A&A 571, A86 (2014)
DOI: 10.1051/0004-6361/201424405

Astronomy & Astrophysics

Analysing weak orbital signals in *Gaia* data

L. B. Lucy

Astrophysics Group, Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2AZ, UK
e-mail: l.lucy@imperial.ac.uk

Received 15 June 2014 / Accepted 22 August 2014

ABSTRACT

Anomalous orbits are found when minimum-χ^2 estimation is applied to synthetic *Gaia* data for orbits with astrometric signatures comparable to the single-scan measurement error (Pourbaix 2002, A&A, 385, 686). These orbits are nearly parabolic, edge-on, and their major axes align with the line-of-sight to the observer. Such orbits violate the Copernican principle (CPr) and as such could be rejected. However, the preferred alternative is to develop a statistical technique that incorporates the CPr as a fundamental postulate. This can be achieved in a Bayesian context by defining a Copernican prior. Pourbaix's anomalous orbits then no longer arise. Instead, the selected orbits have a somewat higher χ^2 but do not violate the CPr. The problem of detecting a weak additional orbit in an astrometric binary with a well-determined orbit is also treated.

Key words. binaries: visual – stars: fundamental parameters – methods: statistical

1. Introduction

With the *Gaia* observatory in orbit at *L*2 and with commissioning underway, astronomers can look forward with increasing confidence to the eventual release of an enormous quantity of high precision astrometric data. Initially, this data will be analysed with the already-existing pipeline software created by the various consortia. The resulting pipeline products will no doubt be entirely satisfactory for the vast majority of observed objects. However, a lesson from earlier large-scale surveys is that a small number of objects at the limit of a survey's range often prove to be of exceptional interest. For such objects, standard reduction techniques may give anomalous and misleading results.

This occurred for the HIPPARCOS mission. As reviewed by Pourbaix (2004) and Perryman (2009, p. 594), orbits fitted to HIPPARCOS data for stars with known spectroscopic orbits led to "discoveries" that were later refuted. As emphasized by Pourbaix (2004), "fitting the noise with an orbital model can have some awful consequences".

This earlier episode suggests that the extraction of orbital parameters from weak orbital signals in *Gaia* data should be investigated. In fact, this is already the subject of an intriguing paper by Pourbaix (2002). He found that min-χ^2 solutions for weak orbits are frequently anomalous – specifically, edge-on and nearly parabolic. In the present paper, the origin of such orbits is explained and a Bayesian technique developed that overcomes this problem.

2. Synthetic data

In this section, synthetic 1D scans of a model astrometric binary are created. In order to focus on orbital parameters, we follow Pourbaix (2002) in assuming that parallactic and proper motions have been subtracted. With regard to notation, previous papers (Lucy 2014a,b; hereafter L14a,b) are followed closely.

2.1. Orbital elements

In contrast to L14a,b, the secondary is here not detected, so the astrometry measures the primary's reflex motion about the system's centre of mass. This motion is parameterized with the Campbell elements $P, e, T, a, i, \omega, \Omega$. Here P is the period, e is the eccentricity, T is a time of periastron passage, i is the inclination, ω is the longitude of periastron, and Ω is the position angle of the ascending node. However, following many earlier investigators – references in L14a – the Thiele-Innes elements are also used, thereby exploiting the resulting linearity in four parameters. Thus, the Campbell vector $\boldsymbol{\theta} \equiv (\boldsymbol{\phi}, \boldsymbol{\vartheta})$, where $\boldsymbol{\phi} = (P, e, \tau)$, and where $\boldsymbol{\vartheta} = (a, i, \omega, \Omega)$ is replaced by the vector $\boldsymbol{\psi}$ whose components are the Thiele-Innes constants A, B, F, G. (Note that in $\boldsymbol{\phi}$, periastron has been replaced by $\tau = T/P$ which by definition $\in (0, 1)$.)

2.2. Model astrometric binary

The model binary has the following elements:

$$P_* = 2.9y \quad e_* = 0.05 \quad \tau_* = 0.4$$
$$a_* = \beta\,\sigma \quad i_* = 40^\circ \quad \omega_* = 150^\circ \quad \Omega_* = 70^\circ. \qquad (1)$$

Note that P_* is less than $t_M = 5y$, the duration of the *Gaia* mission, so that the issue of incomplete orbits (L14a) is not of concern here. Also the semi-major axis a_* is expressed as a dimensionless multiple β of the standard error σ of a single-scan measurement. Thus our ability to detect weak orbits can be investigated by letting $\beta \to 0$.

The eccentricity $e_* = 0.05$ is typical for giant planets in the solar system. But the main reason for such a small value is to highlight the anomaly when nearly parabolic orbits fit the data.

2.3. Observing campaign

A *Gaia*-like observing campaign is defined by t_n, the N times at which the star is scanned, by α_n, the corresponding scanning

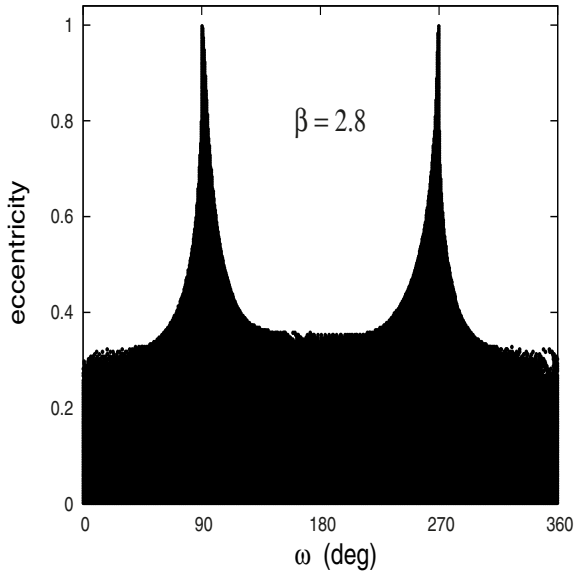

Fig. 1. Feasible domain $\mathcal{D}$ projected on to the (ω, e)-plane. The orbital parameters are given by Eq. (1) with $\beta = 2.8$.

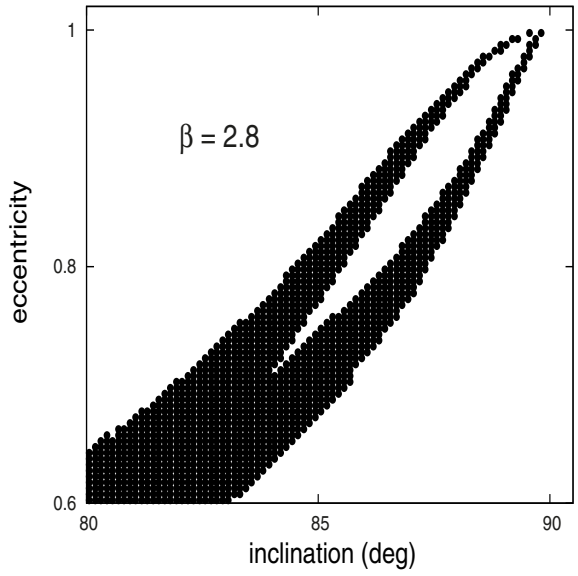

Fig. 2. Feasible domain $\mathcal{D}$ projected on to the (i, e)-plane. The orbital parameters are given by Eq. (1) with $\beta = 2.8$.

angles, and by σ. We take $t_n = t_M z_u$ and $\alpha_n = 2\pi z_u$, where the z_u here and later denote *independent* random numbers $\in (0, 1)$.

Although $\beta = a_*/\sigma$ is the important parameter, we take $\sigma = 40\ \mu$as, the expected accuracy for a single transit at G-band magnitude $\approx$14 (see Fig. 2 in Sozzetti et al. 2014). Note that, in the comprehensive investigation of planet detection with *Gaia* by Casertano et al. (2008), $\sigma = 8\ \mu$as.

From Fig. 1 in Sozzetti et al. (2014), we take $N = 70$ as a representive number of scans during the mission.

2.4. Synthetic scans

Given β, Eq. (1) defines the theoretical orbit. The Cartesian sky coordinates (x_n^*, y_n^*) at t_n can therefore be computed from Eqs. (A.2) of L14a. The corresponding 1D coordinate or abscissa is $s_n^* = s_n(x_n^*, y_n^*)$, where

$$s_n = x_n \cos \alpha_n + y_n \sin \alpha_n \tag{2}$$

Here α_n is the angle between the scanning direction and the x-axis – see Fig. 1 in Pourbaix (2002). A synthetic data set is then

$$\tilde{s}_n = s_n^* + \sigma z_G \tag{3}$$

where the z_G are independent random Gaussian variates sampling $\mathcal{N}(0, 1)$. Note that the χ^2 of the measurement errors is simply

$$\tilde{\chi}^2 = \sum_n z_G^2 \tag{4}$$

The N-dimensional vector $\tilde{\boldsymbol{s}}$ with elements $\tilde{s}_n$ is the data vector from which orbital elements are to be estimated. For a given orbit $\boldsymbol{\theta} \equiv (\boldsymbol{\phi}, \boldsymbol{\psi})$, the goodness-of-fit to $\tilde{\boldsymbol{s}}$ is measured by

$$\chi^2 = \frac{1}{\sigma^2} \sum_n (\tilde{s}_n - s_n)^2 \tag{5}$$

where $s_n = s(t_n, \alpha_n; \boldsymbol{\theta})$.

3. Feasible orbits

In this section, a procedure from L14a is used to explore the likely degradation of extracted orbits as $\beta \to 0$.

3.1. Grid scan

A 3D grid in the $\boldsymbol{\phi}$ variables is set up as follows: the mid-point of grid cell (i, j, k) is $(\log P_i, e_j, \tau_k)$. The grid has 200 constant steps in each of these variables, with ranges (0.0, 1.0) for $\log P$, and (0, 1) for e and τ.

For specified β, a synthetic scan vector $\tilde{\boldsymbol{s}}$ is created as described in Sect. 2.4. Then, at each grid point, the min-χ^2 Thiele-Innes vector $\hat{\boldsymbol{\psi}}$ is computed as described in Appendix A.1. The resulting $\chi^2 = \hat{\chi}^2_{ijk}$.

3.2. Feasible domain $\mathcal{D}$

An orbit $\boldsymbol{\theta}_{ijk} = (\boldsymbol{\phi}_{ijk}, \hat{\boldsymbol{\psi}})$ is deemed to be feasible if

$$\Pr\left(\chi^2 > \hat{\chi}^2_{ijk}\right) > 0.05 \tag{6}$$

and the ensemble of such orbits define the feasible domain(s) $\mathcal{D}$ in $\boldsymbol{\phi}$-space.

For $\beta \gtrsim 10$, the domain $\mathcal{D}$ is a small ellipsoidal volume approximately centred on the exact values $(\log P_*, e_*, \tau_*)$. But as β decreases, $\mathcal{D}$ increases and eventually develops extraordinary topology.

From a sequence of grid scans with $\beta \to 0$, the value $\beta = 2.8$ is found to be such that $\mathcal{D}$ just extends to $e = 1$. Figure 1 illustrates the resulting distortions of $\mathcal{D}$. In this figure, a filled circle is plotted at (ω, e) if Eq. (6) is satisfied, and we see that this projection of the feasible orbits $\boldsymbol{\theta}_{ijk}$ extends far beyond the exact values (150°, 0.05). The most notable features are the two narrow spikes that emerge at $e \sim 0.4$ and reach $e = 1$ at *precisely* $\omega = 90°$ and 270°. Further information about these spikes is provided by other projections of the $\boldsymbol{\theta}_{ijk}$. In Fig. 2, the vectors are projected onto the (i, e)-plane, and this shows that along both spikes $i \to 90°$ as $e \to 1$. Accordingly, if the orbital signal is weak enough ($\beta < 2.8$), an acceptable fit is provided by nearly parabolic, edge-on orbits with $\omega = \pi/2$ or $3\pi/2$. But this

remarkable finding is not original: Pourbaix (2002), in reporting least-squares fits to synthetic 1D scans, found an accumulation of nearly parabolic orbits when $\beta = a_*/\sigma = 1.33$ and noted that such orbits lead to reasonable *apparent* orbits when i and ω are close to $\pi/2$. This serendipitous numerical discovery posed what he called "the puzzling case of almost parabolic orbits".

3.3. Violations of the Copernican principle

When the observed star is at periastron ($t = T$), its Cartesian coordinates are:

$$\begin{aligned} x &= a_{\rm p}\ (\cos\Omega\ \cos\omega - \sin\Omega\ \sin\omega\ \cos i) \\ y &= a_{\rm p}\ (\sin\Omega\ \cos\omega + \cos\Omega\ \sin\omega\ \cos i) \\ z &= a_{\rm p}\ \sin\omega\ \sin i \end{aligned} \tag{7}$$

where the periastron distance $a_{\rm p} = a(1-e)$. Accordingly, if $i = \pi/2$ and $\omega = \pi/2$ or $3\pi/2$, periastron has coordinates $(0, 0, \pm a_{\rm p})$. The major axis is thus aligned with the line-of-sight to the observer, who therefore finds himself in a special location. The observer might then object that this orbit violates the Copernican principle (CPr). But this objection could be raised against a slightly non-circular orbit with the same (i, ω), and an edge-on circular orbit is not particularly objectionable from the standpoint of the CPr. To fully appreciate the CPr violation, the extra fact that Pourbaix orbits (hereafter *P*-orbits) are nearly parabolic must therefore be taken into account.

Consider the astrometric signal's dependence on orbit orientation. Since the maximum elongation of the star from the barycentre is $a(1+e)$, the extreme range for the abscissae s_n is $-a(1+e)$ to $+a(1+e)$. However, when $i = \pi/2$ and $\omega = \pi/2$ or $3\pi/2$, this range shrinks to its minimum, namely $-b$ to $+b$, where $b = a\sqrt{1-e^2}$ is the semi-minor axis Thus, for the Pourbaix solutions, the orbit's *inferred* orientation *and* eccentricity are such that the signal is at a deep minimum. For example, the ratio of the maximum to minimum elongations is $\sqrt{(1+e)/(1-e)} = 14.1$ when $e = 0.99$. This is a large effect, and so the observer would be correct in concluding that P-orbits violate the CPr.

Traditionally, when an analysis leads to a CPr violation, astronomers suspect that some underlying hypothesis H must be wrong. A classic example is Herschel's model of the Milky Way, which violates the CPr because the sun is close to its centre. In this case, the error is Herschel's implicit assumption that interstellar space is transparent.

But note a crucial difference. In these simulations – and in those of L14a – CPr violations arise even though H – Keplerian motion – is rigorously correct. This strongly implies that there must exist a data-analysis technique that includes the CPr ab initio rather than invoking it to pass judgement on a model only after it has been derived.

3.4. Degeneracy

For a single star, the astrometric solution has five parameters: the star's right ascension (RA) and declination at a reference epoch, two components of proper motion, and its parallax. However, because of errors in the $\tilde{s}_n$, this solution has residuals, and so it is likely that the addition of orbital motion (seven parameters) will "improve" the fit – i.e., reduce χ^2. Given their minimal astrometric signatures (Sect. 3.3), P-orbits with $b \lesssim \sigma$ can be added with little effect on the fit. Evidently, a single-star solution is degenerate under the addition of a P-orbit with *arbitrarily large* semi-major axis a so long as the semi-minor axis $b = a\sqrt{1-e^2} \ll \sigma$.

3.5. Imperfect experiments

For reasons beyond the observer's control, experiments in astronomy are often imperfect, yielding data from which a definitive solution cannot be obtained. In double star astronomy, examples are long-period binaries that have only completed a fraction of an orbit since discovery. If a solution is nevertheless attempted, orbits with very different parameters may provide acceptable fits – see Fig. 2 in L14a and references therein. Among these acceptable orbits may be orbits that violate the CPr, as is the case for the nearly parabolic orbit in that figure.

Comparison of the simulations here with those for incomplete orbits in L14a is illuminating. Here and in Pourbaix (2002), we find CPr violations even though the orbit is complete ($P_* < t_M$). This shows that a weak orbital signal suffices for the experiment to be imperfect and to thereby permit solutions that violate the CPr.

When the min-χ^2 elements violate the CPr, we might suspect that there exists a better solution that, despite a higher χ^2, should be preferred because it is consistent with the CPr.

4. A Bayesian prior derived from the CPr

The CPr is now treated as an integral part of Bayesian estimation and not as an a posteriori arbiter of a solution's believability. This is achieved by constructing a Copernican prior.

4.1. Conventional priors

If H denotes the hypothesis and D the data and if I is some relevant information, then by Bayes' theorem (Jaynes 2003, p. 85), the posterior density of H given D and I is

$$\Pr(H|D,I) \ \propto\ \Pr(H|I)\,\Pr(D|H,I) \tag{8}$$

Here $\Pr(D|H,I) \equiv \mathcal{L}(H,I|D)$ is the likelihood, and $\Pr(H|I)$ is the prior probability of H given I.

If there is no information I, the prior reduces to $\Pr(H)$ and so becomes the *subjective* choice of the investigator. This aspect of Bayesian estimation is controversial and much-debated. However, there is little reason to object to current astronomical practice with regard to $\Pr(H)$ since the aim is not to quantify prejudice but to admit ignorance. Thus *flat* priors are typically imposed on the parameters of H. Moreover, the ranges over which these priors are non-zero are chosen to comfortably enclose the intervals within which there is significant likelihood $\mathcal{L}$ and therefore significant posterior density $\Pr(H|D)$. Such priors are *non-informative*.

In the *Gaia* problem, it is tempting to use a variant of this methodology to eliminate CPr-violating orbits. Thus, Fig. 1 suggests a prior on e that is zero for $e > 0.6$. But this would be an ad hoc fix for this particular data set. A Bayesian prior should not depend on, nor be derived from, D.

4.2. A Copernican prior

For the problem under consideration, the symbols H, D and I are defined as follows:

H: the components of the *theoretical* scan vector $\boldsymbol{s}$ are $s_n = s(t_n, \alpha_n; \boldsymbol{\theta})$, the predicted abscissae at the known times t_n and scanning angles α_n for the Keplerian orbit $\boldsymbol{\theta}$.

D: the elements of the data vector $\tilde{\boldsymbol{s}}$ are $\tilde{s}_n$, the measured abscissae at (t_n, α_n).

I: orbits $\boldsymbol{\theta}$ with random orientations and random shifts in epoch are all equally probable *a priori*.

Comments:

(i) information I is the means of incorporating the CPr.
(ii) since orbits are periodic, choosing a random value of $\tau = T/P \in (0, 1)$ is equivalent to a random shift in epoch.
(iii) imposing I is appropriate for orbits *discovered* with *Gaia* but not if a previously-known orbit is targeted.

Given that the hypothesis of Keplerian motion enters via the theoretical vector $\boldsymbol{s}$, the Copernican prior $\Pr(H|I)$ becomes $\pi(\boldsymbol{s}|I)$, the probability density at $\boldsymbol{s}$ when I is taken into account. However, I does not itself suffice to determine this probability density function (PDF). In addition, P, e and a – or their prior distributions – must be specified. We choose the latter option on the grounds of simplicity.

For the bounded quantity e, we assume a uniform prior in (0, 1). For the unbounded positive continuous parameters P and a, it is appropriate (Jaynes 2003, p. 395) to assign equal prior probabilities to equal logarithmic intervals – i.e., Jeffreys' priors.

With these additional assumptions, the Copernican prior is determinate and given by

$$\Pr(H|I) \propto a^{-1}\pi_1(\boldsymbol{s}|I) \tag{9}$$

where $\pi_1(\boldsymbol{s}|I)$ is the PDF at $\boldsymbol{s}$ for orbits scaled to $a = 1''$.

4.3. Accurate treatment

A rigorous calculation of $\pi_1(\boldsymbol{s}|I)$ would proceed as follows: The scale parameter a is set $= 1''$, and e and $\log P$ are randomly chosen in (0, 1) and $(\log P_L, \log P_U)$, respectively. The orbit's orientation (i, ω, Ω) and epoch $\tau = T/P$ are randomly chosen, and the resulting theoretical $\boldsymbol{s}$ computed. These steps are repeated many times, thus generating points that populate the N-dimensional $\boldsymbol{s}$-space with a probability density determined by I and by the prior distributions of e and P. As the sample size $\to \infty$, the result is the desired PDF $\pi_1(\boldsymbol{s}|I)$. However, with $N \sim 70$, this brute-force approach is not feasible. A less rigorous approach must be adopted.

4.4. Approximate treatment

Consider an orbit with $a = 1''$ and eccentricity e. Sampled with random orientations and epochs, the theoretical abscissae s_n will extend over the full permitted range, namely $-(1 + e)$ to $+(1 + e)$. In other words, whatever the values of (t_n, α_n), there will be some combination of orientation (i, ω, Ω) and epoch τ for which these limits are reached. Given that $e \in (0, 1)$, it follows that the s_n-values populate the interval $(-2, +2)$. Since this applies to every component of $\boldsymbol{s}$, the distribution of the vectors $\boldsymbol{s}$ in N-dimensional space is approximately isotropic. Accordingly, most of the information relevant to CPr violations is contained in the distribution of the Euclidean "lengths" of the vectors $\boldsymbol{s}$. We therefore define the statistic ξ given by

$$\xi^2 = N^{-1}\sum_n (s_n/a)^2 \tag{10}$$

With this statistic as the sole basis for assessing CPr violations, the approximate Copernican prior is

$$\Pr(H|I) \propto a^{-1}\,\pi_1(\xi|I) \tag{11}$$

where $\pi_1(\xi|I)$ is the PDF of the lengths ξ for orbits scaled to $a = 1''$. This prior is used in the subsequent tests.

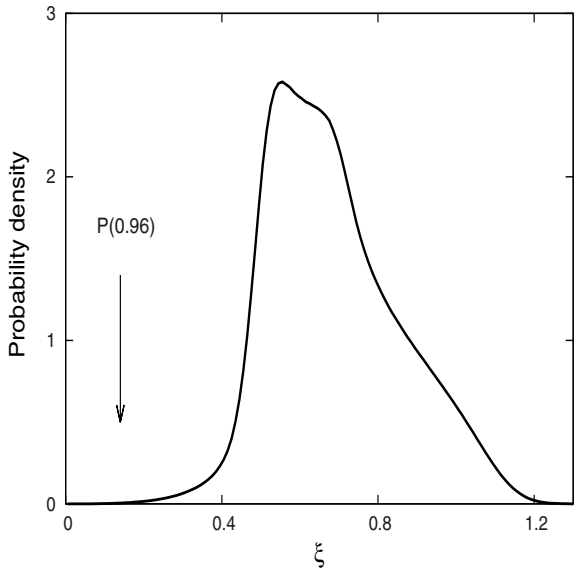

Fig. 3. PDF $\pi_1(\xi|I)$ for astrometric "lengths" ξ defined by Eq. (10). The orbits have random orientations (i, ω, Ω), random epochs τ, random eccentricities $\in (0, 1)$, and random values of $\log P \in (0.0, 1.0)$. The corresponding PDF for P-orbits ($i = \omega = 90°$) with $e = 0.96$ is a near delta function at $\xi = 0.14$.

4.5. Calculation of $\pi_1(\xi|I)$

With the above assumptions, the problem has been reduced to tabulating the 1D function $\pi_1(\xi|I)$. The steps are as follows:

1) The campaign $(t_n, \alpha_n; N)$ is specified (Sect. 2.3).
2) We set $a = 1''$ and choose random values of $e \in (0, 1)$ and $\log P \in (\log P_L, \log P_U)$.
3) A random orientation is selected by taking $\omega = 2\pi z_u$, $\Omega = \pi z_u$, and $\cos i = 1 - 2z_u$.
4) A random epoch is selected by taking $\tau = T/P = z_u$.
5) With the orbit vector $\boldsymbol{\theta}$ determined in steps 2)–4), the coordinates (x_n, y_n) at t_n are computed.
6) From these coordinates, the components of $\boldsymbol{s}$ are given by Eq. (2), and the corresponding length ξ by Eq. (10).
7) Steps 2)–6) are repeated 10^8 times. The resulting histogram of ξ-values gives $\pi_1(\xi|I)$.

Comments:

(i) To eliminate parabolic orbits and to avoid convergence failures when solving Kepler's equation, an upper limit $e = 0.999$ is imposed at step 2).
(ii) Since the exact period is known, we take $P_L = 1y$ and $P_U = 10y$.
(iii) From the 10^8 values of s_n for each n, the maximum and minimum values are derived. These closely approach the expected values ±2 (Sect. 4.4).
(iv) If no assumptions are made about the prior distributions of P and e, then $\pi_1 = \pi_1(\xi; P, e|I)$, thus requiring an extra two dimensions in its tabulation. This is feasible, but the preference here is to investigate the simplest formulation.

The resulting accurate determination of $\pi_1(\xi|I)$ is plotted in Fig. 3. This shows that the astrometric lengths are typically in the interval (0.4, 1.1), and that values $\lesssim 0.3$ are improbable.

The position of a P-orbit in this plot is of interest. The above steps are therefore repeated with the constraints $e = 0.96$ and $i = \omega = 90°$. For this orbit, the range for the s_n is $(-b, +b)$ or $(-0.28, +0.28)$ in units of a. Consistent with this, the PDF is a

near delta function at $\xi = 0.14$, and this location is indicated in Fig. 3. The probability of obtaining an even smaller value is

$$\Pi_1(\xi|I) = \int_0^{\xi} \pi_1(\xi|I)\, \mathrm{d}\xi \tag{12}$$

which gives $\Pi_1(0.14|I) = 1.7 \times 10^{-4}$, showing that P-orbits populate an extremely low probability tail of $\pi_1(\xi|I)$.

5. Bayesian estimation subject to the CPr

Formulae are now developed that allow the Copernican prior to be included in the calculation of posterior densities and credibility intervals.

5.1. Posterior densities

For every orbit in $\boldsymbol{\theta}$-space, there is a theoretical scan vector $\boldsymbol{s}$ corresponding to the scanning campaign (t_n, α_n). From this $\boldsymbol{s}$ and the orbit's a, we can compute ξ from Eq. (10). Then, from ξ, we obtain $\pi_1(\,\xi|\,I)$ by interpolating in the data file plotted in Fig. 3. This procedure results in an ensemble of orbit vectors weighted according to their Copernican priors $a^{-1}\pi_1(\,\xi|\,I)$, and so CPr violations are penalized. From this ensemble, the Bayesian machinery then computes posterior densities by further weighting the orbits in accordance with their goodness-of-fits to the *measured* scan vector $\tilde{\boldsymbol{s}}$.

The posterior density at $(\boldsymbol{\phi}, \boldsymbol{\psi})$ is

$$\Lambda(\boldsymbol{\phi}, \boldsymbol{\psi}|D, I) \propto a^{-1}\, \pi_1(\xi|I)\, \mathcal{L}\,(\boldsymbol{\phi}, \boldsymbol{\psi}|D)\,. \tag{13}$$

Ignoring coefficients independent of $(\boldsymbol{\phi}, \boldsymbol{\psi})$ and assuming normally distributed measurement errors, we have

$$\mathcal{L} \;\propto\; \exp\left(-\frac{1}{2}\,\hat{\chi}^2\right) \times \exp\left(-\frac{1}{2}\,\delta\chi^2\right) \tag{14}$$

where $\hat{\chi}^2(\boldsymbol{\phi})$ is the minimum at $\hat{\boldsymbol{\psi}}$, and $\delta\chi^2$ is the postive increment due to the displacement $\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}$ at fixed $\boldsymbol{\phi}$.

In the absence of the Copernican prior, $\Lambda \propto \mathcal{L} \propto \Pr(\phi) \times \Pr(\psi|\phi)$. The first of these PDFs is sampled at grid points (i, j, k) giving weight factors $\propto \exp(-\chi^2_{ijk}/2)$. The second PDF is randomly sampled as described in Appendix A.4. If $\mathcal{N}_{ijk}$ is the number of random points ψ_ℓ selected in ψ-space at ϕ_{ijk}, then each has weight $\mathcal{N}^{-1}_{ijk}$.

With the Copernican prior included, the PDF Λ given in Eq. (13) is represented by a cloud of discrete orbit vectors

$$\boldsymbol{\theta}_m \equiv (\boldsymbol{\phi}_{ijk}, \boldsymbol{\psi}_\ell) \tag{15}$$

with weights

$$\mu_m = a_m^{-1}\pi_1(\xi_m|I) \times\; \mathcal{N}^{-1}_{ijk}\; \exp\left(-\frac{1}{2}\,\hat{\chi}^2_{ijk}\right) \tag{16}$$

Here m enumerates the random points $\boldsymbol{\psi}_\ell$ across all grid cells (i, j, k).

From these weighted orbits, the posterior mean of a quantity $Q(\boldsymbol{\theta})$ given D *and* I is

$$\langle Q\rangle \;=\; \sum_m \mu_m Q_m \;/\; \sum_m \mu_m \tag{17}$$

and credibility intervals are derived as described in Sect. 4.2 of L14b.

This discrete representation of Λ and the resulting credibility means and intervals become exact as the grid steps $\to 0$ and the $\mathcal{N}_{ijk} \to \infty$.

6. Numerical experiments

The approximate theory (Sect. 4.4) of the Copernican prior is now applied to the model binary defined in Eq. (1).

6.1. Code verification

In the strong-orbit limit, violations of the CPr are not an issue and so, even with the inclusion of the Copernican prior, the posterior means should $\to$ the exact elements given in Eq. (1). To test this, the Bayesian code is used to compute the solution when $\log\beta = 1.5$. The posterior means and *equal-tail* 1σ credibility intervals for the elements are as follows:

$$\begin{aligned}
\log P(y) &= 0.4625^{+0.0011}_{-0.0013} & e &= 0.0529^{+0.0039}_{-0.0039}\\
& & \tau &= 0.384^{+0.012}_{-0.012}\\
\log a/\sigma &= 1.4990^{+0.0033}_{-0.0038} & i &= 40\overset{\circ}{.}6^{+1\overset{\circ}{.}0}_{-1\overset{\circ}{.}0}\\
\omega &= 144\overset{\circ}{.}8^{+4\overset{\circ}{.}9}_{-4\overset{\circ}{.}9} & \Omega &= 69\overset{\circ}{.}6^{+1\overset{\circ}{.}3}_{-1\overset{\circ}{.}3}.
\end{aligned} \tag{18}$$

These results are consistent with expectation: five of the seven credibility intervals include the exact values. Minor deviations occur for τ and ω.

Note that the credibility interval for ω is substantially larger than those for i and Ω. This is a consequence of the small eccentricity, since ω becomes indeterminate as $e \to 0$.

In this strong-orbit regime, $\mathcal{L}$ is sharply peaked in parameter space; consequently, ξ and therefore $\pi_1(\xi|I)$ vary little within the narrow domain of high likelihood. It follows that posterior densities are then largely determined by $\mathcal{L}$, which overwhelms the prior.

6.2. Varying β

Solutions are computed with $\log\beta = -0.6\ (0.05)\ 1.2$, spanning the range from weak to strong orbits. For each β, the elements' posterior means and 1σ credibility intervals are derived as in Sect. 6.1. In addition, for each data vector $\tilde{\boldsymbol{s}}$, the min-χ^2 solution is computed as in Pourbaix (2002). Note that when β changes, so does the random number seed.

In Figs. 4–6, the solution sequences are plotted for $\log a/\sigma$, e, and i. For $\log\beta \gtrsim 0.2$, the credibility intervals are consistent with both the min-χ^2 values and with the exact values. However, at $\log\beta = -0.05$, major disagreements occur. The min-χ^2 value of $\log a/\sigma$ suddenly jumps to 1.14, which is 1.19 dex greater than the exact value. Correspondingly, e jumps to 0.9975 (the highest value allowed by the grid) and i jumps to 89$\overset{\circ}{.}$6. Thus, at $\beta = -0.05$ dex, the min-χ^2 solution is a P-orbit (Sects. 3.2–3.4)

At this same $\beta = -0.05$ dex, the 1σ credibility intervals are $(-0.20, 0.05)$ for $\log a/\sigma$, $(0.06, 0.44)$ for e, and $(52°, 83°)$ for i. Thus, the Bayesian solution with Copernican prior does not undergo a transition into a P-orbit. In fact, the solution remains (marginally) consistent with the exact solution.

For $\log\beta \leq -0.05$, most of the min-χ^2 solutions are P-orbits. But the Bayesian solutions with Copernican prior do not exhibit such strikingly anomalies. Nevertheless, they do eventually ($\log\beta \lesssim -0.3$) become inconsistent with the exact parameters.

6.3. Orbits from noise

Figures 4–6 show that even for extremely weak orbits the Copernican prior has eliminated P-orbits. However, the plotted credibility intervals reveal that when $\log\beta \lesssim -0.3$ the posterior

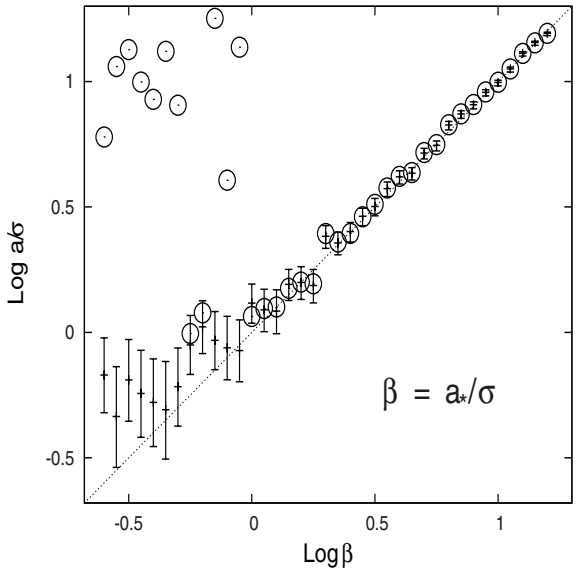

Fig. 4. Sequence of solutions for $\log a/\sigma$. The points with error bars are the posterior means $\langle \log a/\sigma \rangle$ plotted with 1σ credibility intervals. The open circles are the min-χ^2 values. The dotted line is the locus of exact values $\log a_*/\sigma$.

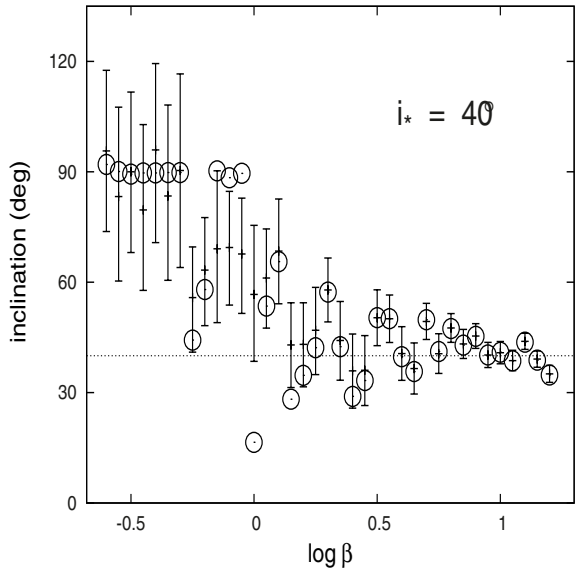

Fig. 6. Sequence of solutions for i. The points with error bars are the posterior means $\langle i \rangle$ plotted with 1σ credibility intervals. The open circles are the min-χ^2 values. Orbits with $i > 90°$ are retrograde. The exact value is $i_* = 40°$.

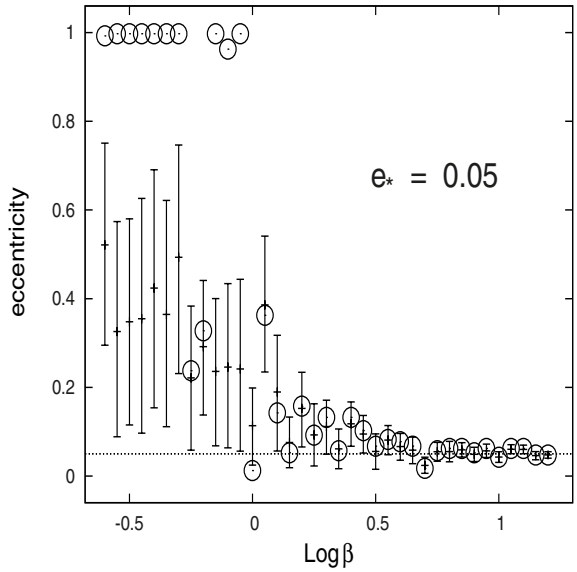

Fig. 5. Sequence of solutions for e. The points with error bars are the posterior means $\langle e \rangle$ plotted with 1-σ credibility intervals. The open circles are the min-χ^2 values. The exact value is $e_* = 0.05$.

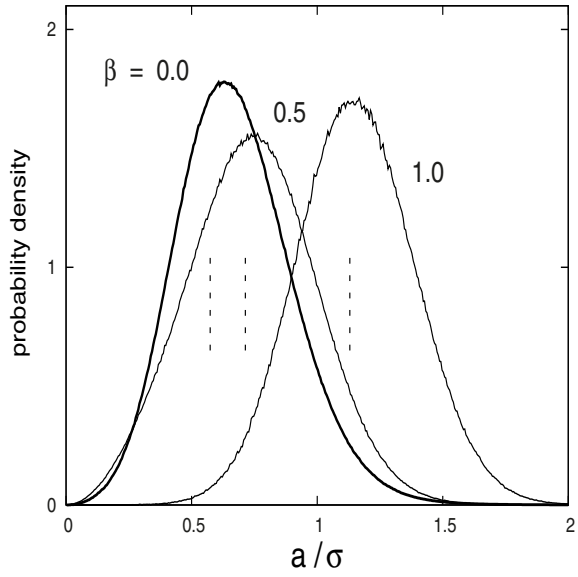

Fig. 7. Posterior densities of a/σ for $\beta = 0.0, 0.5$ and 1.0 for particular realizations of the measurement vectors $\tilde{\boldsymbol{s}}$. The vertical dotted lines indicate the average positions of the corresponding posterior means $\langle a/\sigma \rangle$. Each of these is obtained from 20 independent simulations.

PDFs are systematically displaced from the exact values. This could indicate that the orbital signal is then too weak for detection, a conclusion strongly supported by Fig. 6 which shows that the posterior densities of retrograde and prograde orbits are then about equal.

To investigate this issue further, the code is now used to compute solutions when $a_* = \beta\sigma = 0$. The posterior density of a/σ is plotted in Fig. 7 for a particular realization of the noise vector $\tilde{\boldsymbol{s}}$. For comparison, plots with $\beta = 0.5$ and 1.0 are also included.

In 20 independent repetitions with $\beta = 0$, the range found for $\langle a/\sigma \rangle$ is 0.45 to 0.81, with average $\langle\langle a/\sigma \rangle\rangle = 0.57$. For $\langle e \rangle$ the range is 0.21 to 0.53, with $\langle\langle e \rangle\rangle = 0.37$; and for $\langle i \rangle$ the range is 67° to 113°, with $\langle\langle i \rangle\rangle = 88°$. Since these are consistent with Figs. 4–6 when $\log\beta \lesssim -0.3$, we conclude that the aforementioned systematic displacements simply reflect the code's reponse to data with negligible orbital signal.

The bias in a/σ for $\beta = 0$ evident in Fig. 7 is reminiscent of the bias in the eccentricities of spectroscopic binaries for nearly circular orbits (Lucy & Sweeney 1971). In both cases, bias is the inevitable result of estimating a *non-negative* parameter at or near its zero lower bound. From the values of $\langle\langle a/\sigma \rangle\rangle$ plotted in Fig. 7, the bias of a/σ is typically 0.57, 0.21 and 0.13 at $\beta = 0.0, 0.5$ and 1.0, respectively.

Also of interest when $\beta = 0$ is the frequency of P-orbits for min-χ^2 solutions. From 200 independent simulations, 156 or 78% are P-orbits – i.e., have $e > 0.95, i \approx 90°$ and $\omega \approx 90$ or 270°. Thus, the PDFs of e, i and ω for min-χ^2 solutions when $\beta = 0$ are dominated by near delta functions at the Pourbaix loci.

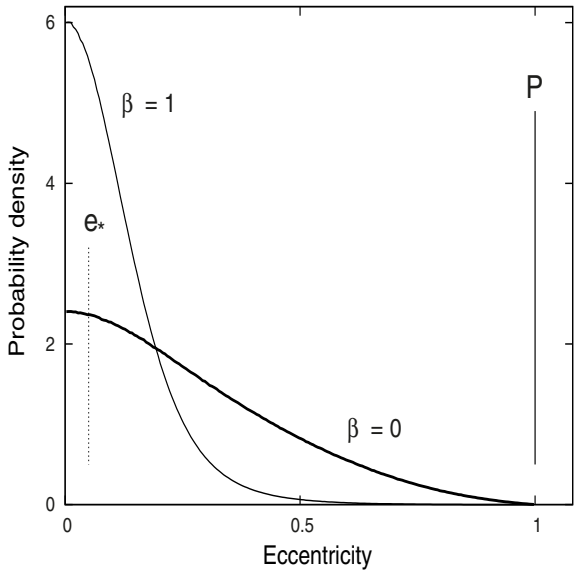

Fig. 8. Posterior PDFs for e when $\beta = 0$ and 1. The Pourbaix peak at $e = 1$ and the exact value $e_* = 0.05$ are indicated.

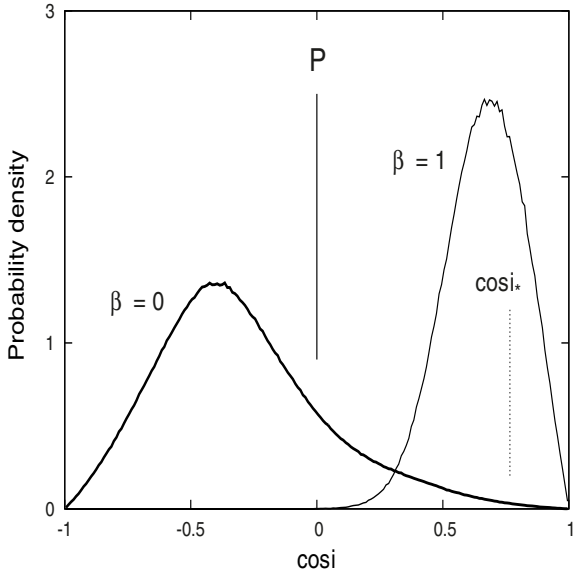

Fig. 9. Posterior PDFs for i when $\beta = 0$ and 1. The Pourbaix peak at $\cos i = 0$ and the exact value $cosi_*$ are indicated.

Figures 8–10 plot the posterior densities of $e, \cos i$ and ω for $\beta = 0$ when the Copernican prior is included. While these plots are free from peaks at the Pourbaix loci, they do show evidence of imperfections that presumably derive from the approximate treatment of the Copernican prior (Sect. 4.4). Ideally, when analysing pure noise, the inferred values of $\cos i$ and ω should be uniformly distributed in $(-1, +1)$ and $(0, 360°)$, respectively. Figures 9, 10 show departures from this ideal.

As in Fig. 7, Figs. 8–10 also include solutions for $\beta = 1$. Figs. 8 and 9 show the emergence of peaks at the exact values of e and $\cos i$, repectively. However, an emerging peak is not evident at ω_* in Fig. 10. This is due to the near indeterminacy of ω when $e \ll 1$ – see Sect. 6.1.

6.4. ξ-probabilities

Given that P-orbits arise when applying a conventional data analysis technique to synthetic *Gaia* data, there is some danger that such orbits will contaminate the huge data bases expected from the *Gaia* mission. On the assumption that this Bayesian procedure cannot feasibly replace the existing pipeline analyses, a less ambitious approach to elimating P-orbits is desirable.

Let $\boldsymbol{\theta}_0$ be the min-χ^2 elements derived from an observed scan vector $\tilde{\boldsymbol{s}}$ and let the corresponding fitted vector be $\boldsymbol{s}_0$. From $\boldsymbol{s}_0$, the astrometric length ξ_0 of the orbit $\boldsymbol{\theta}_0$ is then given by Eq. (10). This scale-free length refers to an orbit with physical parameters (P_0, e_0) observed at epoch τ_0 and orientation $(i_0, \omega_0, \Omega_0)$. We now define p_0 to be the probability that a shorter length ξ would be found with random epochs and orientations but with P and e fixed at their min-χ^2 values. Thus, with steps 3)–6) of Sect. 4.5, we compute

$$p_0 = \Pr(\xi < \xi_0 | P_0, e_0). \tag{19}$$

In Fig. 11, $\log p_0$ is plotted against $\log\beta$ for min-χ^2 orbits. For strong orbits, the values scatter about the exact value = −0.188. But for weak orbits the solutions are the CPr-violating P-orbits with $\log p_0 \sim -4$. Accordingly, if a *Gaia* orbit catalogue were contaminated by P-orbits a cut excluding orbits with $p_0 < -3$ dex would remove them.

Besides P-orbits, other as yet unrecognized anomalies, biases and selection effects may be present in a *Gaia* catalogue. Accordingly, it is worth noting that there are *five* quantities which, in a *perfect catalogue*, are uniformly and independently distributed in $(0, 1)$. These quantities are: $(1 + \cos i)/2$, $\omega/2\pi$, Ω/π, τ and p_0. This statement yields *fifteen* statistical tests that should be applied to a catalogue of *Gaia* orbits.

6.5. Detecting a second companion

A further weak-orbit problem for *Gaia* is that of detecting a second companion (B) when the first (A) is well-determined. This is a goodness-of-fit problem: the presence of B degrades the fit achieved when only A is considered.

To investigate this problem, *Gaia* data is created (Sects. 2.3, 2.4) for a star with invisible companions A and B. Companion A has the elements given in Eq. (1) with $\beta_A = 10$, and B is in a coplanar orbit with $P = 7.2y, e = 0.2, \tau = 0.7$, and a reflex orbit with semi-major axis = $\beta_B\sigma$.

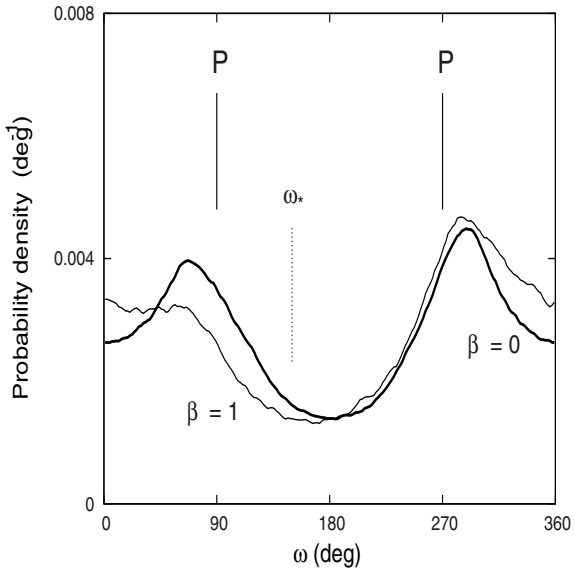

Fig. 10. Posterior PDFs for ω when $\beta = 0$ and 1. The Pourbaix peaks at $\omega = 90, 270°$ and the exact value $\omega_* = 150°$ are indicated.

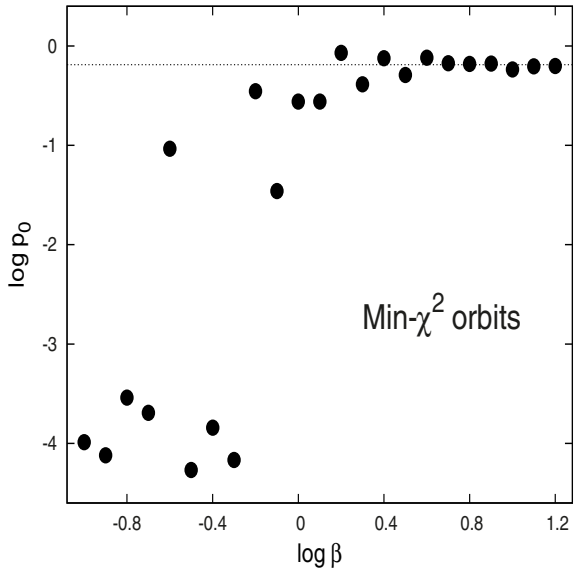

Fig. 11. ξ-probabilities p_0 for a β-sequence of min-χ^2 orbits computed for simulations of the orbit defined by Eq. (1). The dotted line is the exact value $p_* = 0.649$.

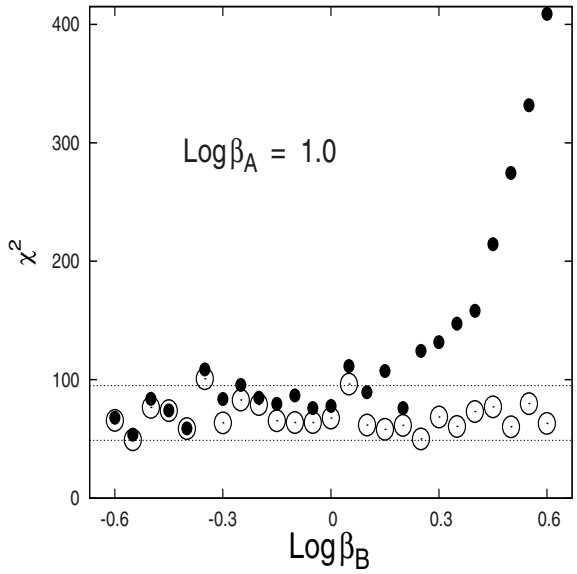

Fig. 12. Detecting a second companion. The filled circles are the posterior means $\langle\chi^2\rangle$ measuring the goodness-of-fits of the Bayesian *single-orbit* solutions to the simulated scan vectors $\tilde{\boldsymbol{s}}$ for the *two-orbit* model. The amplitudes are $\beta_A = 10$ and $\log\beta_B = -0.6(0.05)0.6$. The open circles are the corresponding values of $\tilde{\chi}^2$ given by Eq. (4). The dotted lines are the $\chi^2_{N,\alpha}$ values for $\alpha = 0.025, 0.975$ with $N = 70$.

A 1D sequence of *Gaia* scans is created for this two-companion model with $\log\beta_B = -0.6(0.05)0.6$, and each scan is analysed with the Bayesian code under the assumption of only *one* companion.

For each β_B, the code creates (Sect. 5) a cloud of orbits $\boldsymbol{\theta}_m$ with weights μ_m. The χ^2 of the m-th orbit's fit to the data vector $\tilde{\boldsymbol{s}}$ is

$$\chi^2_m = \hat{\chi}^2_{ijk} + \delta\chi^2_\ell. \tag{20}$$

Now, if the one-companion solution provides a satisfactory fit, then orbits of high weight should have $\chi^2_m \lesssim N$. On the other hand, if the solution is not satisfactory, then these high-weight orbits will have $\chi^2_m > \chi^2_{N,\alpha}$ with $\alpha < 0.05$. These expectations can be reduced to a single measure of goodness-of-fit, namely $\langle\chi^2\rangle$, the posterior mean of χ^2. This is calculated from Eq. (17) with $Q_m = \chi^2_m$.

The values of $\langle\chi^2\rangle$ are plotted against $\log\beta_B$ in Fig. 12. As always with statistical tests, the investigator has discretion as to when he deems a model to be successful. In this case, he is likely to suspect an additional orbit when $\log\beta_B \geq 0.3$. On the other hand, scans with $\log\beta_B \leq 0.1$ are fitted with $\langle\chi^2\rangle \approx \tilde{\chi}^2$, the residuals are therefore consistent with measurement errors and so there is no evidence of an additional orbit.

7. Conclusion

The aims of this paper are twofold. First, to provide a weak-orbit analysis for *Gaia* and, in particular, to investigate the occurrence of the spurious solutions found by Pourbaix (2002). Secondly, to use the *Gaia* problem as a test case for a procedure that incorporates the CPr into the machinery of statistical astronomy.

With regard to spurious solutions, Pourbaix's (2002) finding is confirmed and the puzzle of nearly parabolic, edge-on orbits explained in terms of the near degeneracy (Sect. 3.4) of scan vectors $\boldsymbol{s}$ under the addition of such orbits. Moreover, these orbits are shown to violate the CPr (Sect. 3.3) and do not arise when the CPr is adopted as a fundamental postulate in Bayesian estimation (Sects. 5 and 6).

More generally, incorporating the CPr in statistical analyses may improve solutions derived for imperfect experiments (Sect. 3.5). In addition to poor precision and limited sampling, weather, the seasons, atmospheric opacity and interstellar extinction are among the numerous factors that result in data sets that are less than ideal.

When an astronomer must perforce analyse an imperfect data set, he needs to be aware that supposedly optimum statistical procedures – e.g., min-χ^2 or Bayesian estimation with non-informative priors – can in extreme cases, as with the P-orbits, give anamolous solutions. Moreover, at a more subtle level, even when an anomaly is not immediately evident, the complicated topology (Sect.3.2) of the likelihood function implies that the above "optimum" procedures are unlikely to be so. An investigation of how an estimation procedure can exploit imperfect data should be carried out (Sect. 3.2) and an appropriate prior constructed. Often Copernican considerations with regard to position, epoch or orientation will be crucial.

Acknowledgements. I thank A. H. Jaffe and D. J. Mortlock for helpful discussions on Bayesian methods and the referee for justified criticisms of the original version.

Appendix A: Statistics in Thiele-Innes space

In L14a,b, each observation of the model visual binary yielded *two* measurements $(\tilde{x}_n, \tilde{y}_n)$, the sky coordinates of the secondary's displacement from the primary at time t_n. In this circumstance, minimizing χ^2 to obtain the Thiele-Innes constants $\hat{\psi}_j$ separates into two independent problems, minimizing the x-coordinate contribution to χ^2 to obtain (A, F) and minimizing the y-coordinate contribution to obtain (B, G). This separation results in the considerable simplifications exploited in L14a,b.

However, these simplifications are lost when observing an astrometric binary with a 1D scanning device. On the assumption that the parallactic and proper motion have been subtracted, the measurement at t_n with scanning angle α_n is the observed star's displacement $\tilde{s}_n$ from the binary's barycentre.

A.1. Normal equations

For given orbit $\boldsymbol{\theta} \equiv (\boldsymbol{\phi}, \boldsymbol{\psi})$, the goodness-of-fit criterion $\chi^2(\boldsymbol{\phi}, \boldsymbol{\psi})$ is given by Eq. (5). At fixed $\boldsymbol{\phi}$, the orbit (x, y) is linear in $\boldsymbol{\psi}$. Accordingly, since $s_n = 0$ when $\boldsymbol{\psi} = 0$, the predicted abscissa at t_n is

$$s_n = \sum_j \left(\frac{\partial s}{\partial \psi_j}\right)_n \psi_j. \tag{A.1}$$

Substitution of s_n into Eq. (5) then allows the min-χ^2 solution for the Thiele-Innes vector $\boldsymbol{\psi}$ to be obtained without iteration. The normal equations are

$$\mathcal{A}_{ij}\psi_j = b_i \tag{A.2}$$

where, the curvature matrix,

$$\mathcal{A}_{ij} = \frac{1}{\sigma^2}\sum_n \left(\frac{\partial s}{\partial \psi_i}\right)_n \left(\frac{\partial s}{\partial \psi_j}\right)_n \tag{A.3}$$

and

$$b_i = \frac{1}{\sigma^2}\sum_n \tilde{s}_n \left(\frac{\partial s}{\partial \psi_i}\right)_n . \tag{A.4}$$

The partial derivatives in these equations can be expressed in terms of the elliptical rectangular coordinates $X(E), Y(E)$ via Eqs. (2) and (A.2) of L14a.

The solution of Eq. (A.2) is $\hat{\boldsymbol{\psi}} = (\hat{A}, \hat{B}, \hat{F}, \hat{G})$ and we write $\hat{\chi}^2(\boldsymbol{\phi}) = \chi^2(\boldsymbol{\phi}, \hat{\boldsymbol{\psi}})$.

A.2. Increment in χ^2

At fixed $\boldsymbol{\phi}$, a displacement $\delta\boldsymbol{\psi}$ from $\hat{\boldsymbol{\psi}}$ results in a positive increment $\delta\chi^2$. The abscissa corresponding to this displacement is given by Eq. (A.1). Substitution in Eq. (5) then gives $\chi^2 = \hat{\chi}^2 + \delta\chi^2$. From the quadratic terms in the resulting expression, we obtain

$$\delta\chi^2 = \delta\boldsymbol{\psi}^{'}\,\mathcal{A}\,\delta\boldsymbol{\psi}. \tag{A.5}$$

A.3. Probability density function $p(\boldsymbol{\psi}|\boldsymbol{\phi}, D)$

The distribution of probability at fixed $\boldsymbol{\phi}$ is a quadrivariate normal distribution centred on $\hat{\boldsymbol{\psi}}$. If $\boldsymbol{\Sigma}$ is the covariance matrix, then

$$p = \frac{1}{4\pi^2}\frac{1}{\sqrt{|\boldsymbol{\Sigma}|}}\,\exp\left(-\frac{1}{2}\delta\boldsymbol{\psi}^{'}\boldsymbol{\Sigma}^{-1}\delta\boldsymbol{\psi}\right) \tag{A.6}$$

(James 2006, p. 67). Since $\boldsymbol{\Sigma}^{-1} = \mathcal{A}$, comparison with Eq. (A.6) gives

$$p = \frac{1}{4\pi^2}\frac{1}{\sqrt{|\boldsymbol{\Sigma}|}}\,\exp\left(-\frac{1}{2}\delta\chi^2\right) \tag{A.7}$$

A.4. Random sampling in ψ-space

A random point $\delta\boldsymbol{\psi}_\ell$ sampling $p(\boldsymbol{\psi}|\boldsymbol{\phi}, D)$ is obtained as follows (Gentle 2009, pp. 315–316): The first step is to compute the Cholesky decomposition (Press et al. 1992, pp. 89–91) of $\boldsymbol{\Sigma}$. Thus, we find the lower triangular matrix $\mathbf{L}$ such that

$$\mathbf{L}\mathbf{L}^{'} = \boldsymbol{\Sigma}. \tag{A.8}$$

Now let $\boldsymbol{z}_G$ be a 4D vector whose elements are independent random Gaussian variates sampling $\mathcal{N}(0, 1)$. Then

$$\delta\boldsymbol{\psi} = \mathbf{L}\boldsymbol{z}_G \tag{A.9}$$

is a random displacement from $\hat{\boldsymbol{\psi}}$ satisfying the PDF given by Eq. (A.7)

If we generate $\mathcal{N}$ independent displacements, then the points $\boldsymbol{\psi}_\ell = \hat{\boldsymbol{\psi}} + \delta\boldsymbol{\psi}_\ell$ give us the approximation

$$p(\boldsymbol{\psi}|\boldsymbol{\phi}, D) = \mathcal{N}^{-1}\sum_\ell \delta(\boldsymbol{\psi} - \boldsymbol{\psi}_\ell) \tag{A.10}$$

which is exact in the limit $\mathcal{N} \to \infty$.

References

Casertano, S., Lattanzi, M. G., Sozzetti, A., et al. 2008, A&A, 482, 699
Gentle, J. E. 2009, Computational Statistics (New York : Springer)
James, F. 2006, Statistical Methods in Experimental Physics (Singapore: World Scientific Publishing Co)
Jaynes, E. T. 2003, Probability Theory, The Logic of Science (Cambridge: Cambridge University Press)
Lucy, L. B. 2014a, A&A, 563, A126 (L14a)
Lucy, L. B. 2014b, A&A, 565, A37 (L14b)
Lucy, L. B., & Sweeney, M. A. 1971, AJ, 76, 544
Perryman, M. 2009, Astronomical Applications of Astrometry (Cambridge: Cambridge University Press)
Pourbaix, D. 2002, A&A, 385, 686
Pourbaix, D. 2004, ASPC Conf. Ser., 318, 132
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes, 2nd edn. (Cambridge: Cambridge Univ. Press)
Sozzetti, A., Giacobbe, P., Lattanzi, et al. 2014, MNRAS, 437, 497