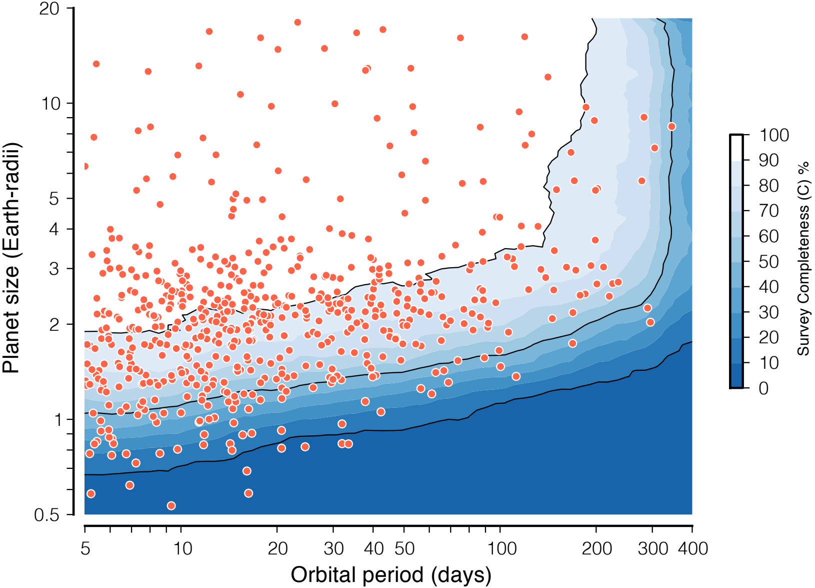
Foreman-Mackey, Hogg and Morton 2014 (ApJ 795, 64) "Exoplanet Population Inference and the Abundance of Earth Analogs from Noisy, Incomplete Catalogs"

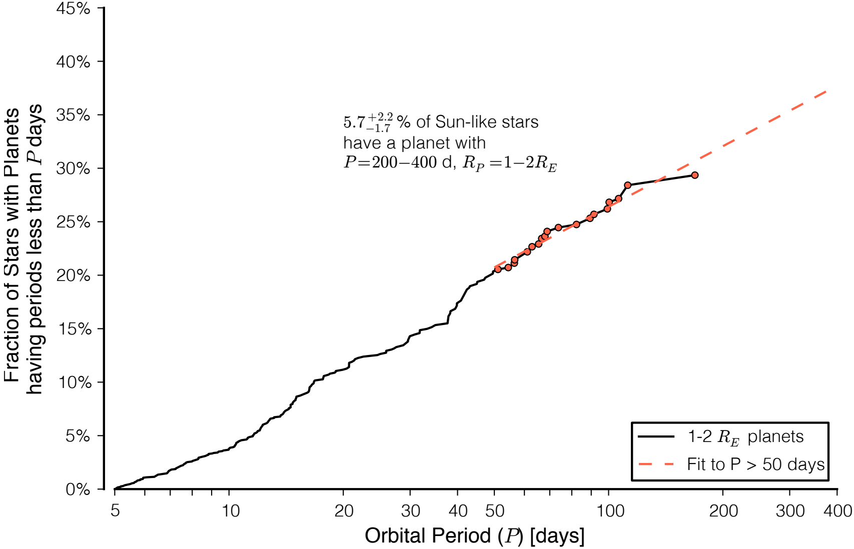
Radek Poleski

17.03.2022

Petigura et al. 2013



Petigura et al. 2013



Assumptions of Foreman-Mackey et al. 2014

- *Conditional independence* – We assume that every object in the catalog is a conditionally independent draw from the observable occurrence rate density.

Assumptions of Foreman-Mackey et al. 2014

- *Conditional independence* – We assume that every object in the catalog is a conditionally independent draw from the observable occurrence rate density.
- *Lack of false positives* – In our inferences, we assume that all of the candidates in the catalog are *true* exoplanets.

Assumptions of Foreman-Mackey et al. 2014

- *Conditional independence* – We assume that every object in the catalog is a conditionally independent draw from the observable occurrence rate density.
- *Lack of false positives* – In our inferences, we assume that all of the candidates in the catalog are *true* exoplanets.
- *Known observational uncertainties* – To apply the importance sampling approximation to the published catalog, we assume that the measurement uncertainties are known and, in this case, Gaussian.

Assumptions of Foreman-Mackey et al. 2014

- *Conditional independence* – We assume that every object in the catalog is a conditionally independent draw from the observable occurrence rate density.
- *Lack of false positives* – In our inferences, we assume that all of the candidates in the catalog are *true* exoplanets.
- *Known observational uncertainties* – To apply the importance sampling approximation to the published catalog, we assume that the measurement uncertainties are known and, in this case, Gaussian.
- *Given empirical detection efficiency* – Petigura et al. (2013) determined the end-to-end detection efficiency of their planet detection pipeline as a function of *true* period and radius by injecting synthetic signals into real light curves and testing recovery.

Assumptions of Foreman-Mackey et al. 2014

- *Conditional independence* – We assume that every object in the catalog is a conditionally independent draw from the observable occurrence rate density.
- *Lack of false positives* – In our inferences, we assume that all of the candidates in the catalog are *true* exoplanets.
- *Known observational uncertainties* – To apply the importance sampling approximation to the published catalog, we assume that the measurement uncertainties are known and, in this case, Gaussian.
- *Given empirical detection efficiency* – Petigura et al. (2013) determined the end-to-end detection efficiency of their planet detection pipeline as a function of *true* period and radius by injecting synthetic signals into real light curves and testing recovery.
- *Smooth rate function* – Throughout our analysis, we make the prior assumption that the occurrence rate density is a smooth function of logarithmic period and radius.

Assumptions of Foreman-Mackey et al. 2014

- *Conditional independence* – We assume that every object in the catalog is a conditionally independent draw from the observable occurrence rate density.
- *Lack of false positives* – In our inferences, we assume that all of the candidates in the catalog are *true* exoplanets.
- *Known observational uncertainties* – To apply the importance sampling approximation to the published catalog, we assume that the measurement uncertainties are known and, in this case, Gaussian.
- *Given empirical detection efficiency* – Petigura et al. (2013) determined the end-to-end detection efficiency of their planet detection pipeline as a function of *true* period and radius by injecting synthetic signals into real light curves and testing recovery.
- *Smooth rate function* – Throughout our analysis, we make the prior assumption that the occurrence rate density is a smooth function of logarithmic period and radius.

Planet parameters and detection efficiency from Petigura et al. (2013)

Goals of Foreman-Mackey et al. 2014

1. Constrain the rate density of small exoplanets orbiting Sun-like stars.

Goals of Foreman-Mackey et al. 2014

1. Constrain the rate density of small exoplanets orbiting Sun-like stars.
2. Place probabilistic constraints on the rate density of Earth analogs $\Gamma_{\oplus}$, defined as *the expected number of planets per star per natural logarithmic bin in period and radius, evaluated at the period and radius of Earth*.

Goals of Foreman-Mackey et al. 2014

1. Constrain the rate density of small exoplanets orbiting Sun-like stars.
2. Place probabilistic constraints on the rate density of Earth analogs $\Gamma_{\oplus}$, defined as *the expected number of planets per star per natural logarithmic bin in period and radius, evaluated at the period and radius of Earth*.

$$\Gamma_{\oplus} = \left.\frac{\mathrm{d}^2 N}{\mathrm{d}\ln P\,\mathrm{d}\ln R}\right|_{R=R_{\oplus},\,P=P_{\oplus}} \tag{1}$$

Goals of Foreman-Mackey et al. 2014

1. Constrain the rate density of small exoplanets orbiting Sun-like stars.
2. Place probabilistic constraints on the rate density of Earth analogs $\Gamma_\oplus$, defined as *the expected number of planets per star per natural logarithmic bin in period and radius, evaluated at the period and radius of Earth*.

$$\Gamma_\oplus = \left.\frac{\mathrm{d}^2 N}{\mathrm{d}\ln P \; \mathrm{d}\ln R}\right|_{R=R_\oplus,\, P=P_\oplus} \tag{1}$$

Rate – indicate the dimensionless expectation value of a Poisson process.
Rate density – a quantity that must be integrated over a finite bin in period and radius to deliver a rate.

Traditional approach

$\Gamma_{\boldsymbol{\theta}}(\boldsymbol{w})$ – occurrence rate density Γ (parameterized by the parameters $\boldsymbol{\theta}$) as a function of the physical parameters $\boldsymbol{w}$ (orbital period, planetary radius, etc.)

Traditional approach

$\Gamma_{\boldsymbol{\theta}}(\boldsymbol{w})$ – occurrence rate density Γ (parameterized by the parameters $\boldsymbol{\theta}$) as a function of the physical parameters $\boldsymbol{w}$ (orbital period, planetary radius, etc.)

$\boldsymbol{w}_k$ – vector of physical parameters describing the planet around target k.

Traditional approach

$\Gamma_{\boldsymbol{\theta}}(\boldsymbol{w})$ – occurrence rate density Γ (parameterized by the parameters $\boldsymbol{\theta}$) as a function of the physical parameters $\boldsymbol{w}$ (orbital period, planetary radius, etc.)

$\boldsymbol{w}_k$ – vector of physical parameters describing the planet around target k.

Model the catalog as a draw from the inhomogeneous Poisson process set by the *observable* rate density $\hat{\Gamma}_{\boldsymbol{\theta}}$:

$$p(\{\boldsymbol{w}_k\}\,|\,\boldsymbol{\theta}) \;=\; \exp\left(-\int \hat{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}\right)\prod_{k=1}^{K}\hat{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{w}_k) \qquad (2)$$

$$\hat{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{w}) \;=\; Q_{\mathrm{c}}(\boldsymbol{w})\,\Gamma_{\boldsymbol{\theta}}(\boldsymbol{w}) \qquad (3)$$

where $Q_{\mathrm{c}}(\boldsymbol{w})$ is the detection efficiency (including transit probability) at $\boldsymbol{w}$.

Occurrence rate density model

$$\Gamma_{\boldsymbol{\theta}}(\boldsymbol{w}) = \begin{cases} \exp(\theta_1) & \boldsymbol{w} \in \Delta_1, \\ \exp(\theta_2) & \boldsymbol{w} \in \Delta_2, \\ \cdots & \\ \exp(\theta_J) & \boldsymbol{w} \in \Delta_J, \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where the parameters θ_j are the log step heights and the bins Δ_j are fixed *a priori*

Including uncertainties of planet parameters

$$p(\{\boldsymbol{x}_k\} \,|\, \boldsymbol{\theta}) \;=\; \int p(\{\boldsymbol{x}_k\} \,|\, \{\boldsymbol{w}_k\})\, p(\{\boldsymbol{w}_k\} \,|\, \boldsymbol{\theta})\, \mathrm{d}\{\boldsymbol{w}_k\} \tag{5}$$

$\{\boldsymbol{x}_k\}$ is the set of all light curves, one light curve $\boldsymbol{x}_k$ per target k (70,000 epochs per target typically)

Definition of hierarchical inference

The values $\{\boldsymbol{w}_k{}^{(n)}\}$ are samples drawn from the posterior probability

$$\boldsymbol{w}_k{}^{(n)} \sim p(\boldsymbol{w}_k \,|\, \boldsymbol{x}_k, \boldsymbol{\alpha}) \tag{6}$$

For target k there are N_k samples. The notation $\boldsymbol{\alpha}$ is a reminder that the catalog was produced under a specific choice of a – probably "uninformative" – *interim prior* $p(\boldsymbol{w}_k \,|\, \boldsymbol{\alpha})$.

Definition of hierarchical inference

The values $\{\boldsymbol{w}_k{}^{(n)}\}$ are samples drawn from the posterior probability

$$\boldsymbol{w}_k{}^{(n)} \quad \sim \quad p(\boldsymbol{w}_k \,|\, \boldsymbol{x}_k,\, \boldsymbol{\alpha}) \tag{6}$$

For target k there are N_k samples. The notation $\boldsymbol{\alpha}$ is a reminder that the catalog was produced under a specific choice of a – probably "uninformative" – *interim prior* $p(\boldsymbol{w}_k \,|\, \boldsymbol{\alpha})$.

Marginalized likelihood:

$$\frac{p(\{\boldsymbol{x}_k\} \,|\, \boldsymbol{\theta})}{p(\{\boldsymbol{x}_k\} \,|\, \boldsymbol{\alpha})} \quad \approx \quad \exp\left(-\int \hat{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}\right) \prod_{k=1}^{K} \frac{1}{N_k} \sum_{n=1}^{N_k} \frac{\hat{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{w}_k^{(n)})}{p(\boldsymbol{w}_k^{(n)} \,|\, \boldsymbol{\alpha})} \tag{7}$$

The data only enter this equation through the posterior constraints provided by the catalog $\{\boldsymbol{w}_k\}$.

Simulations for tests

Period and radius distributions are generated by a separable model

$$\Gamma_{\boldsymbol{\theta}}(\ln P, \ln R) = \Gamma_{\boldsymbol{\theta}}^{(P)}(\ln P)\,\Gamma_{\boldsymbol{\theta}}^{(R)}(\ln R) \tag{8}$$

but fit using the full general model.

Simulations for tests

Period and radius distributions are generated by a separable model

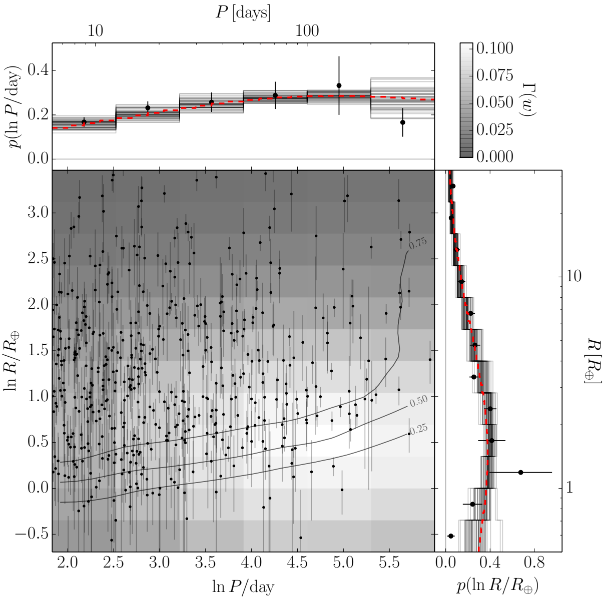
$$\Gamma_{\boldsymbol{\theta}}(\ln P, \ln R) = \Gamma_{\boldsymbol{\theta}}^{(P)}(\ln P)\,\Gamma_{\boldsymbol{\theta}}^{(R)}(\ln R) \tag{8}$$

but fit using the full general model.

The first – Catalog A – is generated assuming a smooth occurrence surface where both distributions are broken power laws.

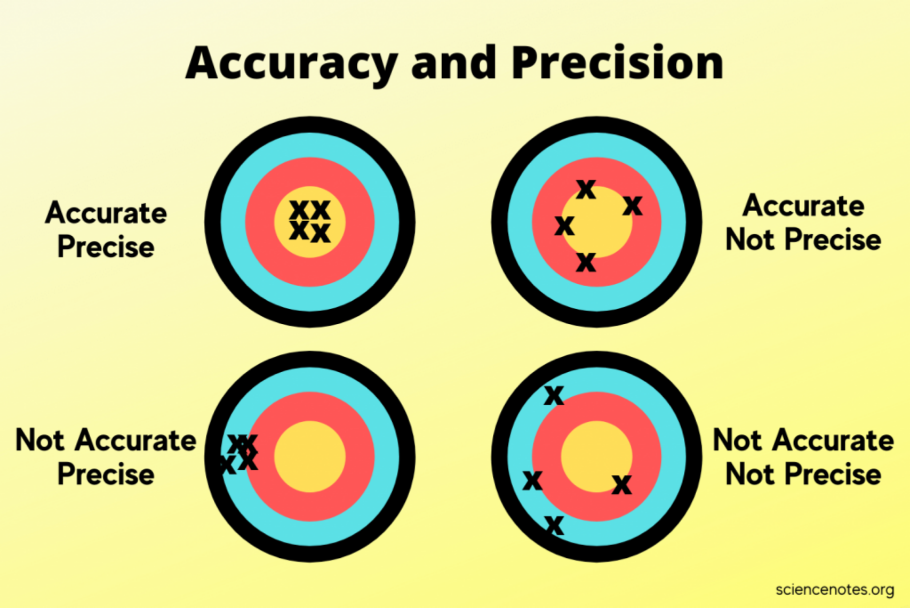
The second – Catalog B – is designed to be exactly the distribution inferred by Petigura et al. (2013) in the range that they considered and then smoothly extrapolated outside that range.

Catalog A – rate density

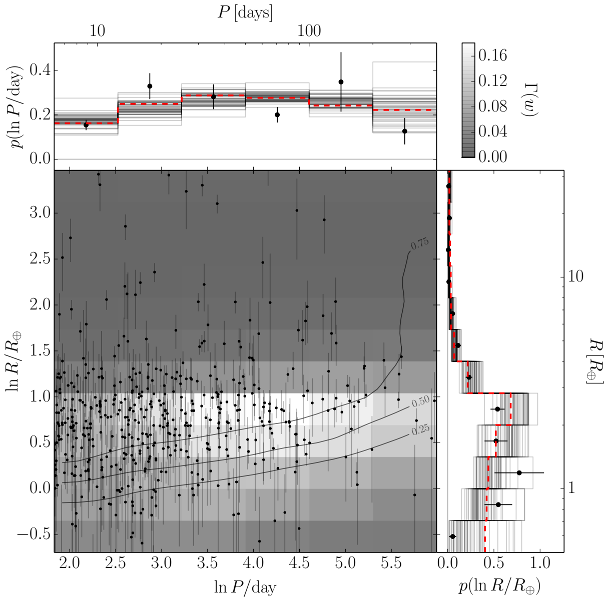


Probabilistic inference gives more precise and more accurate than inverse-detection-efficiency method.

Accuracy vs. precision

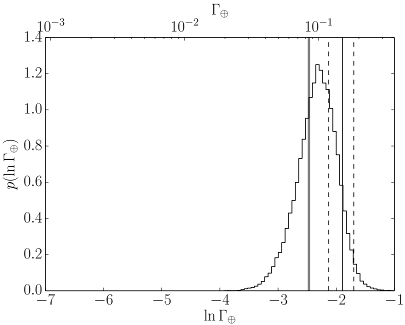


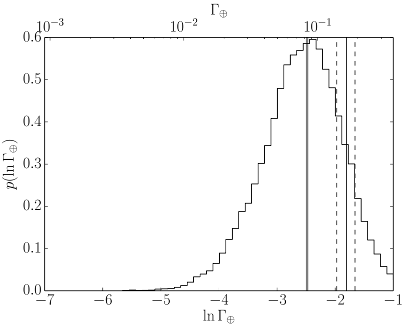
Catalog B – rate density



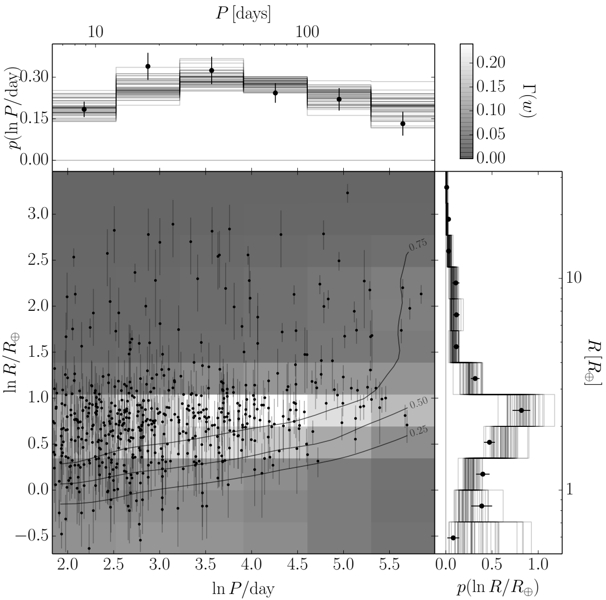
Probabilistic inference is less precise but more accurate

Extrapolated rate density

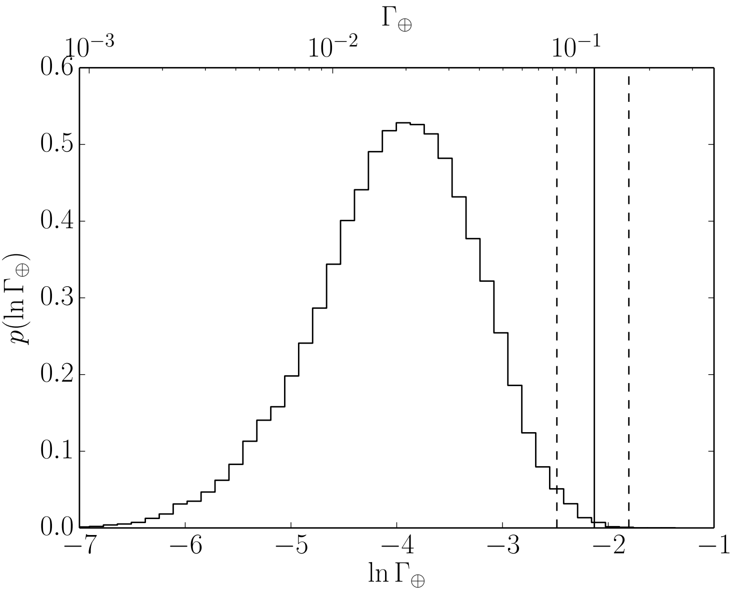




Real data – rate density



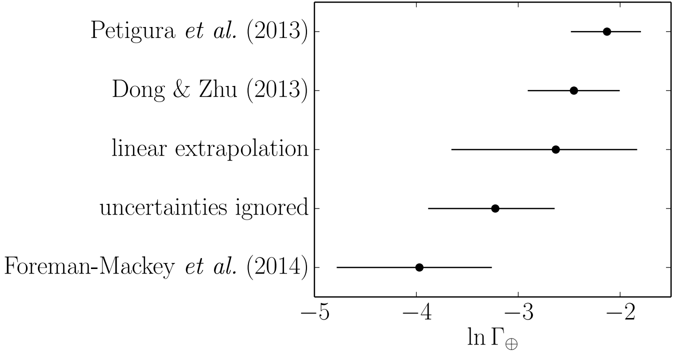
Real data – extrapolated rate density of Earth analogs



Results

$\Gamma_\oplus = 0.019^{+0.019}_{-0.010}\ \mathrm{nat}^{-2}$ – Foreman-Mackey et al. (2014)

$\Gamma_\oplus = 0.119^{+0.046}_{-0.035}\ \mathrm{nat}^{-2}$ – Petigura et al. (2013)



Final note

statistics computed on posterior samplings. For the sake of hierarchical inferences like the method presented here, it would be very useful if the authors of upcoming catalogs also published samples from these distributions along with the value of their prior function evaluated at each sample. In this spirit, we have released the results of this paper as posterior samplings[19] for the occurrence rate density function.

All of the code used in this project is available from http://github.com/dfm/exopop under the MIT open-source software license. This code (plus some dependencies) can be run to re-generate all of the figures and results in this article; this version of the paper was generated with git commit `d56324d` (2014 August 28).

We would like to thank Erik Petigura (Berkeley) for freely sharing his data and code. It is a pleasure to thank Ruth