ everywhere) is unimportant as it does not significantly alter the computation of equation (4). In these cases, it is possible to introduce a truncation by setting $f(x) = 0$ when $f(x) < 0$ that can improve the fit for $f(x) > 0$ (van der Marel & Franx 1993).

When we have discrete velocity measurements $\{x_i\}$, as is common for Milky Way or Local Group studies, these fitting methods are not always easily applicable. One simple method applicable for error-free data is to relate h_3 and h_4 to moments of the data. As shown in Appendix B, for small deviations from Gaussianity, h_3 and h_4 are related to the skewness g and excess kurtosis κ as

$$h_3 = \frac{g}{4\sqrt{3}}, \quad h_4 = \frac{\kappa}{8\sqrt{6}}. \quad (5)$$

For N samples, the corresponding variance in these quantities (for near-Gaussian distributions) are (Kenney & Keeping

1951)

$$\text{Var}(h_3) = \frac{N(N-1)}{8(N-2)(N+1)(N+3)},$$

$$\text{Var}(h_4) = \frac{N(N-1)^2}{16(N-3)(N-2)(N+3)(N+5)}. \quad (6)$$

Typically, however, equation (5) gives quite poor approximations for h_3 and h_4 (in particular h_4 where it is only valid to 10 percent for $|h_4| < 0.01$). In turn, equation (6) gives poor estimates of the expected error and only appears valid using Gaussian-distributed samples of size $\gtrsim 1000$. For smaller samples, following Amorisco & Evans (2012b), the uncertainty in h_3 and h_4 is better approximated as $(2N)^{-1/2}$.

A further method is to bin the data into an approximation of $\mathcal{L}(x)$ and minimize equation (4). This is undesirable, as it generically loses information through smoothing and typically we only have a small numbers of measurements to work with. Sellwood & Gerhard (2020) propose a method for avoiding binning where the (analytic) cumulative distribution function for the Gauss–Hermite series is fitted to the cumulative velocity distribution of simulation data. Such a method does not adapt well when observational uncertainties are considered.

To avoid binning and fully incorporate the measurement uncertainties for a set of samples $\{x_i\}$ measured with uncertainties $\{\sigma_{ei}\}$, we should maximize

$$\prod_i \int_{-\infty}^{\infty} dx' f(x') \mathcal{N}(x_i - x' | \sigma_{ei}), \quad (7)$$

where

$$\mathcal{N}(x | \sigma) = (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2}x^2/\sigma^2}. \quad (8)$$

Generically, the Gauss–Hermite series (and other series expansions) produces negative values of $f(x)$ for some choice of x . This makes the interpretation of $f(x)$ as a PDF awkward. Regions of negative probability are particularly an issue for negative h_4 , where the PDF is zero somewhere for any reasonable choice of h_3 . However, usually the region of negative PDF is small. This deficiency is not simple to fix with the addition of higher order terms. An alternative is to truncate the PDF at the first zero crossing. The roots of the Gauss–Hermite series for $h_3 = 0$ are

$$w = \frac{3}{2} \pm \frac{1}{\sqrt{2}h_4} \sqrt{3h_4^2 - \sqrt{6}h_4}. \quad (9)$$

It is then possible to analytically normalize $f(x)$ between these limits, although the expressions are long-winded. Incorporating these workarounds into a likelihood call significantly increases the cost of a likelihood evaluation and is not easily adapted for $h_3 \neq 0$.

In the next subsection, we will demonstrate how the Gauss–Hermite series can be used when $f(x)$ is a well-defined probability density function, before we move on to presenting an alternative set of models.

2.2 Convolution of the Gauss–Hermite series with observational uncertainties

A sample of points x_i drawn from $f(x)$ but measured with some uncertainty σ_e follow the distribution $f_{\sigma_e}(x)$ given by

$$f_{\sigma_e}(x) = \int_{-\infty}^{\infty} dx' f(x') \mathcal{N}(x - x' | \sigma_e). \quad (10)$$

Given the form of the Gauss–Hermite series, it is possible to compute this convolution analytically. In Appendix C, we provide the details and here quote the final result as

$$f_{\sigma_e}(x) = \frac{\lambda}{\sigma'} \alpha(w') \left\{ 1 + \sum_{n \geq 3} h_n \left(\frac{\sigma}{\sigma'} \right)^n \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{\sigma_e}{\sigma} \right)^{2j} \times \frac{\sqrt{n!}}{j! 2^j \sqrt{(n-2j)!}} H_{n-2j}(w') \right\}, \quad (11)$$

where $w' = (x - V)/\sigma'$ and $\sigma' = \sqrt{\sigma^2 + \sigma_e^2}$. For the truncated series $h_i = 0$ for $i > 4$, we find

$$f_{\sigma_e}(x) = \frac{\lambda}{\sigma'} \alpha(w') \left\{ 1 + \frac{h_3}{(\sigma'/\sigma)^3} \left[H_3(w') + \sqrt{\frac{3}{2}} \left(\frac{\sigma_e}{\sigma} \right)^2 H_1(w') \right] + \frac{h_4}{(\sigma'/\sigma)^4} \left[H_4(w') + \sqrt{3} \left(\frac{\sigma_e}{\sigma} \right)^2 H_2(w') + \sqrt{\frac{3}{8}} \left(\frac{\sigma_e}{\sigma} \right)^4 H_0(w') \right] \right\}. \quad (12)$$

This function can now be used as the likelihood of data with heteroskedastic errors σ_{ei} to recover the underlying h_3 and h_4 moments, provided $f(x)$ is everywhere positive.

Equation (12) can be used to assess the impact of uncertainties on the recovery of the Gauss–Hermite coefficients. We observe for $h_3 = h_4 = 0$, we recover the standard result that the convolution broadens the Gaussian from dispersion σ to σ' . Note that for non-zero h_3 and h_4 , the convolution also introduces contributions from $H_1(w')$ and $H_2(w')$ that we identify as further modifications to the mean and dispersion. However, in the limit of small uncertainties $\sigma_e \ll \sigma$, we find

$$f_{\sigma_e}(x) \approx \frac{\lambda}{\sigma'} \alpha(w') \left(1 + h'_3 H_3(w') + h'_4 H_4(w') \right), \quad (13)$$

where $h'_3 = h_3(\sigma/\sigma')^3$ and $h'_4 = h_4(\sigma/\sigma')^4$. Therefore, small uncertainties do not bias the mean, but bias the dispersion in the expected way: $\sigma'^2 = \sigma^2 + \sigma_e^2$. We remark that h_3 and h_4 are also biased: without accounting for uncertainties h_3 is reduced by a factor of $(1 + (\sigma_e/\sigma)^2)^{-3/2}$ and h_4 by a factor of $(1 + (\sigma_e/\sigma)^2)^{-2}$. In Fig. 1, we show the result of fitting a Gauss–Hermite series with $h_3 = 0.06$ and $h_4 = 0.1$ to a sample of $N \approx 6000$ data points scattered by two choices of uncertainty $\sigma_e/\sigma = 0.2$ and $\sigma_e/\sigma = 0.5$. We see the true parameters are recovered when sampling from the likelihood $\sum_i \ln f_{\sigma_{ei}}(x_i)$ using the EMCEE sampler (Foreman-Mackey et al. 2013). If one fails to account for the uncertainties, we recover the expected biases in h_3 and h_4 (see Amorisco & Evans 2012b, for a similar discussion).

We have demonstrated how the Gauss–Hermite series can be used to model discrete data with heteroskedastic uncertainties, but only when the series is positive definite. Generically, for negative h_4 the series will produce negative wings making its use in a probabilistic framework awkward. In the next section, we will introduce a method to produce everywhere positive-definite PDFs that can be used as replacements for the Gauss–Hermite series.

3 A GENERAL FAMILY OF WEAKLY NON-GAUSSIAN PDFS

Our goal is to construct a family of everywhere positive PDFs, which can model weakly non-Gaussian distributions and also are able to incorporate observational uncertainties. This means that the models must have an easily computable convolution with a Gaussian kernel.

This class of models is very desirable for modelling star-by-star data with heteroskedastic observational uncertainties, as otherwise the likelihood call for each star requires an integration over the uncertainty. We begin with a general scheme for constructing such models before discussing in detail two specific examples useful for modelling positive and negative excess kurtosis.

3.1 Model specification

With the requirement of analytic convolution with uncertainties, it is logical to investigate models that have been constructed by a convolution of a general kernel with a Gaussian. Further Gaussian kernel convolutions can then be simply performed and transform the models into another Gaussian convolution with a slightly modified general kernel. The models are

$$f(w) = b \int_{-\infty}^{\infty} dy K(y) \alpha(y - b(w - w_0)), \quad (14)$$

where $\alpha(y)$ is a unit Gaussian as defined in equation (2) and $K(y)$ is a general kernel normalized to have unit weight. w_0 and b are parameters controlling the mean and variance of the model. To reproduce the unit Gaussian model $\alpha(w)$ when $K(y) = \delta(y)$, we require $b = 1$ and $w_0 = 0$ in this limit, which implies b and w_0 in general depend upon the parameters of the kernel $K(y)$. In the opposite limit where the kernel $K(y)$ is significantly broader than the unit Gaussian, $\alpha(y - b(w - w_0)) \rightarrow \delta(y - b(w - w_0))$ so $f(w) \rightarrow bK(b(w - w_0))$, a scaled and shifted version of the kernel.

As with the Gauss–Hermite series, we have the freedom to introduce an arbitrary shift and scale to redefine the PDF in terms of $x = V + \sigma w$.

3.1.1 Model moments and cumulants

The moments of these models are given by

$$\begin{aligned} \mu_n(c) &= \int_{-\infty}^{\infty} dw (w - c)^n f(w) \\ &= \frac{1}{b^n} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k} (2k - 1)!! \tilde{\mu}_{n-2k}(bc - bw_0), \end{aligned} \quad (15)$$

where $\tilde{\mu}_n(c)$ are the moments of $K(y)$ (we use a tilde to denote properties of $K(y)$). For instance, the mean $\mu \equiv \mu_1(0)$ is given by

$$\mu = \frac{1}{b} \int_{-\infty}^{\infty} dy y K(y - bw_0) = \frac{\tilde{\mu}}{b} + w_0, \quad (16)$$

where $\tilde{\mu} \equiv \tilde{\mu}(0)$. The central moments are defined as $\mu_n \equiv \mu_n(\mu)$, which can be combined in the usual way to find the cumulants. However, the cumulants κ_n are more simply computed via the cumulant generating function $\ln \phi(u)$ defined as

$$\ln \phi(u) = \ln \int_{-\infty}^{\infty} dw e^{iuw} f(w) = \sum_{r=1}^{\infty} \kappa_r \frac{(iu)^r}{r!}. \quad (17)$$

In this way, the r th cumulant is

$$\kappa_r = \frac{1}{i^r} \left. \frac{d^r \ln \phi(u)}{du^r} \right|_{u=0}. \quad (18)$$

From the properties of the Fourier transform, the cumulant generating function of $f(w)$ is given by

$$\ln \phi(u) = iw_0 u - \frac{u^2}{2b^2} + \ln \phi_K(u/b), \quad (19)$$

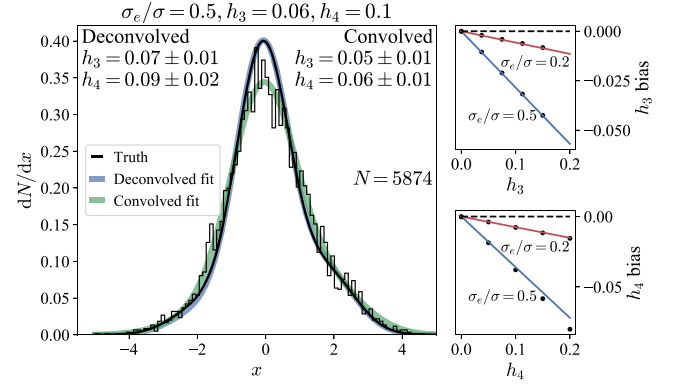


Figure 1. Gauss–Hermite series fitting to a mock distribution: the left-hand panel shows a sample of N draws (the black histogram) from a Gauss–Hermite series with $h_3 = 0.06$ and $h_4 = 0.1$ (the black line). The samples have then been scattered by an uncertainty of 0.5. The blue line is the fit result accounting for uncertainties, whilst green is without accounting for uncertainties. Note that the recovered h_3 and h_4 are higher when the uncertainties are accounted for. This is also depicted in the right-hand panels that show the bias in h_3 and h_4 as a function of these parameters if one fails to account for the uncertainties. Two choices of error are shown (0.2 and 0.5). The lines give the approximations $h_3((1 + (\sigma_e/\sigma)^2)^{-3/2} - 1)$ and $h_4((1 + (\sigma_e/\sigma)^2)^{-2} - 1)$.

where $\phi_K(u)$ is the characteristic function of $K(y)$,

$$\phi_K(u) = \int_{-\infty}^{\infty} dw e^{iuy} K(y). \quad (20)$$

From the cumulants, we define the variance $v \equiv \kappa_2$, the skewness $g \equiv \kappa_3/v^{3/2}$, and the excess kurtosis $\kappa \equiv \kappa_4/v^2$. We note that the variance of $K(y)$, \tilde{v} , is related to the variance of $f(w)$ by

$$v = \frac{1}{b^2}(\tilde{v} + 1), \quad (21)$$

and all higher order cumulants ($r \geq 3$) satisfy $\kappa_r = \tilde{\kappa}_r/b^r$. This demonstrates that the signs of the skewness and the excess kurtosis of the kernel govern, respectively, the signs of the skewness and the excess kurtosis of the model.

3.1.2 Convolution with uncertainties

With such models, the convolution with observational uncertainties of magnitude s is

$$\begin{aligned} f_s(w) &= \int_{-\infty}^{\infty} dw' f(w') \mathcal{N}(w - w'|s) \\ &= b \int_{-\infty}^{\infty} dy K(ty) \alpha(y - b(w - w_0)/t), \end{aligned} \quad (22)$$

where $t^2 = 1 + b^2 s^2$ and $f(w)$ is recovered for $t = 1$ (i.e. $f_0(w) = f(w)$). In principle, if the convolution with $K(y)$ is analytic, the convolution with $K(ty)$ is likely to be analytic. As the convolution is a multiplication in Fourier space, the cumulant generating function of the error-convolved model is given by

$$\ln \phi_s(u) = \ln \phi(u) - \frac{u^2 s^2}{2}. \quad (23)$$

3.1.3 Gauss–Hermite coefficients

The (unnormalized) Gauss–Hermite coefficients of this class of models are given by

$$h_n = \sqrt{4\pi} \int_{-\infty}^{\infty} dw \alpha(w) H_n(w) f(w). \quad (24)$$

Note that in the limit where the kernel K is significantly broader than the unit Gaussian, the Gauss–Hermite coefficients are given by

$$h_n \rightarrow \sqrt{4\pi b} \int_{-\infty}^{\infty} dy K(b(y - w_0)) \alpha(y) H_n(y), \quad (25)$$

i.e. they are the Gauss–Hermite coefficients of $K(b(y - w_0))$. A general expression for the Gauss–Hermite coefficients is found via the generating function of the Hermite polynomials given in equation (A3). We compute the integral

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} \sqrt{n!} h_n = \sqrt{4\pi b} \int_{-\infty}^{\infty} dy K(y) \int_{-\infty}^{\infty} dw \alpha(y - b(w - w_0)) \times \alpha(w) e^{\sqrt{2}wz - z^2/2}. \quad (26)$$

Performing the w integration leaves us with a single integral that is the same as the integral in equation (22) with $t^2 = 1 + b^2$, i.e. $s = 1$. We therefore can express h_n in the compact form

$$h_n = \sqrt{4\pi} \sqrt{\frac{2^n}{n!}} \frac{\partial^n}{\partial z^n} \left[e^{z^2/4} f_1(z) \right]_{z=0}. \quad (27)$$

We conclude that if the error-convolved model has an analytic form, the Gauss–Hermite coefficients and error-deconvolved models can also be expressed analytically. The requirement $h_1 = h_2 = 0$ is equivalent to $f_1(z) \sim e^{-z^2/4} (1 + qz^3)$ for small z and constant q . To match the definition of h_n given in equation (1), we normalize h_n by $h_0 = \sqrt{4\pi} f_1(0)$ such that the normalized $h_0 = 1$. All subsequent references to h_n are to their normalized versions.

3.1.4 Model parametrization

There are two options for how to restrict the kernels (in particular, the parameters b and w_0) when modelling: either we choose to fix the mean $\mu = 0$ and variance $v = 1$, or we fix $h_1 = h_2 = 0$. The first option is analytically simpler as we have shown the moments of $f(w)$ are related to moments of $K(y)$. The latter option follows the parametrization of van der Marel & Franx (1993) who only consider non-zero h_n for $n \geq 3$. In general, h_n are complicated functions of the model parameters that do not have analytic roots, so we can only set $h_1 = h_2 = 0$ approximately. We know the two options are equivalent for narrow kernels when $K(y) = \delta(y)$. In the wide $K(y)$ limit, we can use equation (25) to simplify the mathematics. Our general procedure for setting $h_1 \approx h_2 \approx 0$ is then stitching together the constant variance solution and the wide kernel solution. For symmetric kernels $K(y)$ ($\tilde{\mu}_n(0) = 0$ for odd n) setting $w_0 = 0$ ensures $\mu = 0$ and $h_1 = 0$ for all b .

3.1.5 Half-kernels

The parity of the model $f(w)$ is equal to the parity of the kernel $K(y)$. We will find it useful to build arbitrary kernels from half-kernels,

$$K(y) = \begin{cases} K_+(y) & y \geq 0, \\ K_-(y) & y < 0, \end{cases} \quad (28)$$

such that $f(w) = f_+(w) + f_-(w)$. For a positive half-kernel $K_+(y)$, the corresponding negative half-kernel is given by $K_-(y) = K_+(-y)$. Therefore, by the symmetry properties of $f(w)$ and $K(y)$, we have $f_-(w) = f_+(-w)$. Additionally, the characteristic functions will satisfy the symmetry $\phi_{K_-}(u) = \phi_{K_+}(-u)$. In general, we can stitch together any two half-kernels to build a full kernel. This is a simple way to introduce skewness. There are two possibilities: either we ensure $K_+(0) = K_-(0)$ or $\int_0^{\infty} dy K_+(y) = \int_{-\infty}^0 dy K_-(y) = \frac{1}{2}$. Here we use the latter choice.

4 SPECIFIC CHOICES OF KERNEL

With our general theory established, we now turn to some possible choices of kernel $K(y)$. We note here the similarity between our procedure and the method of Long & Murali (1992) for the construction of barred potentials via convolution with a kernel along the bar's major axis. Williams & Evans (2017) and McGough, Evans & Sanders (2020) showed how flat-topped and cuspy barred potentials could be constructed by this method through convolution with a uniform and Laplace kernel, respectively. Here we will follow those choices to construct two families of PDFs.

4.1 A negative excess kurtosis family – a uniform kernel

We first explore a simple symmetric kernel that produces a family of models with negative excess kurtosis. We later expand this model to incorporate skewness. As the models are symmetric $w_0 = 0$ satisfies both $\mu = 0$ and $h_1 = 0$. Our choice for $K(y)$ is given by a top-hat function

$$K(y) = \begin{cases} \frac{1}{2a} & \text{if } |y| < a, \\ 0 & \text{otherwise} \end{cases} \quad (29)$$

for $a > 0$, which has an error-convolved model given by

$$f_s(w) = \frac{b}{2a} \left[\Phi\left(\frac{bw+a}{t}\right) - \Phi\left(\frac{bw-a}{t}\right) \right], \quad (30)$$

where $t^2 = 1 + b^2 s^2$ and $\Phi(x)$ is the cumulative distribution function for the unit Gaussian PDF:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x dt e^{-t^2/2}. \quad (31)$$

The model has zero mean, and the variance and excess kurtosis are given by

$$v = \frac{1}{b^2} \left(1 + \frac{a^2}{3} \right), \quad \kappa = -\frac{2a^4}{15} \left(1 + \frac{a^2}{3} \right)^{-2}. \quad (32)$$

We observe that there is a simple choice for b such that the variance is unity: $b^2 = (1 + a^2/3)$.

4.1.1 Fixing the Gauss–Hermite coefficients

We now turn to the selection of b such that $h_1 \approx h_2 \approx 0$. As already highlighted, this cannot be done exactly and so only approximate expressions can be derived and choices of fitting functions must be made. For the uniform model, we explicitly detail the procedure we follow, which can be approximately followed for any other choice of model.

In the limit $a \rightarrow \infty$, we require $alb \rightarrow k$ for constant k such that the variance and h_2 remain finite. We introduce a general functional

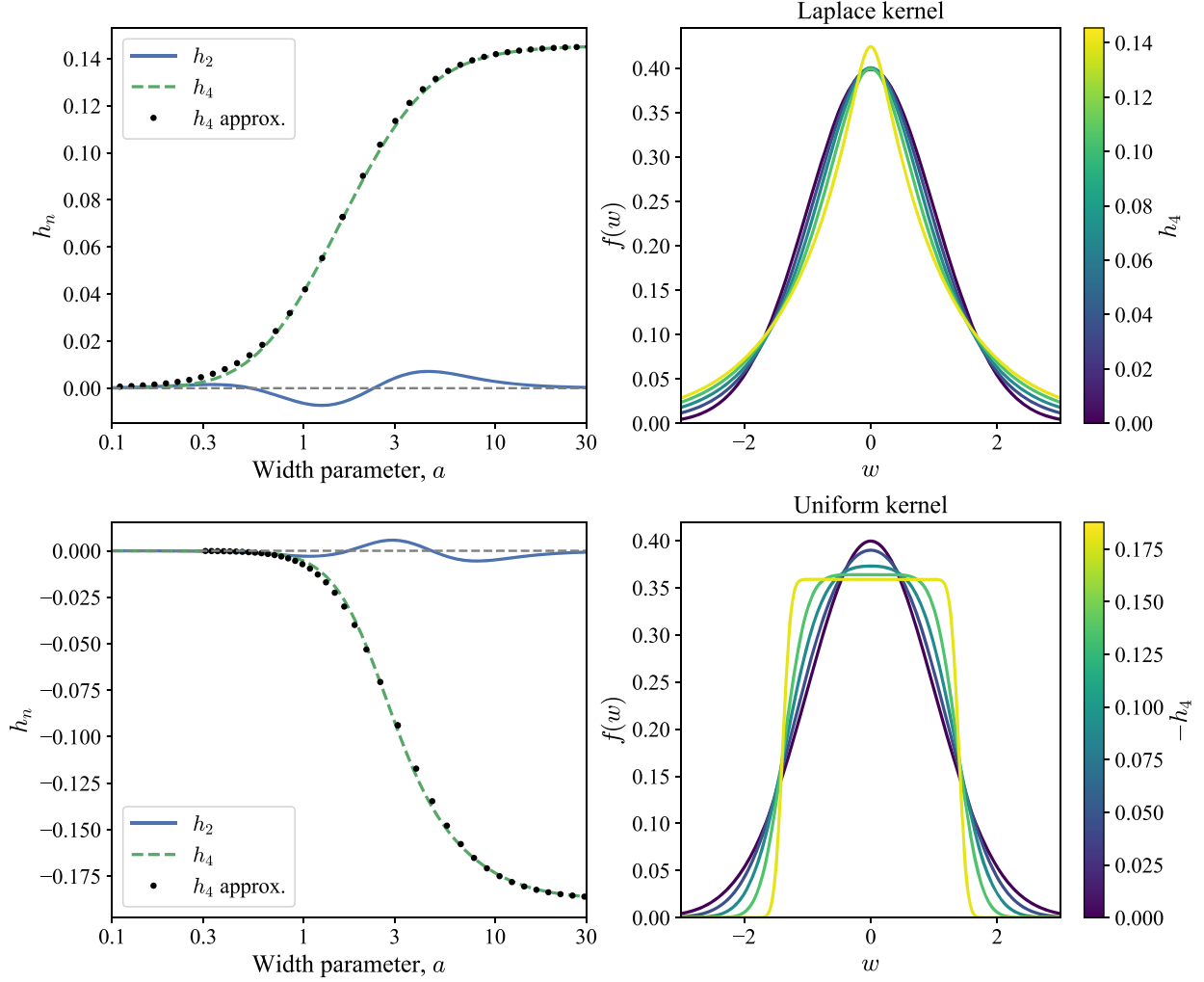


Figure 2. Our two families of PDFs with positive (top) and negative (bottom) excess kurtosis (setting skewness or h_3 to zero). The left-hand panels show the second (blue) and fourth (green) Gauss–Hermite coefficients of the models as a function of the model parameter, a . The model parameters have been selected to minimize variation of h_2 . The right-hand panels show the probability density functions in terms of the scaled coordinate $w = (x - V)/\sigma$ for different choices of h_4 .

dependence of b of the form

$$b^2 = 1 + \left(\frac{a}{k}\right)^2, \quad (33)$$

which encompasses the low (unity variance) and high a limits for different choices of k . To recover the constant variance choice, we can set $k = k_0 = \sqrt{3}$. Using equations (25) and (27), we determine the normalized h_2 and h_4 for large a as

$$\begin{aligned} \lim_{\substack{a \rightarrow \infty \\ a/b \rightarrow k}} h_2 &= \sqrt{\frac{1}{2}} - \frac{2}{\sqrt{\pi}} k e^{-k^2/2} / \operatorname{erf}\left(\frac{k}{\sqrt{2}}\right), \\ \lim_{\substack{a \rightarrow \infty \\ a/b \rightarrow k}} h_4 &= \sqrt{\frac{3}{8}} - \frac{2}{\sqrt{3\pi}} k^3 e^{-k^2/2} / \operatorname{erf}\left(\frac{k}{\sqrt{2}}\right). \end{aligned} \quad (34)$$

We observe that there is a choice of $k = k_\infty$ that sets $\lim_{a \rightarrow \infty} h_2 = 0$ given by $k_\infty = 1.399985\dots$, amazingly close to $7/5$. With this choice, $\lim_{a \rightarrow \infty} h_4 = -0.187777\dots \equiv h_{4,\infty}$. Therefore, to replicate the modelling with a Gauss–Hermite series that imposes the requirement that $h_2 = 0$ we can set $k = k_\infty$. This choice produces $|h_2| \lesssim 0.05$

everywhere. To improve this, we make k a weak function of a that reproduces the constant variance limit $k = k_0$ for small a as

$$k(a) = k_0 - (k_0 - k_\infty) \tanh(a/a_0), \quad (35)$$

where setting $a_0 = 3.3$ minimizes the variation of h_2 to $|h_2| \lesssim 0.006$. We show this set of models in the bottom panels of Fig. 2. We find that a good approximation for a given h_4 is

$$a = a_{h_4} \left(\sqrt{\frac{h_{4,\infty}}{h_4}} - 1 \right)^{-1/2}, \quad (36)$$

where $a_{h_4} = 2$. This fitting function is shown in Fig. 2. In this way, we can parametrize our models in terms of h_4 .

4.1.2 Incorporating skewness

Following the discussion in Section 3.1.5, a generalization of this model that incorporates some skewness is found by using a different

Table 1. Key formulae for the two families of distributions introduced in this paper. $f_s(w)$ is the probability density function of the scaled and shifted coordinate $w = (x - V)/\sigma$ convolved with observational uncertainties σ_e such that $s = \sigma_e/\sigma$. $\Phi(x)$ is the cumulative of the unit normal distribution. Further generic definitions are $w' = w - w_0$, $t = 1 + b^2 s^2$, $a_+ = a + \Delta$, $a_- = a - \Delta$, $\delta = \Delta/a$, $b^2 = 1 + (a/k)^2$, and $k = k_0 - (k_0 - k_\infty)\tanh(a/a_0)$. The parameters a , b , and Δ are chosen as per the ‘auxiliary functions’ and ‘parameters’ columns to ensure $h_1 \approx h_2 \approx 0$ and to match a given h_3 and h_4 . For a given h_3 and h_4 , the quadratic for δ is solved, a is computed using $h_{4,\max}(\delta)$, and finally b and w_0 are found. Code for the computation of all formulae is provided at <https://github.com/jls713/gh.alternative>.

Name	$f_s(w)$	Auxiliary functions	Parameters					
			$h_{4,\infty}$	a_0	a_{h_3}	a_{h_4}	k_0	$k_{\infty,0}$
Uniform	$\frac{b}{2a_+a_-} \left[a_+ \Phi\left(\frac{bw' + a_-}{t}\right) - a_- \Phi\left(\frac{bw' - a_+}{t}\right) - 2\Delta \Phi\left(\frac{bw'}{t}\right) \right]$	$h_{4,\max} = (1 - 4.3\delta^2)h_{4,\infty}$	-0.1878	3.3	2	2	$\sqrt{3}$	$\frac{7}{5}$
		$h_{3,\max} = 0.82\delta$						
Laplace	$\frac{b}{4a_+} \exp\left(\frac{t^2 - 2a_+bw'}{2a_+^2}\right) \operatorname{erfc}\left(\frac{t^2 - a_+bw'}{\sqrt{2}ta_+}\right) + \frac{b}{4a_-} \exp\left(\frac{t^2 + 2a_-bw'}{2a_-^2}\right) \operatorname{erfc}\left(\frac{t^2 + a_-bw'}{\sqrt{2}ta_-}\right)$	$h_{4,\max} = (1 + 2\delta^2)h_{4,\infty}$	0.1454	2.25	1.1	1.6	$\frac{1}{\sqrt{2}}$	1.08
		$h_{3,\max} = 0.37\delta$						

uniform distribution for positive and negative y as

$$K(y) = \begin{cases} \frac{1}{2a_+} & \text{if } 0 < y < a_+, \\ \frac{1}{2a_-} & \text{if } -a_- < y < 0, \\ 0 & \text{otherwise} \end{cases} \quad (37)$$

for $a_+ > 0$ and $a_- > 0$ that produces an error-convolved model of the form

$$f_s(w) = \frac{b}{2a_+a_-} \left[a_+ \Phi\left(\frac{bw' + a_-}{t}\right) - a_- \Phi\left(\frac{bw' - a_+}{t}\right) - 2\Delta \Phi\left(\frac{bw'}{t}\right) \right], \quad (38)$$

where $w' = w - w_0$, $a = \frac{1}{2}(a_+ + a_-)$, and $\Delta = \frac{1}{2}(a_+ - a_-) < a$. We describe numerical calculation of $f_s(w)$ in Appendix F. In Table E2, we provide the cumulants for this model. As expected, Δ controls the skewness of the distribution, which we parametrize in a scale-free way using $\delta = \Delta/a$.

We use equation (25) to find expressions for h_1 and h_2 in the limit of large a and $ab = k$. Adopting the choice of k for $\Delta = 0$, $k = k_\infty = 7/5$, we find $w_0 \approx -\Delta/6b$ approximately sets $h_1 = 0$. We also find that k_∞ must be made a function of $\delta = \Delta/a$ to retain $h_2 = 0$. With these choices, we set

$$k_\infty = \frac{7}{5} \sqrt{1 + \delta^2 + 3\delta^4}, \quad (39)$$

$$w_0 = -\frac{\Delta}{2b} + \frac{\Delta}{3b} \tanh(a/a_0),$$

where $\delta = \Delta/a$. We find that the maximum $|h_4|$ is $h_{4,\max} \approx (1 - 4.3\delta^2)h_{4,\infty}$ and that the maximum h_3 is $|h_{3,\max}| \approx 0.82|\delta|$. We adopt the same functional form for $h_3(a)$ as $h_4(a)$, which means $h_3/h_4 = h_{3,\max}/h_{4,\max}$. For a choice of h_3 and h_4 , we first find δ by solving a quadratic. We then obtain a from equation (36).

The full set of parameters describing the model is given in Table 1.

In Fig. 3, we show the variation of h_n for this model as a function of a and δ . We see how our choices have ensured $h_1 \approx 0$ and $h_2 \approx 0$. δ approximately controls h_3 at fixed a , and a controls h_4 , although there is some variation of h_4 with δ .

4.2 A positive excess kurtosis family – a Laplace kernel

An extension to positive excess kurtosis can be constructed using a Laplace kernel. The exponential naturally produces broader tails than the Gaussian. The choice of kernel is

$$K(y) = \begin{cases} \frac{1}{2a_+} e^{-y/a_+} & \text{if } y > 0, \\ \frac{1}{2a_-} e^{y/a_-} & \text{if } y < 0 \end{cases} \quad (40)$$

for $a_+ > 0$ and $a_- > 0$, which produces an error-convolved distribution of the form

$$f_s(w) = \frac{b}{4a_+} \exp\left(\frac{t^2 - 2a_+bw'}{2a_+^2}\right) \operatorname{erfc}\left(\frac{t^2 - a_+bw'}{\sqrt{2}ta_+}\right) + \frac{b}{4a_-} \exp\left(\frac{t^2 + 2a_-bw'}{2a_-^2}\right) \operatorname{erfc}\left(\frac{t^2 + a_-bw'}{\sqrt{2}ta_-}\right), \quad (41)$$

where $w' = w - w_0$. This model is a sum of exponentially modified Gaussian distributions. Note that the products of exponentials and error functions require careful numerics when the arguments of the exponentials are large and positive. We discuss numerical implementation details in Appendix F.

In Table E2, we provide the cumulants for this model. We follow a similar procedure to the uniform kernel case for restricting the parameters b and w_0 such that $h_1 \approx h_2 \approx 0$. One key difference is that our choice of parametrization of $h_i(a)$ is

$$a = a_{h_i} \left(\frac{h_{i,\max}}{h_i} - 1 \right)^{-1/2}, \quad (42)$$

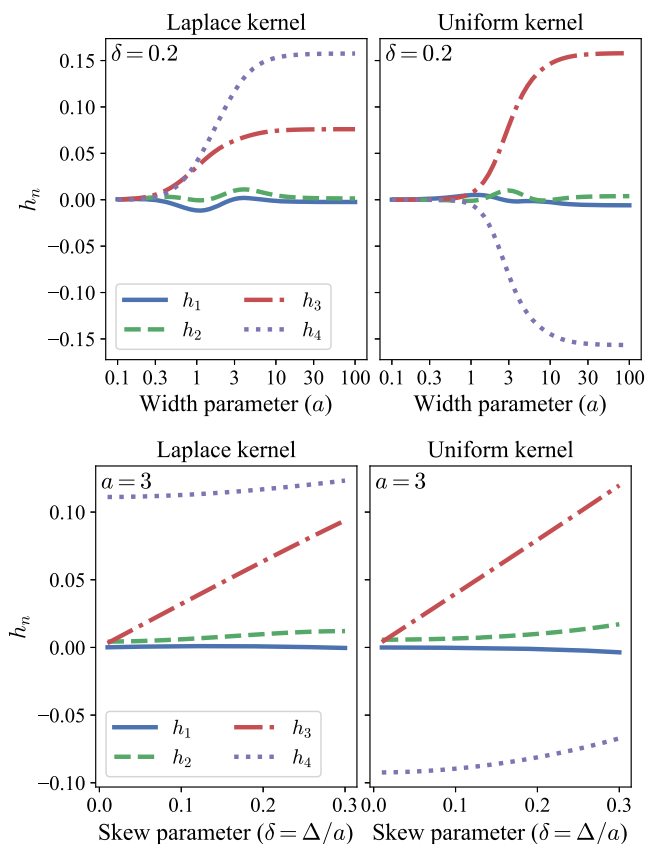


Figure 3. The Gauss–Hermite coefficients, h_n , for our two families of models. The left-hand panels correspond to the positive excess kurtosis family with a Laplace kernel, and the right-hand panels correspond to the negative excess kurtosis family with a uniform kernel. The top panels show the variation of h_n with the width parameter a (at fixed $\delta = 0.2$), and the bottom panels show the variation of h_n with the skewness parameter $\delta = \Delta/a$ (at fixed $a = 3$). Note that $h_1 \approx 0$ and $h_2 \approx 0$ by design to mirror the Gauss–Hermite series from van der Marel & Franx (1993).

with different scale parameters a_{h_i} for h_3 and h_4 . We show the accuracy of this approximation in Fig. 2 along with the PDFs for $h_3 = 0$. The full model parametrization is given in Table 1.

In Fig. 3, we show the variation of h_n for this model as a function of a and δ alongside the equivalent for the uniform kernel. We see our choices have made $h_1 \approx 0$ and $h_2 \approx 0$. h_3 is controlled by δ at fixed a , and varying a varies both h_3 and h_4 together.

4.3 Further choices of kernel

The general procedure we have employed for the uniform and Laplace kernels is applicable for a broader choice of kernels. The sole requirement is that the convolution of the kernel with a Gaussian is analytic. We have considered a number of other kernels with both positive and negative excess kurtosis. In Table E1, we detail a number of half-kernels $K_+(y)$ along with their corresponding error-convolved distributions $f_{s^+}(w)$. Along with the uniform and Laplace kernels, we give expressions for a raised cosine distribution, a Gaussian distribution, and a gamma distribution. In Table E2, we give the cumulants produced by stitching together two half-

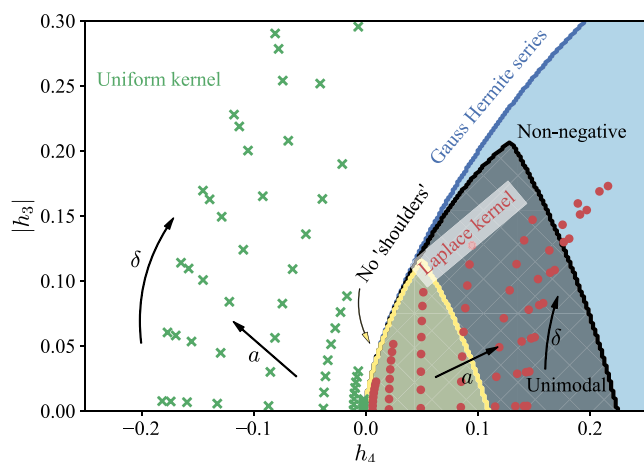


Figure 4. The space of Gauss–Hermite coefficients accessible by our new models: red points show the positive excess kurtosis family formed by the Laplace kernel, whilst green crosses show the negative excess kurtosis family formed from the uniform kernel. The parameters a and $\delta = \Delta/a$ act approximately like polar coordinates in this space. The points are spaced logarithmically in a and linearly in δ (up to a maximum of $\delta = 0.5$ as this is the range over which our approximations are valid). A Gauss–Hermite series can be constructed for all parts of the figure but only in the blue region are the series non-negative everywhere, only within the black hatched region are they unimodal and non-negative, and only within the yellow region do they have a fixed sign of the curvature in the tails.

kernels with different scales (a_{\pm}). The raised cosine distribution produces a negative excess kurtosis family ($\kappa = -0.59$) and can be asymmetricized in the same way as the Laplace family. The Gaussian family produces a family of positive excess kurtosis but non-zero excess kurtosis is always accompanied by skewness. The gamma distribution, which is only defined for positive y , produces a positive excess kurtosis family. We note that the form of the models we have introduced encompasses the well-known Voigt profile that is formed from the convolution of a Cauchy distribution with a Gaussian. It is regularly used in spectroscopy, and can be computed using the real part of the Faddeeva function.

4.4 Comparison with the Gauss–Hermite series

We close our investigation of the new families by comparison to the Gauss–Hermite series. In Fig. 4, we show the range of h_3 and h_4 accessible by our two families of models. We show for comparison the region of (h_3, h_4) space in which the Gauss–Hermite series are non-negative, non-negative and unimodal, and non-negative, unimodal, and with ‘no shoulders’, i.e. the sign of the curvature of the tails does not change. In this space, the parameters (a, δ) behave approximately as polar coordinates. The regions of accessible (h_3, h_4) form wedges highlighting a limitation of our models that the range of available h_3 depends upon h_4 and vice versa. However, the same is true of the non-negative group of Gauss–Hermite models. The unimodal, non-negative Gauss–Hermite series essentially encompasses the range of models accessible by the Laplace kernel, which suggests the Gauss–Hermite is preferable for modelling positive excess kurtosis. However, the advantage of our new family is that non-negativity is built in and does not need to be checked numerically for each model. Additionally, all of our new positive excess kurtosis models have a constant sign for the

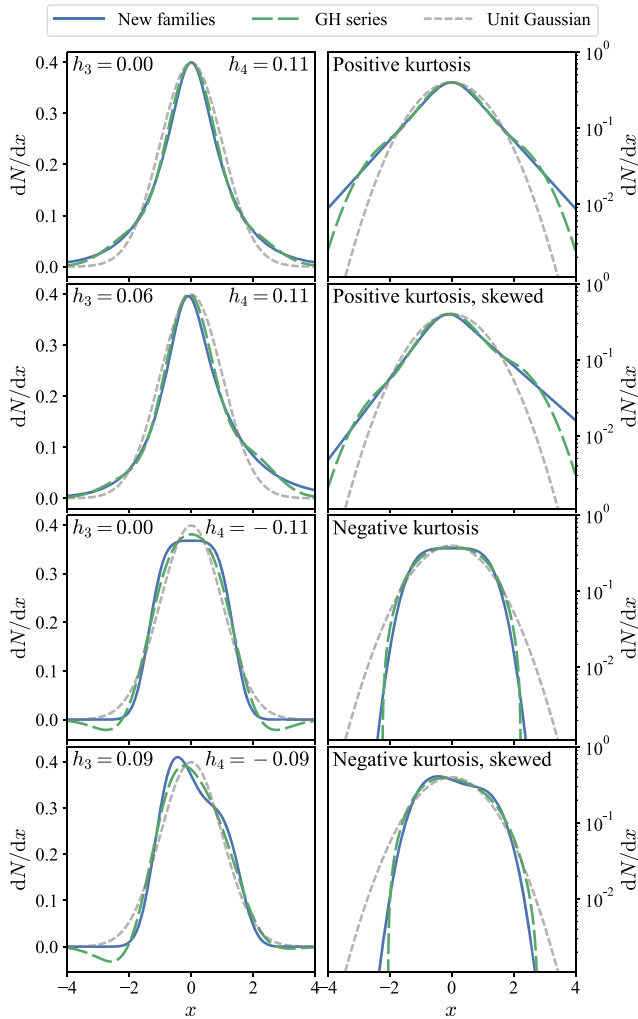


Figure 5. Comparison between the new models and the Gauss–Hermite series for four different choices of h_3 and h_4 (labelled in each row). Each set of panels show the new models in blue solid, the equivalent Gauss–Hermite series in green long-dashed, and a reference unit Gaussian in light grey short-dashed. The left-hand panels have a linear y-axis and right-hand panels logarithmic. The Gauss–Hermite series models are normalized to have unit weight over the region of positivity. The two sets of models match well but for negative excess kurtosis the negative regions of the Gauss–Hermite series are avoided.

curvature in the tails, whilst for smaller positive h_4 values than admitted by the new family the Gauss–Hermite can develop wings or shoulders.

In Fig. 5, we show the Gauss–Hermite series and the new family of models for four choices of h_3 and h_4 . There is a great deal of similarity between the two different models, although importantly the new models avoid the regions of negative probability for negative excess kurtosis. We also note that the new family of positive excess kurtosis models has stronger tails than the Gauss–Hermite series showing the presence of higher order Gauss–Hermite coefficients in their expansion. Another generic feature of our models that does not appear to arise in the Gauss–Hermite series expansion is the presence of a ‘shoulder’ for skewed negative excess kurtosis models: a feature due to our method of stitching together half-kernels.

5 APPLICATION TO DWARF SPHEROIDAL DATA

As an illustration of our approach, we will demonstrate its use on line-of-sight velocity data from dSph galaxies of the Milky Way. dSph galaxies are gas-poor elliptical low stellar mass systems (Mateo 1998). Within Λ cold dark matter theory these objects are anticipated to be highly dark matter dominated. This appears to be borne out by line-of-sight velocity measurements for the dSph galaxies of the Local Group and so make these objects ideal laboratories for investigating the properties of dark matter. The density profile of dark matter within these objects has received considerable attention, in particular the question of whether the central regions are cored or cuspy. The profile shape is a probe of both the effects of baryonic feedback on cold dark matter, or the properties of other dark matter candidates (such as self-interacting dark matter). However, pinning down the density profile is not a simple task due to the well-known mass–anisotropy degeneracy. Solutions to this include the use of proper motions (Massari et al. 2018) or performing analyses on different subpopulations within the dSph galaxies (Walker & Peñarrubia 2011; Amorisco & Evans 2012a). However, it has also been suggested that the velocity distribution contains sufficient information to break the mass–anisotropy degeneracy in its higher order moments, in particular the kurtosis (e.g. Łokas 2002; Richardson & Fairbairn 2013; Read et al. 2019). As highlighted by van der Marel & Franx (1993), kurtosis is a measure of orbital anisotropy with positive excess kurtosis associated with radial orbits. Here we will investigate how our new models can be used to constrain the higher order velocity moments from dSph data. We will focus on the specific example of the Fornax dSph as it has the largest line-of-sight velocity data set.

5.1 Dynamical models of the Fornax dSph

We take the spectroscopic sample of 2633 stars presented in Walker et al. (2009a), who provide line-of-sight velocities with uncertainties for all stars and a metallicity indicator, a pseudo-equivalent width measurement of the Mg triplet, along with its uncertainty for 459 stars. We convert the on-sky positions into projected circularized radius $R_c = \sqrt{1 - \epsilon}(x^2 + y^2/(1 - \epsilon)^2)^{1/2}$, where x/y are the positions along the major/minor axis of Fornax with respect to the dwarf centre. We take the centre of Fornax as $(\alpha, \delta) = (2^{\text{h}}39^{\text{m}}53^{\text{s}}, -34^{\circ}30'32'')$, the position angle as 42.2° , and the ellipticity of $\epsilon = 0.31$ from the King profile fit of Wang et al. (2019). Some targets were observed multiple times by Walker et al. (2009a) from which an inverse-variance-weighted mean line-of-sight velocity and its error are estimated. For the remainder of the stars we use the single velocity measurement along with its error. Pascale et al. (2018) discuss the effect of undetected binarity on the Fornax velocity distributions by comparing samples with and without binaries as determined by repeat observations. They find the distributions are very similar (less than 4 per cent probability the distributions are drawn from different samples) so we conclude binarity is unimportant and do not attempt to correct for it. We correct for the velocity gradient across the dwarf produced by perspective rotation from the bulk proper motion using the equations in Walker, Mateo & Olszewski (2008) and the proper motion measurements from Gaia Collaboration et al. (2018b).

In addition to the spectroscopic data, we use the Fornax surface density profile measurements from Dark Energy Survey (DES) data from Wang et al. (2019) and Moskowitz & Walker (2020). Wang et al. (2019) provide a best-fitting King profile to the data, although

acknowledge all the analytic forms they consider do not exactly match the data. Moskowitz & Walker (2020) provide posterior draws from the parameters of a three-component Plummer-like profile, but with a steeper $\rho \propto r^{-9}$ fall-off. We assume a distance to Fornax of 138 kpc (Mateo 1998; Rizzi et al. 2007), although our results are not particularly sensitive to this choice.

For constructing models of Fornax, we use the AGAMA dynamical modelling framework from Vasiliev (2019). Such an approach has also been taken by Pascale et al. (2018) in fitting density and velocity data on Fornax. To model the stellar distribution, we use the spheroidal distribution functions similar to those developed in work by Williams & Evans (2015) and Posti et al. (2015). They take the form

$$f(\mathbf{J}) = \frac{f_0}{(2\pi)^3} g(\mathbf{J})^{-\Gamma} \exp \left[- \left(\frac{g(\mathbf{J})}{J_{\text{cut-off}}} \right)^\zeta \right],$$

$$g(\mathbf{J}) \equiv g_r J_r + \frac{1}{2}(3 - g_r)(J_z + |J_\phi|). \quad (43)$$

The linear combination of the actions $g(\mathbf{J})$ is set to produce approximately spherical models and g_r controls the degree of radial anisotropy. The models have an exponential cut-off at scale $J_{\text{cut-off}}$ with the option of a flexible inner slope controlled by Γ . For the dark matter, we consider the simple family of density profiles,

$$\rho_{\text{DM}}(r) = \rho_0 (r/r_s)^{-\gamma} (1 + r/r_s)^{-3.2}. \quad (44)$$

For the potential from the stars, we fit the projection of a 3D power law of the form $\rho_{*,0}(1 + (r/r_*)^2)^{-9/2}$ to the Moskowitz & Walker (2020) density profile and assume a V-band stellar mass-to-light ratio of 1 and $L_V = 1.4 \times 10^7 L_\odot$ (Irwin & Hatzidimitriou 1995).

We use three values of the inner dark matter density slope $\gamma = (0, 0.5, 1)$ and by-hand approximately match the density profile and the velocity dispersion profile of Fornax as shown in Fig. 6. The chosen parameters are given in Table 2. As our models are spherical, we always compare to the circularized radius R_c distributions for the data. We also opt to weight the moments calculated from the data by membership probabilities computed in the next subsection and only consider stars within 55 km s^{-1} of the mean velocity of Fornax (approximately five times the dispersion of Fornax).

We observe that the three models have different anisotropy (the cuspy model is tangentially biased, whilst the cored model is more radially biased) that correspond to differences in the excess kurtosis. Although all three models have very similar tracer density profiles and velocity dispersion profiles, the mass density distributions, and hence anisotropy profiles are very different (e.g. Binney & Mamon 1982; Evans, An & Walker 2009). From this admittedly simplistic fitting approach, it seems the kurtosis of the data favours more cored models of Fornax, a conclusion also reached by others (Amorisco, Agnello & Evans 2013; Pascale et al. 2018; Read et al. 2019). However, this is sensitive to how one allocates membership of the dSph – for instance, Łokas (2009) and Amorisco & Evans (2012b) find negative excess kurtosis, possibly due to overly strict thresholds on membership, whilst Breddels & Helmi (2013) find the excess kurtosis is approximately zero.

5.2 Extracting kurtosis profiles

With our dynamical models for Fornax in place, we now wish to extract the kinematic properties as a function of location in the dSph given known uncertainties. For this we use the newly introduced

models. For the data, we avoid applying arbitrary cuts to the data, and instead employ a mixture model for the dwarf plus background Milky Way contaminants. To incorporate more information on dwarf membership, we use the Mg triplet equivalent width measurements from Walker et al. (2009a) in addition to the velocities. When these measurements are unavailable we assign a default value with an uncertainty of 1000 \AA .

5.2.1 Chemo-kinematic mixture model

The likelihood for an individual star at (circularized) radius R with velocity ($v \pm \sigma_v$) and equivalent width ($W \pm \sigma_w$) is

$$\begin{aligned} \mathcal{L}(v, W | \sigma_v, \sigma_w, R) &= p_{\text{mem}}(R) f_{\sigma_v}(v | \mu(R), \sigma(R), h_4(R)) \\ &\quad \times \mathcal{N} \left(W - W_d | \sqrt{\sigma_w^2 + \sigma_{W_d}^2} \right) \\ &\quad + (1 - p_{\text{mem}}(R)) B(v | \sigma_v) \\ &\quad \times \mathcal{N} \left(W - W_o | \sqrt{\sigma_w^2 + \sigma_{W_o}^2} \right), \end{aligned} \quad (45)$$

where $f_{\sigma_v}(v | \mu(R), \sigma(R), h_4(R))$ are the models introduced in this paper and the functions $\mu(R)$, $\sigma(R)$, and $h_4(R)$ are the parameters of the LOSVD as a function of radius (mean, dispersion parameter, and h_4). We assume $h_3 = 0$ everywhere. We use interpolating splines for $\mu(R)$, $\sigma(R)$, and $h_4(R)$ specifying μ_k , σ_k , and h_{4k} at four radii equally spaced in radius percentiles between the 1st and 99th of the data.

For the velocity distribution of the background Milky Way contaminants $B(v | \sigma_v)$, we use GALAXIA (Sharma et al. 2011) to sample stars in a 8 deg^2 field around Fornax and in the magnitude range $18 < V < 20$. To convert the samples into a probability density function, we utilize a Gaussian kernel density estimate with a minimum kernel size of 3 km s^{-1} . For a range of uncertainties σ_v , we use a broader kernel of $\sqrt{(3 \text{ km s}^{-1})^2 + \sigma_v^2}$ and compute the velocity and uncertainty-dependent background model $B(v | \sigma_v)$ for each star. $\mathcal{N}(W - W_i | \sqrt{\sigma_w^2 + \sigma_{W_i}^2})$ are Gaussians for the equivalent width measurements with free means W_i and widths σ_{W_i} . The probability of membership p_{mem} is a function of on-sky location,

$$p_{\text{mem}}(R) = 1 - \frac{n_b}{n_b + n_d \Sigma_d(R) / \Sigma_d(100 \text{ pc})}, \quad (46)$$

where n_b is the on-sky background density from GALAXIA ($0.23 \text{ stars arcmin}^{-2}$), n_d is a free parameter giving the on-sky dwarf central density, and $\Sigma_d(R)$ is the median normalized density profile from Moskowitz & Walker (2020).

We adopt uniform priors on μ_k , $\ln \sigma_k$, and h_{4k} , where we work with a tanh transformation of h_4 to ensure it stays within the required range $-0.188 < h_4 < 0.145$. For the equivalent-width models, we employ uniform priors for W_i and $\ln \sigma_{W_i}$. We use a uniform prior for the logarithm of the dwarf number density, $\ln n_d$. We therefore have a total of 17 parameters: three groups of four spline points ($\{\mu_k\}$, $\{\ln \sigma_k\}$, $\{h_{4k}\}$), two sets of equivalent-width mean and dispersion ($\{W_i\}$, $\{\ln \sigma_{W_i}\}$), and the density n_d . We sample from the product of the likelihood in equation (45) over stars using the EMCEE package (Foreman-Mackey et al. 2013). We focus on the results for $\sigma(R)$ and $h_4(R)$ below.

For the other parameters, we find the mean velocity of Fornax is $(55.3 \pm 0.2) \text{ km s}^{-1}$ (computed by inverse-variance weighting of the spline points), consistent with that reported by McConnachie (2012). We find an indication of radial variation of the mean velocity with the central regions ($\lesssim 100 \text{ pc}$) at a velocity $\sim 2 \text{ km s}^{-1}$ higher.

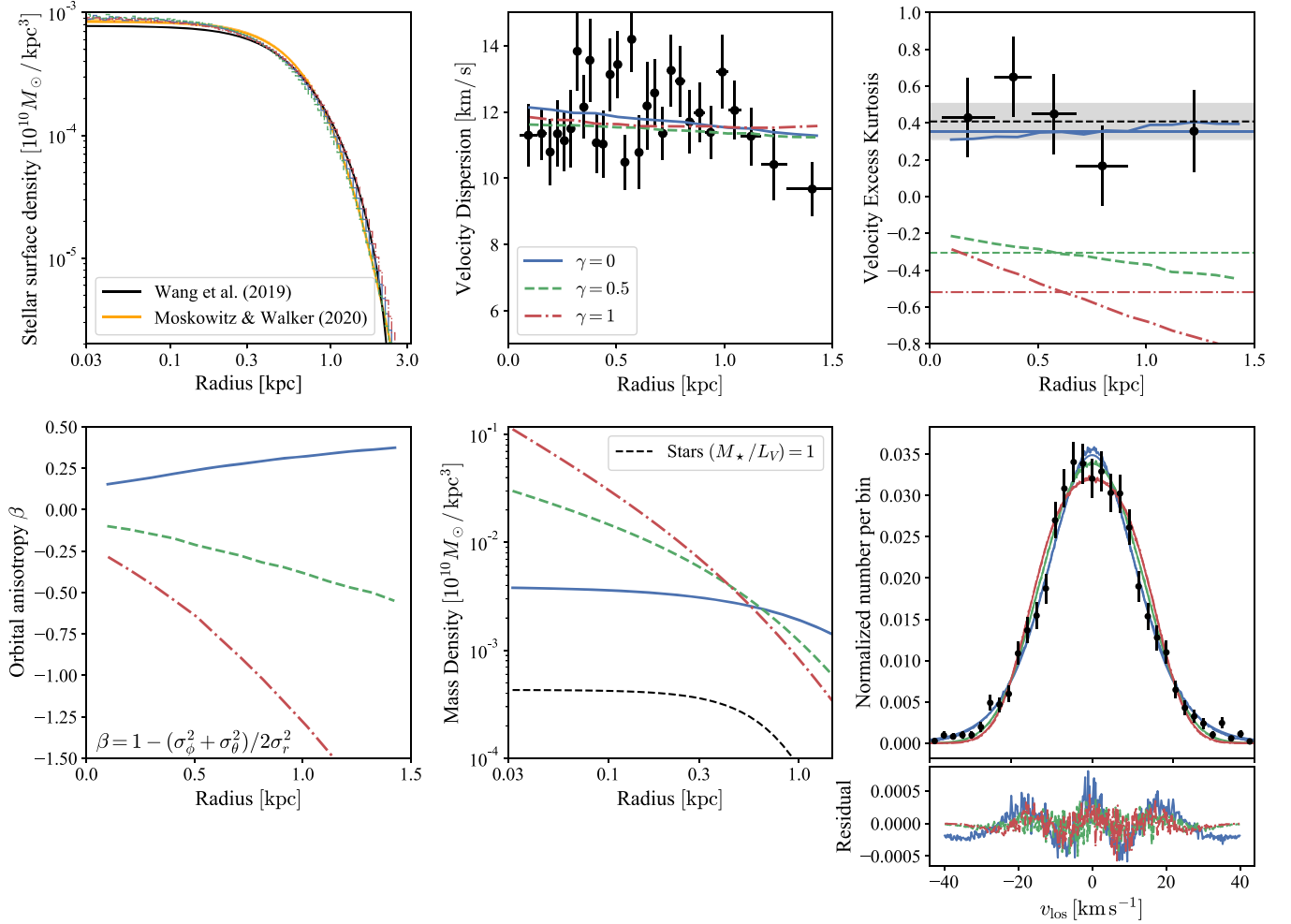


Figure 6. Dynamical models of the Fornax dwarf spheroidal (dSph) galaxy. Data are always shown in black (statistics are weighted using membership probabilities from Section 5.2, x error bars show the bin size and y error bars are from Poisson only, e.g. equation (6), so do not consider measurement uncertainties, although the dispersion is deconvolved by the median uncertainty and the three models are blue solid (cored, $\gamma = 0$), green dashed (weakly cusped, $\gamma = -0.5$), and red dash-dot (strongly cusped, $\gamma = -1$). The top left-hand panel shows the density profiles as a function of radius. For the data, we always use the circularized radius $\sqrt{1 - \epsilon}(x^2 + y^2/(1 - \epsilon^2))^{1/2}$ for flattening $\epsilon = 0.31$ to compare to our spherical models. We also display the posterior distribution of the density fits to the Dark Energy Survey (DES) data from Moskowitz & Walker (2020) and the King fit to the DES data from Wang et al. (2019) scaled to have $L_V = 1.4 \times 10^7 L_\odot$ with $(M_*/L_V) = 1$. The top middle panel shows the velocity dispersion, and top right-hand panel shows the excess kurtosis of the velocity distribution. The horizontal lines are the mean excess kurtosis of the models. The black band is the median and $\pm 1\sigma$ excess kurtosis of the model. In the bottom row, we show the orbital anisotropy (left) and the mass density (middle) along with stellar density. In the bottom right, we show the model velocity distributions (binned histograms) along with fits using the PDFs in this work. The Fornax data are also shown. The models do an excellent job of capturing the kurtosis of the dynamical models, so much so that they are barely distinguishable in the upper panel, but the deviations are shown in the lower panel of residuals.

Table 2. Parameters of our action-based Fornax models. The models differ primarily in their inner dark matter slope γ . ρ_0 and r_s are the normalization and scale radius of the dark matter halo with units $10^{10} M_\odot \text{kpc}^{-3}$ and kpc. f_0 is the normalization of the stellar distribution (with units $10^{10} M_\odot / (\text{kpc km s}^{-1})^{3-\Gamma}$). Γ controls the central density slope of the stars, $J_{\text{cut-off}}$ (with units kpc km s^{-1}) and ζ govern the location and strength of the stellar exponential break, and g_r the radial anisotropy of the stars.

γ	ρ_0	r_s	f_0	Γ	$J_{\text{cut-off}}$	g_r	ζ
0.0	0.0039	4.0	2×10^{-5}	0.0	1.3	1.2	0.6
0.5	0.005	1.3	7.6×10^{-6}	0.0	3.2	1.6	0.8
1.0	0.0038	1.0	3×10^{-6}	0.2	8.6	1.9	1.4

However, all spline points have consistent mean velocities within their uncertainties. We re-ran our models with a mean that does not vary radially and find the conclusions on the variance and kurtosis profiles unchanged. As found by Walker et al. (2009b), Fornax stars are not well separated from the background stars in the Mg equivalent width measurement. For Fornax, we find $W_d = (0.616^{+0.006}_{-0.006}) \text{ \AA}$ and $\sigma_{W_d} = (0.109^{+0.005}_{-0.005}) \text{ \AA}$, whilst for the background distribution we find $W_o = (0.52^{+0.12}_{-0.13}) \text{ \AA}$ and $\sigma_{W_o} = (0.34^{+0.11}_{-0.07}) \text{ \AA}$. Finally, for the central density of Fornax we find $n_d = (20.8^{+3.1}_{-2.5}) \text{ stars arcmin}^{-2}$. This number obviously depends on the magnitude selection of stars so it is more informative to compare relative to the background density, assumed to be $n_b = 0.23 \text{ stars arcmin}^{-2}$. We find $n_d/n_b \approx 90$, in line with the results from Battaglia et al. (2006) (who find $n_d/n_b \approx (72/0.78) = 93$) or Wang et al. (2019) (who find $n_d/n_b \approx 85$).

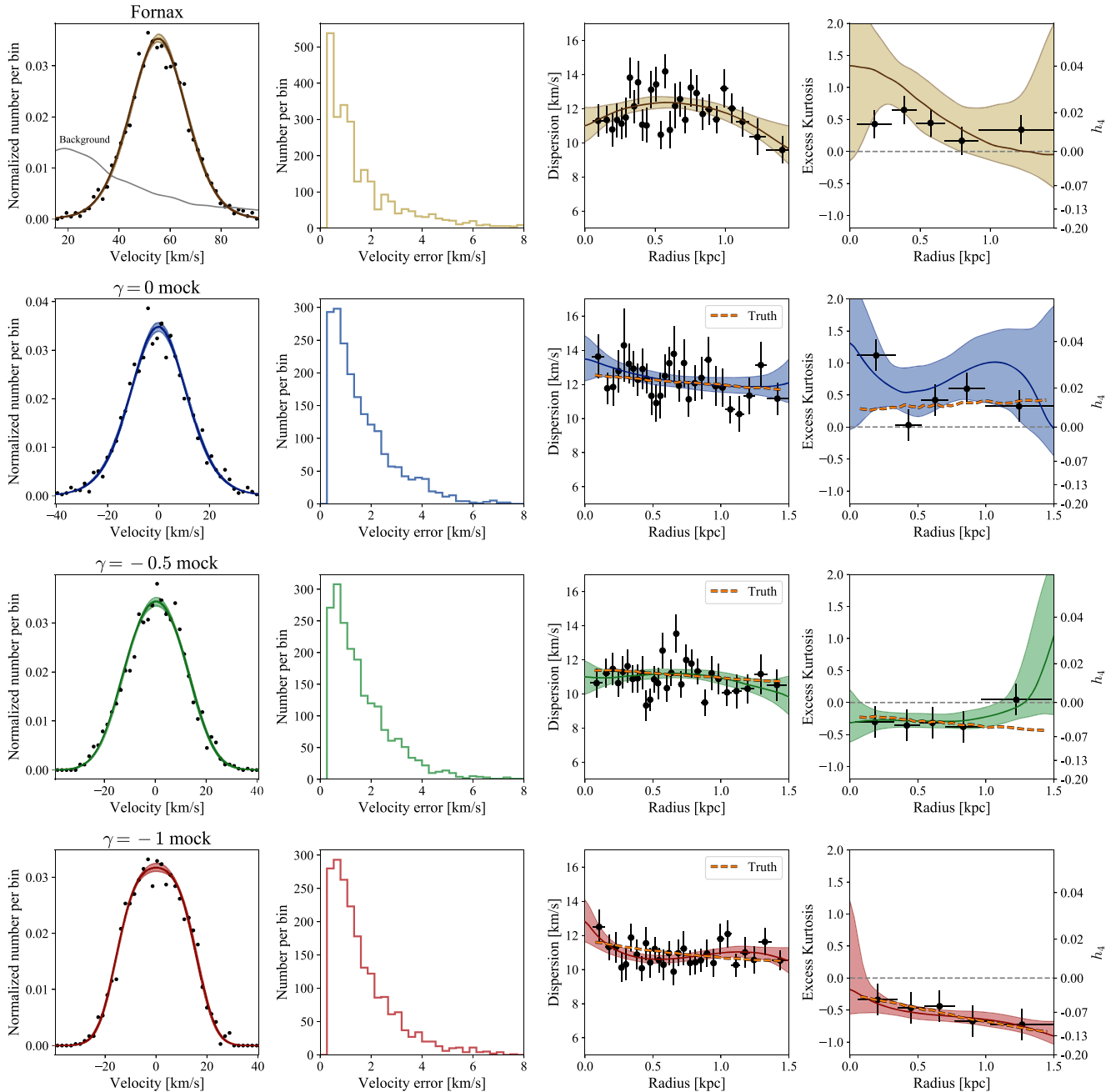


Figure 7. Fits of our family of PDFs to the Fornax data (top row) and three mock Fornax data sets (bottom three rows). The left-hand column shows the velocity distribution of the data (mocks) in black and the median with $\pm 1\sigma$ fits of the models. In the top panel we also show the background GALAXIA (Sharma et al. 2011) model used. The second column shows the distribution of uncertainties. The third and fourth columns show the (membership-weighted) dispersion and excess kurtosis data (black with error bars showing the bin size in x and Poisson errors in y) and fits along with the truth for the mocks in orange.

5.2.2 Results

The results of our procedure for the Fornax data are shown in the top panels of Fig. 7. In agreement with the simple weighted binned estimates (black points), we find that the data favour positive excess kurtosis models. The uncertainties are larger than the binned estimates as they also incorporate uncertainty from the measurement error and membership probabilities. We have also run our fitting using the Gauss–Hermite series for $h_4 > 0$ (where it is positive-definite). The results are shown in Fig. G1. The

recovered h_4 profile is very similar to using the Laplace kernel (see Fig. G2) but the excess kurtosis is smaller, partly because the Gauss–Hermite series does not admit excess kurtosis larger than ~ 0.5 .

To test our procedure, we also attempt to recover the kinematic profiles from our mock dynamical data. For this, we require an approximation to the uncertainty distribution, for which we use a gamma distribution $(\sigma_v - v_e)/v_s \sim \Gamma(1.1, 1)$. v_e is an error floor and v_s governs the width of the tail of the error distribution. We set $v_e = 0.3 \text{ km s}^{-1}$ and $v_s = 1.2 \text{ km s}^{-1}$, which gives a good match to

the error distribution of the Fornax sample. From our mock dynamical models, we select $N_{\text{stars}} = 2000$ stars, assign uncertainties and scatter the velocities by the uncertainty. We do not use a background model and set $p_{\text{mem}} = 1$ for all mock data. We show the results of applying our modelling to the three models in the lower panels of Fig. 7. We find the recovery of the dispersion and excess kurtosis or h_4 is satisfactory.

We run further tests on the recovery of the excess kurtosis of our mock dynamical models with the quality and size of the data set. For speed, we ignore the radial dependence of μ , σ , and h_4 . Additionally, we set a prior lower limit on $\sigma = 0.05 \text{ km s}^{-1}$. We run a set of three experiments varying (i) the number of stars N_{stars} , (ii) the error floor, v_e , and (iii) the scale of the velocity error distribution, v_s . We use default parameters of $N_{\text{star}} = 2000$, $v_s = 1.2 \text{ km s}^{-1}$, and $v_e = 0.3 \text{ km s}^{-1}$ as used in the previous test. We use a fixed random seed to generate the mock samples from the models. The results of these experiments are shown in Fig. 8. We have varied N_{star} from 100 to 10 000 and find that significant negative excess kurtosis can be found with relatively few stars (≈ 200), but for a significant positive excess kurtosis measurement we require $\gtrsim 2000$ stars. We have varied the error floor from $v_e = 0.1$ up to 10 km s^{-1} finding that only for $v_e \lesssim 2 \text{ km s}^{-1}$ can the sign of the excess kurtosis be reliably determined, although again negative excess kurtosis is detectable for $v_e \sim 5 \text{ km s}^{-1}$. Finally, we have varied v_s from 0.1 to 10 km s^{-1} finding its effect on the recovery and size of the uncertainties to be quite weak such that for all of our tests we can fairly reliably distinguish the sign of the excess kurtosis and for $v_s \gtrsim 2 \text{ km s}^{-1}$ the models are distinguishable. Although our results are tailored for Fornax, they can be extended to other dSph galaxies by scaling by the dispersion of Fornax of 11 km s^{-1} .

5.3 Extension to rotation

To conclude our investigation into the use of the new models in the modelling of dSph data, we briefly demonstrate their application to rotating galaxies. Typically dSph galaxies exhibit negligible rotation and most detected rotation can be attributed to perspective effects from the projection of the bulk proper motion (Walker et al. 2008; Amorisco & Evans 2012b). However, if rotation is present, the combined signals of shift in the mean velocity and skewness may prove powerful for its separation from perspective rotation (Amorisco & Evans 2012b).

We modify the dynamical models by multiplying by an odd function of the z component of angular momentum, J_ϕ ,

$$f_{\text{rot}}(\mathbf{J}) = f(\mathbf{J}) (1 + k \text{sgn}(J_\phi)). \quad (47)$$

k controls the degree of rotation and can take values $|k| \leq 1$. We investigate the cored $\gamma = 0$ and cusped $\gamma = 1$ models from the previous section. We choose $k = 0.3$ for the cored model and $k = 0.2$ for the cusped model, and view the models perpendicular to the rotation axis.

In Fig. 9, we show the rotation curves of the two models. The rotation reaches $\sim 3 \text{ km s}^{-1}$ for both models. Such a large rotation signal in Fornax would be obvious so this experiment is purely illustrative. We also display the LOSVDs (where the symmetry allows us to increase the number statistics by multiplying the velocities by the sign of x). We see both distributions are weakly skewed ($g = -0.067$ for the cored model and $g = -0.105$ for the cusped). Both the Gauss–Hermite series and the new models from this paper provide excellent fits to the distributions. In the case of the positive excess kurtosis model the fits are near identical. However, we see for the cusped negative excess kurtosis model the residuals with

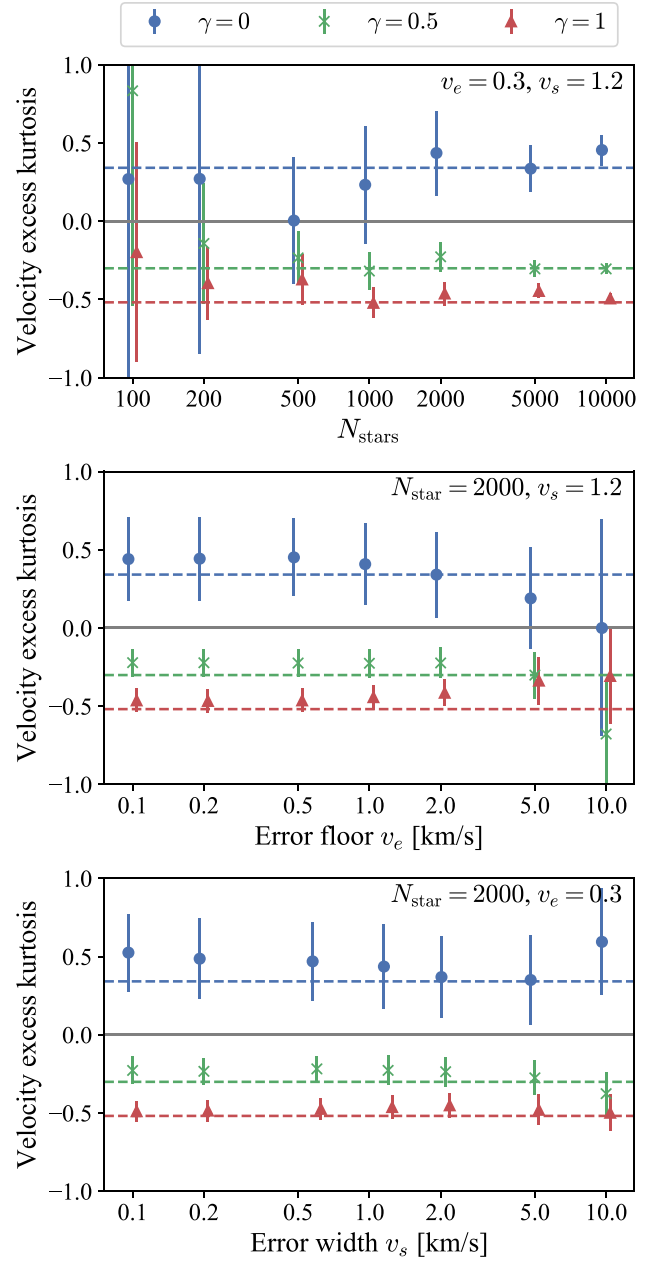


Figure 8. Measurement of velocity excess kurtosis using discrete samples of dynamical dSph models. Three models are inspected (cored = blue dot, weak cusp = green cross, and strong cusp = red triangle). In the top panel the number of tracers is varied; in the central panel the line-of-sight velocity error floor is varied; and in the bottom panel the width of the velocity error distribution varied. The dashed lines show the true values.

respect to our new models are in general smaller and in particular better capture the right-hand wing.

6 APPLICATION TO SPECTRAL FITTING

As discussed in Section 2, the Gauss–Hermite series was originally introduced for the extraction of kinematics from spectra (Gerhard 1993; van der Marel & Franx 1993). Given a library of single stellar population spectra, Cappellari & Emsellem (2004) describe a penalized likelihood algorithm for the simultaneous extraction of the age–metallicity properties of a measured galaxy and the kinematics

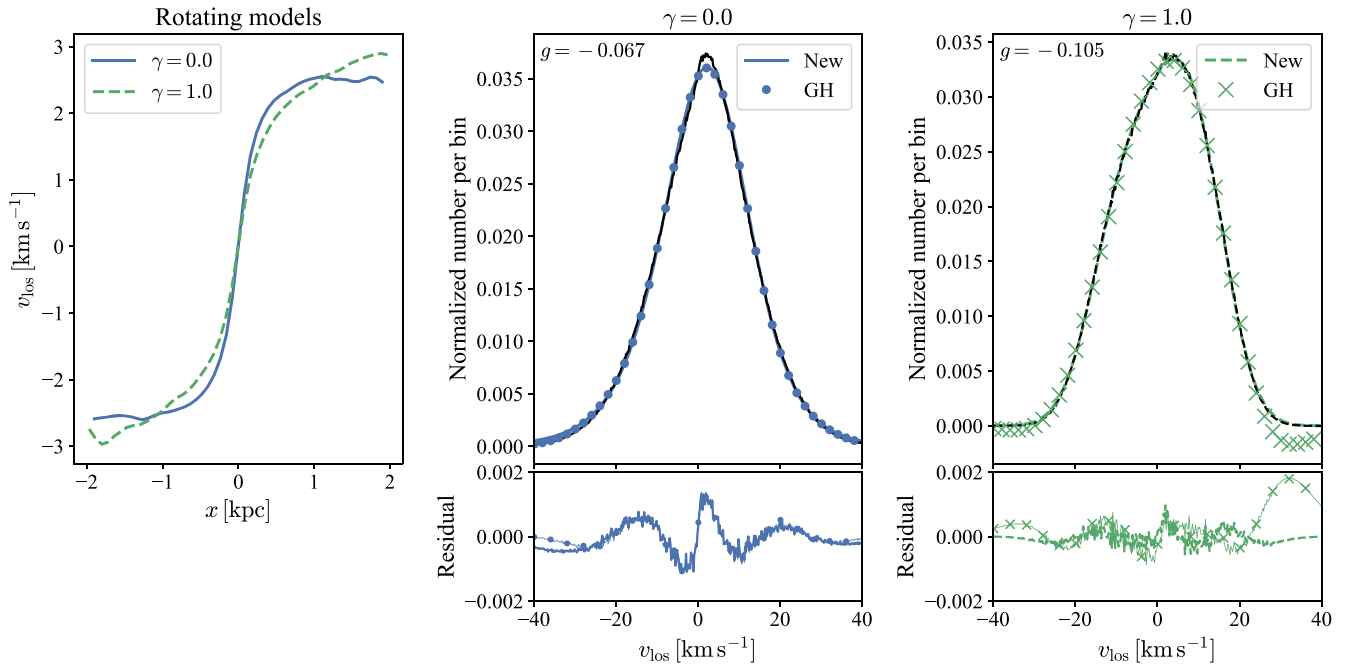


Figure 9. Example of rotating dynamical dSph models. Left-hand panel shows the rotation curve of a cored (blue solid) and cuspy (green dashed) model. x is the on-sky coordinate perpendicular to the rotation axis. The two right-hand panels show the velocity distributions $\text{sgn}(x)v_{\text{los}}$ in black along with fits using the new models of this paper (lines) and the Gauss–Hermite series (with dots and crosses). The skewness g is given in each panel.

via the convolution of the library with a Gauss–Hermite series for the LOSVD. Cappellari (2017) showed how undersampling the LOSVD when the LOSVD dispersion is smaller than half the pixel spacing is problematic for accurate recovery of the kinematics (particularly the mean) but could be circumvented using the analytic Fourier transform of the Gauss–Hermite series. This is now the standard method in the spectral fitting code PPF (Cappellari & Emsellem 2004; Cappellari 2017).

Although in the spectral fitting application the issue of negative wings in the LOSVD is less severe than in the application to discrete data, it still gives rise to unphysical features that could bias the recovery of kinematics. However, the new families of models introduced in this paper provide simple replacements for the Gauss–Hermite series in the spectral fitting procedure. This is because they too have analytic Fourier transforms. In fact, the Fourier transforms are simpler to evaluate than the real-space forms. The Fourier transforms for the two models are

$$\phi_s(u) = \frac{1}{2} e^{i\omega_0 u - u^2/2b^2} \left[\frac{1}{1 - ia_+u/b} + \frac{1}{1 + ia_-u/b} \right] \quad (48)$$

for the Laplace kernel, and

$$\phi_s(u) = \frac{1}{2} e^{i\omega_0 u - u^2/2b^2} \left[\frac{e^{ia_+u/b}}{ia_+u/b} - \frac{e^{ia_-u/b}}{ia_-u/b} \right] \quad (49)$$

for the uniform kernel. We have implemented the models within a version of the PPF code and find that for the included examples the recovery of the kinematics is very similar to using a Gauss–Hermite series.

In Fig. 10, we show examples of a single stellar population with solar metallicity and of age 12.6 Gyr (Vazdekis et al. 2010) convolved with different choices of LOSVD. For all models, we fix the parameter $\sigma = 50 \text{ km s}^{-1}$ but this does not mean the variance of all models are the same. We see that for $h_4 > 0$ the wings of the lines are broader than Gaussian and for $h_4 < 0$ they are narrower.

We observe that when using the Gauss–Hermite series there are prominent bumps in the wings of the lines (particularly visible for $H\alpha$) due to the negative LOSVD but these are avoided when using the models of this paper. It should be noted however that in general the effects and differences are small and the large values of $|h_4|$ chosen for the example have been selected to exaggerate the effect. We also display the Fourier transform of the LOSVD. One feature of the Gauss–Hermite series is that it decays like e^{-u^2} in Fourier space, whereas our negative models oscillate and decay like $1/u$. However, this is unimportant as the template spectra are bandwidth-limited and Nyquist-sampled so we only require the Fourier transform on a fixed grid ($u/\sigma < \pi$) and its behaviour for large u is irrelevant. The disadvantages to our suggested modifications are: (i) one cannot incorporate higher order moments h_5 or h_6 – these are already non-zero for our models. In practice, one rarely uses these higher order moments anyway; (ii) there is only a finite range of h_3 and h_4 accessible by our models. Already h_3 and h_4 are bounded in PPF (default is between -0.3 and 0.3) but the difficulty with our models is in the interdependence of maximum h_3 and h_4 as shown in Fig. 4. One could use a and δ as parameters instead.

7 CONCLUSIONS

We have introduced a new method for constructing PDFs for the modelling of discrete stellar velocity data drawn from weakly non-Gaussian distributions and with heteroskedastic uncertainties. Our study was motivated by the limitation of the commonly used Gauss–Hermite series expansion. Although the Gauss–Hermite series has attractive properties for convolution with measurement uncertainty, a key limitation is that the Gauss–Hermite series does not give well-defined probability density functions – there are regions of negativity particularly for negative excess kurtosis models.

We have demonstrated a generic method for constructing everywhere positive families of probability functions through the

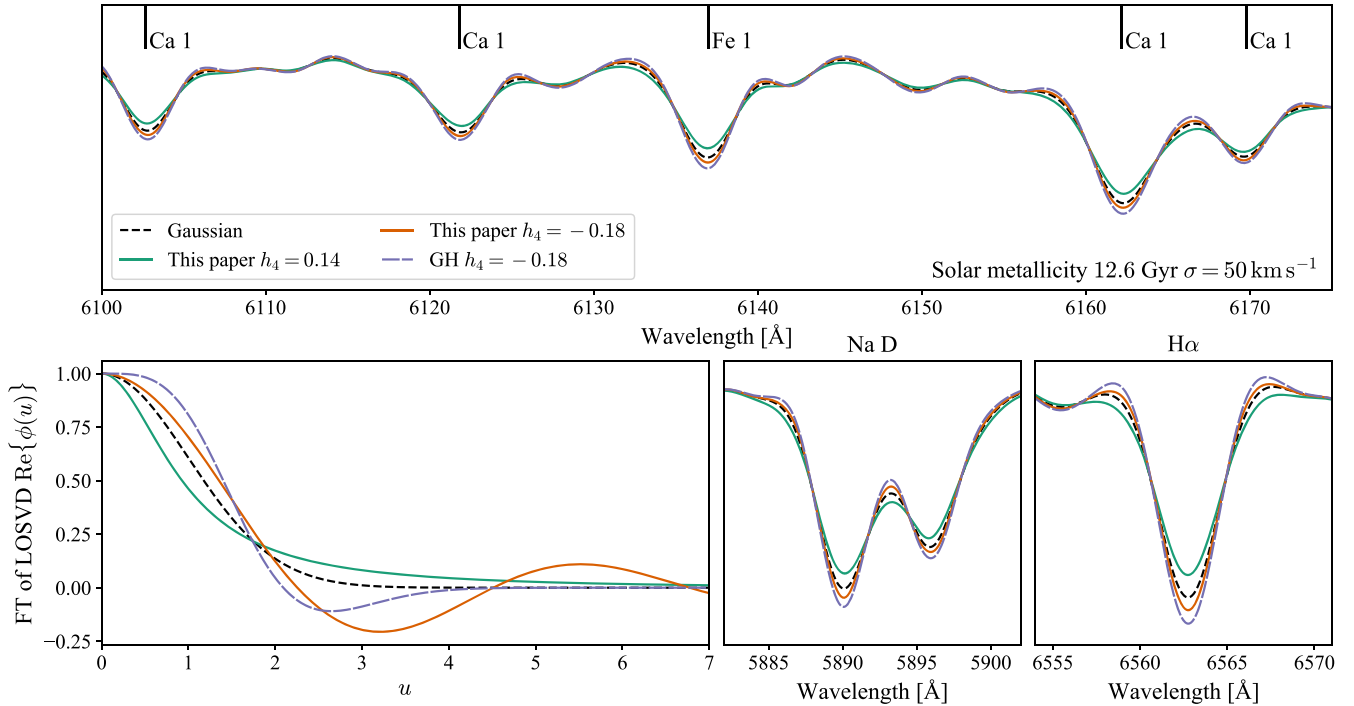


Figure 10. Single stellar population [solar metallicity, 12.6 Gyr convolved with different choices of line-of-sight velocity distribution (LOSVD)]. Four LOSVDs are used, all with identical $\sigma = 50 \text{ km s}^{-1}$ parameters (which does not equal the velocity dispersion except for the Gaussian): (i) Gaussian in black dashed, (ii) solid green is this paper’s model $h_4 = 0.14$, (iii) solid orange is this paper’s model $h_4 = -0.18$, and (iv) Gauss–Hermite series with $h_4 = -0.18$. The top panel and bottom two right-hand panels show example regions of the spectra. The corresponding Fourier transforms of the LOSVDs (important in the computation of the spectra) are given in the bottom left-hand panel. Note the selected h_4 parameters are extreme to exaggerate the effects.

convolution of a Gaussian with a choice of kernel. The resulting PDF inherits the properties (moments and cumulants) of the kernel. We have examined in detail two choices of kernel, Laplace and uniform, which give rise to positive and negative excess kurtosis, respectively, and can be asymmetrized to incorporate skewness. Inspired by the standard implementation of the Gauss–Hermite expansion, we have restricted the models to have approximately $h_1 = h_2 = 0$ such that the models can be used as a direct replacement for the Gauss–Hermite series. However, in theory, the models can be used in greater generality. We provide code for the evaluation of our models at https://github.com/jls713/gh_alternative.

We have demonstrated the use of our new models with two simple applications. The first is to discrete sets of line-of-sight velocity data in galactic modelling. We have shown through a series of dynamical models how non-Gaussianity can break mass–anisotropy degeneracies and how our methods allow for a flexible measurement of the excess kurtosis. This has a ready application to the dwarf galaxies around the Milky Way, for which large data sets of velocities of bright giant stars exist and for which the structure of the dark halo is an open question. For the specific example of the Fornax dSph galaxy, we have shown using a chemo-kinematical mixture model that the data seem to favour positive excess kurtosis velocity distributions and hence indicate cored dark matter density profiles. Additionally, we have tested our modelling procedure on a series of dynamical dSph galaxy models with varying dark matter profiles, finding good recovery of the kinematic profiles and that a cored model well reproduces the observed data. Furthermore, we have investigated how the excess kurtosis measurement depends on the number and uncertainty distribution of stellar tracers. For a Fornax-like dSph, $\gtrsim 2000$ stars with uncertainties $\lesssim 2 \text{ km s}^{-1}$ are sufficient to

measure the sign of the excess kurtosis and measure the mass density slope, although this assumes perfect knowledge of dSph membership. We have also shown how our models are able to capture skewness, which could be a powerful indicator of the subtle effects of rotation and immune to the effects of perspective.

Secondly, although our models were not motivated by spectral fitting applications, their analytic Fourier transforms allow them to be used in spectral fitting algorithms where they could reduce bias produced by negative wings in the LOSVDs of high-quality data.

Our methods for characterizing mildly non-Gaussian velocity distributions are flexible and adaptable, offering significant advantages over the conventional Gauss–Hermite series. The skewness and kurtosis of such distributions contain important clues on the orbits of the tracers and the gravitational potential in which they move. The sizes of such discrete velocity data sets – whether for populations in the Milky Way or its dwarf satellite entourage or for populations in external galaxies like globular clusters and planetary nebulae – have substantially increased in recent years. We envisage that our new methods will play an important role in exploiting these data sets to the full.

ACKNOWLEDGEMENTS

JLS acknowledges support from the Royal Society (URF\R1\191555). This paper made use of MATHEMATICA (Wolfram Research Inc. 2012), NUMPY (van der Walt, Colbert & Varoquaux 2011), SCIPY (Virtanen et al. 2020), MATPLOTLIB (Hunter 2007), SEABORN (Waskom et al. 2017), ASTROPY (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018), EMCEE (Foreman-Mackey et al. 2013), AGAMA (Vasiliev 2019), and PPF (Cappellari 2017). This work has made use of data from the European Space

Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>; Gaia Collaboration et al. 2016, 2018a), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

DATA AVAILABILITY

All data used in this paper are in the public domain: Walker et al. (2009a), Vazdekis et al. (2010), Gaia Collaboration et al. (2018b), Wang et al. (2019), and Moskowitz & Walker (2020).

REFERENCES

- Agnello A., Evans N. W., Romanowsky A. J., Brodie J. P., 2014, *MNRAS*, 442, 3299
- Amorisco N. C., Evans N. W., 2011, *MNRAS*, 411, 2118
- Amorisco N. C., Evans N. W., 2012a, *MNRAS*, 419, 184
- Amorisco N. C., Evans N. W., 2012b, *MNRAS*, 424, 1899
- Amorisco N. C., Agnello A., Evans N. W., 2013, *MNRAS*, 429, L89
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Battaglia G. et al., 2006, *A&A*, 459, 423
- Bender R., Saglia R. P., Gerhard O. E., 1994, *MNRAS*, 269, 785
- Binney J., Mamon G. A., 1982, *MNRAS*, 200, 361
- Binney J., Merrifield M., 1998, *Galactic Astronomy*. Princeton Univ. Press, Princeton, NJ
- Blinnikov S., Moessner R., 1998, *A&AS*, 130, 193
- Breddels M. A., Helmi A., 2013, *A&A*, 558, A35
- Cappellari M., 2017, *MNRAS*, 466, 798
- Cappellari M., Emsellem E., 2004, *PASP*, 116, 138
- Cappellari M. et al., 2011, *MNRAS*, 413, 813
- Debatista V. P., Carollo C. M., Mayer L., Moore B., 2005, *ApJ*, 628, 678
- Emsellem E. et al., 2004, *MNRAS*, 352, 721
- Evans N. W., An J., Walker M. G., 2009, *MNRAS*, 393, L50
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Gaia Collaboration et al., 2016, *A&A*, 595, A1
- Gaia Collaboration et al., 2018a, *A&A*, 616, A1
- Gaia Collaboration et al., 2018b, *A&A*, 616, A12
- Gerhard O. E., 1993, *MNRAS*, 265, 213
- Gradshteyn I. S., Ryzhik I. M., 2007, *Table of Integrals, Series, and Products*, 7th edn. Elsevier/Academic Press, Amsterdam
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Irwin M., Hatzidimitriou D., 1995, *MNRAS*, 277, 1354
- Kendall M., Stuart A., 1977, *The Advanced Theory of Statistics*. Vol.1: Distribution Theory. Charles Griffin & Co., London
- Kenney J., Keeping E., 1951, *Mathematics of Statistics*, Pt. 2, 2nd edn. Van Nostrand, Princeton, NJ
- Kuijken K., Merrifield M. R., 1993, *MNRAS*, 264, 712
- Łokas E. L., 2002, *MNRAS*, 333, 697
- Łokas E. L., 2009, *MNRAS*, 394, L102
- Łokas E. L., Mamon G. A., Prada F., 2005, *MNRAS*, 363, 918
- Long K., Murali C., 1992, *ApJ*, 397, 44
- McConnachie A. W., 2012, *AJ*, 144, 4
- McGough D. P., Evans N. W., Sanders J. L., 2020, *MNRAS*, 493, 2676
- Massari D., Breddels M. A., Helmi A., Posti L., Brown A. G. A., Tolstoy E., 2018, *Nat. Astron.*, 2, 156
- Mateo M. L., 1998, *ARA&A*, 36, 435
- Méndez-Abreu J. et al., 2019, *MNRAS*, 482, L118
- Merrifield M. R., Kent S. M., 1990, *AJ*, 99, 1548
- Moskowitz A. G., Walker M. G., 2020, *ApJ*, 892, 27
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Oldham L. J., Auger M. W., 2016, *MNRAS*, 457, 421
- Olver F. W. J., Olde Daalhuis A. B., Lozier D. W., Schneider B. I., Boisvert R. F., Clark C. W., Miller B. R., Saunders B. V., 2016, NIST Digital Library of Mathematical Functions. Available at: <http://dlmf.nist.gov/>
- Pascale R., Posti L., Nipoti C., Binney J., 2018, *MNRAS*, 480, 927
- Posti L., Binney J., Nipoti C., Ciotti L., 2015, *MNRAS*, 447, 3060
- Price-Whelan A. M. et al., 2018, *AJ*, 156, 123
- Read J. I., Walker M. G., Steger P., 2019, *MNRAS*, 484, 1401
- Richardson T., Fairbairn M., 2013, *MNRAS*, 432, 3361
- Rix H.-W., de Zeeuw P. T., Cretton N., van der Marel R. P., Carollo C. M., 1997, *ApJ*, 488, 702
- Rizzi L., Held E. V., Saviane I., Tully R. B., Gullieuszik M., 2007, *MNRAS*, 380, 1255
- Sellwood J. A., Gerhard O., 2020, *MNRAS*, 495, 3175
- Sharma S., Bland-Hawthorn J., Johnston K. V., Binney J., 2011, *ApJ*, 730, 3
- van der Marel R. P., Franx M., 1993, *ApJ*, 407, 525
- van der Marel R. P., Evans N. W., Rix H. W., White S. D. M., de Zeeuw T., 1994, *MNRAS*, 271, 99
- van der Marel R. P., Cretton N., de Zeeuw P. T., Rix H.-W., 1998, *ApJ*, 493, 613
- van der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22
- van de Sande J. et al., 2017, *ApJ*, 835, 104
- Vasiliev E., 2019, *MNRAS*, 482, 1525
- Vazdekis A., Sánchez-Blázquez P., Falcón-Barroso J., Cenarro A. J., Beasley M. A., Cardiel N., Gorgas J., Peletier R. F., 2010, *MNRAS*, 404, 1639
- Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
- Walker M. G., Peñarrubia J., 2011, *ApJ*, 742, 20
- Walker M. G., Mateo M., Olszewski E. W., 2008, *ApJ*, 688, L75
- Walker M. G., Mateo M., Olszewski E. W., 2009a, *AJ*, 137, 3100
- Walker M. G., Mateo M., Olszewski E. W., Sen B., Woodroffe M., 2009b, *AJ*, 137, 3109
- Wang M. Y. et al., 2019, *ApJ*, 881, 118
- Waskom M. et al., 2017, seaborn: v0.8.1. Zenodo. Available at: <http://citeba.com/how-to-cite/seaborn/>
- Williams A. A., Evans N. W., 2015, *MNRAS*, 448, 1360
- Williams A. A., Evans N. W., 2017, *MNRAS*, 469, 4414
- Wojtak R., Łokas E. L., 2010, *MNRAS*, 408, 2442
- Wolfram Research Inc., 2012, *Mathematica* (Version 9). Available at: <https://www.wolfram.com/mathematica>

APPENDIX A: HERMITE POLYNOMIALS

The convention for the Hermite polynomials employed here follows van der Marel & Franx (1993) such that they are given by the relation

$$\left(-\frac{d}{dy}\right)^l \alpha(y) = \sqrt{l!} H_l(y/\sqrt{2}) \alpha(y). \quad (\text{A1})$$

This is different from the standard $H_n(y)$ defined in e.g. Olver et al. (2016) that are given by $\sqrt{l!2^l} H_l(y)$. The first five Hermite polynomials are

$$H_0(y) = 1, \quad H_1(y) = \sqrt{2}y, \quad H_2(y) = \frac{1}{\sqrt{2}}(2y^2 - 1),$$

$$H_3(y) = \frac{1}{\sqrt{6}}(2\sqrt{2}y^3 - 3\sqrt{2}y), \quad H_4(y) = \frac{1}{\sqrt{24}}(4y^4 - 12y^2 + 3). \quad (\text{A2})$$

The generating function for this definition of the Hermite polynomials is given by

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} \sqrt{n!} H_n(y) = e^{\sqrt{2}yz - z^2/2}, \quad (\text{A3})$$

from which the n th polynomial can be found as

$$H_n(y) = \frac{1}{\sqrt{n!}} \frac{\partial^n}{\partial z^n} \left(e^{\sqrt{2}yz - z^2/2} \right) \Big|_{z=0}. \quad (\text{A4})$$

APPENDIX B: CUMULANTS OF THE GAUSS–HERMITE SERIES

The cumulant generating function of the Gauss–Hermite series given in equation (1) is given by

$$\begin{aligned} \ln \phi_{\text{GH}}(u) &= \ln \int_{-\infty}^{\infty} dx e^{iux} f(x) = \sum_{r=1}^{\infty} \kappa_r \frac{(iu)^r}{r!} \\ &= iVu - \frac{(\sigma u)^2}{2} + \ln \left[1 + \sum_{n \geq 3} i^n h_n H_n(\sigma u) \right]. \end{aligned} \quad (\text{B1})$$

The cumulants are then given by

$$\kappa_1 = V + \tilde{\kappa}_1, \quad \kappa_2 = \sigma^2 + \tilde{\kappa}_2, \quad \kappa_i = \tilde{\kappa}_i, \quad i \geq 3,$$

$$\tilde{\kappa}_r = \sigma^r \frac{d^{r-1}}{du^{r-1}} \left[\left(1 + \sum_{n \geq 3} i^n h_n H_n(u) \right)^{-1} \sum_{n \geq 3} i^{n-r} h_n \frac{dH_n(u)}{du} \right]_{u=0}. \quad (\text{B2})$$

Performing the derivative produces long-winded expressions. For their evaluation, we require the Hermite numbers $H_n(0)$ given by

$$H_n(0) = (-1)^{n/2} \frac{\sqrt{n!}}{(n/2)! \sqrt{2^n}} \quad (\text{B3})$$

for even n and zero for odd n , and the n th derivative of the Hermite polynomials is

$$\left. \frac{d^r H_n(x)}{dx^r} \right|_{x=0} = \sqrt{\frac{2^r n!}{(n-r)!}} H_{n-r}(0) = (-1)^{(n-r)/2} \frac{2^{r-n/2} \sqrt{n!}}{((n-r)/2)!} \quad (\text{B4})$$

for even $n-r$ and zero otherwise. Restricting to $h_n = 0$ for $n > 4$, we find for the first four cumulants (note all κ_r are non-zero),

$$\begin{aligned} \kappa_1 &= V + \sqrt{3}\sigma\lambda h_3 \approx V + \sqrt{3}\sigma h_3, \\ \kappa_2 &= \sigma^2 + \sigma^2\lambda^2(h_4(2\sqrt{6} + 3h_4) - 3h_3^2) \approx \sigma^2(1 + 2\sqrt{6}h_4), \\ \kappa_3 &= \frac{1}{2}h_3\sigma^3\lambda^3 \left(8\sqrt{3} + 12\sqrt{3}h_3^2 \right. \\ &\quad \left. - 3h_4(8\sqrt{2} + 5\sqrt{3}h_4) \right) \approx 4\sqrt{3}h_3\sigma^3, \\ \kappa_4 &= \frac{1}{2}\sigma^4\lambda^4 \left(16\sqrt{6}h_4 - 9h_4^2(8 + 6\sqrt{6}h_4 + 5h_4^2) \right. \\ &\quad \left. + 12h_3^2(15h_4^2 + 8\sqrt{6}h_4 - 8) - 108h_3^4 \right) \approx 8\sqrt{6}h_4\sigma^4, \end{aligned} \quad (\text{B5})$$

where $\lambda = \left(1 + \sqrt{3/8}h_4 \right)^{-1}$ and the approximations for small h_n agree with van der Marel & Franx (1993). Note that for larger values $|h_4|$, the excess kurtosis $\kappa = \kappa_4/\kappa_2^2$ is not a monotonic function of h_4 .

APPENDIX C: CONVOLUTION OF A GAUSS–HERMITE SERIES WITH GAUSSIAN ERRORS

In this appendix, we compute the convolution of a Gauss–Hermite series $f(x)$ with a Gaussian uncertainty distribution $\mathcal{N}(x - x'|\sigma_e)$ as

$$f_{\sigma_e}(x) = \int_{-\infty}^{\infty} dx' f(x') \mathcal{N}(x - x'|\sigma_e). \quad (\text{C1})$$

We work in terms of the scaled and shifted coordinate $w = (x - V)/\sigma$. Dropping the constant factors, we require the convolution

$$f_s(w) = \int_{-\infty}^{\infty} dw' f(w') \mathcal{N}(w - w'|s), \quad (\text{C2})$$

where

$$\mathcal{N}(w|s) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{1}{2}w^2/s^2}, \quad (\text{C3})$$

and $s = \sigma_e/\sigma$. For this computation, we shift to Fourier space. One attractive property of the Gauss–Hermite series is that the terms are eigenfunctions of the Fourier transform, F ,

$$F \left\{ \alpha(w) H_n(w) \right\} \equiv \int_{-\infty}^{\infty} dw \alpha(w) H_n(w) e^{ikw} = \sqrt{2\pi i}^n H_n(k) \alpha(k). \quad (\text{C4})$$

Cappellari (2017) uses this property to rapidly convolve a model spectrum with the Gauss–Hermite series. Each term in the convolution of $\mathcal{N}(w|s)$ with $f(w)$ is given by

$$C_n(w|s) \equiv \int dw' H_n(w') \alpha(w') \mathcal{N}(w - w'|s). \quad (\text{C5})$$

Taking the Fourier transform of the convolution yields

$$F \left\{ C_n(w|s) \right\} = \sqrt{2\pi i}^n H_n(k) \alpha(k) e^{-s^2 k^2/2} = \sqrt{2\pi i}^n H_n(k) \alpha(s'k), \quad (\text{C6})$$

where $s' = \sqrt{1 + s^2}$. The inverse Fourier transform is given by

$$C_n(w|s) = \frac{i^n}{\sqrt{2\pi s'}} \int dk' e^{-ik'w/s'} H_n \left(\frac{k'}{s'} \right) \alpha(k'). \quad (\text{C7})$$

The Hermite polynomials satisfy the following multiplication theorem (equation 18.18.13 of Olver et al. 2016):

$$H_n(\gamma x) = \gamma^n \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} (1 - \gamma^{-2})^j \frac{\sqrt{n!}}{j! 2^j \sqrt{(n-2j)!}} H_{n-2j}(x). \quad (\text{C8})$$

Rewriting $H(k'/s')$ using this series and using the eigenfunction of the Fourier transform property of $H(k')\alpha(k')$ gives an expression for the convolution of each term as

$$C_n(w|s) = \frac{\mathcal{N}(w|s')}{s'^n} \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} s'^{2j} \frac{\sqrt{n!}}{j! 2^j \sqrt{(n-2j)!}} H_{n-2j} \left(\frac{w}{s'} \right). \quad (\text{C9})$$

The sum can be recombined into a single Hermite polynomial as

$$C_n(w|s) = \frac{\mathcal{N}(w|s')}{s'^n} (1 - s'^2)^{n/2} H_n \left(\frac{w}{\sqrt{1 - s'^4}} \right). \quad (\text{C10})$$

This formula can be derived directly from equation (C5) using equation (7.374.8) of Gradshteyn & Ryzhik (2007). Note that the formula is valid for $s > 1$ although it requires numerical care, as both the prefactor and the argument of the Hermite polynomial are imaginary in this case. We give the full result for $f_{\sigma_e}(x)$ in equation (11).

APPENDIX D: OTHER SERIES METHODS

The Gauss–Hermite series bears resemblance to two other series expansions that are more commonly used in probability theory. The Gram–Charlier and Edgeworth series both approximate a general probability density function in terms of its cumulants (see Blinnikov & Moessner 1998, for general derivations). The construction of the two series is similar in that the cumulant generating function

is written as a sum of a Gaussian cumulant generating function (first two terms) plus a truncated cumulant generating function series (for $n \geq 3$):

$$\ln \phi_{\text{approx}}(u) = \ln \phi_{\text{Gaussian}}(u) + \sum_{r \geq 3} \kappa_r \frac{(iu)^r}{r!}. \quad (\text{D1})$$

The approximating probability density function is then the inverse Fourier transform of ϕ_{approx} . The difference between Gram–Charlier and Edgeworth is in how one chooses to truncate the sum. In Gram–Charlier, the sum is truncated at the n th cumulant, whilst for Edgeworth the expansion of the normalized variable x/σ is truncated at the n th power of the standard deviation, σ (making it a truly asymptotic series). The advantage of this approach is the terms in the resulting series are simply related to the cumulants (see Appendix B for the relationship between the Gauss–Hermite coefficients and the cumulants). Up to fourth order, the Gram–Charlier series is given by

$$f(x) = \frac{1}{\sigma} \alpha(w) \left[1 + \frac{\kappa_3}{3! \sigma^3} He_3(w) + \frac{\kappa_4}{4! \sigma^4} He_4(w) \right]. \quad (\text{D2})$$

Correspondingly, the Edgeworth series is given by

$$f(x) = \frac{1}{\sigma} \alpha(w) \left[1 + \frac{\kappa_3}{3! \sigma^3} He_3(w) + \frac{\kappa_4}{4! \sigma^4} He_4(w) + \frac{10\kappa_3^2}{6! \sigma^6} He_6(w) \right]. \quad (\text{D3})$$

Here the Chebyshev–Hermite polynomials $He_n(y)$ are given by

$$\begin{aligned} He_3(y) &= y^3 - 3y, & He_4(y) &= y^4 - 6y^2 + 3, \\ He_6(y) &= y^6 - 15y^4 + 45y^2 - 15. \end{aligned} \quad (\text{D4})$$

For both series, the convolution with observational uncertainties, σ_e , is simple as the series are derived directly from a characteristic function. The convolution is a multiplication in Fourier space such that $\phi_{\text{Gaussian}}(u)$ in equation (D1) has a width $\sigma' = \sqrt{\sigma^2 + \sigma_e^2}$. Equations (D2) and (D3) are then modified as $w \rightarrow w(\sigma/\sigma')$ and $\sigma \rightarrow \sigma'$.

The Gram–Charlier series has very poor convergence properties ($p(x)$ must fall off faster than $e^{-x^2/4}$ for large x), whilst the Edgeworth expansion is designed to converge given any $p(x)$. On the other hand, the Gauss–Hermite series will converge if $p(x)$ satisfies an α -Hölder condition and $\int dx |p(x)|(1 + |x|^{-3/2})$ does not diverge (Blinnikov & Moessner 1998).

We observe that both series methods also necessarily suffer negative probability density regions so are not suitable workarounds for our application.

APPENDIX E: OTHER CHOICES OF KERNEL

In this appendix, we give two tables detailing possible other choices of kernel. Table E1 gives the set of half-kernels, $K_+(y)$, which are non-zero for positive y and their corresponding error-convolved distributions $f_{s+}(w)$, and characteristic function $\phi_{K_+}(u)$. Table E2 gives the cumulants for models formed by stitching together two of these half-kernels for positive and negative y with different scale parameters.

Table E1. Choices of half-kernel $K_+(y)$ (zero for $y < 0$ – all normalized to integrate to $\frac{1}{2}$ except Gamma that is normalized to unity), their error-convolved distribution $f_{s+}(w)$ and their characteristic functions $\phi_{K_+}(u)$. The top two models (Uniform and Cosine) have negative excess kurtosis and the bottom three (Laplace, Gaussian and Gamma) positive. $w' = w - w_0$, $t = 1 + b^2 s^2$, $\mathcal{N}(w|s)$ is a normal distribution with width s and $\Phi(x)$ is the cumulative distribution of the unit normal. ${}_1F_1(a, b, c, x)$ is the confluent hypergeometric function.

Name	$K_+(y)$	$f_{s+}(w)$	$\phi_{K_+}(u)$
Uniform	$\frac{1}{2a}$ if $y < a$,	$\frac{b}{2a} \left[\Phi\left(\frac{bw'}{t}\right) - \Phi\left(\frac{bw' - a}{t}\right) \right]$	$\frac{e^{iau} - 1}{2iau}$
Cosine	$\frac{\pi}{4a} \cos\left(\frac{\pi y}{2a}\right)$ if $y < a$,	$\frac{b\pi}{8a} \left(D(w', w' - a/b) - D(-w', -w' + a/b) \right)^a$	$\frac{\pi}{2} \frac{\pi e^{iau} - 2iau}{\pi^2 - 4a^2 u^2}$
Laplace	$\frac{1}{2a} e^{-y/a}$	$\frac{b}{4a} \exp\left(\frac{t^2 - 2abw'}{2a^2}\right) \text{erfc}\left(\frac{t^2 - abw'}{\sqrt{2}ta}\right)$	$\frac{1}{2 - 2iau}$
Gaussian	$\frac{1}{\sqrt{2\pi a}} e^{-y^2/2a^2}$	$b\mathcal{N}(bw' \sqrt{t^2 + a^2})\Phi\left(\frac{bw'a}{t\sqrt{t^2 + a^2}}\right)$	$e^{-a^2 u^2/2} \Phi(iau)$
Gamma	$\frac{y^{\beta-1} e^{-y/a}}{a^\beta \Gamma(\beta)}$	$\frac{\sqrt{2\beta-3} b}{\sqrt{\pi a} \beta + 1 \Gamma(\beta)} t^{\beta-2} e^{-b^2 w'^2/2t^2}$ $\times \left[at\Gamma\left(\frac{\beta}{2}\right) {}_1F_1\left(\frac{\beta}{2}, \frac{1}{2}, \frac{(t^2 - abw')^2}{2a^2 t^2}\right) \right.$ $\left. + \sqrt{2}(abw' - t^2)\Gamma\left(\frac{1+\beta}{2}\right) {}_1F_1\left(\frac{1+\beta}{2}, \frac{3}{2}, \frac{(t^2 - abw')^2}{2a^2 t^2}\right) \right]$	$\frac{1}{(1-iau)^\beta}$

$${}^a D(w, q) = e^{-\pi^2 t^2/8a^2} e^{i\pi bw/2a} \left[\Phi\left(\frac{i\pi t^2 + 2abw}{2at}\right) - \Phi\left(\frac{i\pi t^2 + 2abq}{2at}\right) \right].$$

Table E2. Cumulants for kernels built from two half-kernels from Table E1 with the requirement $\int_0^\infty dy K_+(y) = \int_{-\infty}^0 dy K_-(y) = \frac{1}{2}$ (except Gamma where we only consider positive y). The width of the positive domain is a_+ and a_- for the negative domain. These quantities can be derived from Table E1 using $\tilde{\kappa}_r = i^{-r} \partial_u^r \ln [\phi_{K_+}(u) + \phi_{K_-}(u)]_{u=0}$. We define the auxiliary variables, $a = \frac{1}{2}(a_+ + a_-)$ and $\Delta = \frac{1}{2}(a_+ - a_-)$. $\tilde{\mu}$ is the mean, \tilde{v} the variance, and $\tilde{\kappa}_3$ and $\tilde{\kappa}_4$ the third and fourth cumulant from which the skewness \tilde{g} and excess kurtosis $\tilde{\kappa}$ are computed as $\tilde{g} = \tilde{\kappa}_3/\tilde{v}^{3/2}$ and $\tilde{\kappa} = \tilde{\kappa}_4/\tilde{v}^2$. Note that all bracketed constants are positive.

Name	$\tilde{\mu}$	\tilde{v}	$\tilde{\kappa}_3$	$\tilde{\kappa}_4$
Uniform	$\frac{\Delta}{2}$	$\frac{a^2}{3} + \frac{\Delta^2}{12}$	$\frac{\Delta a^2}{4}$	$-\frac{1}{120}(16a^4 - 4a^2\Delta^2 + \Delta^4)$
Cosine	$\Delta\left(1 - \frac{2}{\pi}\right)$	$\left(1 - \frac{8}{\pi^2}\right)a^2 + \frac{4}{\pi}\left(1 - \frac{3}{\pi}\right)\Delta^2$	$\frac{2\Delta}{\pi}\left[3\left(\frac{4}{\pi} - 1\right)^2 a^2 + \left(\frac{12}{\pi} - \frac{8}{\pi^2} - 3\right)\Delta^2\right]$	$-\frac{2a^4}{\pi^4}(\pi^4 - 96) + \frac{24a^2\Delta^2}{\pi^4}(112 - 32\pi + 2\pi^2 - \pi^3)$ $-\frac{8\Delta^4}{\pi^4}(12\pi^2 - \pi^3 - 24\pi - 12)$
Laplace	Δ	$2a^2 + \Delta^2$	$2\Delta(6a^2 + \Delta^2)$	$6(2a^4 + 12a^2\Delta^2 + \Delta^4)$
Gaussian	$\sqrt{\frac{2}{\pi}}\Delta$	$a^2 + \left(1 - \frac{2}{\pi}\right)\Delta^2$	$\sqrt{\frac{2}{\pi}}\Delta\left[3a^2 + \left(\frac{4}{\pi} - 1\right)\Delta^2\right]$	$4\Delta^2\left[3\left(1 - \frac{2}{\pi}\right)a^2 + \frac{2}{\pi}\left(1 - \frac{3}{\pi}\right)\Delta^2\right]$
Gamma	$a\beta$	$a^2\beta$	$2a^3\beta$	$6a^4\beta$

APPENDIX F: NUMERICAL IMPLEMENTATION

When fitting the proposed models to data we require accurate computation of the logarithm of the PDF $f_s(w)$. This requires some care and we detail some suggested methods.

F1 Uniform kernel

For the uniform kernel, we require accurate computation of equation (38). We first rewrite this equation as

$$f_s(w) = \frac{b}{2a_+a_-} \left\{ a_+ \left[\Phi\left(\frac{bw' + a_-}{t}\right) - \Phi\left(\frac{bw'}{t}\right) \right] + a_- \left[\Phi\left(\frac{bw'}{t}\right) - \Phi\left(\frac{bw' - a_+}{t}\right) \right] \right\}. \quad (\text{F1})$$

Concentrating on the first part (as similar arguments apply to the second part) we require accurate computation of

$$\ln [\Phi(x + c) - \Phi(x)]. \quad (\text{F2})$$

We first compute

$$\ln \Phi(x) = -\ln 2 + \ln \operatorname{erfc}(-x/\sqrt{2}), \quad (\text{F3})$$

and note the following identity

$$\ln \operatorname{erfc}(x) = \begin{cases} \ln \operatorname{erfc}(x) - x^2 & x > 0, \\ \ln \operatorname{erfc}(x) & x \leq 0, \end{cases} \quad (\text{F4})$$

where $\operatorname{erfcx}(x)$ is the scaled complementary error function $\operatorname{erfcx}(x) \equiv \exp(x^2)\operatorname{erfc}(x)$. An alternative computationally cheaper method to evaluate $\ln \Phi(x)$ uses the LOG_NDTR method implemented in SCIPY that employs a Taylor series expansion for arguments $|x| > 20$: typically this is a factor of 2 faster than evaluating $\ln \operatorname{erfc}(x)$. We then compute the difference of two $\ln \Phi$ calls as

$$\ln \exp(a - b) = a + \ln(1 - \exp(-(a - b))), \quad (\text{F5})$$

where we can evaluate the second term using a LOG1MEXP function. For positive arguments we use the identity $\Phi(-x) = 1 - \Phi(x)$ and instead evaluate $-\Phi(-x) - \Phi(-(x + c))$.

For the second part of equation (F1) we employ a similar procedure: for positive $(bw' - a_+)/t$ we evaluate $\Phi(-(bw' - a_+)/t)$

$-\Phi(-(bw'/t))$. The sum of the two parts can then be performed using a log-add-exp operation: $\ln \exp(a + b) = a + \ln \exp(b - a)$.

F2 Laplace kernel

For the Laplace kernel, we must evaluate equation (41) that is unstable for large positive arguments. An alternative form using the scaled complementary error function $\operatorname{erfcx}(x)$ is

$$f_s(w) = \sqrt{\frac{\pi}{8}} b \alpha \left(\frac{bw'}{t} \right) \left[\frac{1}{a_+} \operatorname{erfcx}\left(\frac{t^2 - a_+ bw'}{\sqrt{2}ta_+}\right) + \frac{1}{a_-} \operatorname{erfcx}\left(\frac{t^2 + a_- bw'}{\sqrt{2}ta_-}\right) \right]. \quad (\text{F6})$$

This expression suffers from its own different numerical issues when the arguments are large and negative. Therefore, for optimum numerical stability we separate out the two terms in the sum and use the form in equation (41) for large negative arguments and equation (F6) for large positive arguments. Note that this occurs for $t^2 - 2abw' > 0$ for one term and $t^2 + 2abw' > 0$ for the other. We compute the logarithm of the two terms using equation (F4) and the similar identity,

$$\ln \operatorname{erfcx}(x) = \begin{cases} \ln \operatorname{erfc}(x) + x^2 & x < 0, \\ \ln \operatorname{erfcx}(x) & x \geq 0. \end{cases} \quad (\text{F7})$$

The sum of the two terms can then be performed using a log-add-exp operation.

APPENDIX G: FORNAX MODELS USING THE GAUSS-HERMITE SERIES

When working with models with positive excess kurtosis, it is possible to use the Gauss-Hermite series as it is always positive definite for $h_3 = 0$ and $h_4 > 0$. In Fig. G1, we provide an alternative fit to the Fornax data using a Gauss-Hermite series instead of the Laplace kernel model introduced in this paper. Note that at equivalent h_4 , the values of kurtosis admitted by the Gauss-Hermite series are smaller than for the Laplace kernel. However, as shown in Fig. G2, both models produce similar dispersion and h_4 profiles.

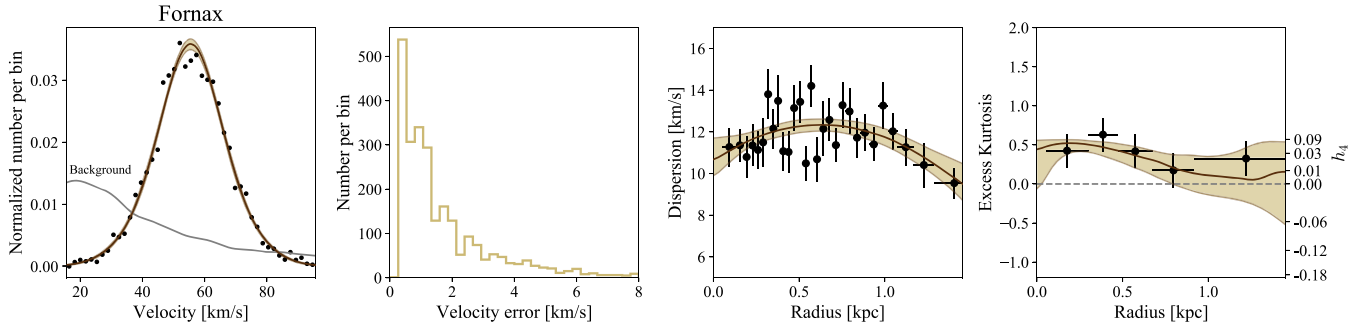


Figure G1. Similar to Fig. 7 but instead using a Gauss–Hermite series for positive excess kurtosis. Note in particular the different relationship between h_4 and the excess kurtosis for these models.

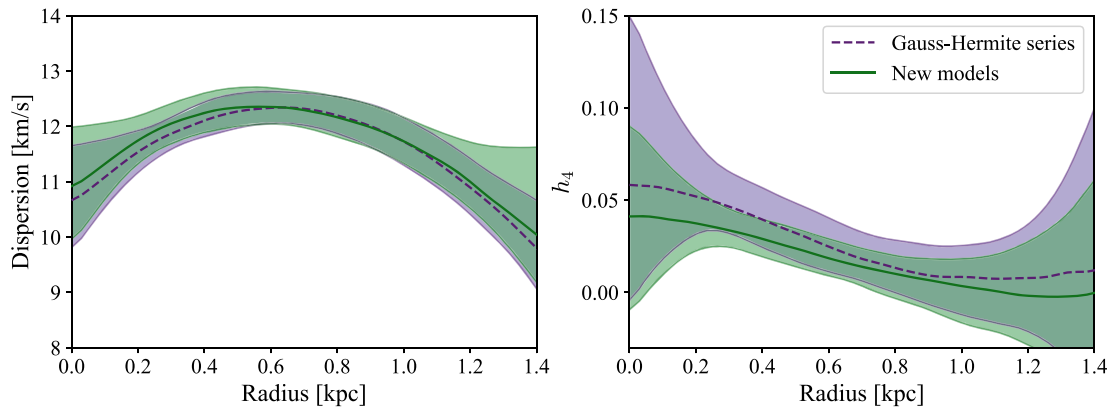


Figure G2. Comparison of the radial dispersion and h_4 profiles using the Gauss–Hermite series (dashed purple) and the Laplace kernel model (green solid) for positive h_4 . The median and $\pm 1\sigma$ bands are shown.

This paper has been typeset from a \LaTeX file prepared by the author.