Bayesian Inference of Galactic Rotation Using Gaia DR2 Data

Colombo Gabriele, Maszkowski Michal, Seghezzi Daniele

April 8, 2025

Abstract

In this report, we present a Bayesian model for the rotational motion of stars in the Milky Way galaxy, using observational data from the Gaia Data Release 2 (DR2) [**GAIADR2**]. We begin by introducing a simplified physical model in which both the stars and the Local Standard of Rest (LSR) follow circular orbits around the Galactic Center (GC) with a constant velocity V_{rot}. The Sun is modeled as moving relative to the LSR with components $U_{\odot}$ (in the x-direction) and $V_{\odot}$ (in the y-direction), all defined within the GC frame. We then extend this model by introducing a random velocity component $v'_{\text{rand}} \sim \mathcal{N}(0, \sigma)$ for each star, to account for intrinsic stellar motion not captured by pure circular rotation.

Using Bayesian inference and applying Markov Chain Monte Carlo (MCMC) sampling via the `emcee` [**EMCEE**] package, we estimate the posterior distributions of the model parameters. The inference is performed using stellar longitude, parallax, and radial velocity measurements, along with their associated uncertainties, all expressed in the Sun-centered reference frame.

For the simpler model, we estimate the parameters (with 95% confidence intervals) as: $V_{\text{rot}} = 211.45^{+0.04}_{-0.04}$ $\text{km}\,\text{s}^{-1}$; $U_{\odot} = 11.638^{+0.005}_{-0.005}$ $\text{km}\,\text{s}^{-1}$; $U_{\odot} = 21.604^{+0.005}_{-0.005}$ $\text{km}\,\text{s}^{-1}$.

In the extended model including the velocity dispersion, we find: $V_{\text{rot}} = 204^{+2}_{-2}$ $\text{km}\,\text{s}^{-1}$; $U_{\odot} = 11.7^{+0.3}_{-0.3}$ $\text{km}\,\text{s}^{-1}$; $U_{\odot} = 21.7^{+0.3}_{-0.3}$ $\text{km}\,\text{s}^{-1}$; $\sigma = 30.6^{+0.2}_{-0.2}$ $\text{km}\,\text{s}^{-1}$.

1 Data

In this work, we utilize stellar data from Gaia Data Release 2 (DR2) [**GAIADR2**]. From the extensive Gaia catalog, we select only stars for which radial velocities v_{rad} were measured relative to the Sun via the Doppler effect. To manage the dataset size efficiently, we apply a random subsampling procedure, restricting the dataset to entries with a random index less than 10^8.

For each selected star, we extract key parameters from Gaia DR2: parallax p and its associated uncertainty σ_p, radial velocity v_{rad} with measurement uncertainty σ_v, and Galactic coordinates — specifically, latitude b and longitude l.

To focus our analysis on stars approximately lying in the Galactic plane, we apply a cut on the latitude: $|b| < 5°$. To ensure data quality, we retain only stars with a relative parallax uncertainty below 20% and a radial velocity uncertainty less than 5 $\text{km}\,\text{s}^{-1}$.

After applying these criteria, we obtain a sample of $\text{N}_{\text{stars}} = 75{,}659$ stars. The upper panel of Figure 1 shows the distribution of their radial velocities with respect to the Sun, plotted as a function of Galactic longitude. For clarity, only values within the range $[-200, 200]$ $\text{km}\,\text{s}^{-1}$ are displayed, although the full dataset includes velocities extending up to approximately 500 $\text{km}\,\text{s}^{-1}$. These extreme values are likely outliers, and a more refined data-cleaning procedure could further enhance the quality of the analysis.

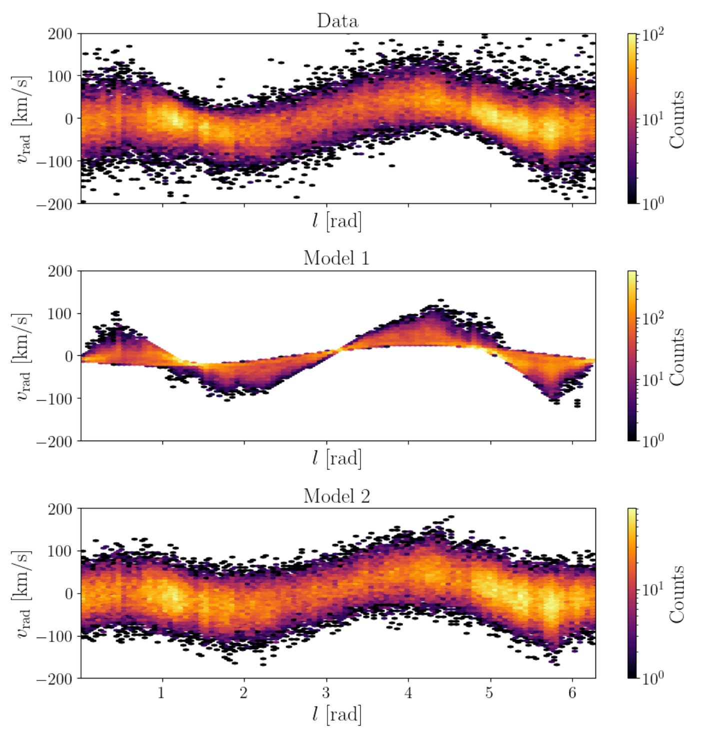

Figure 1: Upper: observed radial velocities in the range $[-200, 200]$ $\text{km}\,\text{s}^{-1}$. Middle: predicted radial velocities from the first model. Lower: simulated radial velocity distribution under the second model.

2 Model

In this section we present the physical and the statistical models we developed for the study of the Galactic kinematics of the Milky Way. Our physical assumptions, presented in the paragraph 2.1, lead to a prediction for the radial motion of stars with respect to the Sun as a function of their longitude and parallax. The prediction, for each star, can then be compared to the direct measurement. In the paragraph 2.2 we formally present our statistical assumptions on the data and the parameters of the two models with their priors and likelihoods.

2.1 Physical Model

In our model, we restrict attention to stellar motion confined to the Galactic plane ($b \approx 0$). We assume that each star moves in a circular orbit around the Galactic Center (GC) with a uniform speed V_{rot}. Fixing the coordinate system as shown in Figure 2, the velocity $\boldsymbol{v}_{\mathrm{s}}$ of a generic star s is given by:

$$\boldsymbol{v}_{\mathrm{s}} = V_{\mathrm{rot}} \begin{pmatrix} \sin\varphi \\ -\cos\varphi \end{pmatrix}, \tag{1}$$

where φ is the angle between the star's position vector and the x-axis, measured from the GC. In the following, primed quantities refer to the Sun's frame of reference, while unprimed quantities are defined in the GC frame. All angles are expressed in radians.

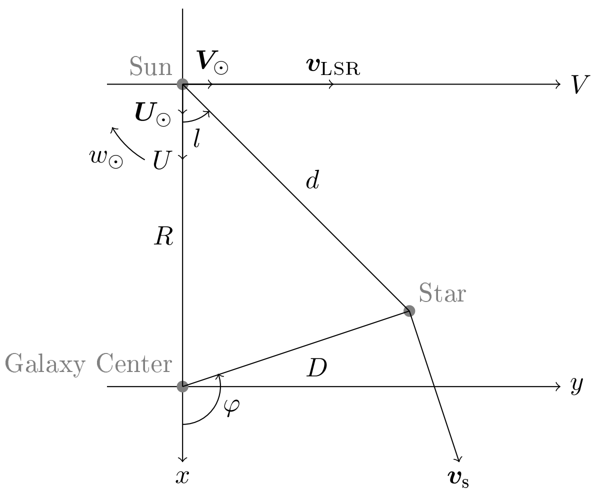

Figure 2: Frames of reference. The angular frequency $w_\odot$ only contributes to the tangential component of the velocities of the stars in the Sun's frame of reference, and is therefore omitted in equation 3.

In this reference frame, the Sun is located at $\varphi = \pi$, at a fixed distance of $R = 8300\,\mathrm{pc}$ from the GC [**GalacticKinematics**]. The Local Standard of Rest (LSR) shares the same rotational velocity as described by Eq. 1. Additionally, the Sun has a peculiar motion with respect to the LSR, with components $U_\odot$ along the x-axis and $V_\odot$ along the y-axis. The total velocity of the Sun is therefore:

$$\boldsymbol{v}_\odot = \begin{pmatrix} 0 \\ V_{\mathrm{rot}} \end{pmatrix} + \begin{pmatrix} U_\odot \\ V_\odot \end{pmatrix}. \tag{2}$$

The velocity of a star in the Sun's frame is then given by:

$$\boldsymbol{v}'_s = \boldsymbol{v}_s - \boldsymbol{v}_\odot. \tag{3}$$

We now project $\boldsymbol{v}'_s$ onto the radial direction relative to the Sun, denoted $\hat{e}'_r$, obtaining the observed radial velocity v'_{rad} of a star with Galactic longitude l:

$$\begin{aligned} v'_{\mathrm{rad}} &= \boldsymbol{v}'_s \cdot \hat{e}'_r \\ &= V_{\mathrm{rot}}\left[\sin\varphi\cos l - (1+\cos\varphi)\sin l\right] - U_\odot \cos l - V_\odot \sin l. \end{aligned} \tag{4}$$

To relate this model to Gaia observations, we must express $\sin\varphi$ and $\cos\varphi$ in terms of l and the parallax p (in milliarcseconds). The distance d (in parsecs) from the Sun to a star is calculated as:

$$d\ [\mathrm{pc}] = \frac{1000}{p\ [\mathrm{mas}]}. \tag{5}$$

Using simple geometric relations (see Figure 2), we find the distance D from the star to the GC, and express $\sin\varphi$ and $\cos\varphi$ as:

$$\begin{aligned} D &= \sqrt{d^2 + R^2 - 2dR\cos l}, \\ \sin\varphi &= \frac{d\sin l}{D}, \\ \cos\varphi &= \frac{d\cos l - R}{D}. \end{aligned} \tag{6}$$

Substituting Eqs. 5–6 into Eq. 4, we obtain the prediction for the radial velocity of star i under the first model, $\hat{v}^{(1)}_{\mathrm{rad},i}$, as a function of its longitude l_i and parallax p_i:

$$\begin{aligned} \hat{v}^{(1)}_{\mathrm{rad},i}(l_i, p_i) = V_{\mathrm{rot}} \sin l_i & \left(\frac{R}{\sqrt{\left(\frac{1000}{p_i}\right)^2 + R^2 - 2\left(\frac{1000}{p_i}\right) R\cos l_i}} - 1 \right) \\ & - U_\odot \cos l_i - V_\odot \sin l_i. \end{aligned} \tag{7}$$

In a more refined second model, we account for the random peculiar motion of stars. We model each star's radial velocity component due to random motion as a Gaussian variable: $v_{\mathrm{rand},i} \sim \mathcal{N}(0, \sigma)$, where σ is the velocity dispersion. The corresponding prediction becomes:

$$\begin{aligned} \hat{v}^{(2)}_{\mathrm{rad},i}(l_i, p_i) = V_{\mathrm{rot}} \sin l_i & \left(\frac{R}{\sqrt{\left(\frac{1000}{p_i}\right)^2 + R^2 - 2\left(\frac{1000}{p_i}\right) R\cos l_i}} - 1 \right) \\ & - U_\odot \cos l_i - V_\odot \sin l_i + v_{\mathrm{rand},i}. \end{aligned} \tag{8}$$

2.2 Statistical Model

In the following we describe the Bayesian inference we have made on the parameters of the two models. We used Monte Carlo Markov Chains (MCMC) implementation provided by the emcee [**EMCEE**] package in Python to get an estimate of the posterior distributions of our models. We provided emcee with the non normalized $\mathbb{P}^*(\theta|\mathcal{D})$ - the numerator of Bayes' theorem (posterior) to sample from

$$\mathbb{P}^*(\theta|\mathcal{D}) = \mathcal{L}(\mathcal{D}|\theta)\mathbb{P}(\theta) \ , \tag{9}$$

where $\mathcal{L}(\mathcal{D}|\theta)$ is the likelihood and $\mathbb{P}(\theta)$ is the prior. In practice, the logarithms of these quantities were used to achieve numerical stability. Moreover, constant factors, not depending on the parameters were neglected, since they do not affect the sampling of the posterior distribution.

In both models, we assumed each measurement and each parameter to be independent of all the others, since there is no a-priori reason to consider them to be correlated. Under this assumption, the likelihood and the prior are a product of individual terms, so the logarithms are given by their sums. Then, we assume each measured value m_{measured} to be the sum of its true value m_{true}, and a random error ϵ coming from a normal distribution centered at 0 with standard deviation σ_{m}, given by the statistical uncertainty provided by GAIA database ($\epsilon \sim N(0, \sigma_{\text{m}})$). Therefore, in mathematical terms: $m_{\text{measured}} = m_{\text{true}} + \epsilon$.

For the first model we have a set of three parameters $\theta_{\mathbf{1}} = (V_{\text{rot}}, U_\odot, V_\odot)$. Neglecting the uncertainties associated to the parallax measurements, the difference between the measure of the radial velocity $v_{\text{rad},i}$ and its prediction $\hat{v}^{(1)}_{\text{rad},i}$ is a random variable extracted from a normal distribution centered in 0, with standard deviation only given by the statistical uncertainty on the measurements of the radial velocity $\sigma_{\text{v},i}$. The log-likelihood of this model is therefore given by the sum of independent terms as:

$$\log \mathcal{L}^{(1)}(\mathcal{D}|\theta_1) = -\frac{1}{2}\sum_{i=1}^{\text{N}_{\text{stars}}} \left[\log(2\pi\sigma^2_{\text{v},i}) + \frac{(v_{\text{rad},i} - \hat{v}^{(1)}_{\text{rad},i})^2}{\sigma^2_{\text{v},i}}\right] . \tag{10}$$

Then, we chose a flat prior for $V_{\text{rot}} \in [0, 500\,\text{km}\,\text{s}^{-1}]$, in order to consider typical values of the rotational motion of stars in spiral barred (Sb) galaxies which are found in the range [144, 330] $\text{km}\,\text{s}^{-1}$[**Schneider2015**]. For $U_\odot$ and $V_\odot$ we chose a Gaussian prior centered in 0, $\log \mathbb{P}^{(1)}(U_\odot) + \log \mathbb{P}^{(1)}(U_\odot) \sim -\frac{U_\odot^2 + V_\odot^2}{v^2_{\text{gal}}}$, assuming the peculiar motion of the Sun to be analogous to a stochastic thermal motion. As a value for v_{gal}, we chose 200 $\text{km}\,\text{s}^{-1}$ since it is the typical scale of stars' velocities in a Sb galaxy.

In our second model, there are four parameters: $\theta_{\mathbf{2}} = (V_{\text{rad}}, U_\odot, V_\odot, \sigma)$. Accounting also for the errors on the parallax measurements, the model prediction $\hat{v}^{(2)}_{\text{rad},i}$ (see equation 8) is a random variable extracted from a normal distribution centered in 0 with variance given by the sum of the variance of the random component σ^2, and the contribution $\left(\frac{\partial \hat{v}^{(2)}_{\text{rad},i}}{\partial \text{p}_i}\right)^2 \sigma^2_{\text{p}_i}$ originating from the error on the parallax measurement, where the derivative can be computed analytically from equation 8. Then, the difference between the measured value of the radial velocity $v_{\text{rad},i}$ and the model prediction $\hat{v}^{(2)}_{\text{rad},i}$ is a random variable extracted from a normal distribution centered in 0 with total variance given by the sum of the variances of the error on the radial velocity measurements, and that on the model prediction, which results in:

$$\sigma^2_{\text{tot},i} = \sigma^2_{\text{v},i} + \left(\frac{\partial \hat{v}^{(2)}_{\text{rad},i}}{\partial \text{p}_i}\right)^2 \sigma^2_{\text{p}_i} + \sigma^2 \ . \tag{11}$$

Under these assumptions, the log-likelihood of this model is:

$$\log \mathcal{L}^{(2)}(\mathcal{D}|\theta_2) = -\frac{1}{2}\sum_i \left\{\log[2\pi(\sigma_i^2 + \sigma^2)] + \frac{(v_{\text{rad},i} - \hat{v}^{(2)}_{\text{rad},i})^2}{\sigma^2_{\text{tot},i}}\right\} . \tag{12}$$

We keep for the first three parameters the same priors decided in the first model, and we assume the fourth parameter σ to be uncorrelated to the others. We then chose for it the non-informative prior of the standard deviation of a Gaussian likelihood, $\log \mathbb{P}(\sigma) = -\log(\sigma)$ [**mackay2003**], since, in our model, it has a similar role.

3 Conclusions

The median values and 95% confidence intervals of the posterior distributions for both models are reported in Table 1. The estimates for $U_\odot$ and $V_\odot$ are consistent across the two models, while the estimates for V_{rot} differ significantly — their respective 95% confidence intervals do not overlap.

The inferred values of $U_\odot$ from both models are in good agreement with results from the literature [**LocalKinematics**], while the estimates of $V_\odot$ are notably higher and incompatible with those reported in the same source. Regarding V_{rot}, our values are slightly lower than the literature estimate [**GalacticKinematics**], but both models yield confidence intervals that partially overlap with the reference value, indicating general consistency.

The posterior distributions for all parameters are presented in Appendix A. Additionally, the middle and lower panels of Figure 1 show the model predictions based on the median posterior values, using the longitude and parallax of the observed stars. For the second model, the effect of stellar random motion was simulated by adding a Gaussian noise term $v_{\text{rand},i} \sim \mathcal{N}(0, \sigma)$, using the inferred value of σ.

Visually, it is evident that the second model provides a better fit to the data, highlighting the importance of incorporating the random motion of stars to accurately model their observed radial velocities.

Parameter	Model 1	Model 2	Literature
V_{rot} [$\text{km}\,\text{s}^{-1}$]	$211.45^{+0.04}_{-0.04}$	204^{+2}_{-2}	225 ± 10 [**GalacticKinem**
$U_\odot$ [$\text{km}\,\text{s}^{-1}$]	$11.638^{+0.004}_{-0.005}$	$11.7^{+0.3}_{-0.3}$	$11.1^{+0.7}_{-0.8}$ [**LocalKinema**
$V_\odot$ [$\text{km}\,\text{s}^{-1}$]	$21.604^{+0.005}_{-0.005}$	$21.7^{+0.3}_{-0.3}$	$12.2^{+0.5}_{-0.5}$ [**LocalKinema**
σ [$\text{km}\,\text{s}^{-1}$]	—	$30.6^{+0.2}_{-0.2}$	—

Table 1: Estimated parameters and their 95% confidence intervals for the two models. The literature value of V_{rot} includes total uncertainty (1σ), combining statistical and systematic contributions as reported in [**GalacticKinematics**].

A Posterior distributions

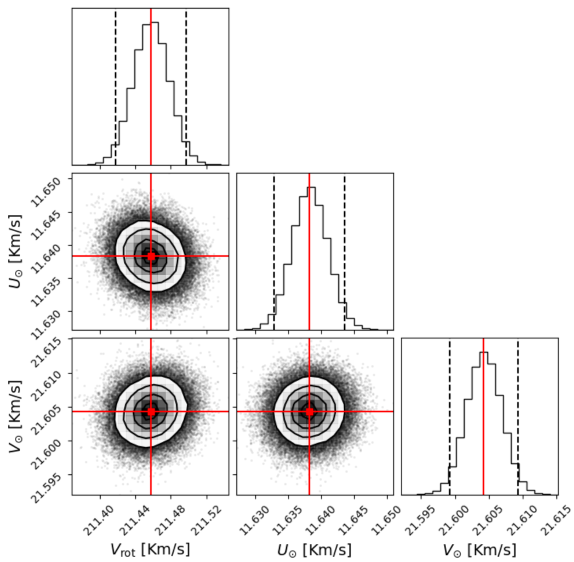

Figure 3: Posterior distribution of the parameters for the first model. Red continuous lines indicate the median values found for the respective parameters, while the black dashed ones delimit their 95% confidence interval.

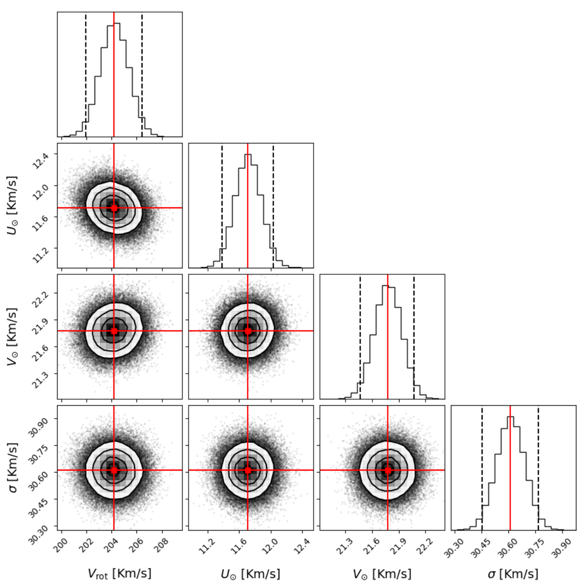

Figure 4: Posterior distribution of the parameters for the second model. Red continuous lines indicate the median values found for the respective parameters, while the black dashed ones delimit their 95% confidence interval.