Draft version October 11, 2018
Typeset using LaTeX **modern** style in AASTeX62

Notes on the Model

Will M. Farr[1, 2]

[1] *Department of Physics and Astronomy, Stony Brook University, Stony Brook NY 11794, USA*
[2] *Center for Computational Astronomy, Flatiron Institute, 162 5th Ave., New York NY 10010, USA*

ABSTRACT

I discuss the contaminated mixture model that we are using.

1. HIERARCHICAL LIKELIHOOD

We assume, following Foreman-Mackey et al. (2014), that the occurrence rate of planets is constant in some set of bins in the period-radius plane, $R_i \leq R < R_{i+1}$, $P_i \leq P \leq P_{i+1}$, $i = 1, \ldots, N_{\mathrm{bin}}$:

$$\frac{\mathrm{d}N}{\mathrm{d}\ln P \mathrm{d}\ln R} = \begin{cases} n_i & R_i \leq R < R_{i+1} \text{ and } P_i \leq P \leq P_{i+1} \\ \tilde{n} & \text{otherwise} \end{cases}. \quad (1)$$

We include a “catch all” bin, where the rate density is $\tilde{n}$ to avoid issues at the edges of the set of bins; with the catch-all bin, we can cut the data set several “sigma” outside the region of interest with bins, and objects that land on the edges will be distributed between the bins and the “catch-all” bin. Note that, for all practical purposes, Kepler measures the period perfectly, so the only uncertainty in any of the measurements is the radius.

We also assume that the contaminated background follows a similar constant-in-bin model,

$$\frac{\mathrm{d}N_{\mathrm{bg}}}{\mathrm{d}\ln P \mathrm{d}\ln R} = b_i \quad R_i \leq R < R_{i+1} \text{ and } P_i \leq P \leq P_{i+1}, \quad (2)$$

but here we do not include the extra bin (since we don’t care about the background-foreground distinction in that bin).

Let us assume that the likelihood is well-approximated by a delta function in period and a log-normal distribution for the radius for each planet detection:

$$p\left(d \mid \ln P, \ln R\right) \propto \delta\left(\ln P_{\mathrm{obs}} - \ln P\right) \frac{1}{\sqrt{2\pi}\sigma_{\ln R}} \exp\left[\frac{\left(\ln R_{\mathrm{obs}} - \ln R\right)^2}{2\sigma_{\ln R}^2}\right]. \quad (3)$$

Then the hierarchical likelihood for a set of planet candidates can be obtained by integrating out the “nuisance” parameters giving the true period and radius of each

will.farr@stonybrook.edu

planet (Mandel et al. 2018; Loredo 2004):

$$p\left(\{d_j\} \mid \{n_i\}, \tilde{n}, \{b_i\}\right) \propto \left[\prod_{j=1}^{N} \int \mathrm{d}\ln P_j \, \mathrm{d}\ln R_j \, p\left(d_j \mid \ln P_j, \ln R_j\right) \right.$$
$$\left. \times \left(\frac{\mathrm{d}N}{\mathrm{d}\ln P_j \mathrm{d}\ln R_j} + \frac{\mathrm{d}N_{\mathrm{bg}}}{\mathrm{d}\ln P_j \mathrm{d}\ln R_j}\right)\right] \exp\left[-N_{\mathrm{ex}} - N_{\mathrm{bg}} - \tilde{N}\right]; \quad (4)$$

where N_{ex} is the expected number of true planet detections within the grid of cells,

$$N_{\mathrm{ex}} = \sum_{i=1}^{N_{\mathrm{bin}}} n_i \Delta \ln P_i \Delta \ln R_i \sum_{k=1}^{N_{\mathrm{stars}}} \left\langle P_{\mathrm{det}}^{(k)} \right\rangle_i, \quad (5)$$

where $\left\langle P_{\mathrm{det}}^{(k)} \right\rangle_i$ is the detection probability for a planet around star k averaged over cell i for a flat population in $\ln P$ and $\ln R$; N_{bg} is the expected number of background observations

$$N_{\mathrm{bg}} = \sum_{i=1}^{N_{\mathrm{bin}}} b_i \Delta \ln P_i \Delta \ln R_i; \quad (6)$$

and $\tilde{N}$ is the expected number of detections in the "catch all" bin

$$\tilde{N} = \tilde{n}\tilde{V}, \quad (7)$$

where $\tilde{V}$ is the volume of the catch-all in $\ln P$-$\ln R$ space. Because the population is assumed to be constant in bins, the integrals in Eq. (4) reduce to

$$p\left(\{d_j\} \mid \{n_i\}, \tilde{n}, \{b_i\}\right) \propto \left[\prod_{j=1}^{N} \tilde{w}_j \tilde{n} + \sum_{i=1}^{N_{\mathrm{bin}}} w_{ji}\left(n_i + b_i\right)\right] \exp\left[-N_{\mathrm{ex}} - N_{\mathrm{bg}} - \tilde{N}\right], \quad (8)$$

where

$$w_{ji} = \int_{R_i \leq R_j < R_{i+1} \cap P_i \leq P_j < P_{i+1}} \mathrm{d}\ln P_j \, \mathrm{d}\ln R_j \, p\left(d_j \mid \ln P_j, \ln R_j\right), \quad (9)$$

are weights assigned to each bin (similarly for $\tilde{w}_j$ for the catch-all bin). The w_{ji} can be pre-computed, as they depend only on the bin boundaries and the observed periods, radii, and radius uncertainties.

2. PRIOR

The foreground and background per-bin rates, n_i and b_i, are degenerate in the likelihood in Eq. (8). We can break the degeneracy using an informative prior on one or both rate distributions. Here we will assume that we have *measured* the background contamination rate (perhaps imperfectly) through some independent process, and incorporate this through an informative prior on the background rate densities b_i.

Following **?**, we choose to estimate the background b_i by using the results of a search over *inverted* Kepler lightcurves using the same pipelines used to produce the catalog

(Coughlin 2017). TODO: more details here. We supply a log-normal prior on b_i based on the observed count of non `FP` inverted-lightcurve candidate events and a relative uncertainty of $1/\sqrt{N}$.

We choose to implement a Gaussian process prior on the (log of the) n_i. We choose a constant-mean, squared exponential covariance GP on the log-rates:

$$\ln \vec{n} \sim N\left(\mu \vec{1}, \mathbf{\Sigma}\right), \tag{10}$$

with

$$\Sigma_{ij} = \sigma^2 \left(1 + \epsilon \delta_{ij}\right) \exp\left[-\frac{\left|\Delta \vec{x}_{ij}\right|^2}{2\lambda^2}\right]. \tag{11}$$

The single-bin variance is $\sigma^2 (1 + \epsilon)$, $\epsilon \ll 1$ is a fractional white-noise component added to the diagonal terms of the matrix for stability, $\Delta \vec{x}_{ij}$ is the displacement vector between the centers of bin i and bin j in the $\ln P$-$\ln R$ plane, and λ is a correlation length scale in this plane[1].

3. CORRELATIONS IN PLANET DETECTION

The model currently assumes that the detection probability for each planet is independent of all the other planets in the same system. This will over-estimate the number of planets per star (since it almost certainly makes second and subsequent planets seem harder to detect than they should be—at least if there are *any* planetary disks out there). I will think about how to modify the detection probability calculation to better reflect the fact that planets occur in disks. That, in itself, would be worth another paper; and a particularly exciting one if we could find evidence for small disk opening angles (i.e. once one planet transits, all the planets transit).

4. ACTUAL DATA

Plots of the actual data go here.

REFERENCES

Coughlin, J. L. 2017, Planet Detection Metrics: Robovetter Completeness and Effectiveness for Data Release 25, Tech. Rep. KSCI-19114-002, NASA Ames Research Center

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, ApJ, 795, 64, doi: 10.1088/0004-637X/795/1/64

Loredo, T. J. 2004, in American Institute of Physics Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, Vol. 735, 195–206

Mandel, I., Farr, W. M., & Gair, J. R. 2018, ArXiv e-prints, arXiv:1809.02063. https://arxiv.org/abs/1809.02063

[1] A natural extension would be to allow the $\ln P$ and $\ln R$ dimensions to have *independent* length scales; or even to impose an arbitrary *metric* matrix that would allow for arbitrary anisotropy in the correlations.