The Intrinsic Distribution and Selection Bias of Long-Period Cometary Orbits

Peter E. JUPP, Peter T. KIM, Ja-Yong KOO, and Paul WIEGERT

A question that arises in the study of cometary orbits is whether or not the directed normals to the orbits are uniformly distributed on the celestial sphere. Previous studies by statisticians have not taken selection effects into account and have tended to reject uniformity. Here a plausible selection mechanism is proposed that gives rise to a one-parameter family of distributions on the sphere. Data on long-period comets are analyzed using this one-parameter family. A nonzero selection effect is detected, and its size is estimated. Subject to this selection effect, uniformity of the directed normals can no longer be ruled out.

KEY WORDS: Directed normal; Quantile–quantile plot; Spherical data; Spherical uniform distribution.

1. INTRODUCTION

The orbits of periodic comets are elliptical, with the sun at one focus. Two important features of a cometary orbit are the *perihelion direction* (the direction from the sun to the point of closest approach of the comet) and the *directed normal* (the direction normal to the orbit, the sense of rotation of the comet being taken into account by, say, a right-hand rule). Note that the perihelion direction and the directed normal are both directions, that is, unit vectors. There are astronomical reasons for believing that observations of cometary orbits are subject to considerable selection effects. Thus interest lies in the *intrinsic distribution* of the orbits, that is, the distribution prior to selection by any observational biases. The aim of this article is to investigate the intrinsic distribution of the directed normals, and, in particular, to test it for uniformity.

Among investigators into the distribution of cometary orbits by statisticians, Mardia (1975) considered the perihelion directions; Watson (1983, pp. 28–32), Hall, Watson, and Cabrera (1987), and Fisher, Lewis, and Embleton (1993, p. 161) considered the directed normals; and Jupp and Mardia (1979), Jupp and Spurr (1983), Jupp (1995), and Mardia and Jupp (2000, pp. 283, 288–289) considered the joint distribution of the directed normals and the perihelion directions. We concentrate here on the directed normals.

One of the earlier in-depth statistical investigations in the astronomical literature was made by Tyror (1957), who studied the distribution of long-period perihelion directions. Later studies by Everhart (1967a,b) and Kresák (1982) gave detailed discussions of possible sources of observational bias, to obtain information about the intrinsic distribution of cometary orbits. However, these authors did not provide an explicit statistical model for selection effects. An extensive investigation and more up-to-date analyses from the astronomical stand-point were made by Wiegert and Tremaine (1999; see also the references cited therein).

It is thought that all comets were formed within the solar system and were subsequently flung out into a large disc by the gravitational action of the planets. This disc was coplanar with the planets, but extended out to tens of thousands of astronomical units. (An astronomical unit is defined as the mean distance from the center of the earth to the center of the sun.) It is thought that comets with large orbits (and hence long periods) then had the planes of their orbits changed randomly by the gravitational effect of passing stars (Weissman 1980). The resulting set of comets constitutes the *Oort cloud* (Oort 1950), a spherical region of space extending out around the solar system to a distance hundreds of times further from the sun than the orbit of Pluto. It is believed that long-period comets may reside in the Oort cloud (or other "reservoirs" in the solar system) for billions of years before becoming visible from the neighborhood of the earth. For astronomical reasons given by Wiegert and Tremaine (1999, p. 84), the best source of information on the Oort cloud is the set of long-period single-apparition comets, and thus these comets have been the subject of a great deal of study. We therefore restrict our attention to this set of comets.

In this article we give a more careful statistical analysis than has been given before, concentrating on the directed unit normals and introducing a plausible selection mechanism that gives rise to a one-parameter family of distributions. The directed normals are observations on S^2, the unit sphere in 3-space. Because the comets are part of the solar system and are affected by the gravitational attraction of the sun and giant planets, it is sensible to use the (heliocentric) *ecliptic coordinate system*, which is indicated in Figure 1. The origin of this system is the sun. The first axis is aligned (traditionally, among astronomers, but arbitrarily) parallel to the *vernal equinox*, which is the direction from the earth to the sun at the spring equinox. This direction lies in the intersection of the *ecliptic plane* (the plane of the earth's orbit round the sun) with the equatorial plane of the earth. The second coordinate axis is orthogonal to the first and lies in the ecliptic plane. The third axis is perpendicular to the first two. *Ecliptic longitude* is measured in the ecliptic plane eastwards from the vernal equinox. *Ecliptic colatitude* is measured from the celestial north pole toward the ecliptic plane.

In Section 2 we carry out an exploratory analysis of the directed normals in terms of ecliptic colatitudes and longitudes.

Peter E. Jupp is Reader, School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews, KY16 9SS U.K. (E-mail: *pej@st-andrews.ac.uk*). Peter T. Kim is Professor, Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, N1G 2W1 Canada (E-mail: *pkim@uoguelph.ca*). Ja-Yong Koo is Professor, Department of Statistics, Inha University, Incheon 402-751, Korea (E-mail: *jykoo@stat.inha.ac.kr*). Paul Wiegert is Research Associate, Department of Physics, Queen's University, Kingston, Ontario, K7L 3N6 Canada (E-mail: *wiegert@astro.queensu.ca*). The authors would like to thank the editor, the associate editor, and the anonymous referees for their careful reviews of this manuscript. In addition, the authors thank Herb Kunze of the Department of Mathematics and Statistics, University of Guelph, for his help in calculating the normalizing constant. Parts of this research have been supported by NSERC (Canada) OGP46204 and by KOSEF (Korea) through The Statistical Research Center for Complex Systems at Seoul National University.

Journal of the American Statistical Association
September 2003, Vol. 98, No. 463, Applications and Case Studies
DOI 10.1198/016214503000000305

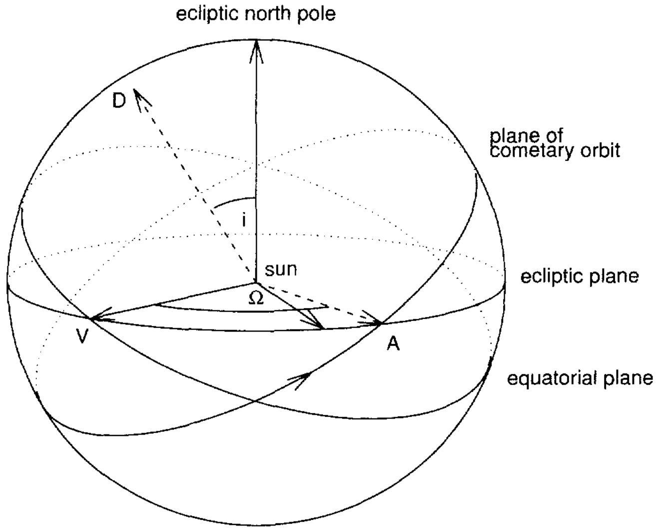

Figure 1. Ecliptic Coordinate System. The origin is at the sun, and the vernal equinox is indicated by V. The solid arrows are the unit coordinate vectors. The point D on the celestial sphere is at the end of the directed unit normal (dashed) to a cometary orbit. The arrow on the great circle given by the plane of the orbit indicates the sense of rotation. The ascending node of the orbit is indicated by A. The angles i and Ω are the inclination of the orbit and the longitude of A.

Equal-area and quantile–quantile (Q–Q) plots indicate that, although longitudinal symmetry is present, the observed distribution of colatitude is inconsistent with the distribution of colatitude given by the spherical uniform distribution or a Watson distribution. In Section 3 these exploratory findings are confirmed by some well-known tests. Q–Q plots of the colatitudes of the directed normals reveal that the distribution of the colatitudes of the directed normals has heavier tails than the distribution of the colatitudes obtained from the spherical uniform distribution but lighter tails than the uniform distribution on $[0, \pi)$. This is consistent with the way in which cometary astronomers sample the sky, because they tend to look mainly near the ecliptic. This motivates a model for the selection process involving a one-parameter family of distributions on the sphere. In Section 4 we estimate the parameter by maximum likelihood. Informal and formal analyses in Section 5 show that, subject to selection according to our model, the data are consistent with a uniform intrinsic distribution.

2. EXPLORATORY DATA ANALYSIS

The data come from the 658 single-apparition long-period cometary orbits found in the catalog of Marsden and Williams (1993). Of these, 315 are *prograde* (i.e., revolve in the same direction as the earth about the ecliptic polar axis, so that the directed normal points to the north of the ecliptic plane), and 343 are *retrograde* (i.e., revolve in the opposite direction, so that the directed normal points to the south of the ecliptic plane).

Catalogs of cometary orbits usually specify each orbit in terms of six measurements. important measurements are the inclination and the longitude of the ascending node. The *inclination*, i, is the angle between the plane of the comet's orbit and the ecliptic plane, with $i = 0$ corresponding to prograde motion in the ecliptic plane (and $i = \pi$ corresponding to retrograde motion in the ecliptic plane). The *longitude of the ascending node*, Ω, is the ecliptic longitude of the point at which the comet crosses the ecliptic in the "upward" direction (toward the ecliptic north pole). The directed normal to the plane of a cometary orbit is given by

$$(\sin i \sin \Omega, -\sin i \cos \Omega, \cos i)',$$

where the superscript "′" denotes transpose. Recently, Wiegert and Tremaine (1999, pp. 87–88) performed some elementary data analysis on i and Ω. In a statistical context, it is useful to transform the coordinates (i, Ω) to the standard spherical polar coordinates (θ, ϕ) on S^2 by

$$\theta = i, \qquad \phi = \Omega - \frac{\pi}{2} \pmod{2\pi},$$

where $\theta \in [0, \pi)$ and $\phi \in [0, 2\pi)$. Thus θ is the *colatitude* measured from the north pole $(0, 0, 1)'$ of S^2, ϕ is the *longitude*, and the directed normal to the plane of the corresponding cometary orbit is

$$(\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)'.$$

Figure 2 gives histograms of these polar coordinates. The histograms reveal some nonuniformity, particularly in the colatitude. This is studied in greater detail in Sections 2.3 and 3.1.

2.1 Equal Area Projection

One method of representing data on S^2 is via an *equal-area projection* onto a disc. The importance of an equal-area projection is that it maps the (spherical) uniform distribution on S^2 to the (ordinary) uniform distribution on the disc. A standard equal-area projection of the unit sphere onto a disc of radius 2 is Lambert's equal-area projection (Watson 1983, p. 21; Mardia and Jupp 2000, pp. 160–151), which maps the point $(\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)'$ in S^2 to the point

$$2\sin(\theta/2)(\cos\phi, \sin\phi)$$

in the plane, where $\theta \in [0, \pi)$ and $\phi \in [0, 2\pi)$. This projection sends the "north pole" to the origin. An alternative equal-area projection maps $(\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)'$ in S^2 to

$$2\sin((\pi - \theta)/2)(\cos\phi, \sin\phi)$$

and so sends the "south pole" to the origin. In the former (latter) case, the upper (lower) hemisphere of S^2 gets mapped into a disc of radius $\sqrt{2}$, and the lower (upper) hemisphere into an annulus with inner and outer radii $\sqrt{2}$ and 2.

Figure 3 plots equal-area projections of the 658 directed normals. Panel (a) is a plot with the north ecliptic pole at the center, and panel (b) is a plot with the south ecliptic pole at the center. Inspection of these plots suggests some bunching near the ecliptic poles, with almost uniform distribution in between.

2.2 Longitudinal Symmetry

Initial assessment of the degree of longitudinal (i.e., circular) symmetry can be carried out graphically. Figure 4(a) uses a Q–Q plot to compare the observed longitude ϕ with that of the uniform distribution on $[0, 2\pi)$. The plot indicates that circular symmetry in ϕ is reasonable, in agreement with the conclusions of Jupp and Spurr (1983), Jupp (1995), and Wiegert and Tremaine (1999, p. 88). In Section 3 we perform a formal test to confirm this.

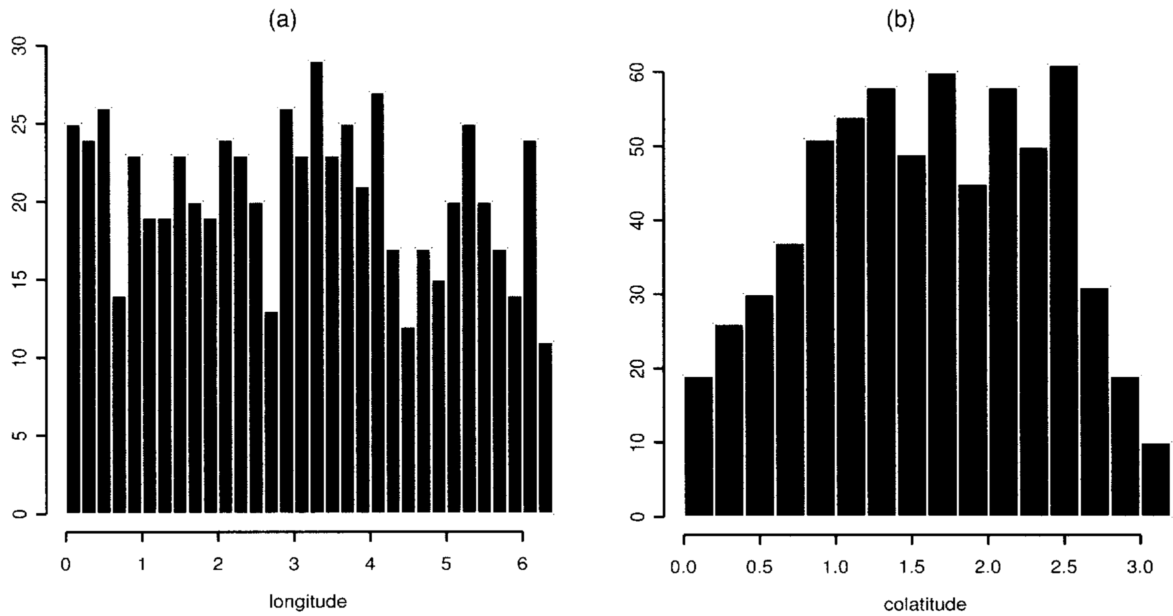

Figure 2. Histograms of Polar Coordinates in Radians: (a) Longitude; (b) Colatitude.

2.3 Classical Models

Traditional parametric analysis of spherical data often involves using the von Mises–Fisher and/or the Bingham distributions. The equal-area plots in Figure 3 suggest that the data have antipodal symmetry. Support for this comes from the similarity of the numbers of observations in prograde and retrograde orbits, together with the circular symmetry of the distribution of longitude. This suggests that it might be appropriate to fit a Bingham distribution. The circular symmetry of the longitude and the clustering near the ecliptic poles indicate that a bipolar Watson distribution (cf. Mardia and Jupp 2000, p. 181) would be a reasonable choice.

If the data come from a bipolar Watson distribution, then $1 - \cos^2\theta$ is distributed approximately as an exponential distribution (cf. Fisher et al. 1993, p. 168). Nevertheless, as displayed graphically in Figure 4(b), the non-linearity in the Q–Q plot of the colatitude strongly suggests a lack of fit. Furthermore, the Q–Q plot in Figure 4(c) of the observed colatitudes of the directed normals against the colatitudes of the spherical uniform distribution, which has density $\sin\theta$ on $[0, \pi)$, indicates a departure from spherical uniformity. In Section 3 we confirm this by a formal test.

The main conclusion of this exploratory data analysis is that there is evidence of spherical nonuniformity, longitudinal uniformity, and clustering of the data at the ecliptic poles, but that a bipolar Watson distribution does not provide a good fit to the data.

3. SELECTION EFFECTS

Formal testing of the long-period comet data supports the findings of Section 2. For testing spherical uniformity, we use

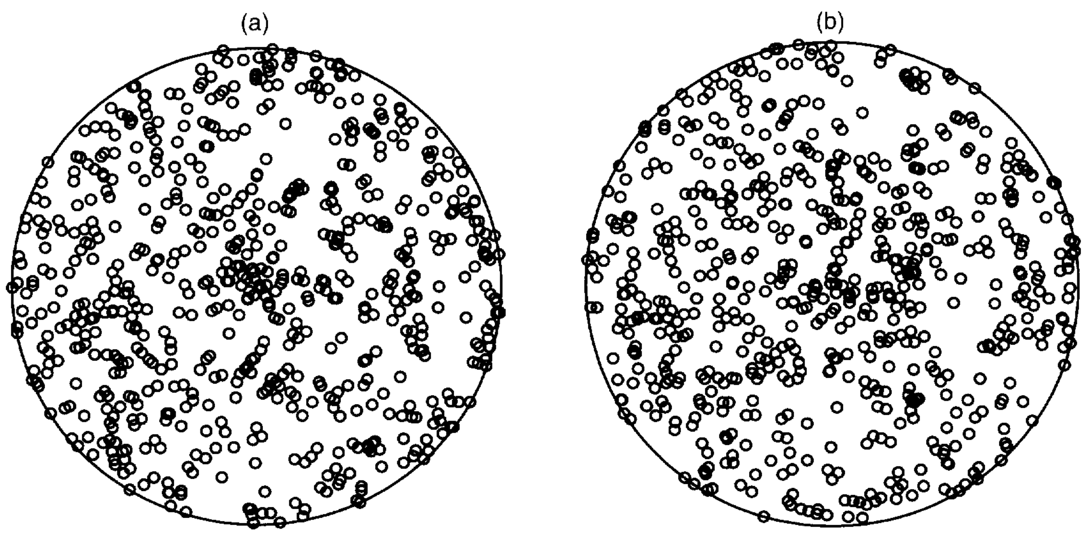

Figure 3. Equal-Area Projections of Directed Normals. (a) View from the north ecliptic pole; (b) view from the south ecliptic pole. Both plots are images of the complete sphere.

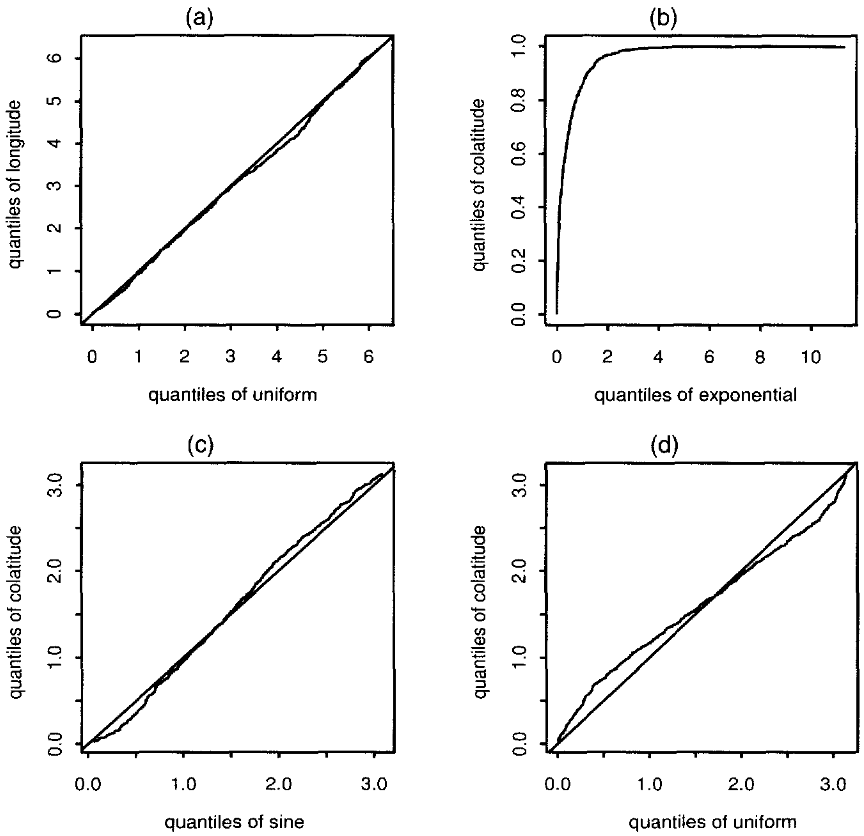

Figure 4. Quantile Plots of Colatitude (θ) and Longitude (ϕ) of the Directed Normals. (a) Longitude against the uniform distribution on $[0, 2\pi)$; (b) $1 - \cos^2\theta$ against an exponential distribution; (c) colatitude against $\sin(\theta)$ (from the spherical uniform distribution); (d) colatitude against the uniform distribution on $[0, \pi)$.

Bingham's (1974) test. Let

$$M_n = \frac{1}{n}\sum_{j=1}^{n} X_j X_j',$$

where $X_i = (\cos\phi_i \sin\theta_i, \sin\phi_i \sin\theta_i, \cos\theta_i)'$ for $i = 1, \ldots, n$. Then, under uniformity,

$$7.5n\{\operatorname{tr} M_n^2 - (2/3)\operatorname{tr} M_n + 1/3\} \sim \chi_5^2 \qquad \text{as } n \to \infty,$$

where χ_ν^2 denotes a chi-squared distribution with ν degrees of freedom and tr denotes the trace of a square matrix (see Giné 1975, pp. 1263–1264; Mardia and Jupp 2000, p. 232). For the comet data,

$$7.5 \times 658\{\operatorname{tr} M_{658}^2 - (2/3)\operatorname{tr} M_{658} + 1/3\} = 29.520$$

and

$$P(\chi_5^2 > 29.520) = .0001,$$

so that, if complications due to sampling bias and other factors are ignored, the null hypothesis of spherical uniformity can be rejected soundly. Note that because the equal-area plots in Figure 3 suggest that the data have antipodal symmetry, it would not be sensible to test uniformity using the Rayleigh test (cf. Mardia and Jupp 2000, p. 207).

To test longitudinal symmetry (i.e., uniformity of ϕ), we use Watson's U^2 test, a test of circular uniformity that is consistent against all alternatives (cf. Mardia and Jupp 2000, pp. 103–105). Applying this test to the longitudes of the directed normals yields

$$U^2 = .073.$$

Because $P(U^2 > .073) \approx .464$ [from Watson's 1961 large-sample asymptotic result quoted as (6.3.37) of Mardia and Jupp 2000, p. 104], longitudinal symmetry cannot be rejected at any reasonable significance level.

3.1 A Selection Mechanism

Inspection of the quantile plot of observed colatitude against the uniform distribution on $[0, \pi)$ displayed in Figure 4(d) shows that the middle portion of the quantile plot is almost linear. Furthermore, comparison with the quantiles of colatitude obtained from the spherical uniform distribution in Figure 4(c) suggests that the distribution of the directed normals lies "between" the uniform distribution on $[0, \pi)$ and the distribution of colatitude obtained from the spherical uniform distribution. This provides a clue for modeling the distribution of the directed normals of the long-period comets.

Comet-seekers do not sample the entire sky evenly when looking for comets. They bias their observations toward the ecliptic plane. Smaller additional biases may arise from the particular properties of individual comets, which may make them easier or harder to detect. For example, a comet that reaches peak brightness when in the daytime sky is much less likely to be observed than one that does so in the nighttime sky. Thus the empirical distribution of long-period single-apparition comets reflects certain biases inherent in comet surveys themselves. Comet seekers searching the solar system for small bodies typically do not point their telescopes randomly into the sky, rather, they bias their search toward an "observational window." The chances of success of any search program are enhanced by looking near the ecliptic, because most asteroids and comets with shorter periods lie near this plane. Although it is not thought that the orbits of long-period single-apparition comets are clustered near the ecliptic, those comets of this kind that spend more time in the vicinity of this observational window are more likely to be discovered, because most search effort will be concentrated there. As a result, it is expected that the observed sample of long-period comets will be biased toward those comets with orbits near the ecliptic, and so with directed normals lying near the ecliptic poles.

As a simple way of quantifying the foregoing notion, we introduce, for each ε in [0, 1], the corresponding *observational window*, A_ε, defined by

$$A_\varepsilon = \{\omega = (\omega_1, \omega_2, \omega_3)' \in S^2 : |\omega_3| \leq \varepsilon\}.$$

Thus A_ε is the band between the two small circles,

$$\left\{\left(\sqrt{1-\varepsilon^2}\cos\phi, \sqrt{1-\varepsilon^2}\sin\phi, \varepsilon\right)' : 0 \leq \phi \leq 2\pi\right\}$$

and

$$\left\{\left(-\sqrt{1-\varepsilon^2}\cos\phi, -\sqrt{1-\varepsilon^2}\sin\phi, -\varepsilon\right)' : 0 \leq \phi \leq 2\pi\right\},$$

which lie on either side of the ecliptic. Note that when $\varepsilon = 1$, the observational window is the entire sphere.

A plausible model for the probability $g(\theta, \phi; \varepsilon)$ that a comet with directed normal at $(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)'$ is observed assumes that this selection probability is proportional to the arc length of the portion of the orbit that is inside the observational window A_ε. This leads to

$$g(\theta, \phi; \varepsilon) = \begin{cases} 1 & \text{if } \sin\theta \leq \varepsilon \\ \dfrac{\arcsin(\varepsilon/\sin\theta)}{\pi/2} & \text{if } \sin\theta > \varepsilon. \end{cases} \quad (1)$$

To see this, consider Figure 5. Let L denote the length of the portion of the cometary orbit that falls within the observational window. Then $L = 4AC$. Furthermore, because the observational window A_ε is defined by $|\sin\theta| \leq \varepsilon$, we have that $BC = \arcsin\varepsilon$. Application of the spherical sine rule to the spherical triangle ABC gives

$$\frac{\sin(L/4)}{\sin(\pi/2)} = \frac{\sin(\arcsin\varepsilon)}{\sin(\theta)},$$

so that

$$\frac{L}{4} = \arcsin\left(\frac{\varepsilon}{\sin\theta}\right).$$

Thus the selection probability is given by

$$\frac{L}{2\pi} = \frac{2}{\pi}\arcsin\left(\frac{\varepsilon}{\sin\theta}\right).$$

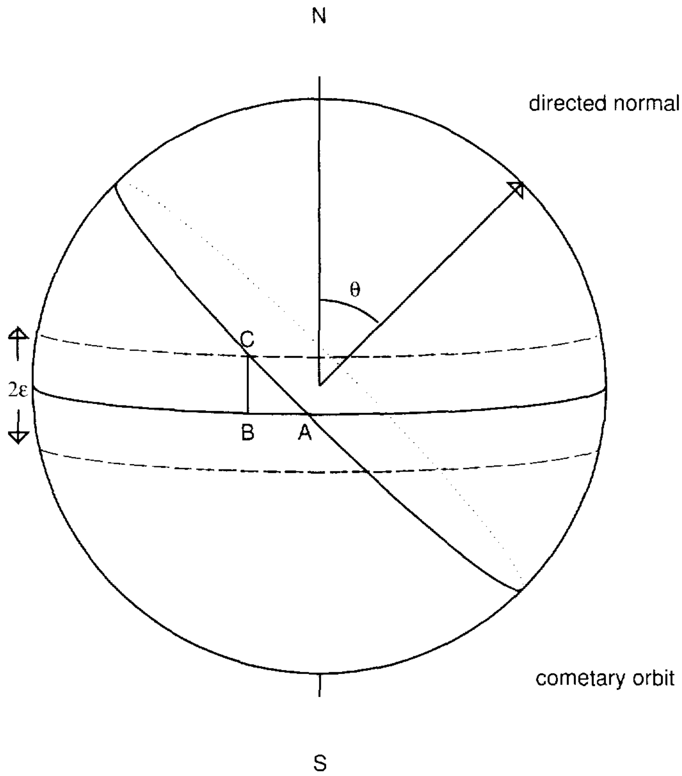

Figure 5. The Length of the Portion of the Path of a Cometary Orbit Inside the Observational Window A_ε of Width 2ε. The points N and S are the north and south ecliptic poles.

Let $f(\theta, \phi)$ denote the probability density with respect to the uniform distribution [i.e., with respect to $(4\pi)^{-1}\sin\theta d\theta d\phi$] of the intrinsic distribution of the directed normals,

$$(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)'.$$

Then the probability density (with respect to the spherical uniform distribution) of the directed normals that are observed is

$$h(\theta, \phi; \varepsilon) = c(\varepsilon)f(\theta, \phi)g(\theta, \phi; \varepsilon), \quad (2)$$

where $c(\varepsilon)$ is a normalizing constant.

Note that (2) implies that the functions $f(\theta, \phi)$ and $g(\theta, \phi; \varepsilon)$ are not jointly identifiable, so that it is not possible to infer both the density $f(\theta, \phi)$ of the intrinsic distribution and the probabilities $g(\theta, \phi; \varepsilon)$ of selection. Nevertheless, if suitable assumptions are made about one of $f(\theta, \phi)$ and $g(\theta, \phi; \varepsilon)$ then it is possible to make inference about the other. In particular, if the intrinsic distribution of directed normals is uniform on the celestial sphere then (2) reduces to

$$h(\theta, \phi; \varepsilon) = \begin{cases} \varepsilon^{-1} & \text{if } \sin\theta \leq \varepsilon \\ \dfrac{2}{\pi\varepsilon}\arcsin\left(\dfrac{\varepsilon}{\sin\theta}\right) & \text{if } \sin\theta > \varepsilon. \end{cases} \quad (3)$$

The fact that, in this case, $c(\varepsilon) = \varepsilon^{-1}$ can be obtained by calculating the derivative of $1/c(\varepsilon)$ with respect to ε and showing, through substitution, that it is unity. Note that when $\varepsilon = 1$ the density (3) reduces to the density of the uniform distribution on the sphere.

It might be argued that our model (2) for the selection probabilities is too simple to be very realistic, first because no account has been taken of the fact that (in accordance with Kepler's second law) the angular velocity of a comet about the earth is not constant (an effect that is particularly pronounced for orbits of high eccentricity), and second because one might expect there to be a mixture over ε of windows A_ε. However, models taking these aspects into account would be considerably more complicated than our simple one-parameter model. In Section 5 we show that our model for the observed distribution [based on intrinsic uniformity together with selection probabilities (2)] fits the data very well. It therefore does not seem worth considering more-complicated models. It should also be mentioned that the observed sample of single-apparition long-period comets has been obtained by a variety of astronomers over a span of centuries, and with no single underlying observational plan.

4. MAXIMUM LIKELIHOOD ESTIMATION OF WINDOW SIZE

Because ε, the window size, is unknown, it must be estimated. One way to do so is by maximum likelihood. From (3), the log-likelihood function of ε based on unit vectors $X_1, \ldots, X_n$ is

$$l(\varepsilon; X_1, \ldots, X_n) = -n \log \varepsilon + \sum_{\sin\theta_j \geq \varepsilon} \log\left(\frac{2}{\pi} \arcsin \frac{\varepsilon}{\sin\theta_j}\right). \tag{4}$$

The maximum likelihood estimate $\hat{\varepsilon}$ of ε is $\hat{\varepsilon} = .84$. Figure 6 plots the log-likelihood function. The interpretation of the maximum likelihood estimate $\hat{\varepsilon} = .84$ in terms of the observational window A_ε is that the colatitude of orbits that can be completely observed ranges from 57° to 123°.

4.1 Bootstrap Confidence Interval

Because the log-likelihood (4) is not differentiable, standard likelihood-based confidence intervals for ε are not appropriate. Consequently, confidence intervals for ε were obtained by bootstrapping, using both percentile bootstrapping and bias-corrected accelerated bootstrapping. (The theory of such intervals can be found in Efron and Tibshirani 1993, chaps. 13–14.) A grid size of 1,000 was used for the interval (0, 1], and the resampling size for the bootstrap was 2,000. Table 1 summarizes the 90% and 95% bootstrap confidence intervals.

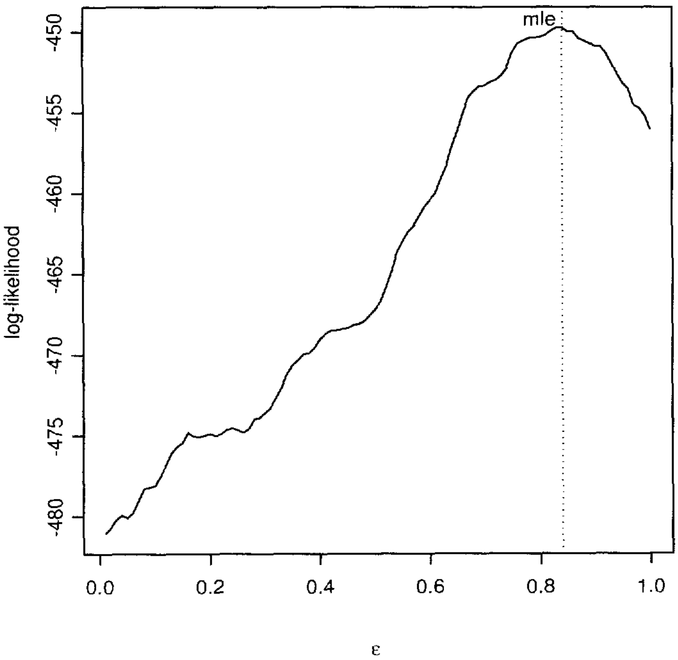

Figure 6. The Log-Likelihood Function in Terms of the Window Size ε.

Table 1. Bootstrap Confidence Intervals for ε

Method	90%	95%
Percentile	(.675,.985)	(.673,.988)
Bias-corrected	(.692,.988)	(.675,.999)

Table 1 shows that the two methods of bootstrapping provide very similar answers. The interesting point is that none of these intervals contains 1.0. Further, an independent Monte Carlo test (using 999 replicate samples) of $\varepsilon = 1$ gave a p value of .002. These results strongly suggest that selection bias exists.

5. THE INTRINSIC DISTRIBUTION OF LONG-PERIOD COMETS

An informal assessment of the goodness of fit of the data to the fitted distribution (3) with $\hat{\varepsilon} = .84$ is given by the plot in Figure 7(a) of the fitted cumulative distribution function of θ. This plot shows also the cumulative distribution functions of the uniform distribution on $[0, \pi)$ (straight line) and of the colatitudes of the spherical uniform distribution (outer curve). Note that the fitted cumulative marginal distribution function of the colatitude resembles that of the colatitudes from the spherical uniform distribution for values near 0 and π, whereas in the middle portion of the domain, the fitted cumulative distribution function is almost linear. Note also that the empirical cumulative distribution function of colatitude lies between the cumulative distribution function of the uniform distribution on $[0, \pi)$ and that of the colatitude obtained from the spherical uniform distribution. Thus the fitted distribution (3) with $\hat{\varepsilon} = .84$ can be regarded as combining features of these two distributions. Figure 7(b) is a Q–Q plot of the empirical quantiles of colatitude against the quantiles of (3) with $\hat{\varepsilon} = .84$. The fit is much better than those to the uniform distribution on $[0, \pi)$ and to the distribution of colatitude obtained from the spherical uniform distribution, which are shown in the Q–Q plots displayed in Figures 4(c) and 4(d). These impressions will now be confirmed by formal goodness-of-fit tests.

We use Watson's U^2 to assess goodness of fit of the colatitude data to three distributions: the uniform distribution on $[0, \pi)$, the distribution of colatitude obtained from the spherical uniform distribution, and the fitted distribution with density (3) and $\hat{\varepsilon} = .84$. The findings are summarized in Table 2. The p values [from Watson's 1961 large-sample asymptotic result, quoted as (6.3.37) of Mardia and Jupp 2000, p. 104] for the uniform distribution on $[0, \pi)$ and the distribution of the colatitudes from the spherical uniform distribution are both negligible. On the

Table 2. U^2 Goodness-of-Fit Test

	Distribution		
	Uniform on $[0, \pi)$	Spherical uniform	(3) with $\hat{\varepsilon} = .84$
U^2	1.79	.52	.12
p value	.00	.00	.18

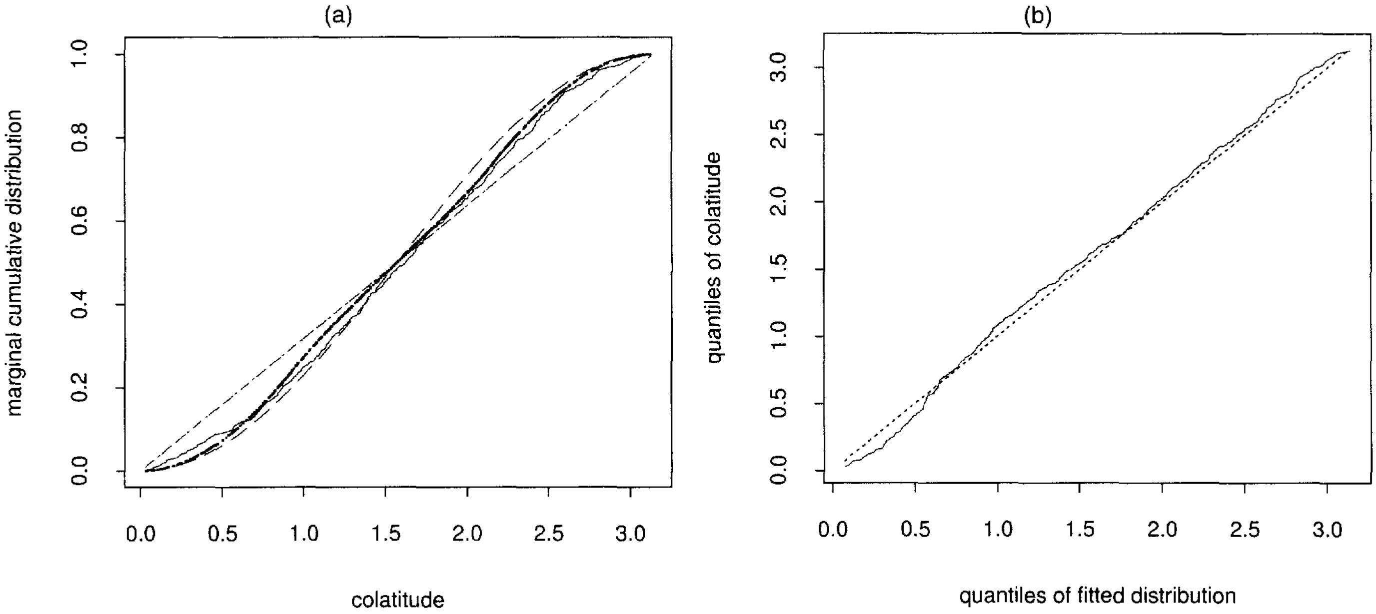

Figure 7. (a) Cumulative Marginal Distribution Function for θ With $\varepsilon = .84$ and (b) Q–Q Plot of the Cumulative Marginal Distribution Function With $\varepsilon = .84$ Against Empirical Quantiles. In (a), the outer lines refer to the uniform and spherical uniform distributions, and the connected line indicates the empirical probabilities.

other hand, the fitted distribution given by the density (3) with $\hat{\varepsilon} = .84$ is a good fit to the observed colatitudes. Combining this test of goodness of fit of the colatitudes of the fitted distribution with the U^2 test of longitudinal symmetry considered in Section 3 gives a goodness-of-fit test of the fitted distribution on the sphere. For our data, the combined p value is .29. Thus, by taking selection into account according to (3) with $\hat{\varepsilon} = .84$, the data are consistent with a uniform intrinsic distribution.

[Received April 2002. Revised May 2003.]

REFERENCES

Bingham, C. (1974), "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics*, 2, 1201–1225.

Efron, B., and Tibshirani, R. (1993), *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Everhart, E. (1967a), "Comet Discoveries and Observational Selection," *Astronomical Journal*, 72, 716–726.

——— (1967b), "Intrinsic Distributions of Cometary Perihelia and Magnitudes," *Astronomical Journal*, 72, 1002–1011.

Fisher, N. I., Lewis, T., and Embleton, B. J. (1993), *Statistical Analysis of Spherical Data*. Cambridge, U.K.: Cambridge University Press.

Giné, E. (1975), "Invariant Tests for Uniformity on Compact Riemannian Manifolds Based on Sobolev Norms," *The Annals of Statistics*, 3, 1243–1266.

Hall, P., Watson, G., and Cabrera, J. (1987), "Kernel Density Estimation With Spherical Data," *Biometrika* 74, 751–762.

Jupp, P. E. (1995), "Some Applications of Directional Statistics to Astronomy," in *New Trends in Probability and Statistics, Vol. 3*, eds. E. M. Tiit et al. Utrecht: VSP, 123–133.

Jupp, P. E., and Mardia, K. V. (1979), "Maximum Likelihood Estimation for the Matrix von Mises–Fisher and Bingham Distributions," *The Annals of Statistics*, 7, 599–606.

Jupp, P. E., and Spurr, B. D. (1983), "Sobolev Tests for Symmetry of Directional Data," *The Annals of Statistics*, 11, 1225–1231.

Kresák, Ĺ. (1982), "Comet Discoveries, Statistics and Observational Selection," in *Comets*, ed. L. Wilkening. Tucson, AZ: University of Arizona Press, pp. 56–82.

Mardia, K. V. (1975), "Statistics of Directional Data" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 37, 349–393.

Mardia, K. V., and Jupp, P. E. (2000), *Directional Statistics*. Chichester, U.K.: Wiley.

Marsden, B. G., and Williams, G. V. (1993), *Catalogue of Cometary Orbits* (8th ed.). Cambridge, MA: Minor Planet Center, Smithsonian Astrophysical Observatory.

Oort, J. H. (1950), "The Structure of the Cloud of Comets Surrouding the Solar System, and a Hypothesis Concerning Its Origin," *Bulletin of the Astronomical Institute of the Netherlands*, XI(408), 91–110.

Tyror, J. G. (1957), "The Distribution of the Directions of Perihelion of Long-Period Comets," *Monthly Notices of the Royal Astronomical Society*, 117, 369–379.

Watson, G. S. (1961), "Goodness-of-Fit Tests on a Circle," *Biometrika*, 48, 109–114.

Watson, G. S. (1983), *Statistics on Spheres*. New York: Wiley.

Weissman, P. R. (1980), "Stellar Perturbations of the Cometary Cloud," *Nature*, 288, 242–243.

Wiegert, P., and Tremaine, S. (1999), "The Evolution of Long-Period Comets," *Icarus*, 137, 84–121.