Draft version March 4, 2021
Typeset using LaTeX **modern** style in AASTeX63

# Millions of radial velocity orbits from Gaia*

Daniel Foreman-Mackey[1] and TBD
[1] *Center for Computational Astrophysics, Flatiron Institute, New York, NY*

ABSTRACT

This is a paper about unresolved binaries in Gaia RVs.

*Keywords:* Astrostatistics (1882) — Binary stars (154) — Radial velocity (1332)

## 1. INTRODUCTION

By the end of the *Gaia* Mission, it will discover SOME LARGE NUMBER of exoplanet and multiple star systems based on time resolved astrometry and radial velocities of many targets. In the meantime, we only have static measurements of the radial velocity, but it turns out that there is still information in the existing public facing catalog to place constraints on the orbital parameters of multiple systems using the *Gaia* data. It has been previously demonstrated that the published radial velocity and astrometric "errors" can be used a proxy for multiplicity or data quality. In this paper, however, we demonstrate that it is possible to use the reported radial velocity errors and a probabilistic model to place constraints on the radial velocity amplitude and, in some cases, some other properties of the orbit. These measurements are useful for many applications, including *(a)* discovering black holes, *(b)* vetting transiting exoplanet discoveries, *(c)* measuring the masses of a large sample of eclipsing binaries, *(d)* quantify the binary fraction across the H–R diagram, and *(e)* informing constraints on exoplanet formation and evolution theory, to name a few.

## 2. THE BASIC IDEA

Comment on RV error vs sample variance.

## 3. ESTIMATING THE PER-TRANSIT RADIAL VELOCITY PRECISION

A key element of our analysis is that we have a reasonably accurate estimate of the per-transit radial velocity measurement uncertainty. The *Gaia* pipeline does not release a public estimate of this, but it can be estimated from the data in the public

Corresponding author: Daniel Foreman-Mackey
dforeman-mackey@flatironinstitute.org

* Git hash: dd88ceb6db05e75bf23bfc3e75de6d70bf7d747f

catalog. To make this estimate, the key assumption that we make is that the RV uncertainty depends only on a target's (reddened) $G_{\mathrm{BP}}-G_{\mathrm{RP}}$ color and apparent $G$-band magnitude $m_{\mathrm{G}}$. While this is certainly not the full story, we discuss several validation experiments below and argue that this is not an overly restrictive assumption given the precision of our measurements.

*A toy model*—To build intuition, let's start with a simplified toy model that we will extend below. If we have a catalog of $N$ targets where only the sample variance ${s_n}^2$ of the RV time series and the number of observations $T_n$ of target $n$. For our initial toy model, let's also assume that we know that the only source of noise is the per-transit measurement uncertainty, which we will assume to be Gaussian. In this case, the generative model is

$$x_{n,t} \sim \mathcal{N}(\mu_n,\,\sigma^2) \tag{1}$$

$${s_n}^2 = \frac{1}{T_n - 1}\sum_{t=1}^{T_n}(x_{n,t} - \langle x_{n,t}\rangle)^2 \tag{2}$$

where $x_{n,t}$ are the individual RV measurements for target $n$ at time $t$ and

$$\langle x_{n,t}\rangle \equiv \frac{1}{T_n}\sum_{t=1}^{T_n} x_{n,t} \quad . \tag{3}$$

In this model, our data are the empirical sample variances ${s_n}^2$ and we want to infer the underlying variance $\sigma^2$ that is shared by all targets $n$. In other words, we must evaluate the likelihood function

$$\mathcal{L}(\sigma^2;\,\{{s_n}^2\}) = p(\{{s_n}^2\}\,|\,\sigma^2) \tag{4}$$

$$= \prod_{n=1}^{N}\int p({s_n}^2\,|\,x_{n,t})\,p(x_{n,t}\,|\,\sigma^2)\,\mathrm{d}x_{n,t} \quad . \tag{5}$$

From this equation we can see that the statistic $X_n$

$$X_n = \frac{(T_n - 1)\,{s_n}^2}{\sigma^2} \tag{6}$$

is chi-squared distributed with $T_n - 1$ degrees of freedom. Therefore, Equation 4 becomes

$$\mathcal{L}(\sigma^2;\,\{{s_n}^2\}) = \prod_{n=1}^{N}\left|\frac{\partial X_n}{\partial {s_n}^2}\right| p(X_n\,|\,\sigma^2)$$

$$= \prod_{n=1}^{N}\frac{T_n - 1}{\sigma^2}\,p(X_n\,|\,\sigma^2) \tag{7}$$

where $p(X_n\,|\,\sigma)$ is a chi-squared density with $T_n - 1$ degrees of freedom. In this equation, the Jacobian term is important since $\sigma^2$ is a parameter of the model.

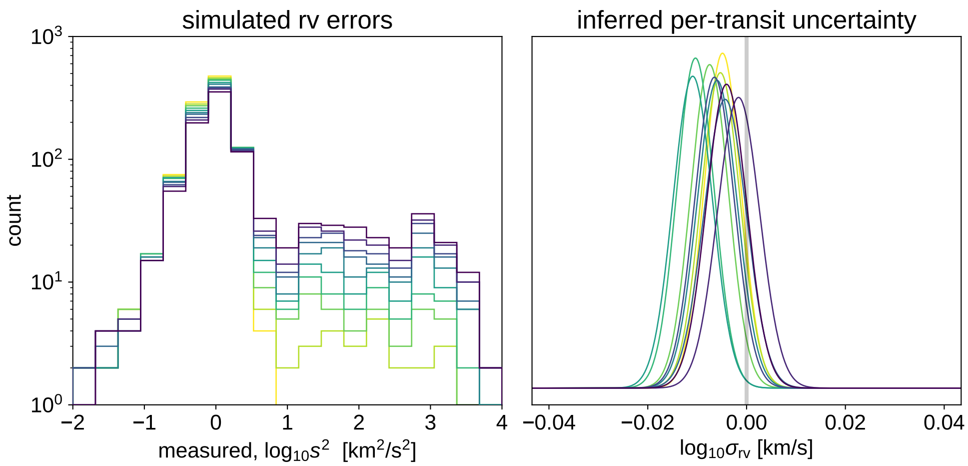


**Figure 1.** This is a cool figure.

*A more realistic model*—In the previous discussion we assumed that the *only* source of noise contributing to the RV sample variance measurement was the individual RV measurement uncertainties. This isn't going to be true when working with real data since there will by other sources of excess noise, both systematic and astrophysical; in fact, that's the whole point of this paper! To handle this, we add $N$ new parameters ${\delta_n}^2$ to our model, which we will add to the baseline uncertainty $\sigma^2$ when computing the expected variance for a target $n$. With this change, our model is nearly identical to Equation 7, with $\sigma^2$ replaced with $\sigma^2 + {\delta_n}^2$

$$\mathcal{L}(\sigma^2,\,\left\{{\delta_n}^2\right\};\,\left\{{s_n}^2\right\}) = \prod_{n=1}^{N} \frac{T_n - 1}{\sigma^2 + {\delta_n}^2}\, p(X_n \,|\, \sigma^2 + {\delta_n}^2) \quad . \tag{8}$$

This is a very flexible model with more parameters than data points, but it is still possible to perform parameter estimation.

## 4. BULK RADIAL VELOCITY INFERENCE

## 5. ARTISANAL RADIAL VELOCITIES

## 6. VALIDATION

## 7. DISCUSSION

This is a paper. See Figure 1 and Figure 2.


## ACKNOWLEDGMENTS

The authors would like to thank the Astronomical Data Group at Flatiron for listening to every iteration of this project and for providing great feedback every step of the way.


*Facilities:* *APOGEE*, *Gaia*

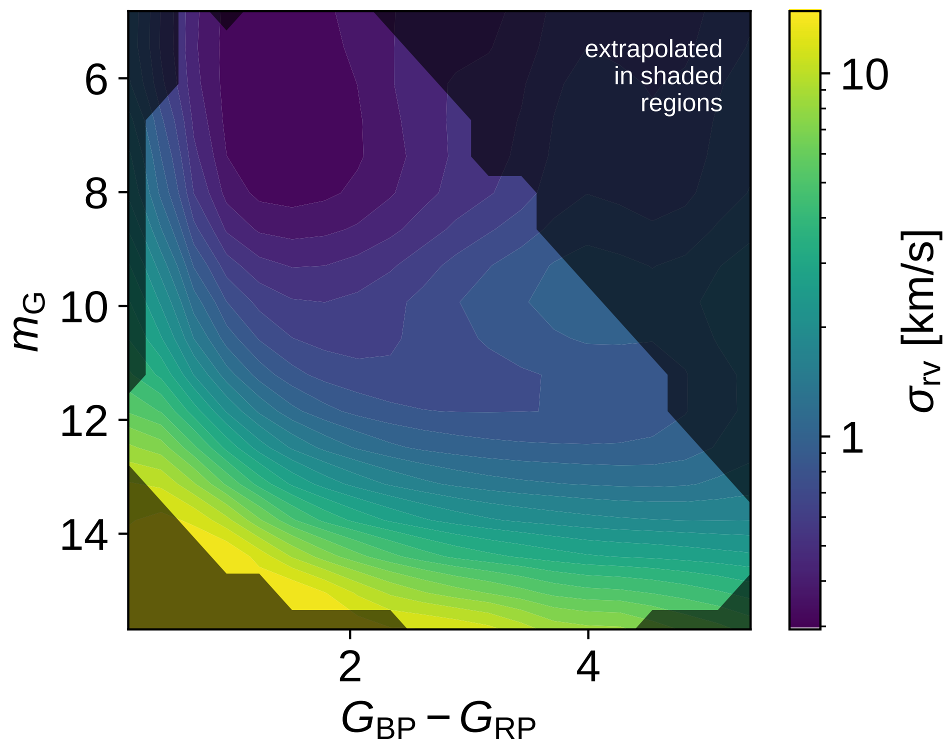


**Figure 2.** This is another cool figure.

*Software:* AstroPy (Astropy Collaboration et al. 2013, 2018), JAX (Bradbury et al. 2018), NumPy (Harris et al. 2020), Matplotlib (Hunter 2007), SciPy (Virtanen et al. 2020)

# APPENDIX

## A. PROBABLY SOME FANCY MATH

## REFERENCES

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, JAX: composable transformations of Python+NumPy programs, 0.2.9. http://github.com/google/jax

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2