

1 Purpose

The primary objective of the EPRV3 Evidence Challenge is to compare different algorithms and implementations for performing model comparison. We will provide ~ 6 simulated radial velocity (RV) datasets, each with zero to three planets. Participants will be asked to compute quantitative estimates of Bayesian evidence (also known as the marginalized likelihood) for each model and the estimated uncertainty in each evidence, using whatever computational methods and simplifying assumptions that they choose. Participants are encouraged to apply multiple methods to analyze the same datasets and models.

The evidences for two models (and the same dataset) can be combined to compute the posterior odds ratio for any pair of n -planet to $(n + 1)$ -planet models. Any groups that are unable to estimate the evidence (e.g., using non-Bayesian methods), will be asked to provide estimates of quantities that they consider analogous to the posterior odds ratio. Of course, one should always provide an estimate of the uncertainties in the results that they report.

We will compare results from different groups and different methods during the breakout sessions of the EPRV3 meeting in August 2017. We hope that this exercise will lead to a better understanding of the advantages and disadvantages of various algorithms, implementations and approaches. We are optimistic that this will lead to a paper describing the results, lessons learned, and recommendations for future studies.

2 Detailed Model Specification

In order to facilitate quantitative comparisons, it is important that each group use a standardized set of assumptions for the physical and statistical models. Therefore, we describe the process used to generate the datasets in detail. If you believe it is important to be provided with additional information, please let the organizers know as soon as possible.

2.1 Dataset Characteristics

We will provide ~ 6 simulated datasets with anywhere from zero to three planets in each dataset. The datasets are generated with a set of consistent properties:

- Each dataset is a simple timeseries, including the time of observation (t_i), “measured” radial velocity (v_i), and a measurement uncertainty (σ_i).
- Number of observations: $n_{\text{obs}} = 200$
- Observing baseline: ~ 600 days
- Each dataset includes between zero and three planets (inclusive), as described below.
- Each dataset includes a single velocity offset and correlated, Gaussian noise to represent stellar activity.

2.2 Physical Model

In each dataset, the radial velocity of the star ($v(t)$) is computed via n-body integrations using Newtonian gravity, one star and zero to three planets. The simulated measurements are the sum of the line-of-sight velocity of the star and Gaussian noise. While the full model formally includes mutual planetary interactions, we fully expect that it would be well-described by the linear superposition of zero to three Keplerian orbits plus a constant velocity offset and a noise term. For the sake of computational efficiency, we restrict the range of injected planet orbital periods to between 10 and 2,400 days. Note that for the purpose of computing evidences, we ask that you use a prior ranging from orbital periods ranging from 1 to 10^4 days.

2.3 Statistical Model

2.3.1 Likelihood

Each simulated data point is generated according to

$$v_i = v_{\text{pred}}(t_i|\theta) + \epsilon_i, \quad (1)$$

where $v_{\text{pred}}(t_i|\theta)$ is the velocity predicted at time t_i by a model parameterized by θ , and ϵ_i is the perturbation to the measurement due to noise. $\epsilon \sim \text{Normal}(\mathbf{0}, \Sigma)$, i.e., the noise vector, ϵ is drawn from a multivariate process with covariance matrix Σ . Therefore, the appropriate likelihood ($\mathcal{L}(\theta) = p(\mathbf{v}|\theta)$) is a multi-variable normal distribution, centered on the predictions of the model (parameterized by θ),

$$\log \mathcal{L}(\theta) = -\frac{1}{2}(\mathbf{v} - \mathbf{v}_{\text{pred}}(\theta))^T \Sigma^{-1}(\mathbf{v} - \mathbf{v}_{\text{pred}}(\theta)) - \frac{1}{2} \log |\det \Sigma| - \frac{n_{\text{obs}}}{2} \log(2\pi), \quad (2)$$

where \mathbf{v} is the vector of v_i 's, and $\mathbf{v}_{\text{pred}}(\theta)$ is the vector of velocities predicted by the model with parameters θ . Throughout this document, we use \log to denote the natural logarithm, not base 10.

The Gaussian noise is correlated from one observation to the next. The covariance matrix, Σ , is given by

$$\Sigma_{i,j} = K_{i,j} + \delta_{i,j} (\sigma_i^2 + \sigma_j^2), \quad (3)$$

where $K_{i,j}$ is a quasi-periodic kernel, $\delta_{i,j}$ is the Kronecker delta, σ_i^2 is the estimated measurement uncertainty for the i th observation, and σ_j^2 is the amplitude of an additional, unknown noise term (often casually referred to as RV “jitter”). For the quasi-periodic kernel, we assume

$$K_{i,j} = \alpha^2 \exp \left[-\frac{1}{2} \left\{ \frac{\sin^2[\pi(t_i - t_j)/\tau]}{\lambda_p^2} + \frac{(t_i - t_j)^2}{\lambda_e^2} \right\} \right], \quad (4)$$

where the hyperparameters are fixed at the following values:

- $\alpha = \sqrt{3}$ m/s,
- $\lambda_e = 50.0$ days,
- $\lambda_p = 0.5$ (unitless), and
- $\tau = 20.0$ (days).

2.3.2 Priors

When computing the Bayesian evidence, we ask that each group adopt a common set of priors described below, so as to enable direct comparisons of their results. We assume a prior that factorizes in terms of each planet's orbital period (P_i), RV semi-amplitude (K_i), eccentricity (e_i), argument of pericenter (ω_i) and mean anomaly at epoch (M_i), as well as the white-noise terms (σ_J) and RV offset (C).

- For each planet's orbital period, a truncated Jeffreys prior, $p(P) dP = \frac{dP}{P} \times \frac{1}{\log(P_{\max}/P_{\min})}$ for $P_{\min} \leq P \leq P_{\max}$. For the primary analysis, we will assume $P_{\min} = 1.25$ day and $P_{\max} = 10^4$ days for each of the planets. For an alternative analysis, we will provide specific values of $P_{\min,i}$ and $P_{\max,i}$ for each planet and dataset to use when computing the evidence in a file named `prior_bounds_000?.txt` (see §2.3.3).
- For each planet's RV semi-amplitude, a truncated modified Jeffreys prior, $p(K) dK = \frac{dK}{K_0(1+K/K_0)} \times \frac{1}{\log(1+K_{\max}/K_0)}$ for $0 < K \leq K_{\max}$, where $K_0 = 1$ m/s and $K_{\max} = 999$ m/s.
- For each planet's eccentricity, a truncated Rayleigh distribution, $p(e) de = \frac{e de}{\sigma_e^2} \exp\left(-\frac{e^2}{2\sigma_e^2}\right) / \left[1 - \exp\left(-\frac{e_{\max}^2}{2\sigma_e^2}\right)\right]$ from $0 \leq e < e_{\max} = 1$ and zero for $e \geq e_{\max} = 1$, where $\sigma_e = 0.2$.
- For each planet's argument of pericenter, $p(\omega) d\omega = \frac{d\omega}{2\pi}$ from $0 \leq \omega < 2\pi$ radians.
- For each planet's mean anomaly, $p(M) dM = \frac{dM}{2\pi}$ from $0 \leq M < 2\pi$ radians.
- For the additional white-noise term, $p(\sigma_J) d\sigma_J = \frac{d\sigma_J}{\sigma_{J,0}(1+\sigma_J/\sigma_{J,0})} \times \frac{1}{\log(1+\sigma_{J,\max}/\sigma_{J,0})}$ for $0 < \sigma_{J,0} \leq \sigma_{J,\max}$, where $\sigma_{J,0} = 1$ m/s and $\sigma_{J,\max} = 99$ m/s.
- For the RV velocity offset, $p(C) dC = \frac{dC}{2C_{\max}}$ from $-C_{\max} \leq C \leq C_{\max}$, where $C_{\max} = 1,000$ m/s.

Remember that \log refers to the natural logarithm. Thus, the combined prior for a given model (i.e., number of planets, N_{pl}) is

$$p\left(\{P_i, K_i, e_i, \omega_i, M_i\}_{i=1..N_{pl}}, \sigma_J, C\right) = p(\sigma_J)p(C) \prod_{i=1}^{N_{pl}} p(P_i)p(K_i)p(e_i)p(\omega_i)p(M_i). \quad (5)$$

2.3.3 Two Sets of Priors for Orbital Periods

The primary analysis will be based on the priors described above, i.e., $P_{\min} = 1.25$ day and $P_{\max} = 10^4$ days for each of the planets. Note that even for a very well-behaved dataset (i.e., one dominant posterior mode if we assume $P_1 < P_2 < P_3$), the posterior would have $N_{pl}!$ modes corresponding to the $N_{pl}!$ permutations of the planets' orbital periods, when using the primary prior. Therefore, any groups making the assumption that $P_1 < P_2 < P_3$ would need to multiply their estimate of the evidence by $N_{pl}!$, so as to correspond to the common prior which does not impose an order restriction.

Based on preliminary results reported at the EPRVIII breakout sessions, we noticed that when analyzing some datasets and assuming multiple planets, different groups sometimes focus their exploration of parameter space on different regions, particularly in terms of the orbital periods. This made it difficult to compare the methods for estimating the evidence. Therefore, the participants decided that we will also report results for each data set using a second choice of priors for orbital period that force all groups to explore the same regions of parameter space in orbital period. That is we will specify different values of $P_{\min,i}$ and $P_{\max,i}$ for each planet and each dataset to use when computing the evidence in a file named `prior.bounds_000?.txt`. The first column is a 'P' to indicate we are specifying the prior for orbital period, the second column is the integer index for which planet, the third column is $P_{\min,i}$, and the fourth column is $P_{\max,i}$. Note that these do not necessarily bound true orbital parameters used to generate the datasets. These merely represent a common set of not-obviously-terrible period ranges to facilitate more direct comparison of numerical methods.

2.3.4 Prior for over models

Remember, the primary purpose is to compare properties of different estimates of the evidence, and not to figure out how many planets are present in the system. Therefore, participants should submit their estimate for the evidence for each of the 0-planet, 1-planet, 2-planet, and 3-planet models, assuming that is the correct model (i.e., not including a prior for the number of planets), *regardless of the number of planets that they believe are in the system*.

In case some participants perform non-Bayesian estimates, it may be useful to have something that can be compared between Bayesian and non-Bayesian estimates. For purposes of comparisons to analyses that can not report the marginalized likelihood, we may compare the posterior odds ratio to whatever they provide that they think is analogous to a posterior odds ratio. For the purpose of estimating posterior odds ratios, we will use the following prior for the number of planets (N_{pl}),

$$p(N_{pl}) = \alpha^{N_{pl}}, \quad \text{for } N_{pl} = 1, 2, 3 \quad (6)$$

$$1 - \sum_{i=1}^3 \alpha^i \quad \text{for } N_{pl} = 0, \quad (7)$$

and set $\alpha = \frac{1}{3}$. Any participants submitting non-Bayesian estimates should take this into consideration, so that they can calibrate their estimates appropriately.

3 File Formats

See the README.md in the data subdirectory for a description of the input data file format.

We also provide specifications for format of output files to facilitate rapid comparison of the results.