Distances and velocities from parallaxes and proper motions - Doing it with just 1D sampling

Coryn A.L. Bailer-Jones
Max Planck Institute for Astronomy, Heidelberg
http://www.mpia.de/homes/calj
This version: 2024-02-12

1 Abstract

The kinegeometric posterior is a non-Gaussian 3D PDF in distance and velocity (Bailer-Jones 2023; hereafter paper VI). In paper VI I computed summary statistics of this posterior – specifically the quantiles of the marginal distributions – by sampling it in 3D using MCMC. Here I show that a 3D sampling can be replaced by a 1D sampling of the marginal distance posterior. This probably requires far fewer samples to achieve the same accuracy, so is likely to be an order of magnitude faster.

2 Notation

Both the proper motion and velocity are 2D vectors. For brevity I sometimes write $\mu = (\mu_\alpha, \mu_\delta)$ and $v = (v_\alpha, v_\delta)$. I write the proper motion in RA as μ_α rather than $\mu_{\alpha*}$, but it has angular units per unit time, i.e. it is "corrected" for declination. $\mathcal{N}_x(m_x, \Sigma_x)$ denotes a (generally multivariate) Gaussian in x with mean m_x and covariance Σ_x. The subscript kg will denote (kinegeometric) posteriors.

3 Summary of the procedure

1. Sample the 1D marginal distance posterior $P^*_{\rm kg}(r \mid \varpi, p)$ $(\equiv g(r))$ in equation 11b to provide a set of N_g samples $\{r_i\}$.
2. For each sample r_i, the *conditional* posterior velocity PDF $P^*_{\rm kg}(v \mid r_i, \varpi, \mu, p)$ $(\equiv f(r_i; v))$ in equation 13 is a 2D Gaussian in velocity with mean and covariance given by equations 9 (using also equations 4 and 2). The average of these N_g Gaussians (equation 12c) gives the *marginal* 2D velocity posterior. We don't explicitly average the distributions. Instead, draw n samples of (v_α, v_δ) from each of the N_g functions $f(v; r_i)$. A value of $n = 1$ is probably sufficient. These samples represent the 2D marginal velocity posterior, from which we can compute summary statistics for velocity.
3. Compute the covariances between the distance and the two velocity components using equation 15d.

4 Details

A noise-free proper motion μ (angular velocity) is related to the distance r and transverse velocity v via

$$\frac{v}{\mathrm{km\,s^{-1}}} = k \, \frac{r}{\mathrm{kpc}} \frac{\mu}{\mathrm{mas\,yr^{-1}}} \quad \text{with} \quad k = 4.740471 \, . \tag{1}$$

The unnormalized 3D posterior PDF over distance and velocity is a product of a 3D Gaussian likelihood in parallax ϖ and proper motion (μ_α, μ_δ), and the prior. The mean and covariance of the 3D likelihood are

$$m_3 = \left[m_\varpi, m_\mu\right]^\mathsf{T} = \left[\tfrac{1}{r}, \tfrac{v_\alpha}{kr}, \tfrac{v_\delta}{kr}\right]^\mathsf{T} \tag{2a}$$

$$\Sigma_3 = \begin{bmatrix} \Sigma_{\varpi\varpi} & \Sigma_{\varpi\mu} \\ \Sigma_{\mu\varpi} & \Sigma_{\mu\mu} \end{bmatrix} = \begin{bmatrix} \sigma_\varpi^2 & \sigma_\varpi \sigma_{\mu_\alpha} \rho_{\varpi\mu_\alpha} & \sigma_\varpi \sigma_{\mu_\delta} \rho_{\varpi\mu_\delta} \\ \sigma_\varpi \sigma_{\mu_\alpha} \rho_{\varpi\mu_\alpha} & \sigma_{\mu_\alpha}^2 & \sigma_{\mu_\alpha} \sigma_{\mu_\delta} \rho_{\mu_\alpha\mu_\delta} \\ \sigma_\varpi \sigma_{\mu_\delta} \rho_{\varpi\mu_\delta} & \sigma_{\mu_\alpha} \sigma_{\mu_\delta} \rho_{\mu_\alpha\mu_\delta} & \sigma_{\mu_\delta}^2 \end{bmatrix} \tag{2b}$$

where in equation 2a m_μ is the two-element vector with the velocities, and in equation 2b $\Sigma_{\varpi\varpi}$ is the scalar in the top-left, $\Sigma_{\mu\mu}$ is the 2×2 matrix in the bottom-right, and the other two are the two-element vectors.

I factorize the prior into a product of (1) a 2D Gaussian prior in velocity conditional on the distance (and direction p), $P(v_\alpha, v_\delta \mid r, p)$, with mean $m_\tau(r,p)$ and covariance $\Sigma_\tau(r,p)$ (τ for τransverse), and (2) a 1D non-Gaussian prior over distance, $P(r \mid p)$. In paper VI the mean and covariance of the velocity prior were fit using a mock catalogue as a function of HEALpixel; the distance prior is the generalized gamma distribution from Bailer-Jones et al. (2021; hereafter paper V).

Using Bayes' theorem and then factorizing, the unnormalized posterior is[1]

$$P^*_{\mathrm{kg}}(r, v_\alpha, v_\delta \,|\, \varpi, \mu_\alpha, \mu_\delta, \Sigma_3, p) \tag{3a}$$

$$= \underbrace{P(\varpi, \mu_\alpha, \mu_\delta \,|\, r, v_\alpha, v_\delta, \Sigma_3)}_{\text{likelihood}} \; \underbrace{P(v_\alpha, v_\delta \,|\, r, p) \; P(r \,|\, p)}_{\text{prior}} \tag{3b}$$

$$= P(\mu_\alpha, \mu_\delta \,|\, \varpi, r, v_\alpha, v_\delta, \Sigma_2) P(v_\alpha, v_\delta \,|\, r, p) \; \underbrace{P(\varpi \,|\, r, \cancel{v_\alpha, v_\delta}, \Sigma_{\varpi\varpi}) P(r \,|\, p)}_{\text{geometric posterior}} \tag{3c}$$

where Σ_2 will be computed below, and the dependence on the velocity in the parallax likelihood (the penultimate term in the last line) drops out due to conditional independence. The missing normalization constant is $1/P(\varpi, \mu)$.

Step 1: First term of equation 3c, $P(\mu \,|\, \varpi, r, v, \Sigma_2)$.
The first of the four terms on the right side of equation 3c is the 2D likelihood in proper motion conditioned on the parallax, for given v and r. The conditional of a Gaussian is

[1] For brevity I exclude symbols to represent the various parameters of the priors, but these can be considered represented by the direction p.

also a Gaussian, $\mathcal{N}_\mu(m_2, \Sigma_2)$, with mean and covariance[2] computed from equation 2 as

$$m_2 = m_\mu + \underbrace{\Sigma_{\mu\varpi}\Sigma_{\varpi\varpi}^{-1}(\varpi - m_\varpi)}_{X_\mu} \tag{4a}$$

$$\Sigma_2 = \Sigma_{\mu\mu} - \Sigma_{\mu\varpi}\Sigma_{\varpi\varpi}^{-1}\Sigma_{\varpi\mu} \; . \tag{4b}$$

The term in the exponential of this Gaussian (dropping the factor of $-1/2$) can be rewritten (using equation 2a for m_μ) as

$$\begin{aligned} &\left[\mu - \left(\frac{v}{kr} + X_\mu\right)\right] \Sigma_2^{-1} \left[\mu - \left(\frac{v}{kr} + X_\mu\right)\right]^\mathsf{T} \\ &\quad = (v - kr[\mu - X_\mu])(k^2r^2\Sigma_2)^{-1}(v - kr[\mu - X_\mu])^\mathsf{T} \end{aligned} \tag{5}$$

both of which are independent of v. This has the form of a Gaussian in v, $\mathcal{N}_v$, but we need to be careful with the normalization constant: $\mathcal{N}_\mu \, d\mu = \mathcal{N}_v \, dv = \mathcal{N}_v \, kr \, dv$. Therefore

$$P(\mu \,|\, \varpi, r, v, \Sigma_2) = kr \, \mathcal{N}_v(kr[\mu - X_\mu], kr\Sigma_2) \; . \tag{6}$$

We have transformed from a Gaussian in μ to a Gaussian in v, which is possible because $m_\mu \propto v$ (equation 2a). The factor of kr arises from the change of variables (it gets the units right). Note that if the proper motion and parallax are uncorrelated (i.e. the likelihood factorizes), then $\Sigma_{\mu\varpi} = 0$ (from equation 2b) in equation 4a and so this PDF becomes independent of the parallax.

Step 2: First two terms of equation 3c, $P(\mu \,|\, \varpi, r, v, \Sigma_2)P(v \,|\, r, p)$.
We now see this as a product of two Gaussians in v

$$\begin{aligned} &\frac{kr}{\sqrt{\det(2\pi k^2r^2\Sigma_2)\det(2\pi\Sigma_\tau)}} \exp\left[-\frac{1}{2}(v - kr[\mu - X_\mu])(k^2r^2\Sigma_2)^{-1}(v - kr[\mu - X_\mu])^\mathsf{T}\right] \times \\ &\qquad \exp\left[-\frac{1}{2}(v - m_\tau)\Sigma_\tau^{-1}(v - m_\tau)^\mathsf{T}\right] \; . \end{aligned} \tag{7}$$

Recall that both m_τ and Σ_τ are functions of distance (and direction). By expanding and

[2] The new covariance is the inverse of the sum of the inverse covariances. The new mean is the inverse covariance weighted sum of the means, multipled by the new covariance. See `https://statproofbook.github.io/P/mvn-cond.html`.

then completing the square, we can write this[3] as a *non-normalized* Gaussian in v

$$Q(r,\varpi,\mu)\,\mathcal{N}_v(m_v,\Sigma_v) \quad \text{where} \tag{8a}$$

$$Q(r,\varpi,\mu) = \frac{kr}{\sqrt{\det(2\pi[k^2r^2\Sigma_2+\Sigma_\tau])}} \times \tag{8b}$$

$$\exp\left[-\frac{1}{2}(kr[\mu-X_\mu]-m_\tau)(k^2r^2\Sigma_2+\Sigma_\tau)^{-1}(kr[\mu-X_\mu]-m_\tau)^{\mathsf{T}}\right] \tag{8c}$$

$$= kr\,\mathcal{N}_{m_\tau}(kr[\mu-X_\mu],k^2r^2\Sigma_2+\Sigma_\tau) \tag{8d}$$

and

$$m_v = ([k^2r^2\Sigma_2]^{-1}+\Sigma_\tau^{-1})^{-1}\,([k^2r^2\Sigma_2]^{-1}kr(\mu-X_\mu)+\Sigma_\tau^{-1}m_\tau) \tag{9a}$$

$$\Sigma_v = ([k^2r^2\Sigma_2]^{-1}+\Sigma_\tau^{-1})^{-1}\,. \tag{9b}$$

$Q(r,\varpi,\mu)$ also depends on the covariances of the data and the mean and covariance of the velocity prior, but it does not depend on v. Note that $Q\mathcal{N}_v$ has units of $krv^{-1}\times v^{-1} = \mu^{-1}v^{-1}$, which correctly retains the units of the original term.

Step 3: The 3D joint distance–velocity posterior.
We can now write the kinegeometric posterior (expression 3a) as

$$P^*_{\text{kg}}(r,v\,|\,\varpi,\mu,\Sigma_3,p) = \underbrace{\mathcal{N}_v(m_v,\Sigma_v)}_{f(r;v)}\,\underbrace{Q(r,\varpi,\mu)\,P(\varpi\,|\,r,\Sigma_{\varpi\varpi})P(r\,|\,p)}_{g(r)} \tag{10a}$$

$$= P^*_{\text{kg}}(v\,|\,r)\,P^*_{\text{kg}}(r) \tag{10b}$$

which defines $f(r;v)$ and $g(r)$. That these are equal to $P^*_{\text{kg}}(v\,|\,r)$ and $P^*_{\text{kg}}(r)$ respectively will be shown below. Note that both of these functions depend on ϖ and μ.

Step 4: 1D marginal distance posterior.
Integrating equation 10 over v we get the marginal posterior in distance. As the v-dependent term is just a normalized Gaussian in v, this integral is unity.

$$P^*_{\text{kg}}(r\,|\,\varpi,\mu,\Sigma_3,p) = \int P^*_{\text{kg}}(r,v\,|\,\varpi,\mu,\Sigma_3,p)\,dv \tag{11a}$$

$$= Q(r,\varpi,\mu)\,\underbrace{P(\varpi\,|\,r,\Sigma_{\varpi\varpi})\,P(r\,|\,p)}_{\text{geometric posterior}} \equiv g(r)\,. \tag{11b}$$

All terms depend on the distance. This is a (complicated) 1D function of distance which we can sample with a Monte Carlo method.

[3]See `https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.113.6244` (the matrix cookbook). This can be written in an alternative way involving just one inverse according to `https://math.stackexchange.com/questions/157172/product-of-two-multivariate-gaussians-distributions`, but I've not found a confirmation/derivation of this.

Step 5: 2D marginal velocity posterior.
Integrating equation 10 over r we get the marginal posterior in velocity

$$\begin{aligned} P^*_{\rm kg}(v \,|\, \varpi, \mu, \Sigma_3, p) &= \int P^*_{\rm kg}(r, v \,|\, \varpi, \mu, \Sigma_3, p)\, dr && (12a) \\ &= \int f(r; v)\; g(r)\, dr && (12b) \\ &\simeq \frac{1}{N_g} \sum_{i=1}^{N_g} f(r_i; v) \quad \text{where} \quad r_i \in g(r)\;. && (12c) \end{aligned}$$

This is a Monte Carlo integral: we draw N_g distance samples[4] from $g(r)$ and then take the average of f computed at these.[5] Each $f(r_i; v)$ is a Gaussian distribution in v at the particular r_i. Thus the (unnormalized) marginal posterior PDF in velocity is just a sum of Gaussians (which is not generally a Gaussian). To sample these, we draw n samples from each of the N_g Gaussians. $n = 1$ may well suffice, for example if the spread of the velocity components ($f(r_i; v)$) is larger than their typical widths.[6] This set of samples – each is a vector (v_α, v_δ) – characterizes the 2D marginal velocity posterior. From these we can then compute summary statistics, such as 1D quantiles on the marginal velocities, or the 2D mean and covariance of the 2D distribution.[7]

We see that the conditional velocity posterior at a given distance r is

$$P^*_{\rm kg}(v \,|\, r) \;=\; \frac{P^*_{\rm kg}(v, r)}{P^*_{\rm kg}(r)} \;=\; \mathcal{N}_v(m_v, \Sigma_v) \;\equiv\; f(r; v) \tag{13}$$

by diving equation 10a by equation 11b. This is a 2D Gaussian with mean and covariance from equation 9 evaluated at the specified distance.

Step 6: Covariance between distance and velocity.
We are almost done. In addition to summary statistics on the two marginal distributions, we should also report the correlation coefficients between the distance and each velocity component. These are often strongly correlated, because for a given proper motion an increase in the distance can be compensated for by a decrease in the velocity.

[4]Unfortunately the samples we draw here are not the same as those we need for computing the distance posterior in equation 11b.

[5]In the formal definition g is normalized

$$\frac{\int g(r) f(r)\, dr}{\int g(r)\, dr} \;=\; \langle f(r) \rangle \;\simeq\; \frac{1}{N} \sum_{i=1}^{N} f(r_i)$$

but this is not required in practice because the set of samples is the same whether drawn from a normalized or an unnormalized distribution.

[6]Strictly we should make N draws, and for each draw first select at random which of the N_g Gaussians we drawn from. But for large N and N_g these procedures will be statistically equivalent.

[7]If $n > 1$ then we need to first combine the n velocity from each velocity component at each r_i before computing variances and covariances, otherwise we will get an incorrect contribution from the intra-component variation.

These correlation coefficients we can estimate from the distance samples, as follows.[8] First we compute the expectation value of v under the posterior (omitting many dependencies; steps explained below):

$$\begin{aligned}
E[v] &= \iint P^*_{\mathrm{kg}}(r,v)\, v\, dv\, dr & \text{(14a)}\\
&= \iint P^*_{\mathrm{kg}}(v\,|\,r)\, v\, P^*_{\mathrm{kg}}(r)\, dv\, dr & \text{(14b)}\\
&\simeq \frac{1}{N_g}\sum_{i=1}^{N_g} \int P^*_{\mathrm{kg}}(v\,|\,r_i)\, v\, dv \quad \text{where} \quad r_i \in P^*_{\mathrm{kg}}(r)\ . & \text{(14c)}\\
&\simeq \frac{1}{N_g}\sum_{i=1}^{N_g} E[v\,|\,r_i]\ . & \text{(14d)}
\end{aligned}$$

In going from line 14b to line 14c I approximate the integral over r with a sum over the Monte Carlo samples already drawn from $P^*_{\mathrm{kg}}(r)$ (equation 11b). The integral in line 14c is just the expectation value of the velocity posterior at $r = r_i$, which I write as $E[v\,|\,r_i]$. This is the mean of the conditional velocity posterior in equation 13, which is equation 9a evaluated at $r = r_i$.

Taking a similar appoach, we compute the 2-element covariance between r and v as

$$\begin{aligned}
\Sigma_{rv} &= \iint P^*_{\mathrm{kg}}(r,v)(r - E[r])(v - E[v])\, dr\, dv & \text{(15a)}\\
&= \iint P^*_{\mathrm{kg}}(r)(r - E[r])\, dr\ P^*_{\mathrm{kg}}(v\,|\,r)(v - E[v])\, dv & \text{(15b)}\\
&\simeq \frac{1}{N_g}\sum_{i=1}^{N_g} (r_i - E[r]) \int P^*_{\mathrm{kg}}(v\,|\,r_i)(v - E[v])\, dv \quad \text{where} \quad r_i \in P^*_{\mathrm{kg}}(r) & \text{(15c)}\\
&\simeq \frac{1}{N_g}\sum_{i=1}^{N_g} (r_i - E[r])(E[v\,|\,r_i] - E[v]) & \text{(15d)}
\end{aligned}$$

where $E[r] \simeq \frac{1}{N_g}\sum_{i=1}^{N_g} r_i$. In going from line 15b to line 15c I again approximate the integral over r with a sum over the Monte Carlo samples already drawn from $P^*_{\mathrm{kg}}(r)$. We can then take $(r_i - \overline{r})$ outside of the v integral, because this integral is now computed conditioned on each individual r_i. In line 15c this v integral is the expectation value of $(v - E[v])$ at $r = r_i$, which is written as such in line 15d. $E[v]$ is given by equation 14d (this of course only has to be computed once).

Note that we have only used the N_g distance samples. We do not use the velocity samples explicitly.

We can calculate the variance–covariance of the two velocity components in a similar way, and the variance of the distance directly from the distance samples. With those we can then compute the correlation coefficients.

[8] Once we have the set of velocity samples, the calculations shown here are equivalent to the obvious approach of calculating the covariance from the two sets of velocity and distance samples about their means.

5 Conclusion

To compute the 3D distance–velocity posterior we need to draw N_g samples from the marginal 1D distance posterior, as well as the same number of draws from N_g 2D Gaussians. In paper VI I used 5200 posterior evaluations (with emcee) to compute the 3D posterior, including burn-in. In paper V we used 550 posterior evaluations (with Metropolis) to compute the 1D distance posterior, including burn-in. Assuming the two types of posterior evaluation take the same time, and that the Gaussian velocity sampling and the various 2×2 matrix products takes essentially zero time, then the method presented here should given a speed-up by a factor of nearly ten.

My thanks go to Rene Andrae for discussions and assistance.

References

Bailer-Jones C.A.L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, *Estimating distances from parallaxes. V. Geometric and photogeometric distances to 1.47 billion stars in Gaia Early Data Release 3*, AJ 161, 147 [paper V]

Bailer-Jones C.A.L., 2023, *Estimating distances from parallaxes. VI. A method for inferring distances and transverse velocities from parallaxes and proper motions demonstrated on Gaia Data Release 3*, AJ, in press [paper VI]