
# Joint inference from parallax and proper motions

prepared by: C.A.L. Bailer-Jones
reference: GAIA-C8-TN-MPIA-CBJ-081
issue: 1
revision: 0
date: 2017-03-14
status: Issued

## Abstract

I investigate the estimation of distance and transverse velocity (as speed and direction) from Gaia parallax and proper motions, using posterior sampling. The method is illustrated using TGAS data. My goal is not to suggest producing such estimates for the data releases, but rather to help educate users in their interpretation of the data.

# Document History

| Issue | Revision | Date | Author | Comment |
|---|---|---|---|---|
| 1 | 0 | 2017-03-14 | CBJ | Issued |

# Contents



# 1 Introduction

Determining the distance to a star from a noisy parallax is a non-trivial inference problem. This arises from the highly nonlinear transformation between these two quantities, coupled with the fact that distances cannot be negative and the existence, in general, of other constraints on plausible distances. I previously explored this problem and showed how it can be addressed in the usual Bayesian framework.[1] Here I extend the methodology to the three-dimensional inference of distance and transverse velocity (represented as transverse speed and direction) from the parallax and two proper motions.

# 2 Definitions and transformation

The data vector is the parallax, proper motion in right ascension, and proper motion in declination, written as the column vector

$$\mathbf{x} = (\varpi,\ \mu_{\alpha^*},\ \mu_\delta)^\mathsf{T} \tag{1}$$

[1] Bailer-Jones C.A.L., 2015, *Estimating distances from parallaxes*, PASP 127, 994.

with units mas, mas yr$^{-1}$, and mas yr$^{-1}$ respectively. All of these quantities can be positive, negative, or zero. The corresponding $3 \times 3$ covariance matrix for these three measurements is $C_x$. These are all provided by AGIS. The parameter vector we want to infer is the distance, tangential speed, and direction of travel (increasing anticlockwise from North), written as

$$\boldsymbol{\theta} \,=\, (r,\, v,\, \phi)^{\mathsf{T}} \tag{2}$$

with units pc, km s$^{-1}$, and radians respectively. Distance is positive and tangential speed is non-negative.

Given the true parameters, the *noise-free* prediction of the data vector $\mathbf{x}$ is given by the simple geometrical transformation

$$\mathbf{m} \,=\, \left( \frac{10^3}{r},\, \frac{10^3}{c_2}\frac{v\sin\phi}{r},\, \frac{10^3}{c_2}\frac{v\cos\phi}{r} \right)^{\mathsf{T}} \tag{3}$$

where $c_2 = \mathrm{AU\,km^{-1}\,yr^{-1}} = 4.74047$. The inverse transformation, to give the *nominal parameters* in terms of $\mathbf{x}$, is

$$\boldsymbol{\theta}_{\mathrm{nom}} \,=\, \left( \frac{10^3}{\varpi},\, c_2\frac{\sqrt{\mu_{\alpha^*}^2+\mu_\delta^2}}{\varpi},\, \arctan\left(\frac{\mu_{\alpha^*}}{\mu_\delta}\right) \right)^{\mathsf{T}} \tag{4}$$

Note, however, that because the data are noisy, this inverse transformation does *not* generally give the most suitable estimate for the parameters. We must instead use inference, as explained in the next section.

Although I will sample the posterior, it can still be informative to compute the first order Taylor expansion estimate of the variance in the parameters. This is given by

$$C_{\boldsymbol{\theta}} \,=\, J\, C_x\, J^{\mathsf{T}} \tag{5}$$

where

$$J \,=\, \left(\frac{\partial\boldsymbol{\theta}}{\partial\mathbf{x}}\right) \,=\, \begin{pmatrix} \frac{\partial r}{\partial\varpi} & \frac{\partial r}{\partial\mu_{\alpha^*}} & \frac{\partial r}{\partial\mu_\delta} \\ \frac{\partial v}{\partial\varpi} & \frac{\partial v}{\partial\mu_{\alpha^*}} & \frac{\partial v}{\partial\mu_\delta} \\ \frac{\partial\phi}{\partial\varpi} & \frac{\partial\phi}{\partial\mu_{\alpha^*}} & \frac{\partial\phi}{\partial\mu_\delta} \end{pmatrix} \,=\, \begin{pmatrix} -10^3\frac{1}{\varpi^2} & 0 & 0 \\ -c_2\frac{\mu}{\varpi^2} & c_2\frac{\mu_{\alpha^*}}{\varpi\mu} & c_2\frac{\mu_\delta}{\varpi\mu} \\ 0 & \frac{\mu_\delta}{\mu^2} & -\frac{\mu_{\alpha^*}}{\mu^2} \end{pmatrix} \tag{6}$$

is the Jacobian matrix, with $\mu^2 = \mu_{\alpha^*}^2 + \mu_\delta^2$. The Jacobian is evaluated at the measured data.

# 3 Inference

AGIS reports astrometric measurements and their corresponding covariance matrix. I interpret this here as a multidimensional Gaussian likelihood in the measurements. For the parallaxes and proper motions this is

$$P(\mathbf{x}|\boldsymbol{\theta}) \;=\; \frac{1}{(2\pi)^{3/2}|C_x|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\theta})^{\mathsf{T}} C_x^{-1}(\mathbf{x}-\boldsymbol{\theta})\right]. \tag{7}$$

The 3D posterior over the parameters $\boldsymbol{\theta}$ is

$$P(\boldsymbol{\theta}|\mathbf{x}) \;=\; \frac{1}{Z}\,P(\mathbf{x}|\boldsymbol{\theta})P(\boldsymbol{\theta}) \tag{8}$$

where $Z$ is a normalization constant and $P(\boldsymbol{\theta})$ is a prior. I use here a separable prior

$$P(\boldsymbol{\theta}) \;=\; P(r)P(v)P(\phi) \tag{9}$$

with

$$P(r) \;=\; \begin{cases} \dfrac{1}{2L^3}\, r^2 e^{-r/L} & \text{if } \; r>0 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$P(v) \;=\; \begin{cases} \dfrac{1}{B(\alpha,\beta)} \left(\dfrac{v}{v_{\max}}\right)^{\alpha-1} \left(1-\dfrac{v}{v_{\max}}\right)^{\beta-1} & \text{if } \; 0\le v\le v_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$P(\phi) \;=\; \frac{1}{2\pi} \tag{12}$$

for $L \ge 0$, $v_{\max} > 0$, $\alpha, \beta > 0$ and $B(\alpha,\beta)$ is the beta function. The prior over distance is the exponentially decreasing space density prior introduced in Bailer-Jones (2015), with length scale $L$ (I use 500 pc in the experiments in the next section, although this could be adapted according to the line-of-sight: see CBJ-080). The prior over speed is a beta distribution, which is only non-zero between 0 and $v_{\max}$. The shape of the distribution is controlled by the two parameters $\alpha$ and $\beta$. An example is shown in Figure 1. In the experiments which follow I adopt $v_{\max} = 750\,\mathrm{km\,s^{-1}}$, which seems reasonable for Milky Way stars. The prior over the angle $\phi$ is uniform. These priors are just adopted here for the sake of illustration.

# 4 Examples

The posterior does not have a simple form, so I characterize it via Monte Carlo sampling.

I initially tried the Metropolis algorithm with a multivariate Gaussian proposal distribution. The covariance of this proposal distribution I set to $C_{\boldsymbol{\theta}}$ (equation 5), or a fraction thereof. I initialized

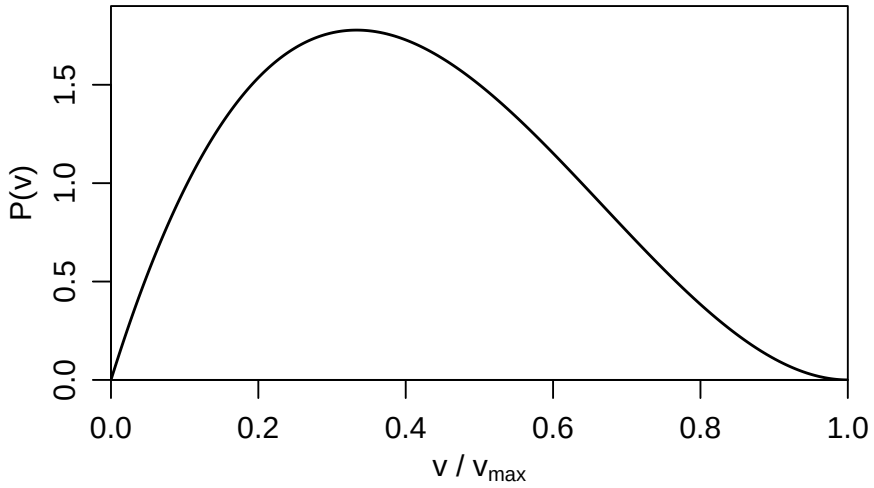



FIGURE 1: A beta distribution prior for the transverse speed $v$ with $\alpha = 2$ and $\beta = 3$.

at $\boldsymbol{\theta}_{\rm nom}$ (equation 4), run the chain for $10^4$ samples with a burn-in of $10^2$ and then thinned the chains by a factor of 10. The results were generally poor, in the sense that the chains looked very unsettled.

Much better performance was obtained with the emcee algorithm. In what follows I initialize the walkers over a narrow uniform distribution defined to always respect the parameter boundary conditions. I use 200 walkers and sample for 1000 iterations (preceded by an additional burn-in of 500 iterations). I thin the resulting chain by a factor of 20 in the walkers and 10 in the iterations, leaving a total of 1000 samples. I use a periodic boundary condition on $\phi$ to ensure that the walkers don't diverge to very large positive or negative values. This is achieved by wrapping the proposed angle to the range 0 to $2\pi$ (by taking the modulus).[2] The acceptance rate was generally between 0.5 and 0.6.

Results from using emcee for 10 different stars (two per page) are shown on pages 7 to 11. For each star there are nine panels, with the Tycho name indicated in the top right (along with a running number). The three panels in the left column show the evolution of the chains for each of the three parameters. The units are pc, km s$^{-1}$, and radians. The three panels in the central column show the one-dimensional (marginalized) posteriors as a density estimation or histogram. In these panels, the solid red vertical line shows the median of the posterior, the two dashed red lines the 5% and 95% percentiles. The blue vertical line here shows the nominal parameters ($\boldsymbol{\theta}_{\rm nom}$). The panels in the right column show the two-dimensional marginal posteriors as individual points.

The "native" range (that used inside emcee) of the angle $\phi$ is 0 to $2\pi$, and this is shown in the chains panel (left column). But if folding $\phi$ to $-\pi$ to $+\pi$ gives a smaller range – difference between max($\phi$) and min($\phi$) – then I use this instead in the one- and two-dimensional posterior

[2] Although it may at first seem that a more complicated procedures is required, it turns out not to be the case.

plots (central and right columns), and also use this to compute the summary statistics of the posterior. Note that computing statistics on a variable with periodic boundary conditions is ambiguous if it covers more than half the range, which here means more than $\pi$. The median, for example, is generally different depending on whether we represent the angle in the range $(0, 2\pi)$ or $(-\pi, +\pi)$ (and it doesn't just differ by $2\pi$). The same goes for the other quantiles, the mean, and the other moments. Of course if the direction is uncertain by more than $\pi$, there isn't much point summarizing it anyway.

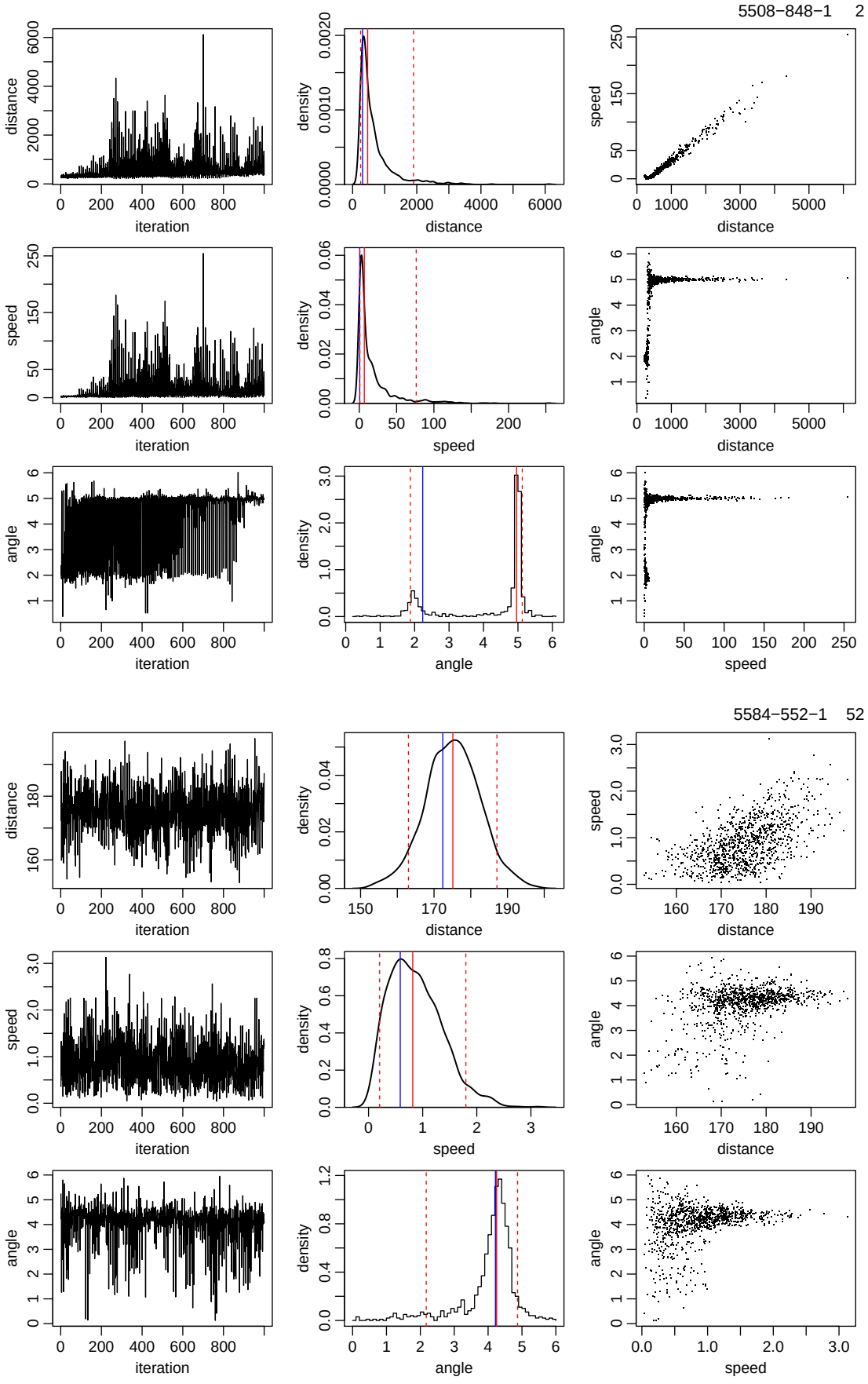
In many cases we see that the marginal one-dimensional posteriors for the parameters are simple, unimodal profiles, implying a well-defined estimate. We also see in many cases a strong correlation between the distance and speed. This is not surprising, as decreasing the parallax to the stars keeping the proper motions fixed would increase both the distance and the transverse speed. The correlation coefficient between distance and speed as estimated from element $(1, 2)$ of the covariance matrix $C_{\boldsymbol{\theta}}$ is sometimes also very large (e.g. 0.98).

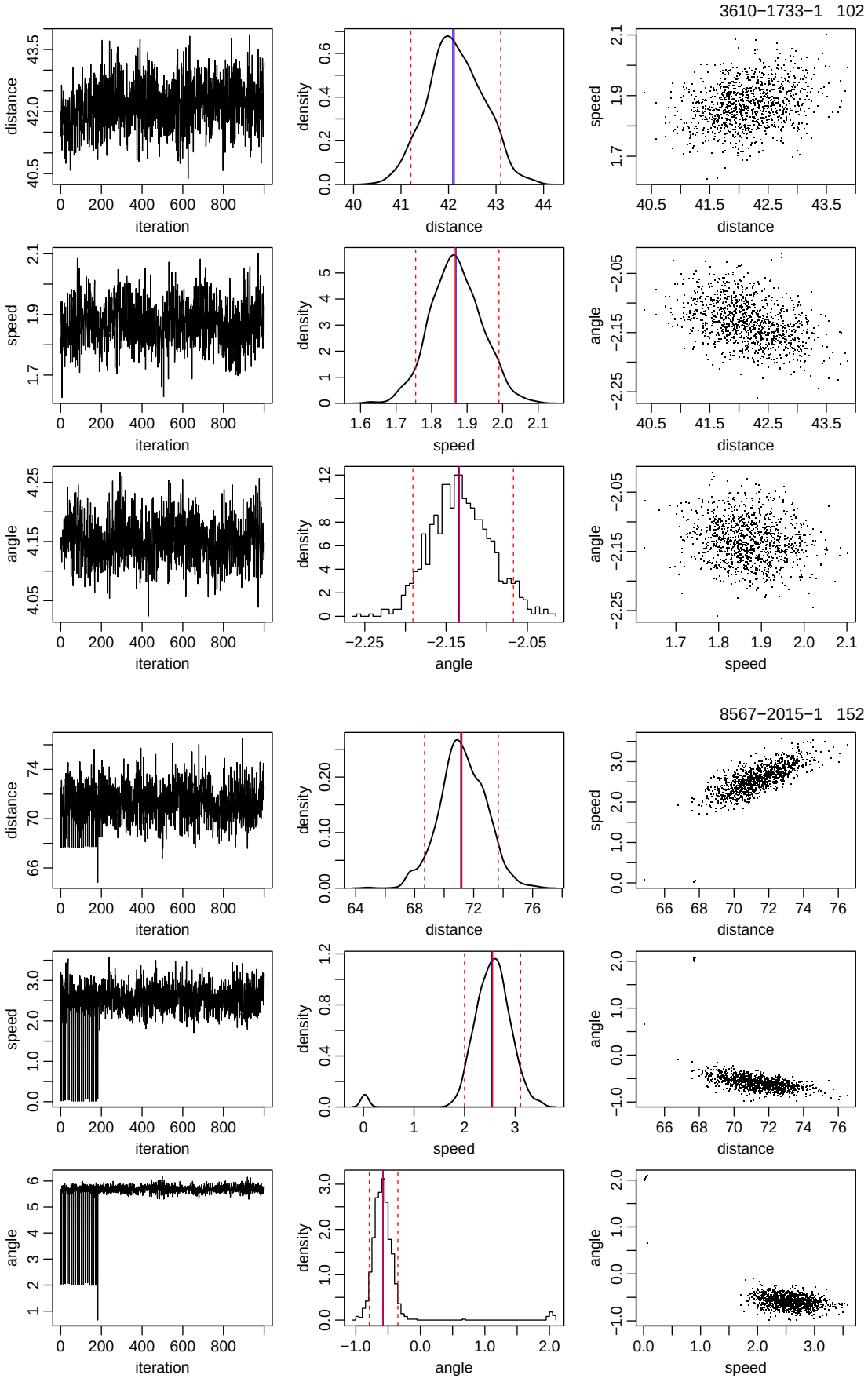
Table 2 summarizes the data and posteriors for these examples. I use the median (of each marginal posterior) as an estimator of the posterior, because it is easier to compute than the mode and more robust than the mean. I (therefore) use the quantiles at 0.05 and 0.95 to define the equal-tailed 90% confidence interval to represent the uncertainty. The correlation coefficients are computed from the posterior samples.

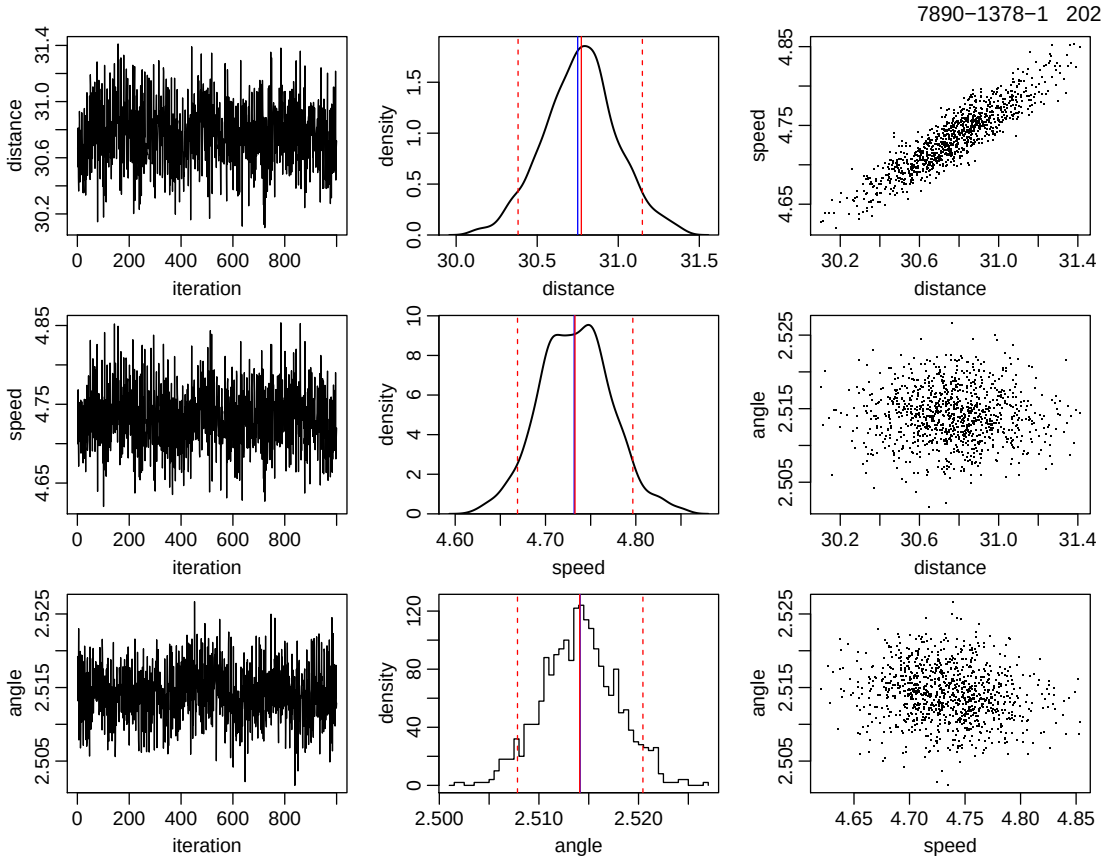
# 5 Conclusions

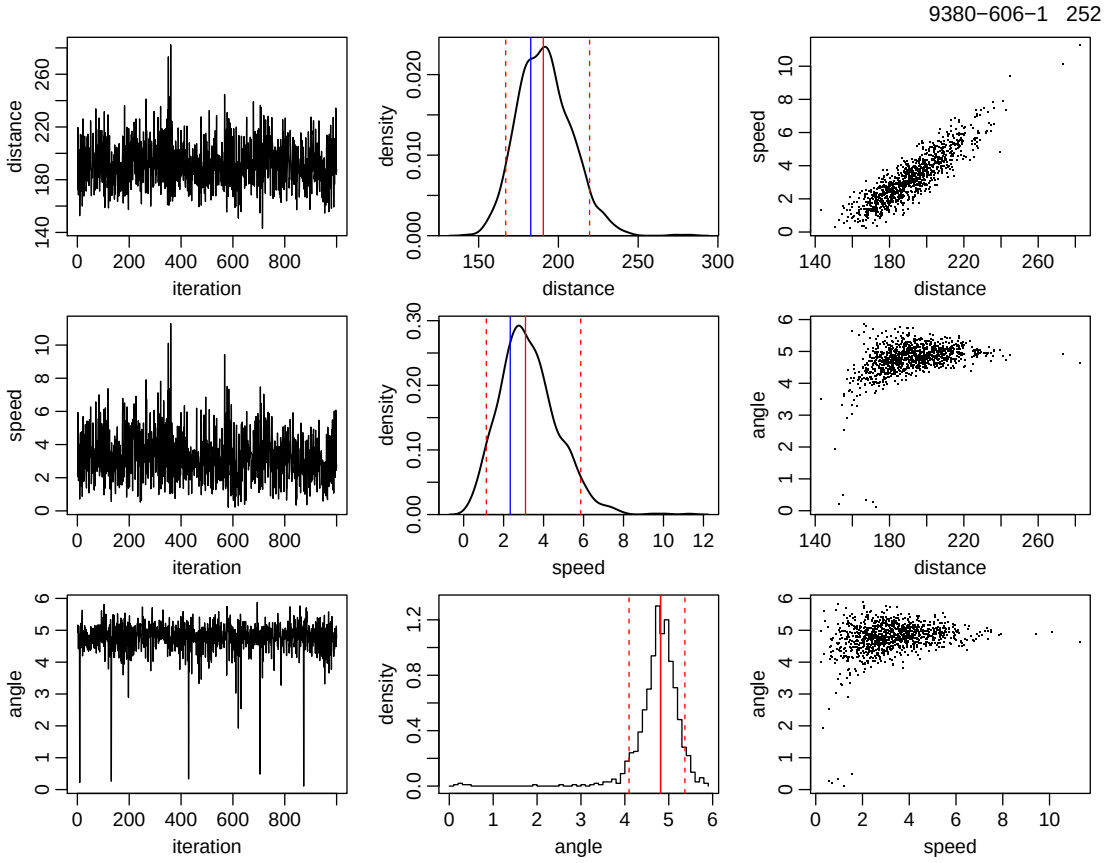
I have shown how to infer the probability distribution over three physical parameters from three astrometric measurements. One could of course extend this to infer all six phase space parameters by including in the data also the two-dimensional position (RA, Dec) and the radial velocity. In that case a more useful target coordinate system for the parameters is Galactic coordinates, which therefore also depends on the assumed phase space coordinates of the Sun in the Galactic system.

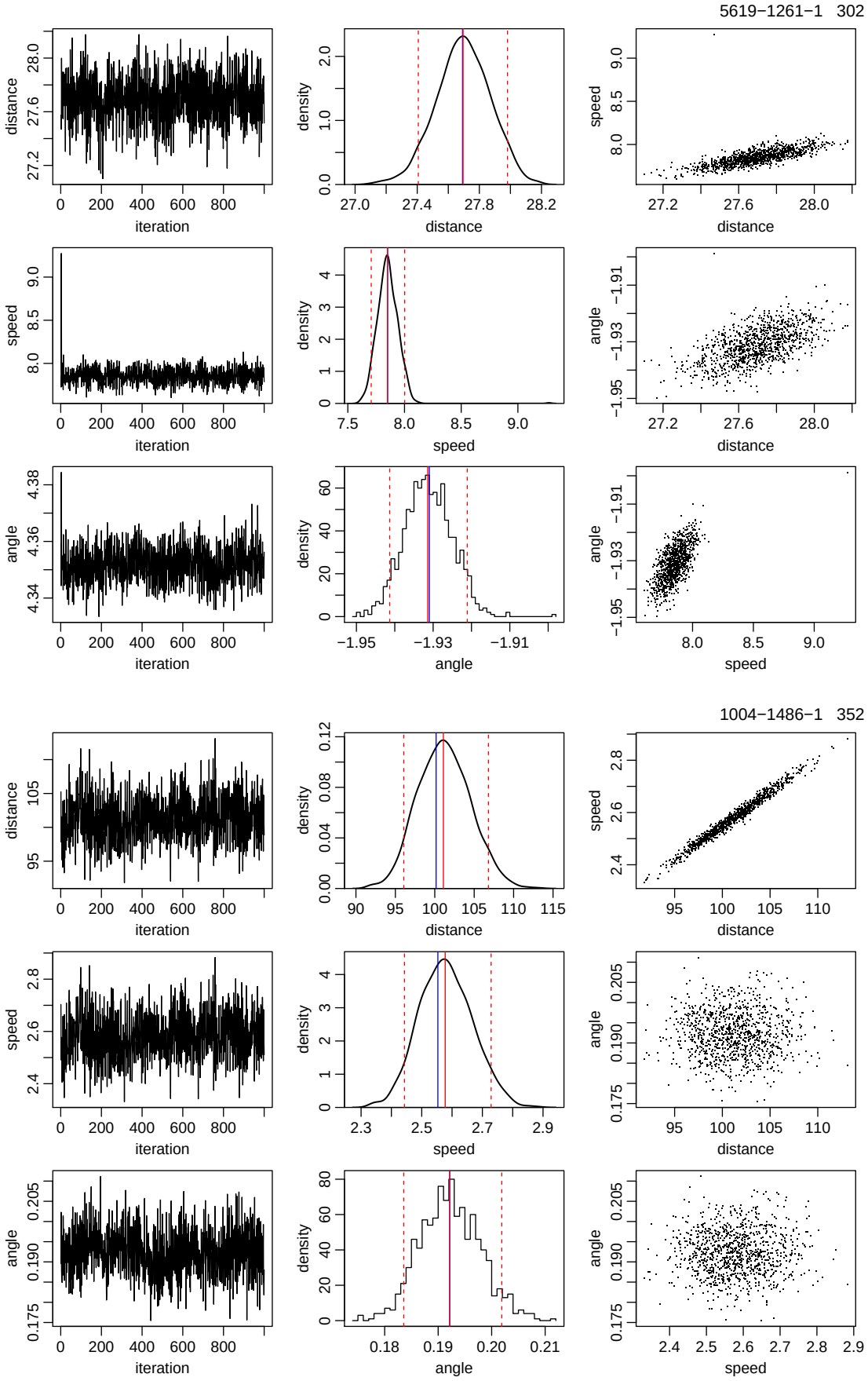
Inferring velocities is much more specialized than distance estimation, so I am not recommending that anything like the above be done to produce data products for the data releases. But this exercise does give some idea of how precisely velocities and distances can be estimated from given astrometric data and their uncertainties. As such as transformation is non-trivial, and the results not necessarily intuitive, this can be a useful aid, or educational tool.


5508−848−1 2
distance
density
speed
angle
iteration
5584−552−1 52


3610−1733−1 102
distance
density
speed
angle
iteration
8567−2015−1 152


7890−1378−1 202
distance
iteration
density
distance
speed
distance
speed
iteration
density
speed
angle
distance
angle
iteration
density
angle
angle
speed


9380−606−1 252
distance
iteration
density
distance
speed
distance
speed
iteration
density
speed
angle
distance
angle
iteration
density
angle
angle
speed


5619−1261−1 302
distance
density
speed
angle
iteration
1004−1486−1 352

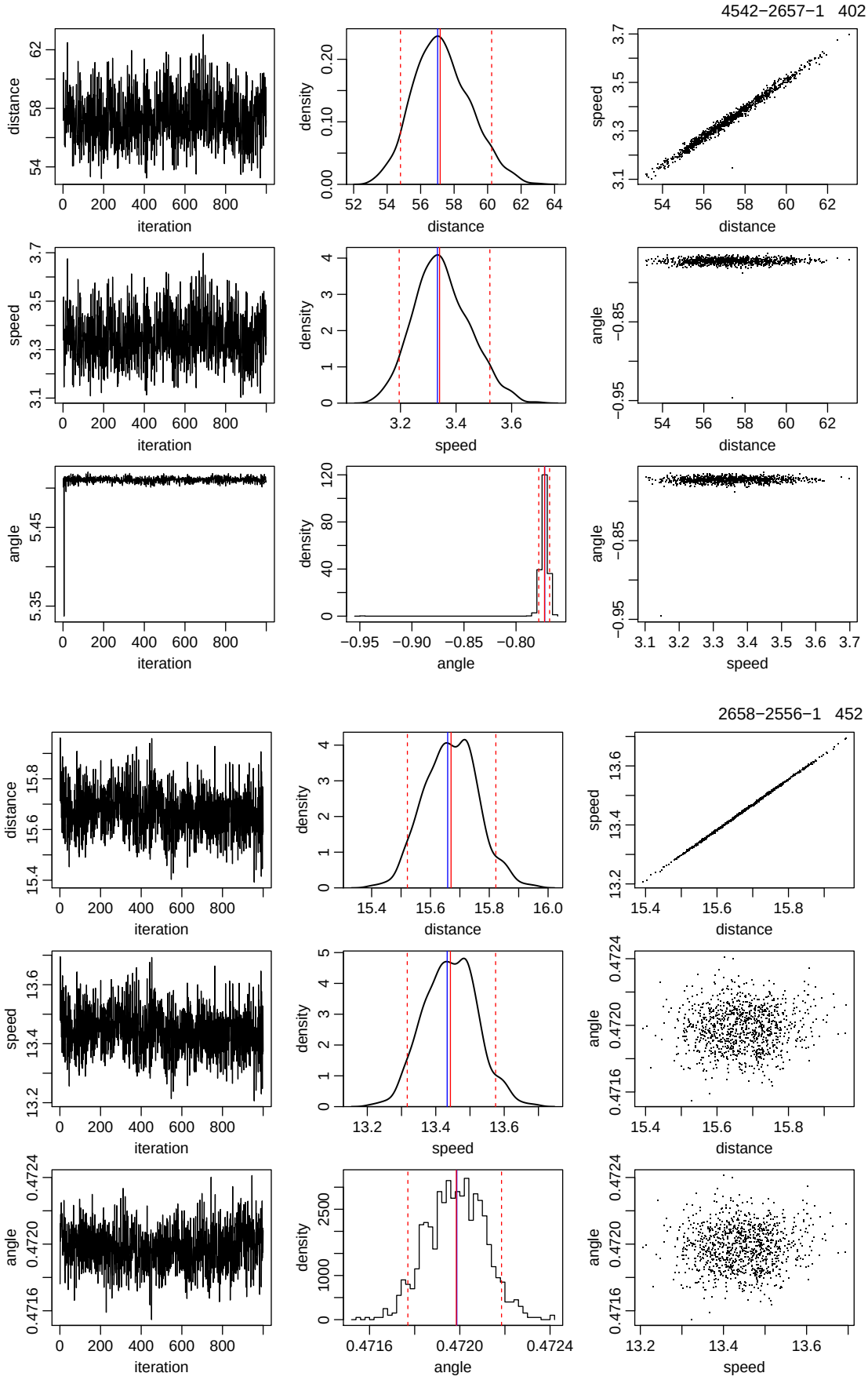

4542−2657−1 402
distance
density
speed
angle
iteration
2658−2556−1 452

TABLE 2: Summary of data and resulting posteriors. Columns 2–4 are the measured data, columns 5–7 are their corresponding standard uncertainties, columns 8–10 are the median of the resulting posteriors, columns 11–12, 13–14, and 15–16 are the 5% and 95% percentiles for the three parameters (respectively), and columns 17–19 are the three correlation coefficients (also computed from the posterior samples).

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tycho ID | $\varpi$ | $\mu_{\alpha^*}$ | $\mu_\delta$ | $\sigma_\varpi$ | $\sigma_{\mu_{\alpha^*}}$ | $\sigma_{\mu_\delta}$ | $r$ | $v$ | $\phi$ | $r_{05}$ | $r_{95}$ | $v_{05}$ | $v_{95}$ | $\phi_{05}$ | $\phi_{95}$ | $\rho(r,v)$ | $\rho(r,\phi)$ | $\rho(v,\phi)$ |
| | mas | mas/yr | mas/yr | mas | mas/yr | mas/yr | pc | km/s | rad | pc | pc | km/s | km/s | rad | rad | | | |
| 5508-848-1 | 3.237 | 0.312 | -0.244 | 0.739 | 2.297 | 0.800 | 469.7 | 6.71 | 4.96 | 250.0 | 1899.8 | 0.81 | 76.28 | 1.88 | 5.12 | 0.99 | 0.39 | 0.33 |
| 5584-552-1 | 5.801 | -0.631 | -0.341 | 0.248 | 0.687 | 0.362 | 175.1 | 0.82 | 4.25 | 163.0 | 187.1 | 0.20 | 1.80 | 2.17 | 4.87 | 0.55 | 0.36 | 0.29 |
| 3610-1733-1 | 23.759 | -7.918 | -5.003 | 0.332 | 0.347 | 0.349 | 42.1 | 1.87 | -2.13 | 41.2 | 43.1 | 1.76 | 1.99 | -2.19 | -2.07 | 0.27 | -0.41 | -0.16 |
| 8567-2015-1 | 14.060 | -4.110 | 6.315 | 0.283 | 1.015 | 0.757 | 71.2 | 2.56 | -0.58 | 68.7 | 73.7 | 2.00 | 3.11 | -0.79 | -0.35 | 0.73 | -0.49 | -0.75 |
| 7890-1378-1 | 32.521 | 19.057 | -26.278 | 0.237 | 0.140 | 0.101 | 30.8 | 4.73 | 2.51 | 30.4 | 31.1 | 4.67 | 4.80 | 2.51 | 2.52 | 0.90 | -0.02 | -0.15 |
| 9380-606-1 | 5.474 | -2.679 | 0.272 | 0.446 | 1.357 | 1.132 | 190.5 | 3.10 | 4.81 | 167.0 | 219.5 | 1.13 | 5.86 | 4.10 | 5.37 | 0.89 | 0.40 | 0.26 |
| 5619-1261-1 | 36.112 | -55.970 | -21.080 | 0.231 | 0.483 | 0.291 | 27.7 | 7.85 | -1.93 | 27.4 | 28.0 | 7.71 | 8.00 | -1.94 | -1.92 | 0.68 | 0.58 | 0.65 |
| 1004-1486-1 | 9.983 | 1.027 | 5.278 | 0.326 | 0.032 | 0.028 | 101.1 | 2.58 | 0.19 | 96.1 | 106.8 | 2.44 | 2.73 | 0.18 | 0.20 | 0.99 | -0.01 | 0.02 |
| 4542-2657-1 | 17.539 | -8.606 | 8.832 | 0.472 | 0.042 | 0.041 | 57.2 | 3.34 | -0.77 | 54.8 | 60.2 | 3.20 | 3.52 | -0.78 | -0.77 | 0.99 | 0.03 | 0.09 |
| 2658-2556-1 | 63.863 | 82.280 | 161.185 | 0.384 | 0.022 | 0.023 | 15.7 | 13.44 | 0.47 | 15.5 | 15.8 | 13.32 | 13.57 | 0.47 | 0.47 | 1.00 | 0.05 | 0.05 |