Model-Independent Estimates of Dark Matter Distributions

Xiao WANG, Matthew WALKER, Jayanta PAL, Michael WOODROOFE, and Mario MATEO

A new nonparametric method is described to estimate the distribution of mass within spherical galaxies. The problem of estimating the mass, $M(r)$, within radius r is converted into a problem of estimating a regression function nonparametrically, subject to shape restrictions. We represent the restrictions by the interception of quadratic cones and use the second-order cone programming to estimate the unknown parameters. We establish asymptotic results that are used to construct confidence intervals on $M(r)$. We apply the technique to new kinematic data for four dwarf galaxies. Results indicate that dark matter dominates the stellar kinematics of these systems at all radii.

KEY WORDS: Dark matter; Dwarf galaxies; EM algorithm; Second-order cone programming; Wicksell's problem.

1. INTRODUCTION

The dark matter problem in modern astronomy dates to Zwicky's observation that the velocity dispersion of galaxies in the Coma cluster is too large for the cluster to remain bound by its luminous (e.g., stellar) mass (Zwicky 1933, 1937). Later work established similar mass deficits in other galaxy clusters, binary galaxies, and individual galaxies. Today, the standard cosmological model invokes nonbaryonic dark matter not only to hold galaxies and galaxy clusters together, but also to form them in the first place, and sets the Universe's ratio of dark to luminous mass at approximately 5 to 1 (Bennett et al. 2003). It is now thought that all galaxies reside within dark matter "halos" that extend well beyond the luminous galactic material. Pending direct detection in experiments using particle accelerators, we can learn about dark matter only from astronomical observations. The 15 dwarf spheroidal (dSph) galaxies that orbit the Milky Way are particularly useful objects to study. First, the stellar kinematics of at least some dSph's suggest extremely large ratios of dark to luminous mass (Aaronson 1983); this simplifies modeling, because the contribution of stars to the mass budget is negligible. Second, dSph's are the smallest galaxies thought to contain dark matter and so provide important constraints regarding the clustering properties of dark matter at small scales. Finally, dSph's are relatively nearby and thus observationally accessible. These qualities combine to make dSph's ideal objects with which to study the spatial distribution of dark matter at small scales.

The distribution of mass within a galaxy determines the joint distribution of stellar position and velocity. The statistical problem is how to recover the distribution of mass from a sample of stellar positions and velocities. Figure 1 shows the projected positions of red giant stars—assumed to have the same spatial distribution as the overall stellar populations—in four dSph galaxies: Carina, Fornax, Sculptor, and Sextans. We have measured the line-of-sight velocities for a subset of these stars. Considering a single dwarf spheroidal galaxy, let $\mathbf{X} = (X_1, X_2, X_3)$ denote position within the galaxy, so normalized that $\mathbf{0} = (0, 0, 0)$ is the center, and let $r = \|\mathbf{X}\| := \sqrt{X_1^2 + X_2^2 + X_3^2}$ be distance from the center. Suppose that the density of (total) mass is of the form $\rho(\mathbf{X}) = \rho_0(r)$. If ρ is known, then the mass within r units from the center is obtained from $M(r) = 4\pi \int_0^r t^2 \rho(t)\, dt$. It is easy to see that $M(r) \sim 4\pi\rho(0)r^3/3$ as $r \to 0$, where $\rho(0)$ is the central mass density.

Two data sets are available for mass estimation. First, Irwin and Hatzidimitriou (1995) provided very large data sets consisting of counts of number of stars whose projections fall into circular annuli. We observe a sample of stars with projected radii, $S_1, \ldots, S_l$, and counts, N_k, of the number of stars for which $r_{k-1} < S < r_k$, where $0 = r_0 < r_1 < \cdots < r_l$ divide the stars into bins and $S^2 = X_1^2 + X_2^2$. Figure 2 displays the projected counts of stars from four dSph galaxies: Carina, Fornax, Sculptor, and Sextans. The second data set is the velocity data, which we have acquired from observations with the Magellan/Clay Telescope at Las Campanas Observatory in Chile (Walker et al. 2006). Let $\mathbf{V} = (V_1, V_2, V_3)$ denote the three-dimensional velocity of a star. The component of each star's velocity along the line of sight is measured from the Doppler shift of absorption lines in the stellar spectra. Along with the known projected positions on the plane orthogonal to the line of sight, and with proper choice of coordinates, these observables become X_1, X_2, and V_3. Because stars must be targeted individually for spectroscopic observations, velocity data sets contain measurements for fewer stars than those included in imaging data sets. Walker et al. (2006) provided a complete description of data acquisition and reduction procedures, which yield a sample of (X_1, X_2, U_3, σ), where U_3 is the observed line-of-sight velocity with measurement error and σ is the standard error.

Our goal in this article is to develop further a new nonparametric method to estimate the mass distributions in nearby dwarf spheroidal galaxies. In previous work, Wang, Woodroofe, Walker, Mateo, and Olszewski (2005) and Walker et al. (2006) developed a monotone quadratic spline method to estimate dark matter distributions. But because $M(r)$ has a cubic behavior as $r \to 0$, the quadratic fit necessarily has unphysical asymptotic behavior at the galaxy center. In this article we improve on the former approach in the following aspects:

a. A new nonparametric maximum likelihood method is used to analyze the star counts data, which improves the original Wicksell's estimator (Walker et al. 2006).

Xiao Wang is Assistant Professor, Department of Mathematics and Statistics, University of Maryland, Baltimore, MD 21250 (E-mail: *wangxiao@umbc.edu*). Matthew Walker is Postdoctoral Research Associate, Institute of Astronomy, University of Cambridge, Cambridge CB3 OHA, U.K. (E-mail: *walker@ast.cam.ac.uk*). Jayanta Pal is Postdoctoral Fellow, SAMSI, Research Triangle Park, NC 27606 (E-mail: *jpal@samsi.info*). Michael Woodroofe is L. J. Savage Professor, Department of Statistics, University of Michigan, Ann Arbor, MI 48109 (E-mail: *michaelw@umich.edu*). Mario Mateo is Professor, Department of Astronomy, University of Michigan, Ann Arbor, MI 48109 (E-mail: *mmateo@umich.edu*). The authors thank the editor, an associate editor, and two referees for their valuable comments and constructive suggestions.

Journal of the American Statistical Association
September 2008, Vol. 103, No. 483, Applications and Case Studies
DOI 10.1198/016214508000000652

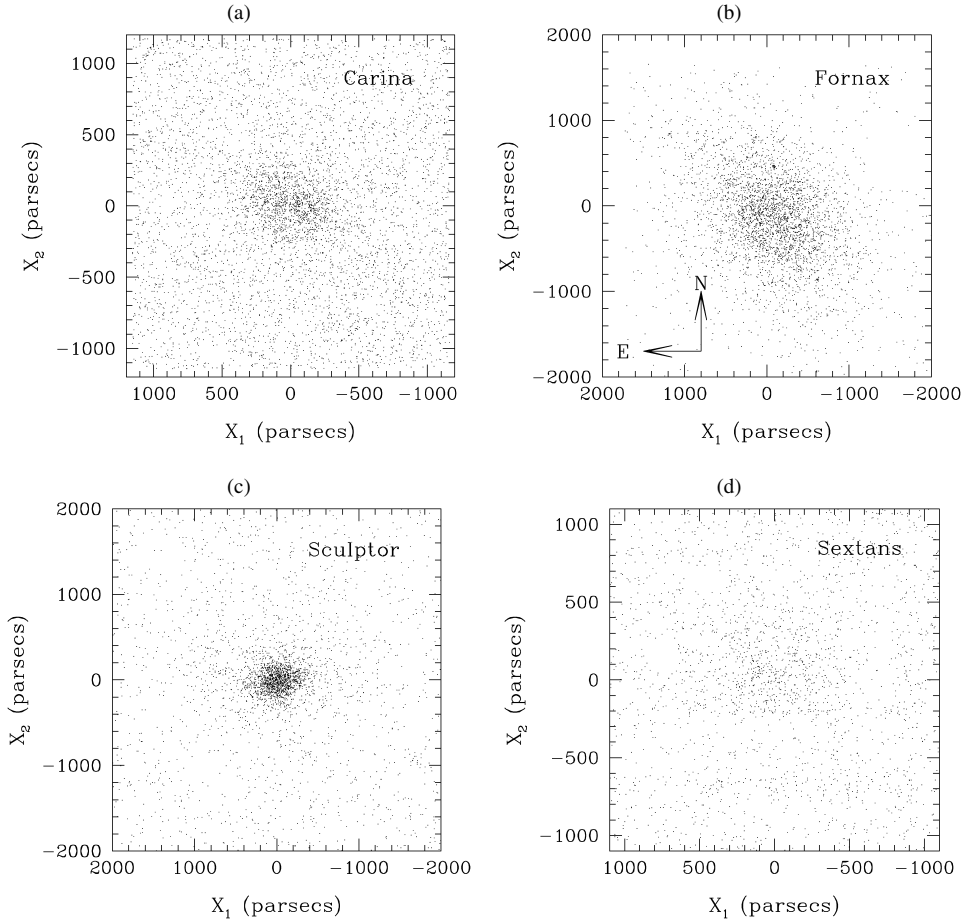

Figure 1. The projected positions of red giant stars from four dSph galaxies: (a) Carina, (b) Fornax, (c) Sculptor, and (d) Sextans.

b. The new estimation procedure combines cubic spline techniques with the isotonic methods to obtain an estimate of the mass distribution with the correct asymptotic behavior.

c. Asymptotic results are established and used to construct confidence intervals for the estimators.

We illustrate the process on simulated data and on new kinematic data obtained for four dSph galaxies.

The article is organized as follows. Section 2 introduces the Jeans Equation and uses the monotone cubic spline to estimate the mass distribution. Section 3 presents the detailed algorithms, shape restrictions, and complications that arise when dealing with astronomical observations. Section 4 establishes the asymptotic distribution of the estimators of the spline coefficients. Section 5 reports a Monte Carlo simulation study conducted to validate the method. Section 6 applies our method to estimating the mass distribution of four dwarf spheroidal galaxies. Section 7 ends with a discussion.

2. JEANS EQUATION AND MONOTONE CUBIC SPLINE ESTIMATION

In general, the Jeans Equations [eqs. (4-21), (4-24), and (4-27) of Binney and Tremaine (1987)] provide a robust description of the mass distribution of any collisionless gravitational system. Their general form permits any number of mass components, including dark matter. Suppose that the distribution of $\mathbf{X}$ and $\mathbf{V}$ has a density that depends only on the Euclidean norms $\|x\|$ and $\|v\|$, so that the joint density of $\mathbf{X}$ and $\mathbf{V}$ is $f(\mathbf{x}, \mathbf{v}) = f_0(r, v)$, where $v^2 = v_1^2 + v_2^2 + v_3^2$. In the Jeans Equations, an important role is played by the velocity dispersion, for which the statistical notation is $\sqrt{\mu(r)}$ and

$$\mu(r) = E(\|\mathbf{V}\|^2 | \mathbf{X} = \mathbf{x}). \tag{1}$$

For an isotropic velocity distribution, $\mu(r) = 3E(V_3^2|\mathbf{X} = \mathbf{x}) = \phi(r^2)/f(r)$, where $\phi(r) = \int v_3^2 f_0(\sqrt{r}, v)\, d\mathbf{v}$ and $f(r) = \int f_0(r, v)\, d\mathbf{v}$. In the presence of spherical symmetry and velocity isotropy, the relevant form of the Jeans Equation is

$$M(r) = -\frac{r^2 \mu(r)}{G} \frac{d}{dr} \log[f(r)\mu(r)], \tag{2}$$

where G is the gravitational constant.

It is convenient to work with squared projected radius $Y = S^2 = X_1^2 + X_2^2$, as was done by Groeneboon and Jongbloed (1995) and Hall and Smith (1988). The joint density of $(Y, \mathbf{V})$ can be written as

$$g_{Y,\mathbf{V}}(y, \mathbf{v}) = \pi \int_y^\infty \frac{f_0(\sqrt{z}, v)}{\sqrt{z - y}}\, dz. \tag{3}$$

Now let $\psi(y) = \int v_3^2 g_{Y,\mathbf{V}}(y, \mathbf{v})\, d\mathbf{v}$, so that $\psi(y)/g_Y(y) = E(V_3^2|Y = y)$, and similarly for ϕ. We also define the important function

$$\Psi(y) = \int_y^\infty \frac{\psi(t)\, dt}{\sqrt{t - y}}. \tag{4}$$

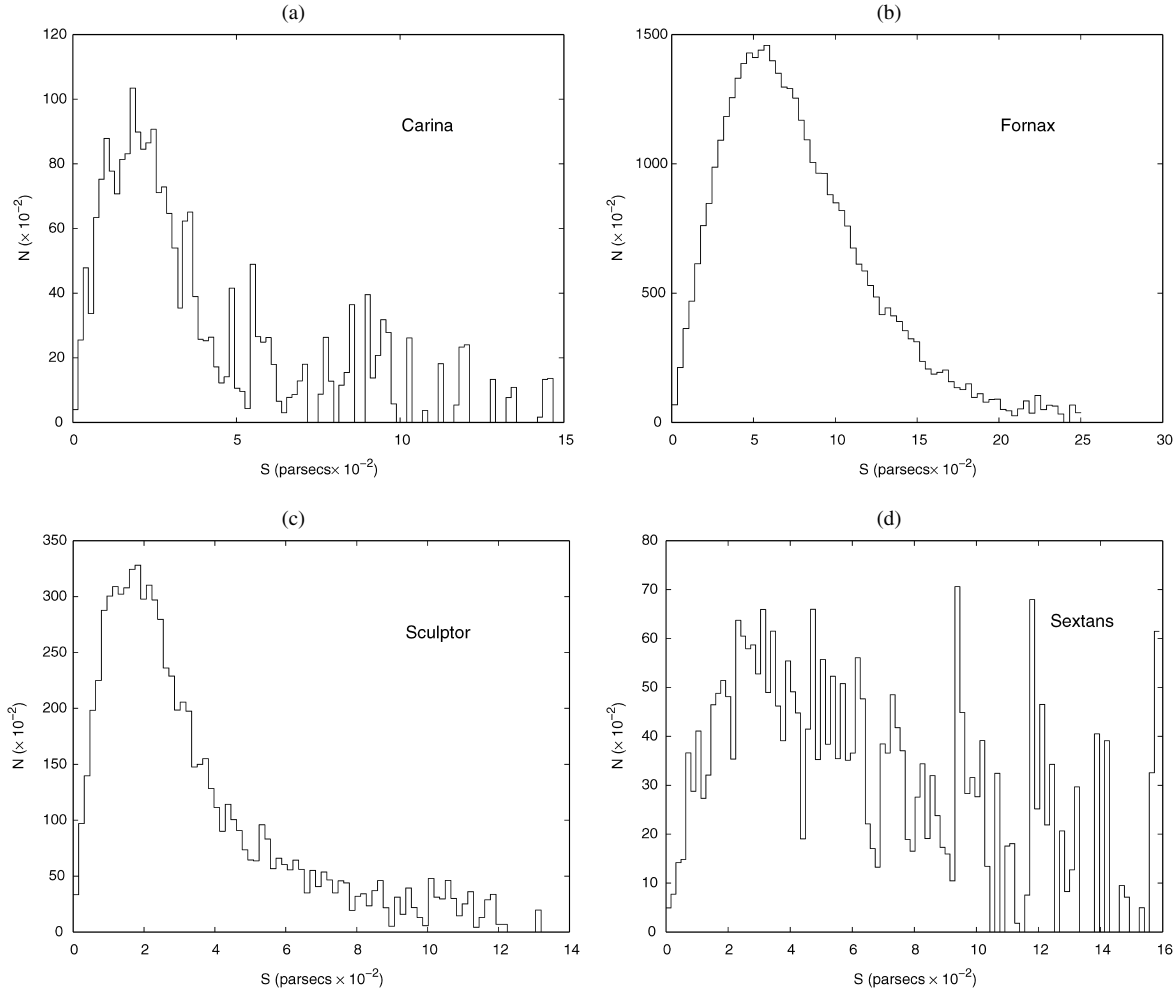

Figure 2. The projected counts of stars from four dSph galaxies: (a) Carina, (b) Fornax, (c) Sculptor, and (d) Sextans.

Then, reversing the orders of integration and recognizing a beta integral, we can write $\Psi(y) = \pi^2 \int_y^\infty \phi(z)\,dz$ and $\phi(z) = -\frac{1}{\pi^2}\Psi'(z)$. Thus the expression for $M(r)$ from the Jeans Equation (2) can be rewritten as

$$M(r) = \frac{2r^3}{G\pi^2}\frac{\Psi''(r^2)}{f(r)}. \tag{5}$$

Estimating $M(r)$ from (5) clearly requires estimates of the functions f and Ψ. We use the projected count data to estimate f (see Sec. 3.1). Let $(X_{i,1}, X_{i,2}, V_{3,i}), i = 1, \ldots, n$, denote a sample from the velocity data. Then

$$\Psi_n^{\#}(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{V_{3,i}^2 I(Y_i > z)}{\sqrt{Y_i - z}} \tag{6}$$

is an unbiased estimator of $\Psi(z)$, $E\Psi_n^{\#}(z) = \Psi(z)$ for each z. Unfortunately, $\Psi^{\#}$ does not satisfy the shape restrictions; $\Psi^{\#}(z)$ is unbounded as z approaches any of the Y_i. It is decidedly neither monotone nor convex (see Fig. 7 in Sec. 5). For these reasons, $\Psi^{\#}$ may be called the *naive* estimator.

This naive estimator can be improved by imposing physically justified shape restrictions on $M(r)$. Whereas Wang et al. (2005) approximated M with a quadratic spline, we now consider the cubic spline, which gives the correct asymptotic behavior as $r \to 0$. Let $0 = r_0 < r_1 < \cdots < r_m < \infty$ be knots and suppose that

$$M(r) = \sum_{i=1}^{m}\beta_i [r - r_{i-1}]_+^3, \tag{7}$$

where $\beta_1, \ldots, \beta_m$ are unknown coefficients and $x_+^3 = [\max(0, x)]^3$. We have $M(r) \sim \beta_1 r^3$ as $r \to 0$; thus $3\beta_1/4\pi$ provides an estimate of the central mass density with the correct asymptotic behavior at the galaxy center. Because $M(r)$ is a nondecreasing function and is bounded when $r \to \infty$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$ is located in a restricted space $\mathcal{K}$. The knots selection and constrained space $\mathcal{K}$ are discussed in Section 3.2. If we solve for Ψ'' from the Jeans Equation (5), then we can write

$$\Psi''(r^2) = \frac{G\pi^2 f(r) M(r)}{2r^3} = \sum_{i=1}^{m-1}\beta_i \gamma_i(r^2),$$

where $\gamma_i(r^2) = \frac{G\pi^2 f(r)}{2r^3}[r - r_{i-1}]_+^3$. We then solve for $\Psi(y)$ in terms of the β_i to get

$$\Psi(y) = \int_y^\infty (t - y)\Psi''(t)\,dt = \sum_{i=1}^{m-1}\beta_i \Gamma_i(y), \tag{8}$$

where $\Gamma_i(y) = \int_y^\infty (t - y)\gamma_i(t)\,dt$. We obtain the estimate of $\boldsymbol{\beta}$ by minimizing the integral of $[\Psi_n^{\#} - \Psi]^2$ or, equivalently, min-

imizing $\kappa = \frac{1}{2}\int_0^\infty \Psi(t)^2\,dt - \int_0^\infty \Psi_n^{\#}(t)\Psi(t)\,dt$. We also have the matrix notation $\kappa = \frac{1}{2}\boldsymbol{\beta}^T\mathbf{Q}\boldsymbol{\beta} - \boldsymbol{\beta}^T z$, where the elements of $\mathbf{z}$ are given by $z_i = \int_0^\infty \Psi_n^{\#}(t)\Gamma_i(t)\,dt$ and the elements of $\mathbf{Q}$ are given by $q_{ij} = \int_0^\infty \Gamma_i(t)\Gamma_j(t)\,dt$. Thus

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathcal{K}} \frac{1}{2}\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{z}. \tag{9}$$

3. ALGORITHMS AND COMPLICATIONS

3.1 Estimating f

Recall that $f(r) = \int f_0(r, \mathbf{v})\,d\mathbf{v}$. In addition to spherical symmetry and velocity isotropy, we also assume that the density of the stars decays as we move away from the center of the galaxy. In particular, f_0 decreases with r, as does f. Wang et al. (2005) followed the method of Wicksell (2006) to obtain an estimator of f, but this estimator does not satisfy the shape restriction, leading to negative estimate of density function at some points. Because we have observations, N_k, of the number of stars for which $r_{k-1} < S < r_k$ in the kth annulus, estimating f amounts to an inverse problem that can be solved as follows.

The goal is to estimate $f(\cdot)$ from the grouped count data. In the current discussion, a nonparametric maximum likelihood estimator (NPMLE) for f is sought in the class of nonincreasing, piecewise constant densities with knots at $r_0, r_1, \ldots, r_l$, that is,

$$f(r) = a_k \quad \text{if } r_{k-1} \le r < r_k,$$

where $a_1 \ge a_2 \ge \cdots \ge a_l$. The likelihood function for $(a_1, a_2, \ldots, a_l)$ is then

$$\prod_{k=1}^{l}\left[\int_{r_{k-1}}^{r_k} g(s)\,ds\right]^{N_k},$$

where $g(s)$ denotes the density of S,

$$g(s) = s\int_s^\infty \frac{4\pi r f(r)\,dr}{\sqrt{r^2 - s^2}}.$$

Rather than having the grouped count of the projected data S_i, suppose that we observe the counts for R_i, that is, we obtain $M_k = \#(r_{k-1} \le R_i < r_k)$, $k = 1, \ldots, l$. Then the likelihood turns out to be

$$\prod_{k=1}^{l}\left[\int_{r_{k-1}}^{r_k} 4\pi r^2 f(r)\,dr\right]^{M_k} = \prod_{k=1}^{l}\left[4\pi a_k\frac{r_k^3 - r_{k-1}^3}{3}\right]^{M_k}.$$

Maximizing the likelihood is equivalent to maximizing the function $\sum_{k=1}^{l} M_k \log a_k$ subject to the constraints

$$a_1 \ge a_2 \ge \cdots \ge a_l \tag{10}$$

and

$$\frac{4\pi}{3}\sum_{k=1}^{l} a_k(r_k^3 - r_{k-1}^3) = \int_{r_0}^{r_l} 4\pi r^2 h(r)\,dr = 1. \tag{11}$$

This is an isotonic regression problem with the optimum solution $(a_1^*, \ldots, a_l^*)$. It is easy to see that

$$a_k^* = \min_{i<k}\max_{j>k}\frac{M_i + \cdots + M_j}{w_i + \cdots + w_j}, \tag{12}$$

where $w_k = 4\pi(r_k^3 - r_{k-1}^3)/3$ (see Robertson, Wright, and Dykstra 1987).

Because we do not observe the count data for the R_i, we treat this as a missing-data problem. Missing-data problems have been fairly abundant in the literature, and one of the popular techniques is to use an EM algorithm. This was proposed by Dempster and Laird (1977), and later discussed in detail by Wu (1983).

We proceed to compute the M_k using the EM algorithm, replacing the M_k by $E(M_k|N_1, \ldots, N_l)$ at each step. To do this, we observe that

$$\begin{aligned} p_{kd} &= P(x_{k-1} < R < x_k, x_{d-1} < S < x_d) \\ &= 4\pi\int_{x_{k-1}}^{x_k}\int_{x_{d-1}}^{x_d} I_{(s,\infty)}(r)h(r)\frac{rs}{\sqrt{r^2 - s^2}}\,dr\,ds. \end{aligned}$$

Thus

$$p_{kd} = \begin{cases} \dfrac{4\pi}{3}a_k[(x_k^2 - x_{d-1}^2)^{1.5} - (x_k^2 - x_d^2)^{1.5} \\ \quad - (x_{k-1}^2 - x_{d-1}^2)^{1.5} + (x_{k-1}^2 - x_d^2)^{1.5}] & \text{if } k > d \\ \dfrac{4\pi}{3}a_k[(x_k^2 - x_{k-1}^2)^{1.5}] & \text{if } k = d \\ 0 & \text{if } k < d. \end{cases} \tag{13}$$

Moreover,

$$E[M_k|N_1, \ldots, N_l] = \sum_{d=1}^{l} N_d\frac{p_{kd}}{p_{+d}}. \tag{14}$$

Let $\mathbf{P} = (p_{kd})_{k,d=1}^{l}$ and $\Delta = \mathrm{diag}(\mathbf{1}^t\mathbf{P})$, so that $E[\mathbf{M}|\mathbf{N}] = \mathbf{P}\Delta^{-1}\mathbf{N}$. Then the algorithm works follows:

Step 1. Start with initial values $(a_1^0, \ldots, a_m^0) = \mathbf{a}^0$. Let $i = 0$.
Step 2. Compute $\hat{M}_k^i = E_{\mathbf{a}^i}[M_k|N_1, \ldots, N_m]$ from (14).
Step 3. Compute $u^i = \sum_{k=1}^{m}\hat{M}_k^i \log a_k^i$. If $|u^i - u^{i-1}| < \epsilon$, stop and declare the estimates as $\mathbf{a}^i$.
Step 4. Compute $\mathbf{a}^{i+1} = (a_1^*, \ldots, a_m^*)$ from (12) with $\hat{M}_k^i$. Let $i = i + 1$ and start from step 2.

Figure 3 shows the estimated three-dimensional distributions, f, of stars for these four galaxies. Because the imaging sample sizes are so large (tens of thousands of stars), there is generally very little statistical uncertainty in the estimate of f compared with the estimate of μ. Thus in our analysis, f is treated as known.

3.2 Knot Selection and Shape Restrictions for $M(r)$

The number of knots in (7) is chosen as the number of steps in the NPMLE $\hat{f}$ of f, and the location of the knots is selected as the places where the jumps occur. In our study, the number of knots ranges from 10 to 25, and the locations are a subset of $(r_0, r_1, \ldots, r_l)$ from Irwin and Hatzidimitriou (1995). Furthermore, because the monotonicity restriction does not allow "wiggles," the fit is less sensitive to knot choices. Figure 4 shows the monotone regression spline fits for a logistic function with sample size 50. The knots are chosen uniformly over the range, and the numbers of knots are 10, 15, and 30. Figure 4(d) superimposes fits from (a)–(c). It suggests that the fits are robust to knot choices within a wide range of numbers of knots.

In what follows we discuss the constraints for the spline coefficient. One natural constraint is that $M(r)$ remains bounded as

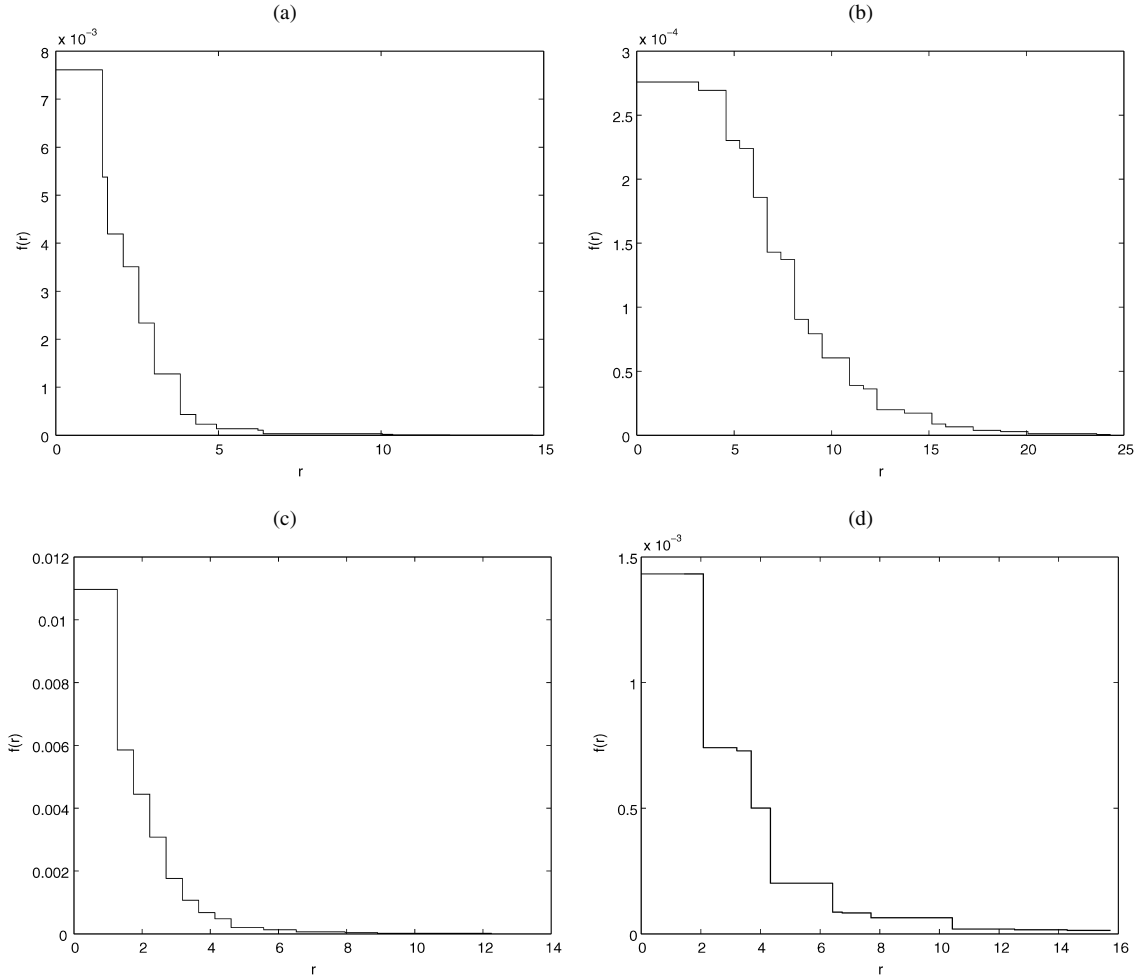

Figure 3. The estimated three-dimensional distributions of stars for four dSph galaxies: (a) Carina, (b) Fornax, (c) Sculptor, and (d) Sextans.

$r \to \infty$. For $r > r_{m-1}$, we have that $M(r) = (\sum_{i=1}^m \beta_i)r^3 - (3\sum_{i=1}^m \beta_i r_{i-1})r^2 + (3\sum_{i=1}^m \beta_i r_{i-1}^2)r - \sum_{i=1}^m \beta_i r_{i-1}^3$. Thus this constraint requires that $\beta = (\beta_1, \ldots, \beta_m)' \in \mathcal{D}$, where

$$\mathcal{D} = \left\{ \beta \in R^m : \sum_{i=1}^m \beta_i = 0, \sum_{i=1}^m \beta_i r_{i-1} = 0, \sum_{i=1}^m \beta_i r_{i-1}^2 = 0 \right\}. \tag{15}$$

We also require that $M(r)$ be a nondecreasing function of r or, equivalently, that the derivative $M'(r)$ be nonnegative. Thus $\beta \in \mathcal{C}$, where

$$\mathcal{C} = \left\{ \boldsymbol{\beta} \in R^m : \sum_{i=1}^m \beta_i (r - r_{i-1})_+^2 \geq 0, \text{ for all } r > 0 \right\}. \tag{16}$$

The constrained set $\mathcal{C}$ is defined by an infinite number of linear inequalities and thus is a convex set. But $M'(r)$ is a piecewise quadratic function, and finding a general condition for a quadratic function to be nonnegative on an interval is surprisingly difficult. Wang and Li (2008) derived a necessary and sufficient condition for any quadratic function that is nonnegative on [0, 1]: The second-order polynomial $g(x) = a_2x^2 + a_1x + a_0$ satisfies $g(x) \geq 0$ for all $x \in [0, 1]$ if and only if $a_0 \geq 0$, $a_0 + a_1 + a_2 \geq 0$, and there exists a constant $\delta \geq 0$ such that $4a_0(a_2 + \delta) \geq (a_1 - \delta)^2$.

Write $c_i(z) = M'((r_i - r_{i-1})z + r_{i-1})$, for which $z \in [0, 1]$, $i = 1, \ldots, m$. Let $C_i = \{\boldsymbol{\beta} : c_i(z) \geq 0 \text{ for } z \in [0, 1]\}$. Thus $\mathcal{C} = \bigcap_{i=1}^m C_i$. Specifically, when $\beta \in C_i$, the constraint for $\boldsymbol{\beta}$ is $c_i(z) = a_{i2}(\boldsymbol{\beta})z^2 + a_{i1}(\boldsymbol{\beta})z + a_{i0}(\boldsymbol{\beta}) \geq 0$ for all $z \in [0, 1]$, where a_{i2}, a_{i1}, and a_{i0} are linear functions of $\boldsymbol{\beta}$. For example, if $\boldsymbol{\beta} \in C_2$, then $a_{i2} = \beta_1 + \beta_2$, $a_{i1} = -2\beta_2 r_1$, and $a_{i0} = \beta_2 r_1^2$. Thus we have

$$C_i = \big\{ \boldsymbol{\beta} : a_{i0}(\beta) \geq 0, 4a_{i0}(\boldsymbol{\beta})[a_{i2}(\boldsymbol{\beta}) + \delta_i] \geq [a_{i1}(\boldsymbol{\beta}) - \delta_i]^2 \text{ for some } \delta_i \geq 0 \big\}. \tag{17}$$

Furthermore, galaxies in dynamic equilibrium have mass density, $\rho(r)$, that is a nonincreasing function of r. A necessary and sufficient condition for $\rho(r) \propto M'(r)/r^2$ to be decreasing is that $\beta \in \mathcal{E}$, where

$$\mathcal{E} = \left\{ \boldsymbol{\beta} : \sum_{i=1}^{k-1} \beta_i r_{i-1}(r_k - r_{i-1}) \leq 0, \sum_{i=1}^{k} \beta_i r_{i-1}(r_{k+1} - r_{i-1}) \leq 0, \text{ for } k = 1, \ldots, m-1 \right\}, \tag{18}$$

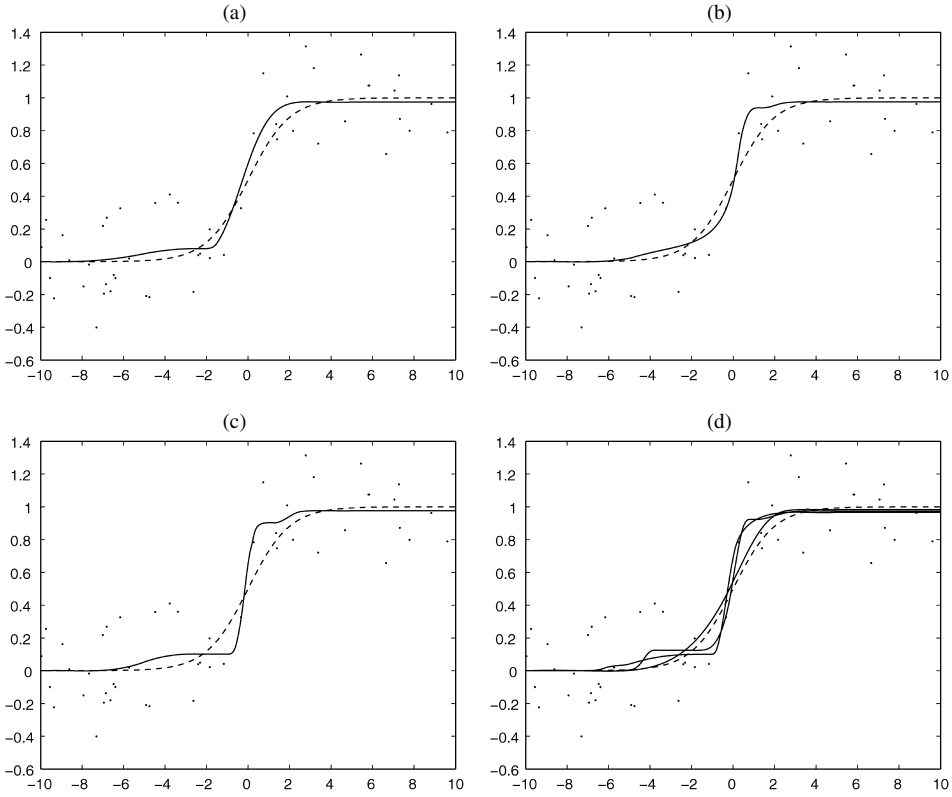

Figure 4. Monotone regression spline fits for a logistic function with sample size 50. The knots are chosen uniformly over the range, and the number of knots are (a) 10, (b) 15, and (c) 30.

which are linear constraints of β. Thus the constraint imposed on β is that $\beta \in \mathcal{C} \cap \mathcal{D} \cap \mathcal{E}$, which is a convex set.

3.3 Second–Order Cone Programming

The optimization problem (9) can be solved efficiently by second-order cone programming (SOCP). SOCP addresses the problem of minimizing a linear objective function over the second-order cones. A typical SOCP is given by

$$\begin{aligned} &\min \mathbf{c}'\boldsymbol{\beta}, \\ &\text{s.t.} \quad \mathbf{A}\boldsymbol{\beta} = \mathbf{b}, \\ &\qquad \boldsymbol{\beta} \in \mathcal{K}, \end{aligned} \tag{19}$$

where $\boldsymbol{\beta}$ is the unknown parameter, $\mathbf{c}$ and $\mathbf{b}$ are known vectors, and $\mathbf{A}$ is the known matrix. The second-order cone $\mathcal{K}$ is defined as $\mathcal{K}^1 \times \mathcal{K}^2 \times \cdots \times \mathcal{K}^k$, which is the direct product of k cones (Kuo and Mittelmann 2004). $\mathcal{K}^i$ is one of the following three cones:

a. $R^+ = \{\boldsymbol{\beta} \geq 0, \boldsymbol{\beta} \in \mathbb{R}^n\}$

b. Quadratic cone: $\mathcal{K}_q = \{\boldsymbol{\beta} \in \mathbb{R}^n : \beta_1^2 \geq \sum_{j=2}^n \beta_j^2, \beta_1 \geq 0\}$

c. Rotated quadratic cone: $\mathcal{K}_r = \{\boldsymbol{\beta} \in \mathbb{R}^n : 2\beta_1\beta_2 \geq \sum_{j=3}^n \beta_j^2, \beta_1 \geq 0, \beta_2 \geq 0\}$.

Note that the constraints C_i in (17) have exactly the form of a rotated quadric cone, $2u_0^i u_1^i \geq (u_2^i)^2$, where $u_0^i = 2a_{i0}(\beta)$, $u_1^i = 2[a_{i2}(\beta) + \delta_i]$, $u_2^i = a_{i1}(\beta) - \delta_i$, and u_0^i, u_1^i, and u_2^i are all linear functions of the unknown parameters.

The problem of minimizing (9) under the constraints is equivalent to

$$\begin{aligned} &\min \boldsymbol{\nu} - \mathbf{y}'\mathbf{N}\boldsymbol{\beta}, \\ &\text{s.t.} \quad \boldsymbol{\beta} \in \mathcal{C} \cap \mathcal{D} \cap \mathcal{E}, \\ &\qquad \mathbf{R}\boldsymbol{\beta} - \mathbf{t} = 0, \\ &\qquad 2\boldsymbol{\nu}\boldsymbol{\omega} \geq \|\mathbf{t}\|^2, \boldsymbol{\omega} = 1, \end{aligned} \tag{20}$$

where $\mathbf{Q} = \mathbf{R}'\mathbf{R}$. After introducing the new variables $\boldsymbol{\nu}$, $\mathbf{t}$, and $\boldsymbol{\omega}$, we convert this problem to a conic programming problem. The problem is to minimize a linear function of the unknown parameters, which are now $(\boldsymbol{\nu}, \mathbf{t}, \boldsymbol{\omega}, \boldsymbol{\beta}')'$, subject to the second-order cone constraints. A detailed algorithm for this has been given by Tsuchiya (1999). Although it is a special case of semidefinite programming (SDP), the SOCP can be solved more efficiently than SDP using the primal-dual interior-point method. A worst-case theoretical analysis shows that the number of iterations required to solve a problem grows at most as the square root of the problem size (Nesterov and Todd 1997), whereas numerical experiments indicate that the typical number of iterations ranges between 5 and 50 almost independent of the problem size (Lobo, Vandenberghe, Boyd, and Lebret 1998). Various software packages for solving the conic programming are available, including MOSEK, SeDumi, and SDPT3.

3.4 Complications

Two complications must be overcome before the procedure can be applied: Velocities are observed with error, and selection effects may be present. For a sample $(X_{i,1}, X_{i,2}, V_{i,3})$ of

projected positions and line-of-sight velocities, $V_{i,3}$ is observed with error. Thus we observe $U_i = V_{i,3} + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are random errors with known standard deviations $\sigma_1, \ldots, \sigma_n$. If we replace $V_{i,3}^2$ with $U_i^2 - \sigma_i^2$, (6) still produces an unbiased estimator of Ψ, and the procedure described in the previous section can proceed with only that minimal change.

Selection effects are more complicated. Suppose that the probability of observing a star with projected position (x_1, x_2) can depend (only) on x_1 and x_2, say $w_0(x_1, x_2)$. Then the joint probability density of X_1, X_2, and $\mathbf{V}$ for stars in the sample becomes

$$g^*(x_1, x_2, \mathbf{v}) = w(y) g_0(y, \mathbf{v}), \tag{21}$$

where $y = x_1^2 + x_2^2$, $w(y) = \frac{1}{c}\int_{-\pi}^{\pi} w_0[\sqrt{y}\sin(\phi), \sqrt{y} \times \cos(\phi)]\, d\phi$, and c is a normalizing constant. If an astronomer can specify the function w, then

$$\Psi^{\#}(t) = \frac{1}{n} \sum_{i:Y_i > t} \frac{U_i^2 - \sigma_i^2}{w(Y_i)\sqrt{Y_i - t}} \tag{22}$$

provides an unbiased estimator of Ψ and an estimation ν_r, and M can proceed as described in the previous section. If the astronomer cannot specify w_0, then w can be estimated from the sample. Integrating over $\mathbf{v}$ in (21) shows that the probability density of X_1 and X_2 for stars in the sample is $g^*(x_1, x_2) = w(y) g_0(y) = g_0^*(y)$, say; thus $w(y) = g_0^*(y)/g_0(y)$. If f_0 is known (from, e.g., imaging studies), then $g(x_1, x_2) = g_0(y)$, where $g_0(y) = \int_y^{\infty} \frac{f_0(z)\, dz}{\sqrt{z - y}}$. Furthermore, g_0^* can be estimated from the stars in the sample; for example, using a kernel estimator, $\hat{g}_0^*(y) = \frac{1}{nh}\sum_{i=1}^n K(\frac{y - Y_i}{h})$, where K is a kernel and $h > 0$ is a bandwidth. Then $\hat{w}(y) = \hat{g}_0^*(y)/g_0(y)$ provides an estimator of w that can be substituted into (22).

4. ASYMPTOTIC RESULTS

In this section we derive the asymptotic distribution of the $\hat{\beta}$, which can be used for constructing confidence intervals of the estimators. Assume that the mass function $M(r)$ is linear in β and has the form (7), where the knots $r_0, \ldots, r_{m-1}$ are known. Let β_0 be the true parameter.

Because space $\mathcal{D}$ includes three equality constraints such that $\mathbf{B}\beta = 0$, where

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ r_0 & r_1 & \cdots & r_{m-1} \\ r_0^2 & r_1^2 & \cdots & r_{m-1}^2 \end{bmatrix},$$

the only free parameter is $\boldsymbol{\omega} = (\beta_1, \ldots, \beta_{m-3})'$, and $(\beta_{m-2}, \beta_{m-1}, \beta_m)'$ is the linear transformation of $\boldsymbol{\omega}$. Let $\mathbf{B}_1$ be the first $m - 3$ columns of B, and let $\mathbf{B}_2$ be the last three columns of $\mathbf{B}$. Then, under constraint $\mathcal{D}$, $\boldsymbol{\beta} = \mathbf{T}\boldsymbol{\omega}$, where $\mathbf{T} = [\mathbf{I}_{(m-3)\times(m-3)}; -\mathbf{B}_2^{-1}\mathbf{B}_1]'$. Thus, because $\boldsymbol{\beta} \in \mathcal{C} \cap \mathcal{D} \cap \mathcal{E}$, $\boldsymbol{\omega}$ belongs to

$$\Omega = \Bigg\{ \boldsymbol{\omega} \in R^{m-3} : \boldsymbol{\beta} = \mathbf{T}\boldsymbol{\omega}, \sum_{i=1}^{m} \beta_i (r - r_{i-1})_+^2 \geq 0, \sum_{i=1}^{k-1} \beta_i r_{i-1}(r_k - r_{i-1}) \leq 0, \sum_{i=1}^{k} \beta_i r_{i-1}(r_{k+1} - r_{i-1}) \leq 0, \text{ for } k = 1, \ldots, m-1 \Bigg\}.$$

Thus the estimator $\hat{\boldsymbol{\omega}}$ is given by

$$\begin{aligned} \hat{\boldsymbol{\omega}} &= \arg\min_{\boldsymbol{\omega}\in\Omega} \frac{1}{2}\boldsymbol{\omega}'\tilde{\mathbf{Q}}\boldsymbol{\omega} - \boldsymbol{\omega}'\tilde{\mathbf{z}} \\ &= \arg\min_{\boldsymbol{\omega}\in\Omega} (\mathbf{u} - \boldsymbol{\omega})'\tilde{\mathbf{Q}}(\mathbf{u} - \boldsymbol{\omega}), \end{aligned} \tag{23}$$

where $\tilde{\mathbf{Q}} = \mathbf{T}'\mathbf{Q}\mathbf{T}$, $\tilde{\mathbf{z}} = \mathbf{T}'\mathbf{z}$, and $\mathbf{u} = \tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{z}}$.

Note that the matrix $\mathbf{Q}$ depends only on f and is a constant matrix, because f is treated as known. Let $\Psi_2^{\#}$ be defined as

$$\begin{aligned} \Psi_2^{\#}(z) &= \int_0^z (z - y)\Psi^{\#}(y)\, dy \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{U_i^2 - \sigma_i^2}{w(Y_i)} \\ &\quad \times \left[\frac{4}{3}(Y_i - z)^{3/2} I(Y_i > z) - Y_i^{3/2} + 2zY_i^{1/2}\right]. \end{aligned}$$

Then the element of $\mathbf{z}$ can then be computed numerically as

$$\begin{aligned} z_i &= \int_0^{\infty} \Psi^{\#}(t)\Gamma_i(t)\, dt \\ &= \int_0^{\infty} \Psi_2^{\#}(z)\gamma_i(z)\, dz = \frac{1}{n}\sum_{l=1}^{n} h_i(Y_l, U_l), \end{aligned} \tag{24}$$

where h_i can be computed explicitly. By the central limit theorem, under some regularity conditions, we have the asymptotic distribution for $\mathbf{z} = (z_1, \ldots, z_m)'$,

$$\sqrt{n}(\mathbf{z} - \mathbf{Q}\boldsymbol{\beta}_0) \to^D \mathrm{N}(0, \boldsymbol{\Sigma}), \quad \text{as } n \to \infty,$$

where the elements for $\boldsymbol{\Sigma}$ are given by $\Sigma_{ij} = \mathrm{cov}[h_i(Y_1, U_1), h_j(Y_1, U_1)]$. In practice, $\boldsymbol{\Sigma}$ can be approximated by sample covariance matrix. Thus we have

$$\sqrt{n}(\mathbf{u} - \boldsymbol{\omega}_0)' \to^D \mathrm{N}(0, \tilde{\boldsymbol{\Sigma}}), \quad \text{as } n \to \infty, \tag{25}$$

where $\tilde{\boldsymbol{\Sigma}} = (\mathbf{T}'\mathbf{Q}\mathbf{T})^{-1}\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}(\mathbf{T}'\mathbf{Q}\mathbf{T})^{-1}$.

Consider the tangent cone, Ξ, to Ω at the point $\boldsymbol{\omega}_0$. Loosely speaking, the cone Ξ is a local approximation to the Ω near $\boldsymbol{\omega}_0$. The tangent cone Ξ has the following approximation properties: $\inf_{\boldsymbol{\xi}\in\Xi} \|(\boldsymbol{\omega} - \boldsymbol{\omega}_0) - \boldsymbol{\xi}\| = o(\|\boldsymbol{\omega} - \boldsymbol{\omega}_0\|)$ for all $\boldsymbol{\omega} \in \Omega$ and $\inf_{\boldsymbol{\omega}\in\Omega} \|(\boldsymbol{\omega} - \boldsymbol{\omega}_0) - \boldsymbol{\xi}\| = o(\|\boldsymbol{\xi}\|)$ for all $\boldsymbol{\xi} \in \Xi$. Following theorem 3.3 of Shapiro (1989), (23), and (25), we have the asymptotic distribution for $\hat{\omega}$.

Proposition 1. Assume that model (7) holds and that the matrix $\tilde{\boldsymbol{\Sigma}}$ is positive definite. Then $\sqrt{n}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)'$ converges in distribution to the minimizer $\boldsymbol{\xi}(\mathbf{Z})$ of $(\mathbf{Z} - \boldsymbol{\xi})'\tilde{\mathbf{Q}}(\mathbf{Z} - \boldsymbol{\xi})$ over the cone Ξ, where $\mathbf{Z}$ follows $\mathrm{N}(0, \tilde{\boldsymbol{\Sigma}})$.

If $\boldsymbol{\omega}_0$ is an interior point of Ω, then $\Xi = R^{m-3}$, and $\boldsymbol{\xi}(\mathbf{Z})$ has distribution $\mathrm{N}(0, \tilde{\boldsymbol{\Sigma}})$. When $\boldsymbol{\omega}_0$ is on the boundary of Ω, Ξ is defined by the linear inequalities corresponding to the constraints active at $\boldsymbol{\omega}_0$. We may need to solve a semi-infinite programming (SIP) problem to obtain the distribution of $\boldsymbol{\xi}(\mathbf{Z})$. We plan to tackle this problem in future work. In this article we assume that $\boldsymbol{\omega}_0$ is the interior point of Ω, which implies that the true mass function is a strictly increasing function and the mass density function is a strictly decreasing function. By simulating the random samples from the asymptotic distribution of $\hat{\boldsymbol{\omega}}$, we can construct the confidence interval for the estimators of mass distribution or velocity dispersion.

5. SIMULATION

In this section we use Monte Carlo simulation to illustrate and validate our approach. Consider the Plummer model, a parametric model used by astronomers. In this model the joint distribution of $(\mathbf{X}, \mathbf{V})$ is

$$f_0(r, v) = \frac{c_0}{b^5}\left[\frac{b}{\sqrt{1+(1/3)r^2}} - \frac{1}{2}v^2\right]_+^{7/2}, \qquad (26)$$

where c_0 is a constant and b is a parameter related to the velocity dispersion through $E(V^2|R=r) = b/[2\sqrt{1+r^2/3}]$. For this model, $f(r) = 1/[4\pi\sqrt{3}(1+r^2/3)^{5/2}]$. We simulate the projected count data and velocity data from this model. For each of 1,000 simulated data sets, we estimate $M(\cdot)$ and $\mu(\cdot)$.

In each Monte Carlo realization, we first draw a sample of 100, 000 positions from the Plummer model with $b = 200$. Figure 5(a) shows simulated projected counts for 100,000 stars in equally spaced bins ranging from $0 = r_0 < r_1 < \cdots < r_{25} = 12$ from the Plummer model. Figure 6 superimposes 50 randomly selected estimates of f from the simulated data sets. As we can see, the variance of the estimator is small because of the large sample size. In all figures, the dotted line denotes the true function.

Figure 5(b) shows 500 simulated (S, V_3) pairs from the same model. The naive estimator, $\Psi^{\#}$, may be computed from these data using (6). Figure 7 shows this, along with the improved estimator, $\hat{\Psi}$. The need for the improved estimator for Ψ is clear because, as expected, $\Psi^{\#}$ is a highly irregular function that reacts strongly to the presence of individual data points. For each of the simulated star count and velocity data sets, we estimate $M(\cdot)$ and $\mu(\cdot)$. Although the selection function is $\omega(y) = 1$ in the simulations, we use the estimated $\hat{\omega}$ in our analysis to obtain $M(\cdot)$ and $\mu(\cdot)$, and this selection function can be estimated consistently. Furthermore, the effect of variability in estimating f is quite small. In Figure 8(a), the solid line is the mean of estimated mass over all 1,000 simulations. The dashed–dotted lines

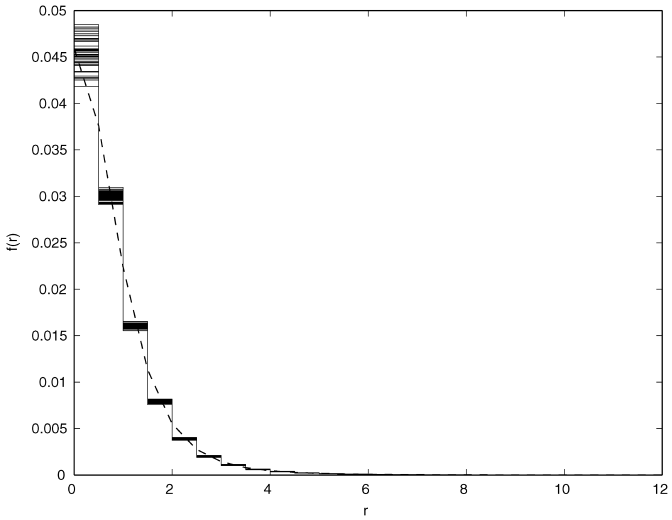

Figure 6. The 50 estimated $f(r)$'s from the projected counts data superimposed together.

are 95% pointwise confidence bands. The true functions are shown in the dashed lines. In Figure 8(b), 50 randomly selected estimators of the mass distribution are superimposed to give a better sense of the variability in the estimator. In Figure 9(a), the solid line is the mean of estimated μ over all 1,000 simulations, and the dashed–dotted lines are 95% pointwise confidence bands. In Figure 9(b), 50 randomly selected estimators of μ are superimposed.

Figure 8(a) ($r > 10$) merits some comment, because it illustrates an important feature of the problem in general and our method in particular: Estimates become unreliable at the boundary points, and shape-restricted methods are especially prone to such problems (e.g., Woodroofe and Sun 1993). The data also become sparse at large values of r. Determining exactly where reliability ends is difficult. From our simulation experience, the estimates are reliable within the radius enclosing 95% of the data points.

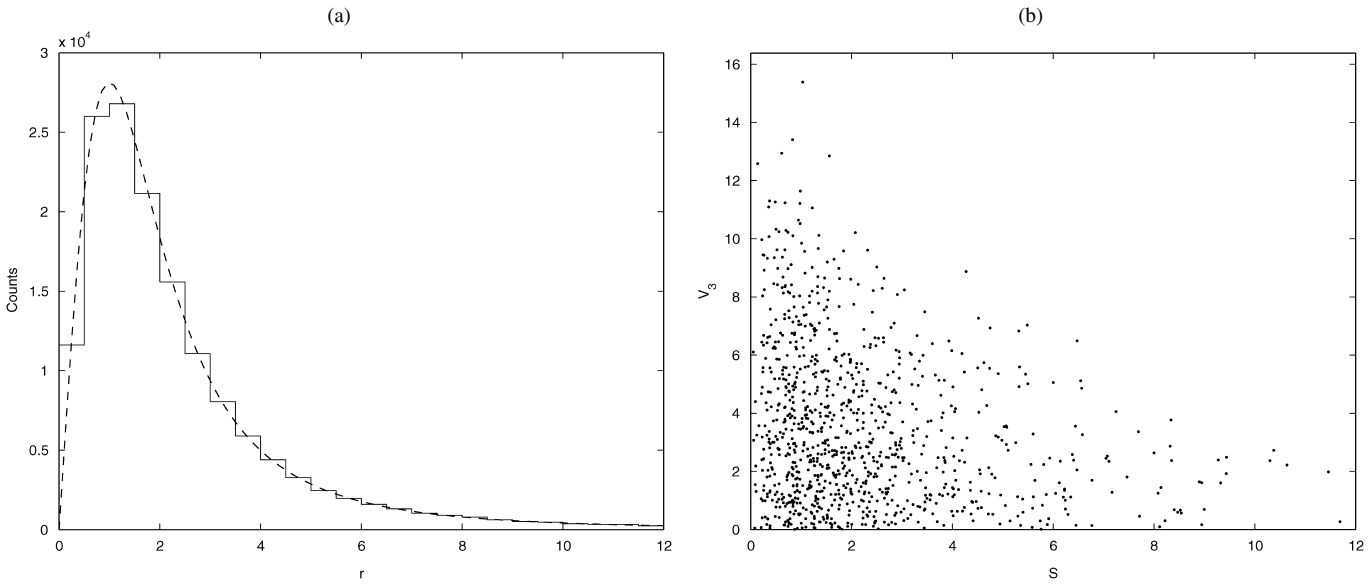

Figure 5. The simulated projected counts for 100,000 stars from the Plummer model (a) and the simulated 500 (S, V_3) pairs from the same Plummer model (b).

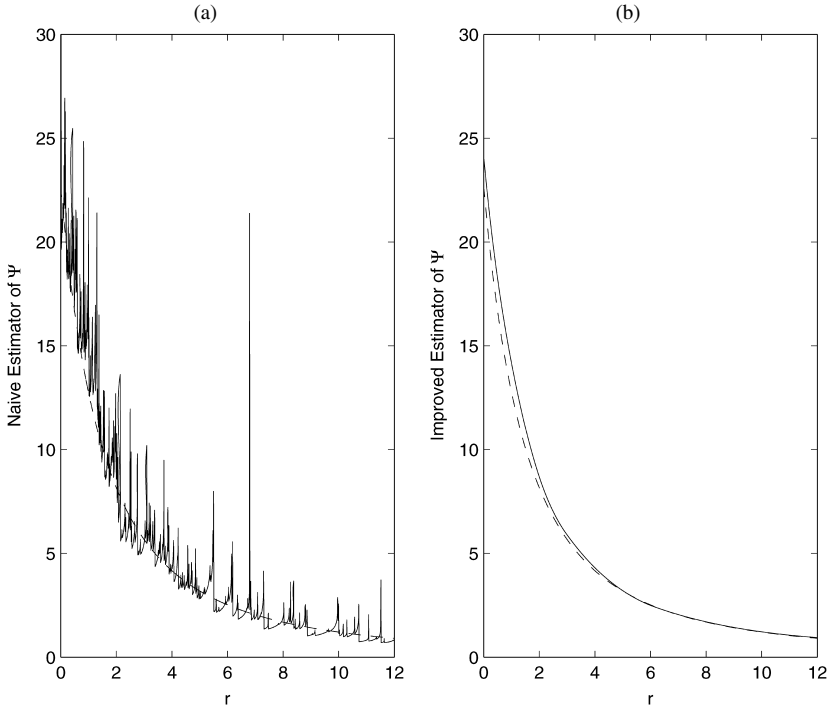

Figure 7. The naive estimator $\Psi^{\#}$ as defined by (6) (a) and the improved estimator $\hat{\Psi}$ derived from imposing the shape restrictions on the form of $M(r)$ (b). (——, the estimator; − − − −, the true distribution calculated from the underlying Plummer model.)

6. APPLICATIONS: DWARF SPHEROIDAL GALAXIES

We operate on real velocity data to estimate the mass distributions of four dSph galaxies: Carina, Fornax, Sculptor, and Sextans. We have used the Michigan/MIKE Fiber System at the Magellan 6.5-m telescope to acquire spectra from more than 5,000 stars. Figure 10 maps the positions of observed stars. A comparison with Figure 1 gives a sense of the selection effect, for which we compensate using (22). We measure the line-of-sight velocity of each star from the Doppler shift of spectral absorption lines. Walker et al. (2006) provided a complete description of data acquisition and reduction procedures, which yield a sample of (X_1, X_2, U_3, σ). After we identify and remove interloping Milky Way stars projected along the line of sight, the final samples contain measurements for 759, 2,017, 988, and 460 stars belonging to the dSph galaxies Carina, Fornax, Sculptor, and Sextans. Figure 11 displays the distribution of S, U_3 pairs in each galaxy. The errors associated with U_3 have median value $\sigma = 2$ km s^{-1}.

The solid curves in Figure 12(a)–(d) give the estimates of $\mu(r)$ calculated from the cubic spline model. For comparison, data points give conventional parametric estimates of $\mu(r)$, in which the velocity data are binned with radius, a normal

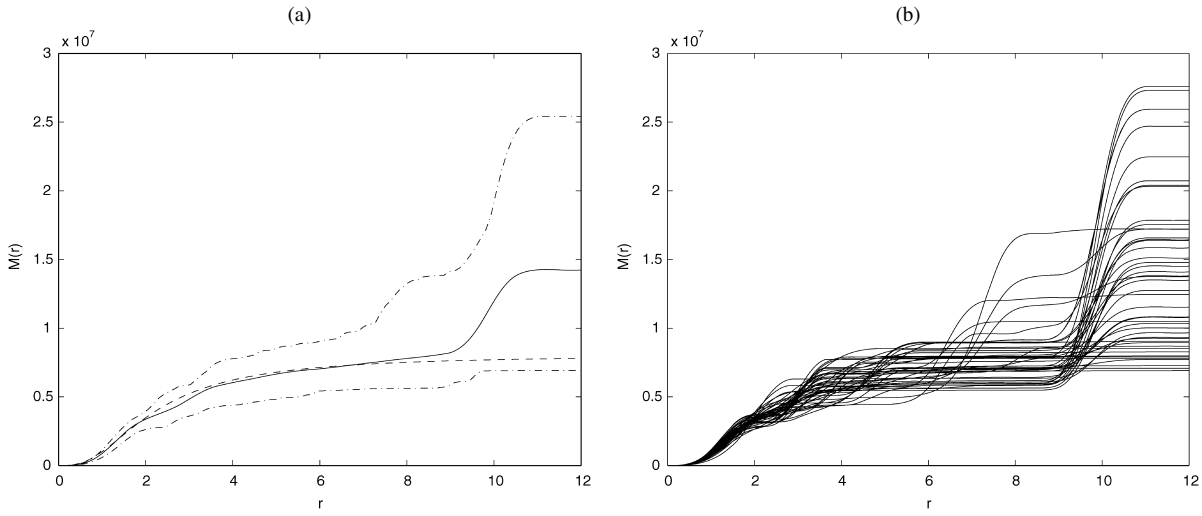

Figure 8. (a) The mean of estimated mass over all 1,000 simulations (——), 95% pointwise confidence bands (– · – · –), and the true functions (− − −). (b) The 50 randomly selected estimators of the mass distribution superimposed together.

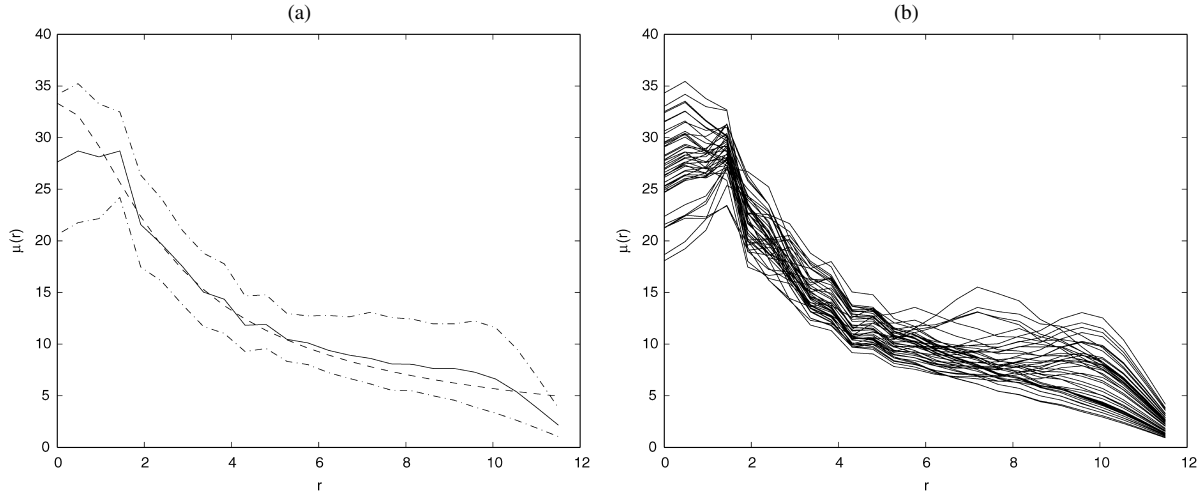

Figure 9. (a) The mean of estimated μ's over all 1,000 simulations (——), 95% pointwise confidence bands (– · – · –), and the true functions (– – –). (b) The 50 randomly selected estimators of μ superimposed together.

velocity distribution is assumed, and the plotted points indicate the maximum likelihood variance within each bin (Walker et al. 2006). Figure 12(f) shows the associated estimates of $M(r)$. Figure 12(e) displays mass density distributions, related to the mass profiles by $\rho(r) = M'(r)/4\pi r^2$. The primary virtue of the cubic spline method lies in the fact that the estimated mass density converges on a maximum value as $r \to 0$; in contrast, the earlier quadratic spline model of Wang

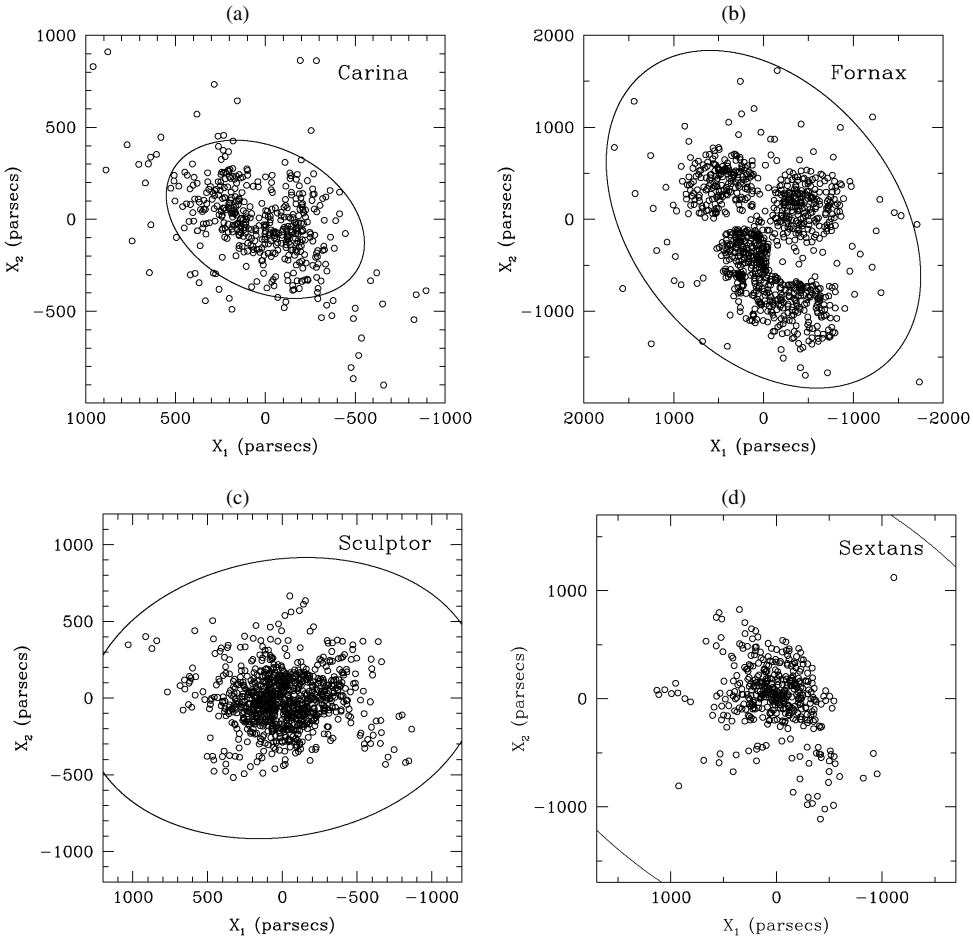

Figure 10. The projected positions of red giant stars for which we have obtained spectra and measured the line-of-sight velocity, V_3: (a) Carina, (b) Fornax, (c) Sculptor, and (d) Sextans. Ellipses indicate the approximate projected distance at which the stellar density falls to background level.

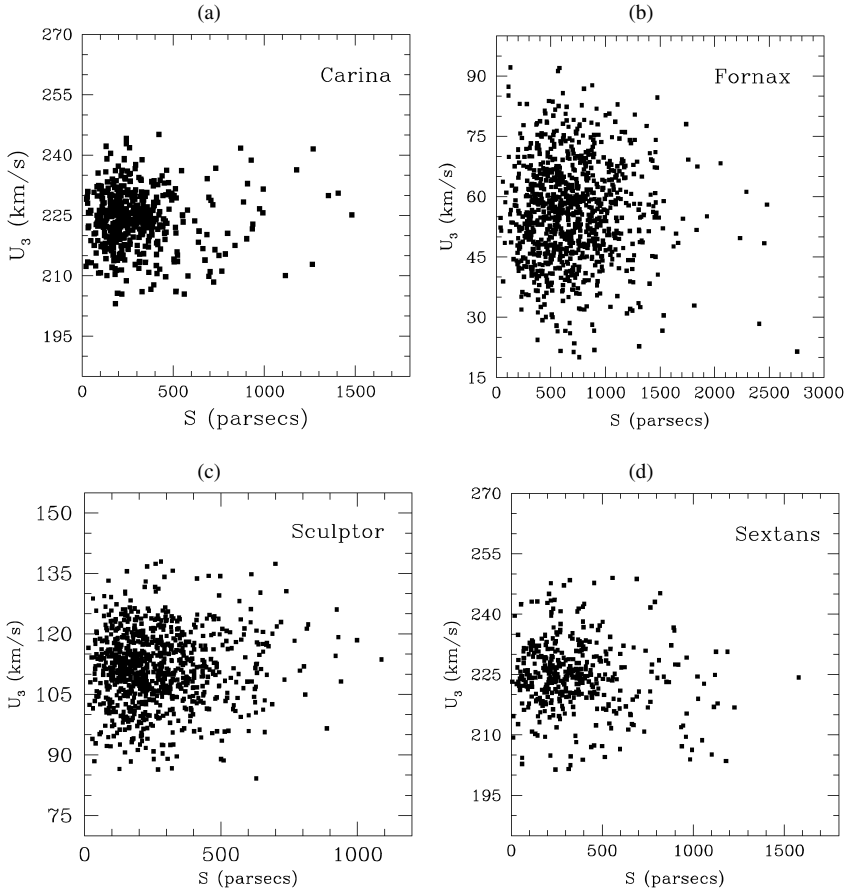

Figure 11. S, U_3 pairs in four dSph's, from data obtained at the Magellan telescopes: (a) Carina, (b) Fornax, (c) Sculptor, and (d) Sextans. Velocity measurements have median error $\sigma = 1.8$ km s^{-1} for 68% confidence intervals.

et al. (2005) is unphysical because it gives infinite central density.

Figure 12(g) indicates the mass-to-light ratio within each radius. The ratio of enclosed mass to luminosity (where luminosity is electromagnetic energy emitted per unit time), this quantity measures the prevalence of dark matter. When stated in solar units (M_{sun}/L_{sun}), $M/L \gg 1$ indicates a significant dark matter component. The M/L profiles shown in Figure 12 are calculated from our nonparametric estimates of $M(r)$ and $f(r)$, where $f(r)$ is proportional to the luminosity profile, $L(r)$, and we assume that the stars are Sun-like, having $M/L = 1$. The estimated profiles indicate that dSph kinematics are dominated by dark matter at all radii and that the proportion of dark matter increases steadily beyond $r > 500$ parsecs. We are not yet constraining the total mass of these systems; future data sets will determine whether the observed behaviors extend to even larger radii.

6.1 Comparison With Parametric Cosmological Models

A general family of parametric models for spherical mass distributions is given by the profile of Hernquist (1990), in which the mass density varies with radius as

$$\rho(r) = \rho_0 (r/r_0)^{-\gamma} \left[1 + (r/r_0)^{\alpha}\right]^{(\gamma-3)/\alpha}. \qquad (27)$$

The mass density has slope proportional to γ at small radius and proportional to -3 at large radius. The parameter α controls the sharpness of the transition. N-body simulations of massive, nonrelativistic (or "cold"), self-gravitating dark matter particles predict that dark matter halos of all mass scales follow a "universal" profile (Navarro, Frenk, and White 1997; NFW hereinafter) with $\alpha = 0$ and $\gamma = 1$. Whereas the N-body simulations are able to reproduce the observed large-scale structure (galaxy clusters and superclusters) of the universe with high success, the small galaxies generally show shallower ($\gamma \sim 0$) density profiles at small radii (e.g., McGaugh, Rubin, and de Block 2001). It is of great interest to establish whether our dSph velocity data are consistent with the "cuspy" NFW profile in which $\gamma \sim 1$ or instead favor shallower, or "cored," inner mass density profiles with $\gamma \sim 0$.

The solid curves in Figure 13 represent our nonparametric estimates of the mass profile for each dSph and identify regions of 95% confidence. Overplotted with dashed curves are profiles corresponding to the NFW models with parameters chosen to fit our empirical velocity dispersion profiles (data points in Fig. 12) in the least squares sense. We find that the best-fitting NFW models generally are more massive than our cubic spline estimates at $r \leq 100$ parsecs; nevertheless, the best-fitting NFW and nonparametric profiles all indicate approximately the same mass ($\sim 10^8 M_\odot$) at the outer limit of the data ($r \sim 1$ kiloparsec), where $M_\odot$ is the mass of the sun. We also see that the cubic spline estimates have mass density profiles that are constant interior to $r \sim 100$ parsecs. This behavior favors the presence

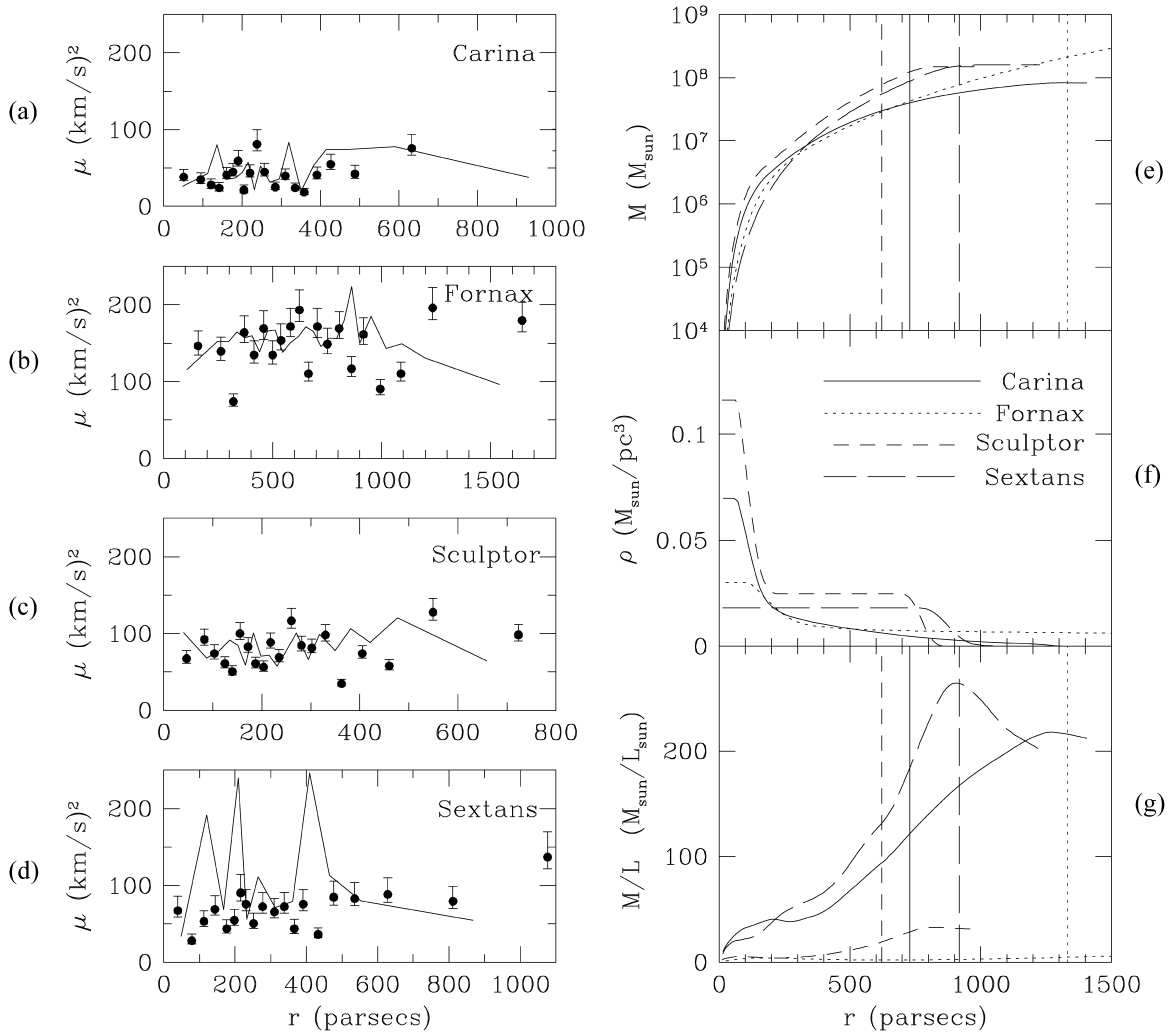

Figure 12. (a)–(d) Velocity dispersion profiles for four dSph galaxies [(a) Carina, (b) Fornax, (c) Sculptor, and (d) Sextans], calculated from real data. Data points indicate conventional parametric estimates of $\mu(r)$ (——, nonparametric estimates calculated from the cubic spline model; $\cdots$, the 95% confidence region). (e) Estimates of the mass density. (f) Estimates of the mass profile $\hat{M}(r)$ confidence intervals shown in Figure 13. (g) Cumulative mass-to-light ratio from $\hat{M}(r)$ and $\hat{f}(r)$. Values M/L $\gg 1$ indicate a dominant dark matter component.

of cores [$\gamma \sim 0$ in (27)] rather than the cusps ($\gamma \sim 1$) predicted by the NFW models.

6.2 Comparison With Quadratic Spline Estimates

Compared with the best-fitting NFW models, the cubic spline mass estimates have less mass at radii $r \leq 100$ parsecs, but their steeper slopes result in similar estimates of the masses enclosed at $r \sim 1$ kiloparsec. The mass density profiles corresponding to the cubic spline estimates have significantly shallower inner slopes, with ρ attaining a finite, constant value as $r \to 0$. To what extent does the implied disagreement of our cubic spline estimates with cosmological NFW models at small radius depend on the details of the estimation?

Figure 14 again compares nonparametric mass profiles with the best-fitting NFW profiles, but considering nonparametric estimates obtained using both the new cubic spline method and the quadratic spline method of Wang et al. (2005). Figure 14(a) displays nonparametric profiles obtained with the quadratic (dotted lines) and cubic (dashed lines) splines, as well as the NFW profiles (solid black lines). Figure 14(b) shows the difference between the profiles, $[M_{NFW}(r) - M_{nonpar}(r)]$, as a fraction of the NFW mass. The quadratic spline estimates generally agree with the NFW profiles to within a factor of 2 over most of the sampled region. At a small radius, the mass density profiles of the quadratic estimates imply $\gamma \sim 1$. This behavior makes sense if one considers that the mass density, $\rho(r) = M(r)/(4/3\pi r^3)$, is proportional to r^{-1} when $M(r) \propto r^2$ and becomes constant when $M(r) \propto r^3$. Thus the nonparametric estimate of the inner ($r \leq 100$ parsecs) mass density profile is sensitive to the order of the adopted spline. But for a given dSph, the NFW, quadratic and cubic estimates of of the mass at a large radius all agree to within a factor of 2. The outer estimates ($r > 500$ parsecs) are driven primarily by the data and provide useful constraints on dSph masses.

7. DISCUSSION

We have presented a new method for estimating the distribution of mass in a spherical galaxy. The problem of estimating the mass, $M(r)$, is converted into a problem of estimating a regression function nonparametrically subject to shape restric-

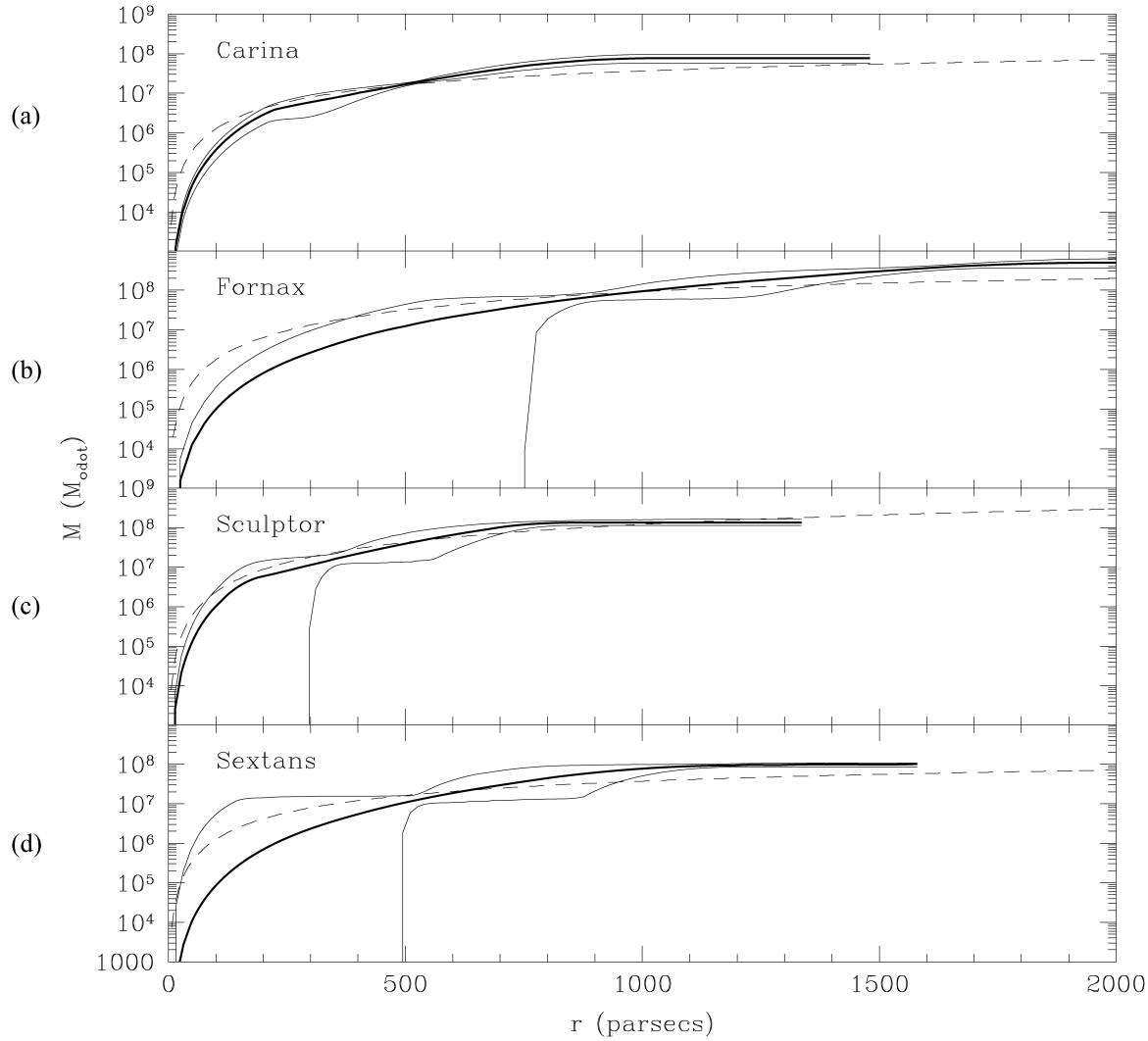

Figure 13. Comparison of nonparametric mass estimates to parametric cosmologic models: (a) Carina, (b) Fornax, (c) Sculptor, and (d) Sextans. In each panel the thick, solid line is the nonparametric mass estimate, $\hat{M}(r)$. The thin solid lines enclose the 95% confidence region. The dashed line is the mass profile of the NFW dark matter halo, with parameters chosen to best reproduce the empirical velocity dispersion profiles.

tions. The monotonicity of $r^3\Psi''(r^2)/f(r)$ in (5) provides useful information, and we model $M(r)$ by a simple cubic spline. We represent the restrictions by the interception of quadratic cones and use the SOCP to estimate the unknown parameters. The method provides the correct asymptotic behavior of $M(r)$ when r is close to 0. Our analysis indicates that dark matter dominates the stellar kinematics of these dSph systems. Wang and Woodroofe (2007) used a kernel method to estimate the velocity dispersion and then to estimate $M(r)$, but the resulting estimators do not satisfy the shape restriction. Imposing the shape restrictions in that analysis complicates the algorithm and leaves a quadratic programming problem at the end.

Much remains to be done in this area. The data that we use in this project may well be subject to various selection effects; for example, some regions of a galaxy may be sampled more or less intensively than others. In our analysis we introduce a selection function, which can be estimated if it is unknown. The introduction of a selection function suggests that it may be possible to use our formalism to provide guidance on how to select stars optimally within a galaxy to determine the underlying mass distribution. Some broader, open-ended goals are to relax the spherical symmetry assumption to include a selection function in the model and to find large-sample approximations to the sampling distributions of the estimators. Wang, Woodroofe, Pal, Walker, and Mario (2007) included an anisotropy constant in the model and extended the isotropic case to include some anisotropy. But this model is computationally complicated. Future research will focus on finding a more convenient model.

[Received June 2007. Revised October 2007.]

REFERENCES

Aaronson, M. (1983), "Accurate Radial Velocities for Carbon Stars in Draco and Ursa Minor: The First Hint of a Dwarf Spheroidal Mass-to-Light Ratio," *Astrophysical Journal Letters*, 266, 11–15.

Alizadeh, F., and Goldfarb, D. (2003), "Second-Order Cone Programming," *Mathematical Programming*, 95, 3–51.

Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Spergel, D. N., Tucker, G. S., Wollack, E., Wright, E. L., Barnes, C., Greason, M. R., Hill, R. S., Komatsu, E., Nolta, M. R., Odegard, N., Peirs, H. V., Verde, L., and Weiland, J. L. (2003), "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results," *The Astrophysical Journal Supplement Series*, 148, 1–27.

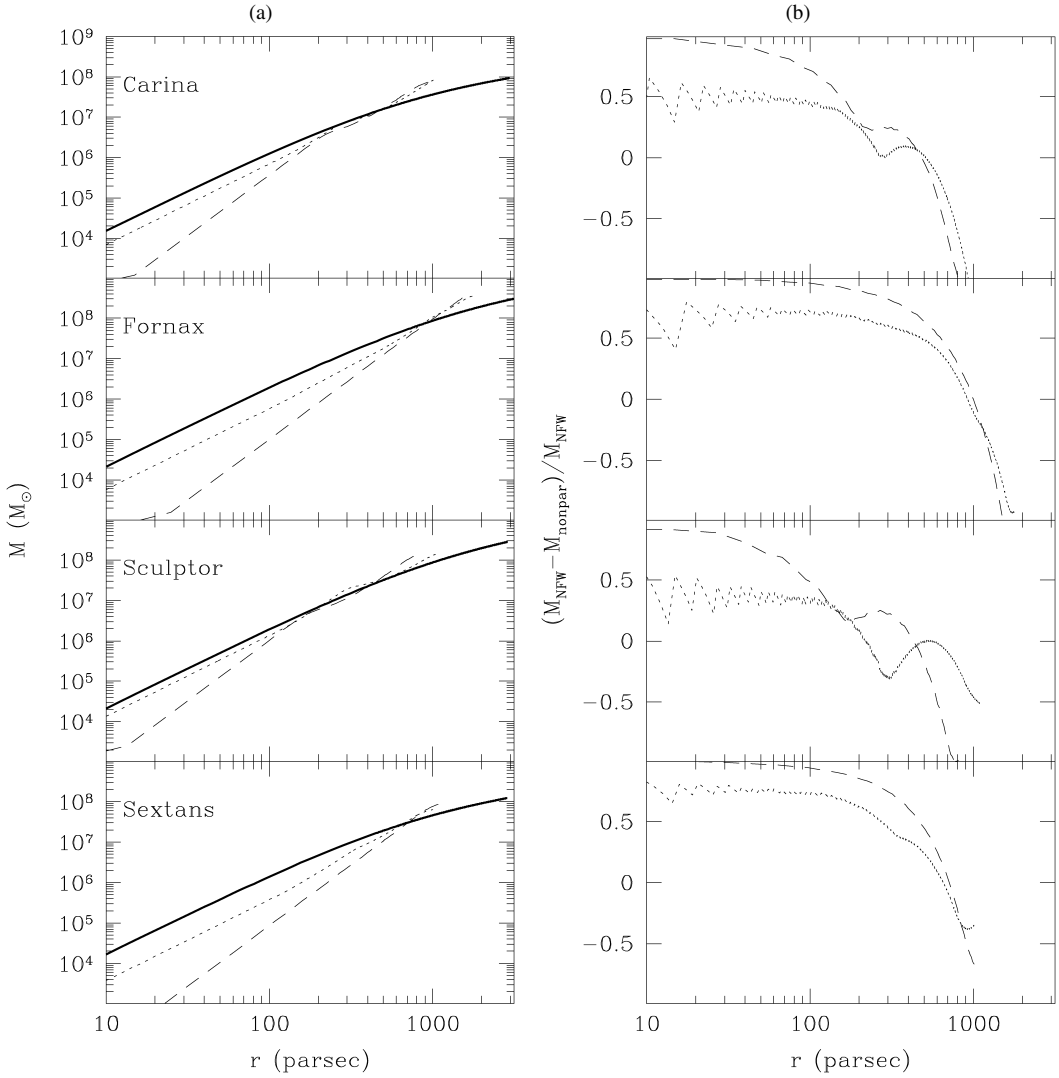

Figure 14. Comparison of parametric (motivated by cosmological simulations) and nonparametric mass profiles. (a) Best-fitting NFW profiles (——) and nonparametric mass estimates using quadratic ($\cdots$) and cubic ($- - -$) splines. (b) Mass difference between NFW and nonparametric ($\cdots$, quadratic splines; $- - - -$, cubic splines) profiles, as a fraction of the NFW mass.

Ben-Tal, A., and Nemirovski, A. (2001), *Lectures on Modern Convex Optimization*, Philadelphia, SIAM.

Binney, J., and Tremaine, S. (1987), *Galactic Dynamics*, Princeton, NJ: Princeton University Press.

Dempster, A. P., and Laird, N. M. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Groeneboom, P., and Jongbloed, G. (1995), "Isotonic Estimation and Rates of Convergence in Wicksell's Problem," *The Annals of Statistics*, 23, 1518–1542.

Hall, P., and Smith, R. L. (1988), "The Kernel Method for Unfolding Sphere Size Distributions," *Journal of Computational Physics*, 74, 409–421.

Hernquist, L. (1990), "An Analytical Model for Spherical Galaxies and Bulges," *Astrophysical Journal*, 356–359.

Irwin, M., and Hatzidimitriou, D. (1995), "Structural Parameters for the Galactic Dwarf Spheroidals," *Royal Astronomical Society*, 277, 1354–1378.

Kuo, Y.-J., and Mittelmann, H. D. (2004), "Interior Point Methods for Second-Order Cone Programming and OR Applications," *Computational Optimization and Applications*, 28, 255–285.

Lobo, M., Vandenberghe, L., Boyd, S., and Lebret, H. (1998), "Applications of Second-Order Cone Programming," *Linear Algebra and Its Applications*, 284, 193–228.

Mateo, M. (1998), "Dwarf Galaxies of the Local Group," *Annual Review of Astronomy and Astrophysics*, 36, 435–506.

Mateo, M., Olszewski, E. W., Vogt, S. S., and Keane, M. J. (1998), "The Internal Kinematics of the Leo I Dwarf Spheroidal Galaxy: Dark Matter at the Fringe of the Milky Way," *Astronomical Journal*, 116, 2315–2327.

McGaugh, S. S., Rubin, V. C., and de Block, W. J. G. (2001), "High-Resolution Rotation Curves of Low Surface Brightness Galaxies, I. Data," *The Astronomical Journal*, 122, 2381–2395.

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1997), "A Universal Density Profile From Hierarchical Clustering," *Astrophysical Journal*, 490, 493–508.

Nesterov, Y., and Todd, M. (1997), "Self-Scaled Barries and Interior Point Methods for Convex Programming," *Mathematics of Operation Research*, 22, 1–42.

Robertson, T., Wright, F., and Dykstra, R. (1987), *Order-Restricted Statistical Inference*, New York: Wiley.

Shapiro, A. (1989), "Asymptotic Properties of Statistical Estimators in Stochastic Programming," *The Annals of Statistics*, 17, 841–858.

Tsuchiya, T. (1999), "A Convergence Analysis of the Scaling-Invariant Primal-Dual Path-Following Algorithms for Second-Order Cone Programming," *Optimization Methods and Software*, 11/12, 141–182.

Walker, M. G., Mateo, M., Olszewski, E. W., Bernstein, R. A., Wang, X., and Woodroofe, M. (2006), "Internal Kinematics of the Fornax Dwarf Spheroidal Galaxy," *Astronomical Journal*, 131, 2114–2139.

Wang, X., and Li, F. (2008), "Isotonic Smoothing Spline Regression," *Journal of Computational and Graphical Statistics*, 17, 21–37.

Wang, X., and Woodroofe, M. (2007), "Kiefer–Wolfowitz Comparison Theorem for Wichsell's Problem," *The Annals of Statistics*, 35, 1559–1575.

Wang, X., Woodroofe, M., Pal, J., Walker, M., and Mario, M. (2007), "Nonparametric Estimation of Dark Matter Distributions," in *Statistical Challenges in Modern Astronomy IV*, Vol. 371, eds. G. J. Babu and E. D. Feigelson, Provo, UT: Astronomical Society of the Pacific, pp. 268–279.

Wang, X., Woodroofe, M., Walker, M., Mateo, M., and Olszewski, E. (2005), "Estimating Dark Matter Distributions," *Astrophysical Journal*, 626, 145–158.

Wicksell, S. D. (1925), "The Corpuscle Problem," *Biometrika*, 17, 84–99.

Woodroofe, M. B., and Sun, J. (1993), "A Penalized Maximum Likelihood Estimate of f(0+) When f Is Nonincreasing," *Statistica Sinica*, 3, 501–515.

Wu, C. F. J. (1983), "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, 11, 95–103.

Zwicky, F. (1933), "Die Rotverschiebung von Extragalaktischen Nebeln," *Helvetica Physica Acta*, 6, 110–127.

——— (1937), "On the Masses of Nebulae and of Clusters of Nebulae," *Astrophysical Journal*, 86, 217.