Exoplanet Population Inference

A Tutorial

Dan Foreman-Mackey

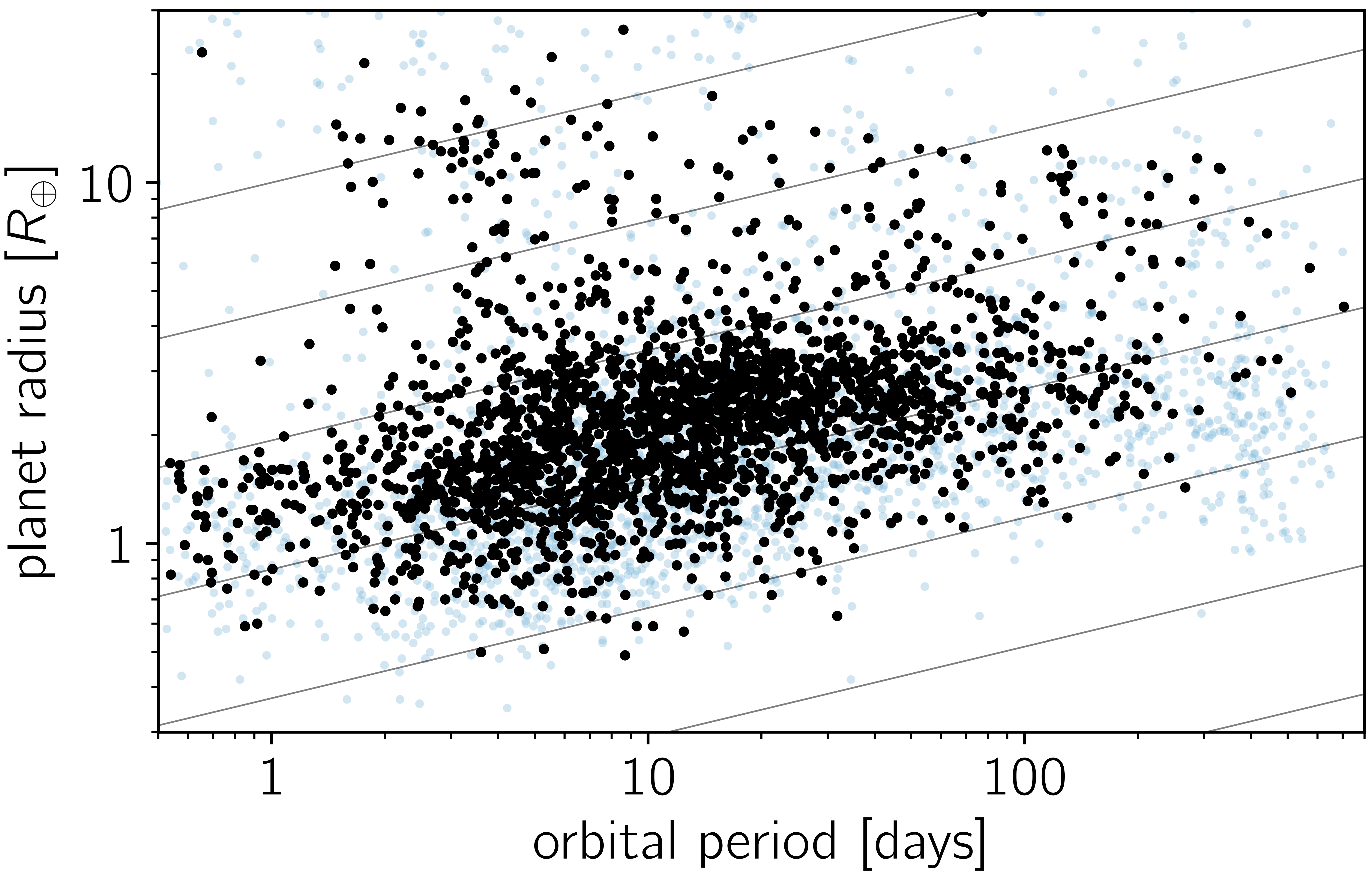
CCA@Flatiron // dfm.io

Today I'll mostly talk about transiting exoplanets*.

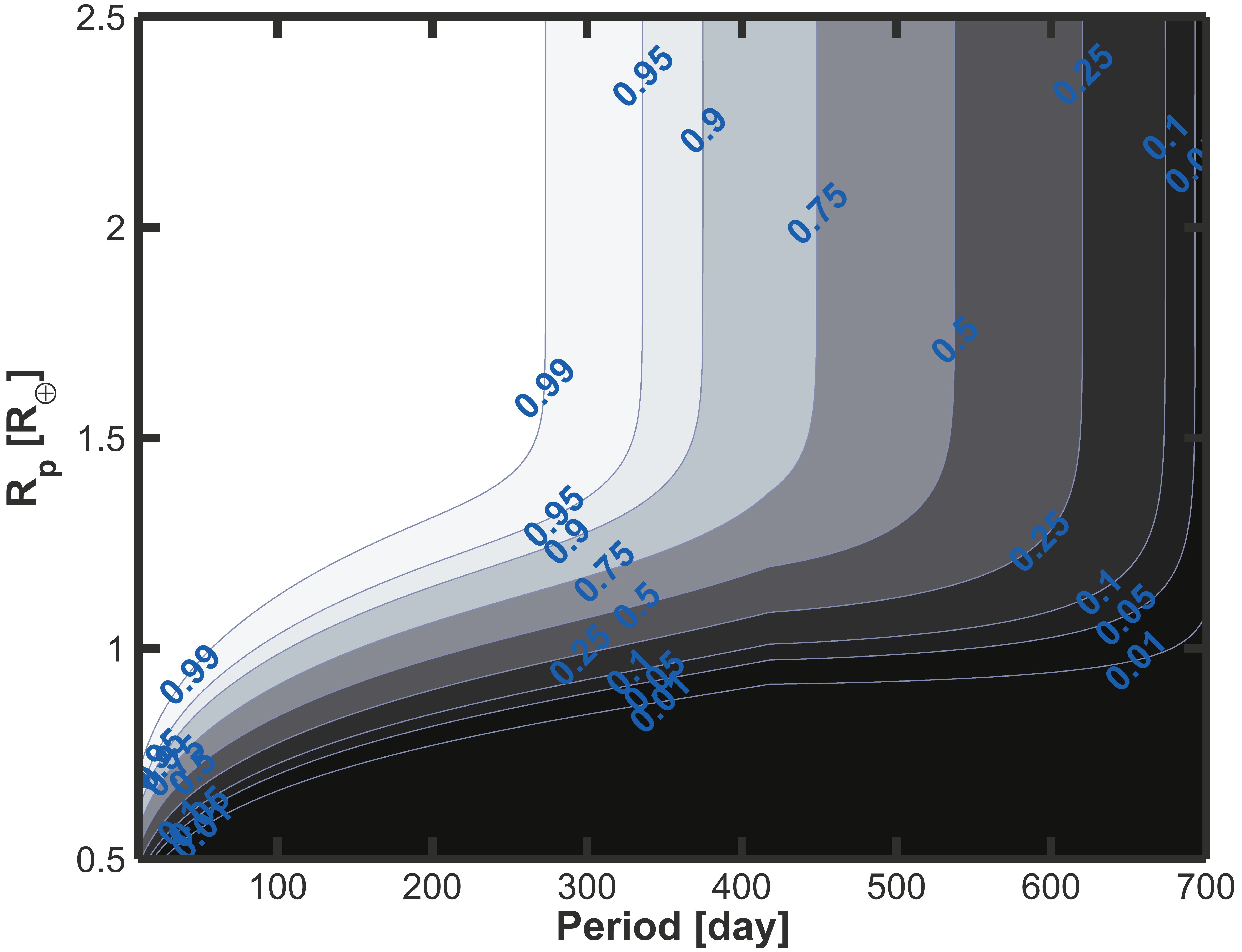
The methods can apply more broadly.

* this is what I know about and work on!

Exoplanet population inference



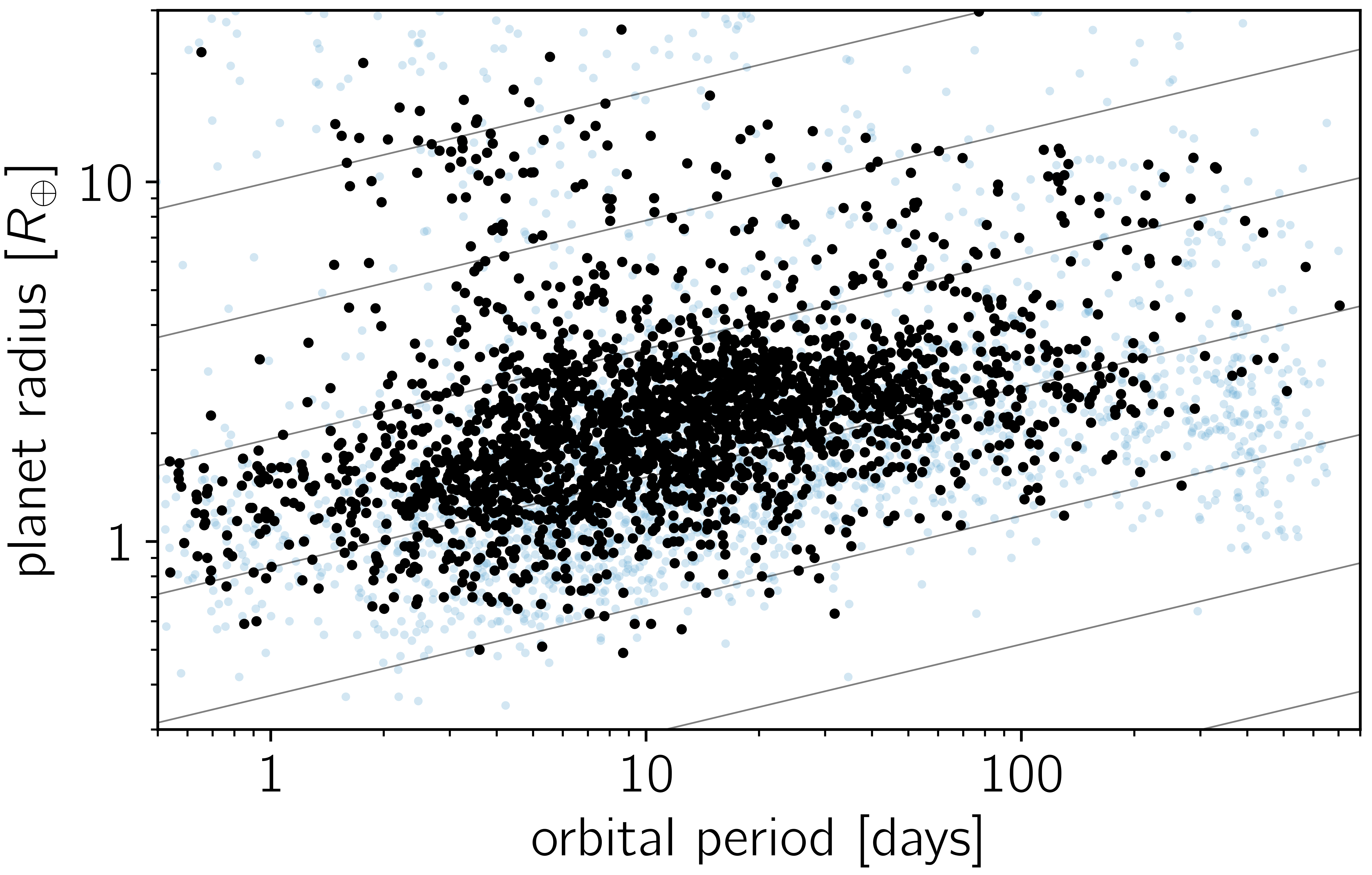
data: NASA Exoplanet Archive



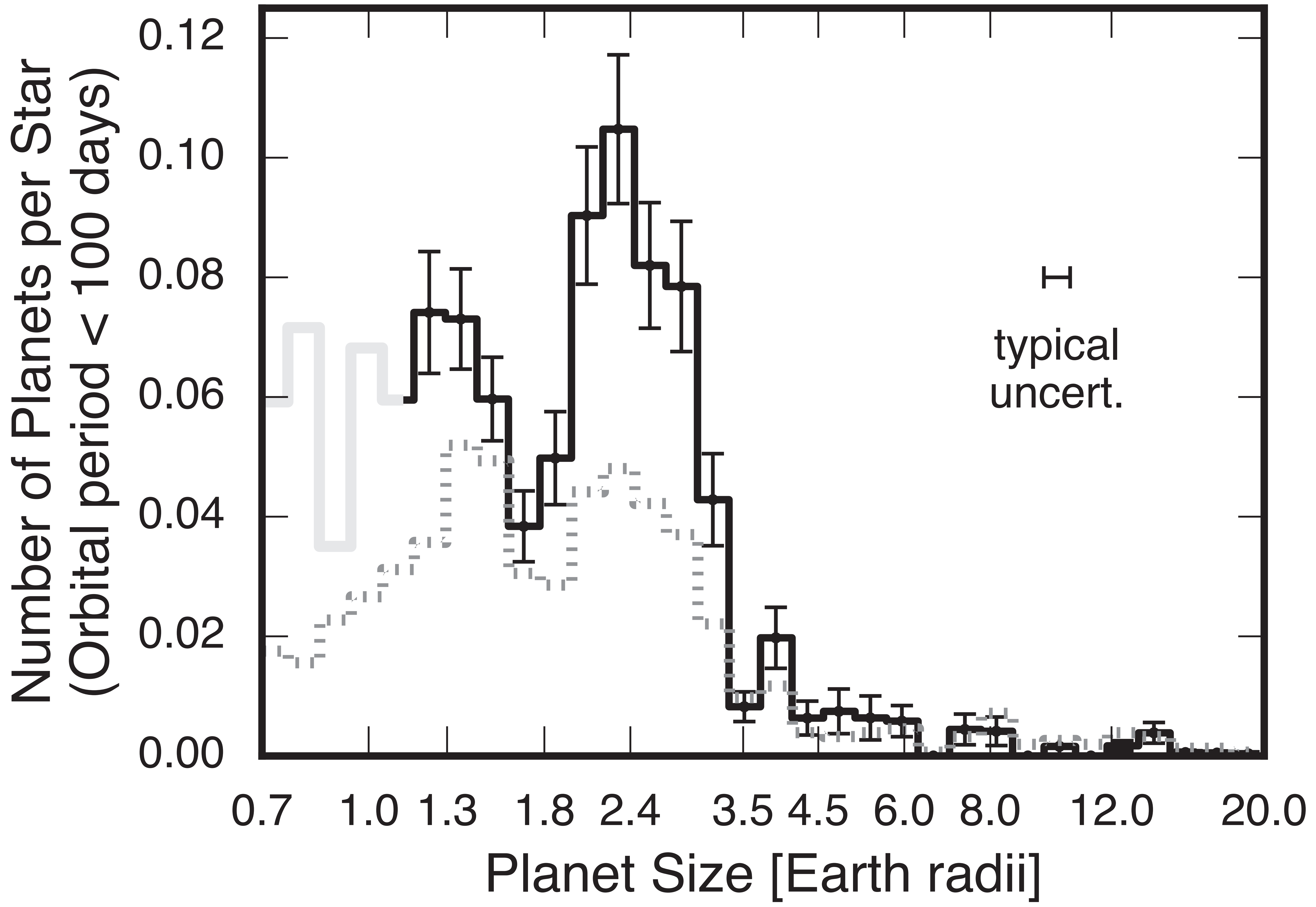
Burke, Christiansen et al. (2015)

Take these catalogs and get the physics of planet formation and evolution.

That's **hard**.



data: NASA Exoplanet Archive



Fulton & Petigura (2018)

2

What is an occurrence rate?

1

The expected number of planets per star.

2

The fraction of stars with planets.

The expected number of planets per star per unit planet property.

4

etc.

None of these definitions is **inherently better** than the others.

But. They are all **different**.

They have different **units**.

They all depend on a specific *(often unstated)* **definition of "planets"**.

So. It can be hard to compare and understand how they relate.

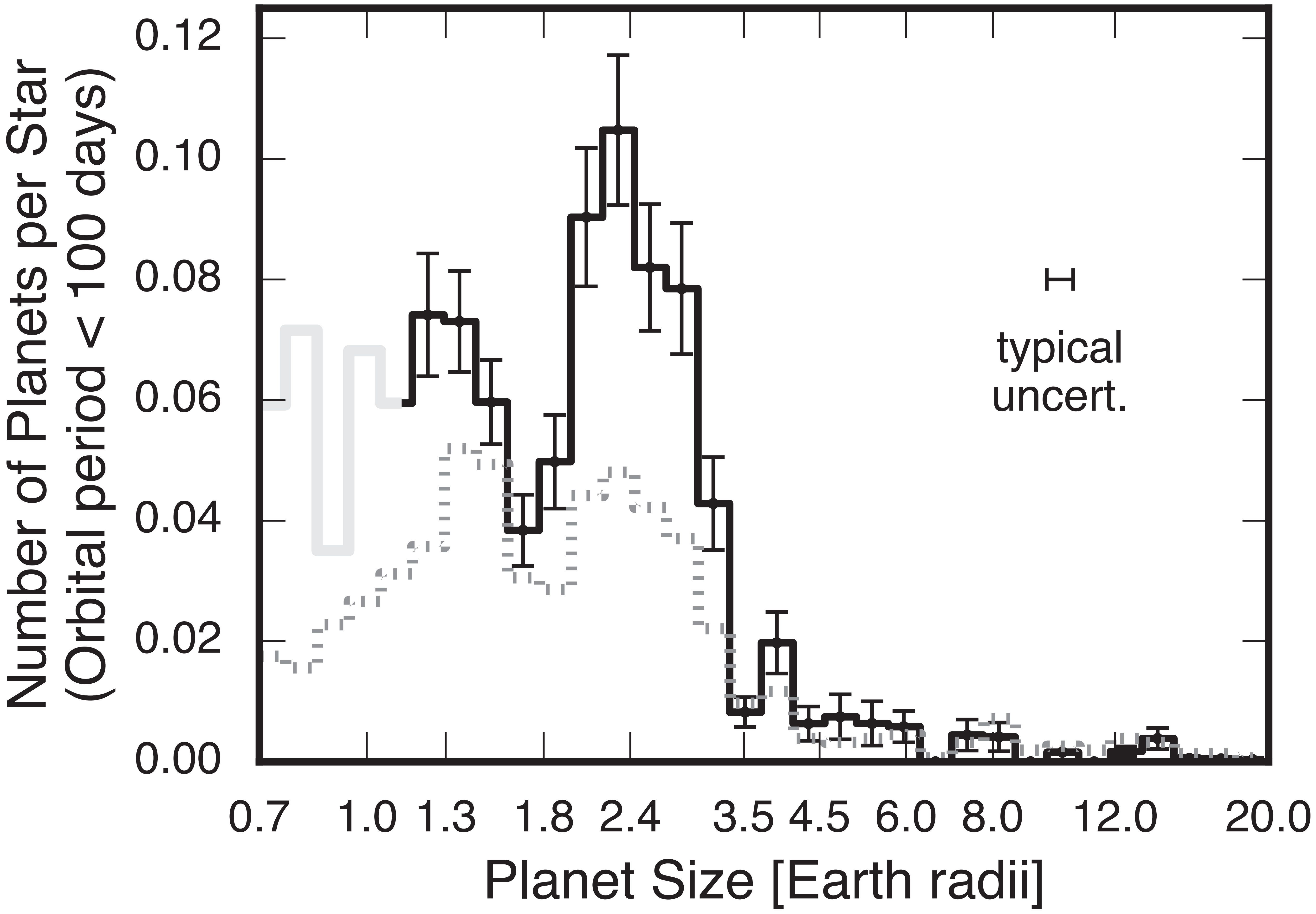
Them: * **"The occurrence rate is 10%."**

* including me and others in the room

Them: * "The occurrence rate is 10%."

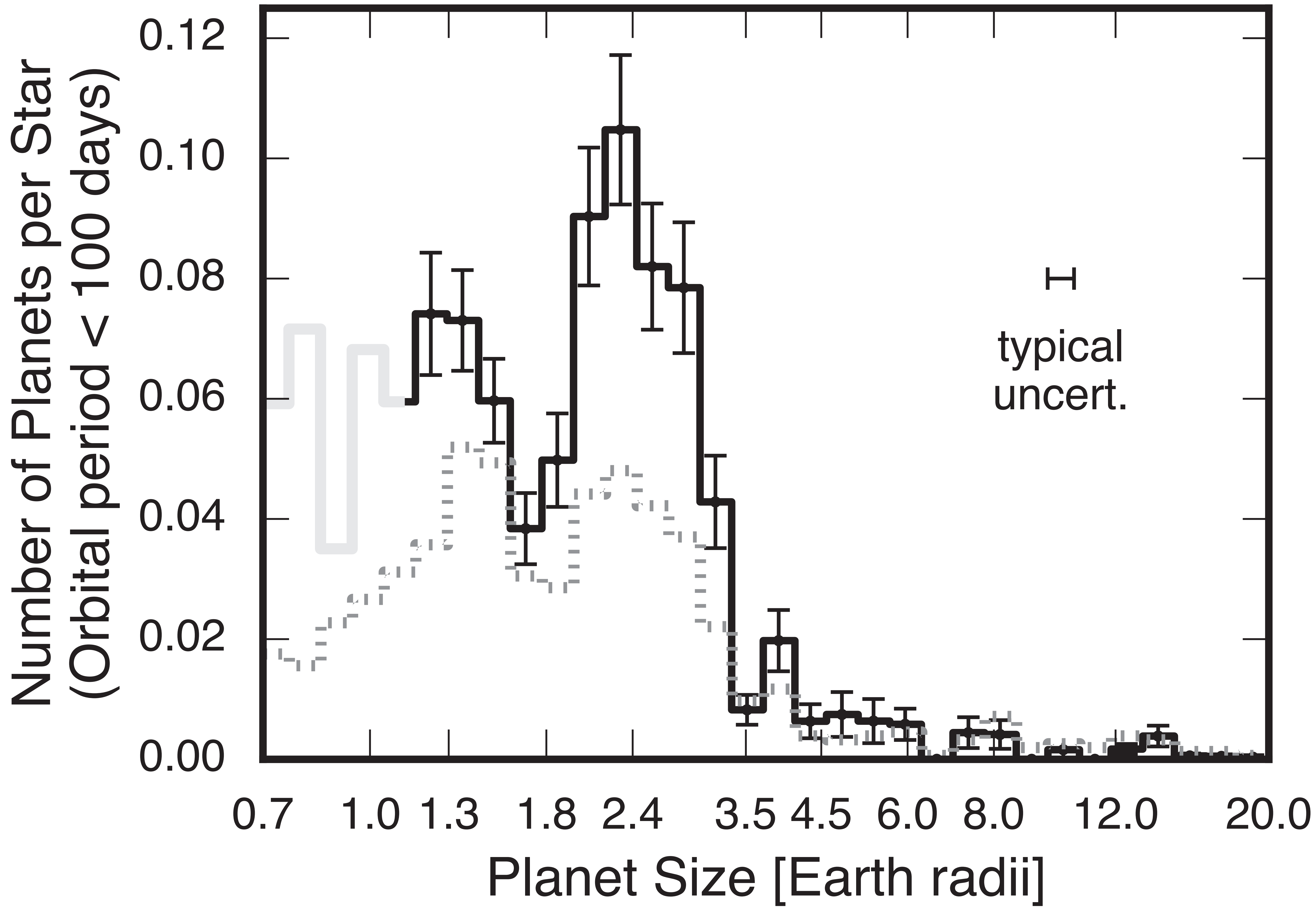
Y'all: "what does it all mean?!?1?"

* including me and others in the room



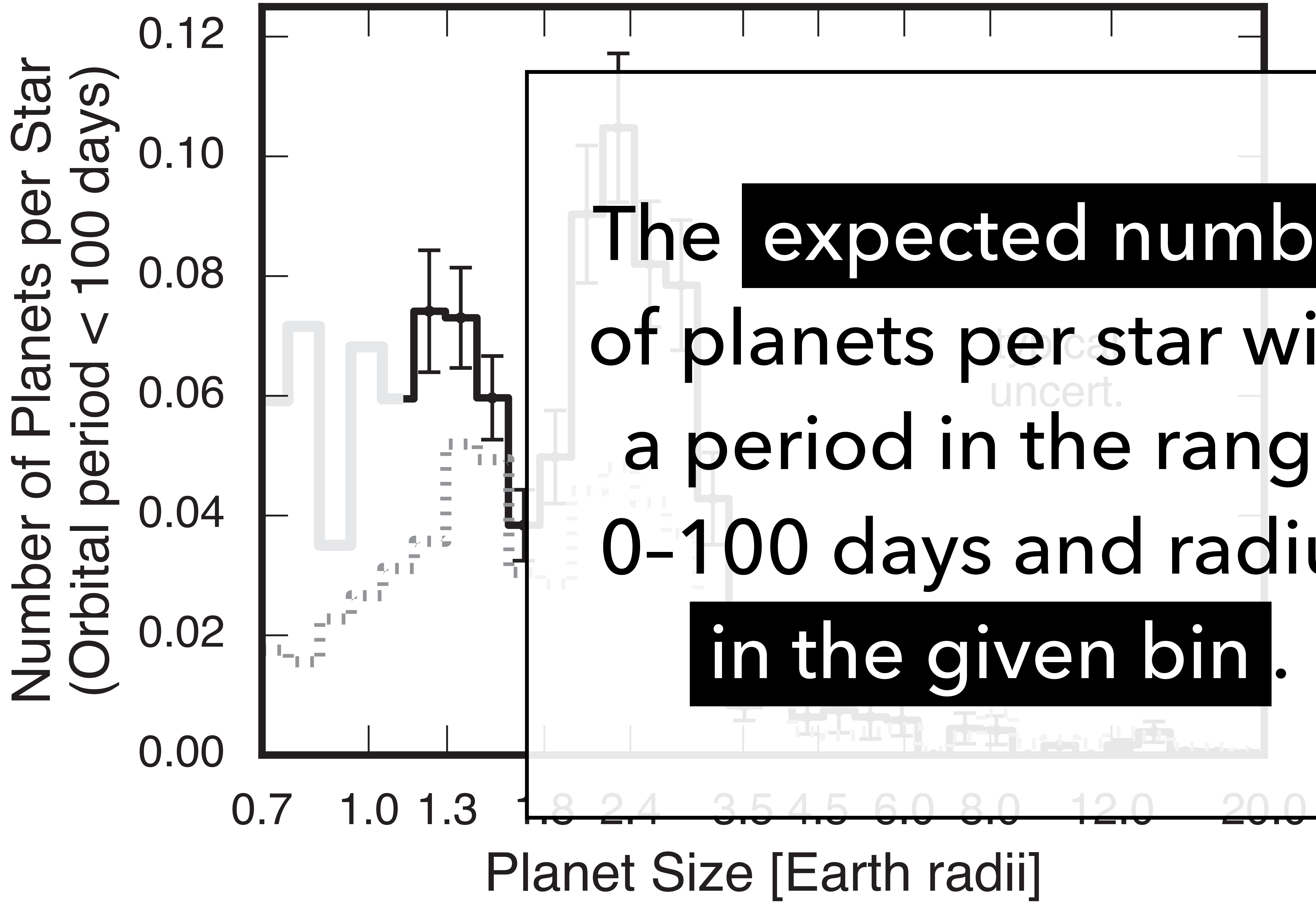
Fulton & Petigura (2018)

what do these numbers mean?



Fulton & Petigura (2018)

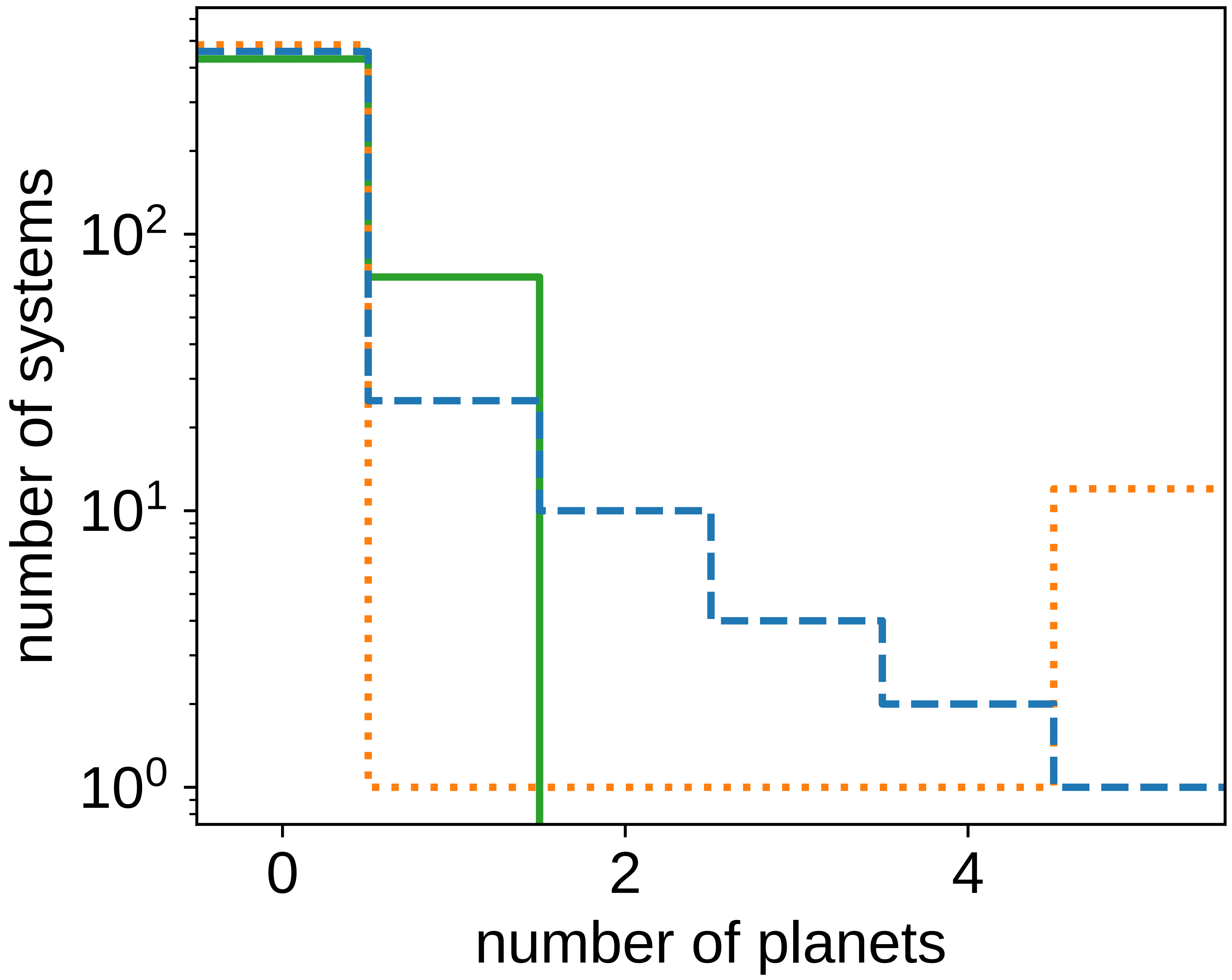
what do these numbers mean?

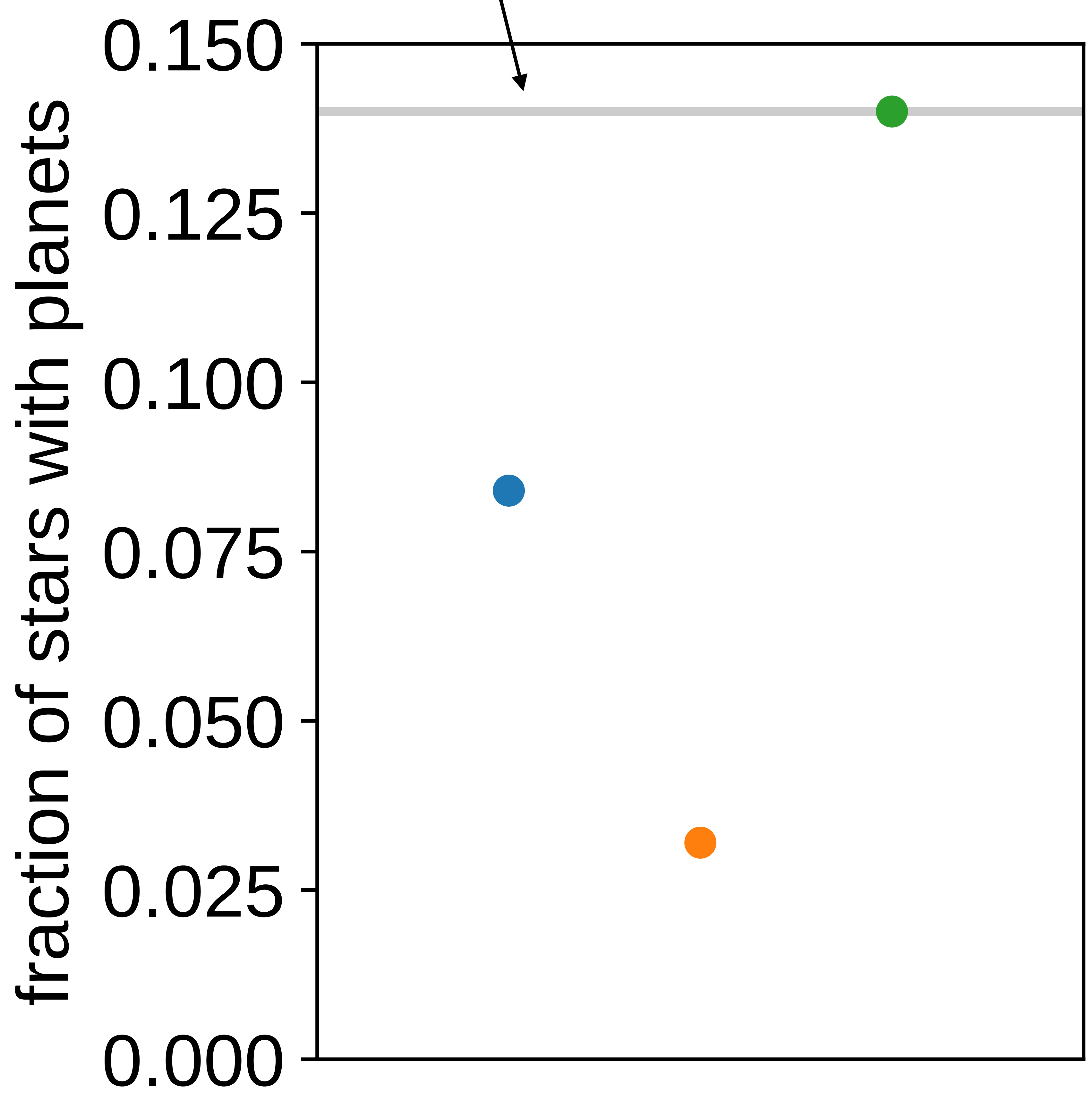


The **expected number** of planets per star with a period in the range 0–100 days and radius **in the given bin**.

Fulton & Petigura (2018)

expected number of planets per star





Simulations

github.com/dfm/exostar19

How to estimate an occurrence rate?

1. **Inverse detection efficiency**

2. **Probabilistic modeling**

3. **Approximate Bayesian Computation**

Inverse detection efficiency

$$N_{\text{expect}} = \frac{1}{N_{\text{tot}}} \sum_{j=1}^{N} \frac{1}{P_{\text{det}}(x_j)}$$

Note: don't do this!

2

Probabilistic modeling

$$N_{\text{expect}} = \arg\max_{N_{\text{expect}}} p(N_{\text{obs}},\, \{x_j\} \,|\, N_{\text{expect}},\, N_{\text{tot}})$$

3

Approximate Bayesian Computation

3 Approximate Bayesian Computation

1. **Inverse detection efficiency**
2. **Probabilistic modeling**
3. **Approximate Bayesian Computation**

1. **Inverse detection efficiency**

$\approx$

2. **Probabilistic modeling**

$=$

3. **Approximate Bayesian Computation**

1. **Inverse detection efficiency**
2. **Probabilistic modeling**
3. **Approximate Bayesian Computation**

$P(q_j)$ true number of planets

want

have

observed number of planets n_j, x_j the properties of the planets *and the star* observed number of planets

true number of planets

$$P(n_j \mid x_j, q_j)$$

the properties of the planets *and the star*

Start with either zero or one planet(s).

There are four options.

observed number of planets	true number of planets: $q_j = 0$	1
$n_j = 0$	1	$1 - P_{det}(x_j)$
1	0	$P_{det}(x_j)$

value of $P(n_j \mid x_j, q_j)$

But. We don't know the true number of planets.

Marginalize!

$$P(n_j \mid x_j) = \sum_{q_j \in \{0, 1\}} P(q_j)\, P(n_j \mid x_j, q_j)$$

$$
\begin{aligned}
P(n_j \mid x_j) &= \sum_{q_j \in \{0,1\}} P(q_j)\, P(n_j \mid x_j, q_j) \\
&= Q\, P(n_j \mid x_j, q_j = 1) + (1 - Q)\, P(n_j \mid x_j, q_j = 0)
\end{aligned}
$$

$$P(n_j \,|\, x_j) = \sum_{q_j \in \{0,1\}} P(q_j)\, P(n_j \,|\, x_j,\, q_j)$$

$$= Q\, P(n_j \,|\, x_j,\, q_j = 1) + (1 - Q)\, P(n_j \,|\, x_j,\, q_j = 0)$$

this is the parameter that we want to fit for!

But. We don't know the properties of the unobserved planets.

Marginalize!

systems with no planets

$$P(n_j = 0) = \int p(x_j)\, P(n_j = 0 \,|\, x_j)\, \mathrm{d}x_j$$

$$= 1 - Q \int p(x_j)\, P(n_j = 1 \,|\, x_j,\ q_j = 1)\, \mathrm{d}x_j$$

$$= 1 - Q\, P_0$$

systems with detected planets

$$P(n_j = 1) = p(x_j)\, P(n_j = 1 \,|\, x_j)$$

$$= p(x_j)\, Q\, P(n_j = 1 \,|\, x_j,\ q_j = 1)$$

systems with no planets

$$
\begin{aligned}
P(n_j = 0) &= \int p(x_j)\, P(n_j = 0 \,|\, x_j)\, \mathrm{d}x_j \\
&= 1 - Q \int p(x_j)\, P(n_j = 1 \,|\, x_j,\, q_j = 1)\, \mathrm{d}x_j \\
&= 1 - Q\, P_0
\end{aligned}
$$

detection probability

systems with detected planets

$$
\begin{aligned}
P(n_j = 1) &= p(x_j)\, P(n_j = 1 \,|\, x_j) \\
&= p(x_j)\, Q\, P(n_j = 1 \,|\, x_j,\, q_j = 1)
\end{aligned}
$$

Put it all together.

An exercise for the reader…

the fraction of stars with observed planets

$$Q = \frac{N_1}{N_0 + N_1} \frac{1}{P_0}$$

the occurrence rate

the fraction of stars with observed planets

$$Q = \frac{N_1}{N_0 + N_1} \frac{1}{P_0}$$

the occurrence rate

$$P_0 = \int p(x_j)\, P(n_j = 1 \,|\, x_j,\, q_j = 1)\, \mathrm{d}x_j$$

the detection probability averaged over the distribution of planet and stellar properties

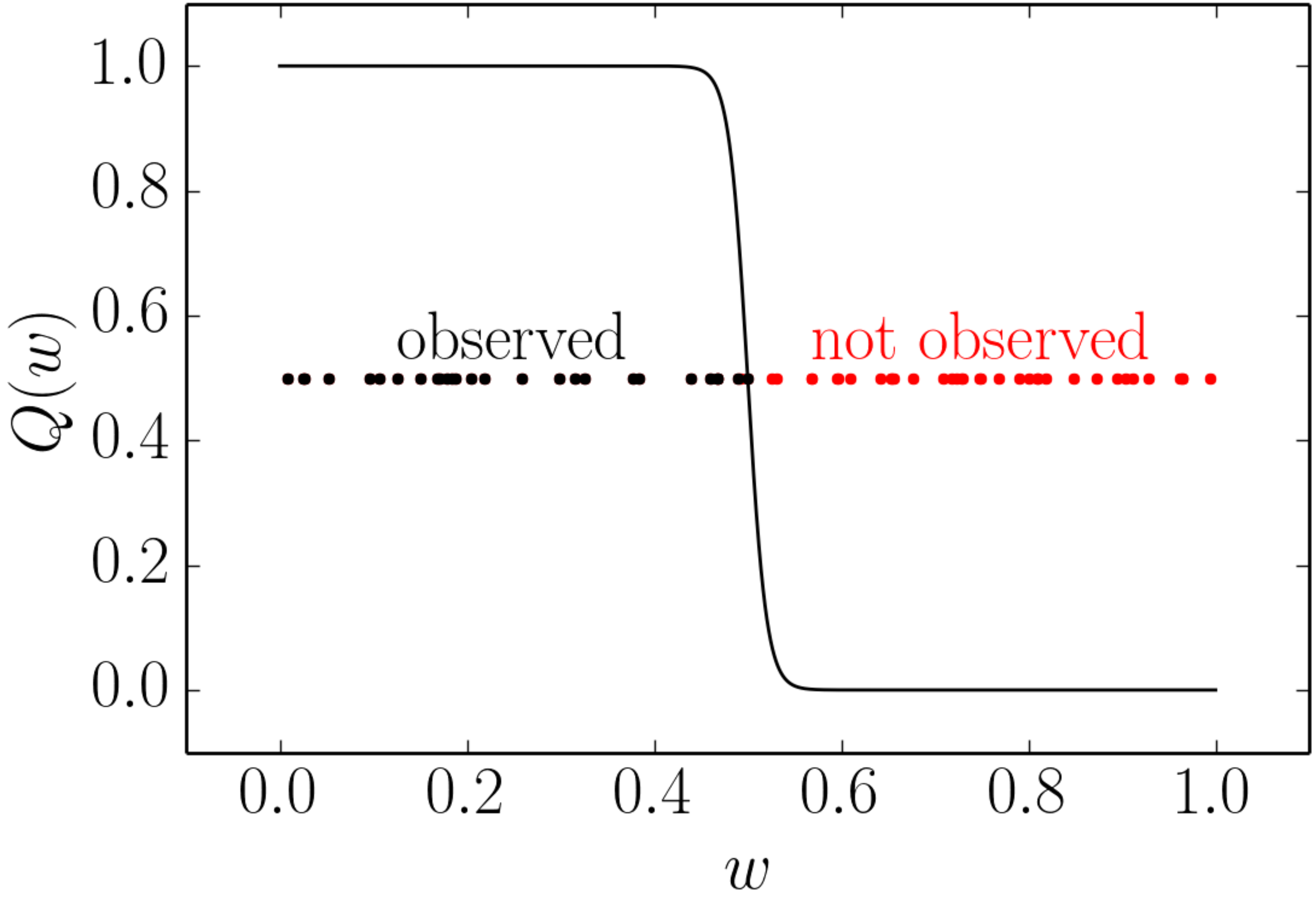
the fraction of stars with observed planets

$$Q = \frac{N_1}{N_0 + N_1}\,\frac{1}{P_0} \neq \frac{1}{N_0 + N_1}\sum_{j=1}^{N_1}\frac{1}{P_j}$$

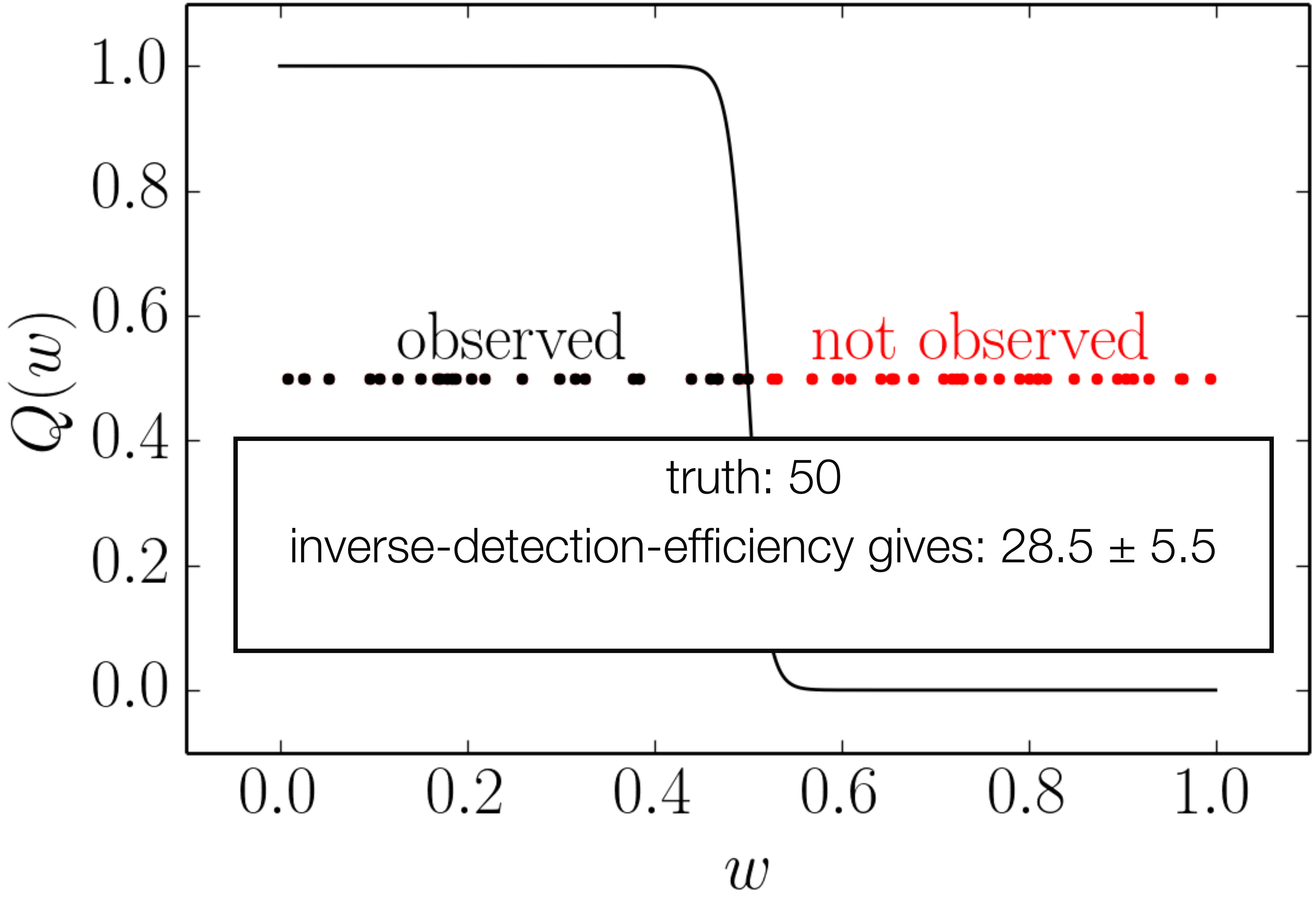
the occurrence rate

$$P_0 = \int p(x_j)\,P(n_j = 1 \mid x_j,\, q_j = 1)\,\mathrm{d}x_j$$

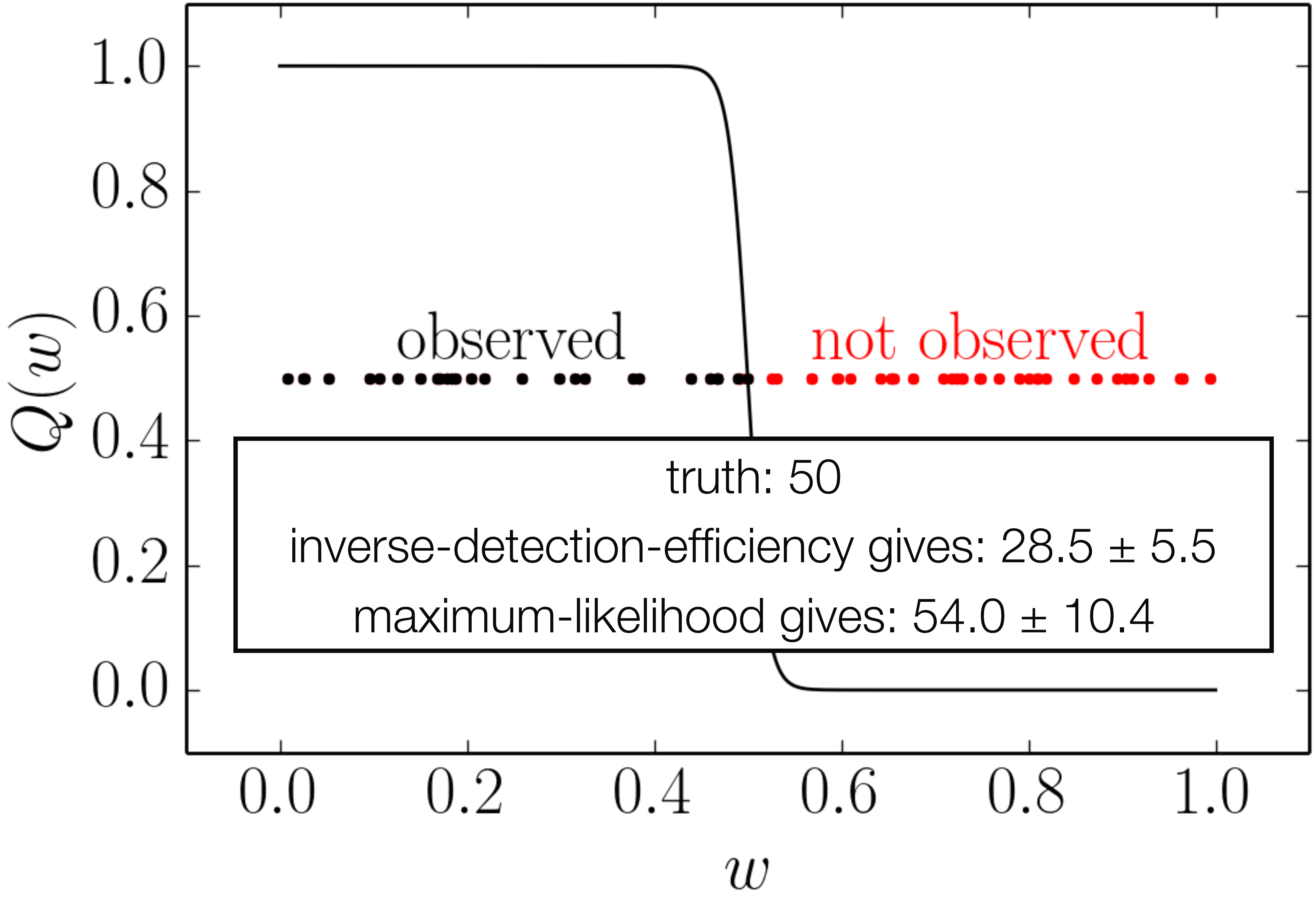
the detection probability averaged over the distribution of planet and stellar properties



see: dfm.io/posts/histogram1



see: dfm.io/posts/histogram1



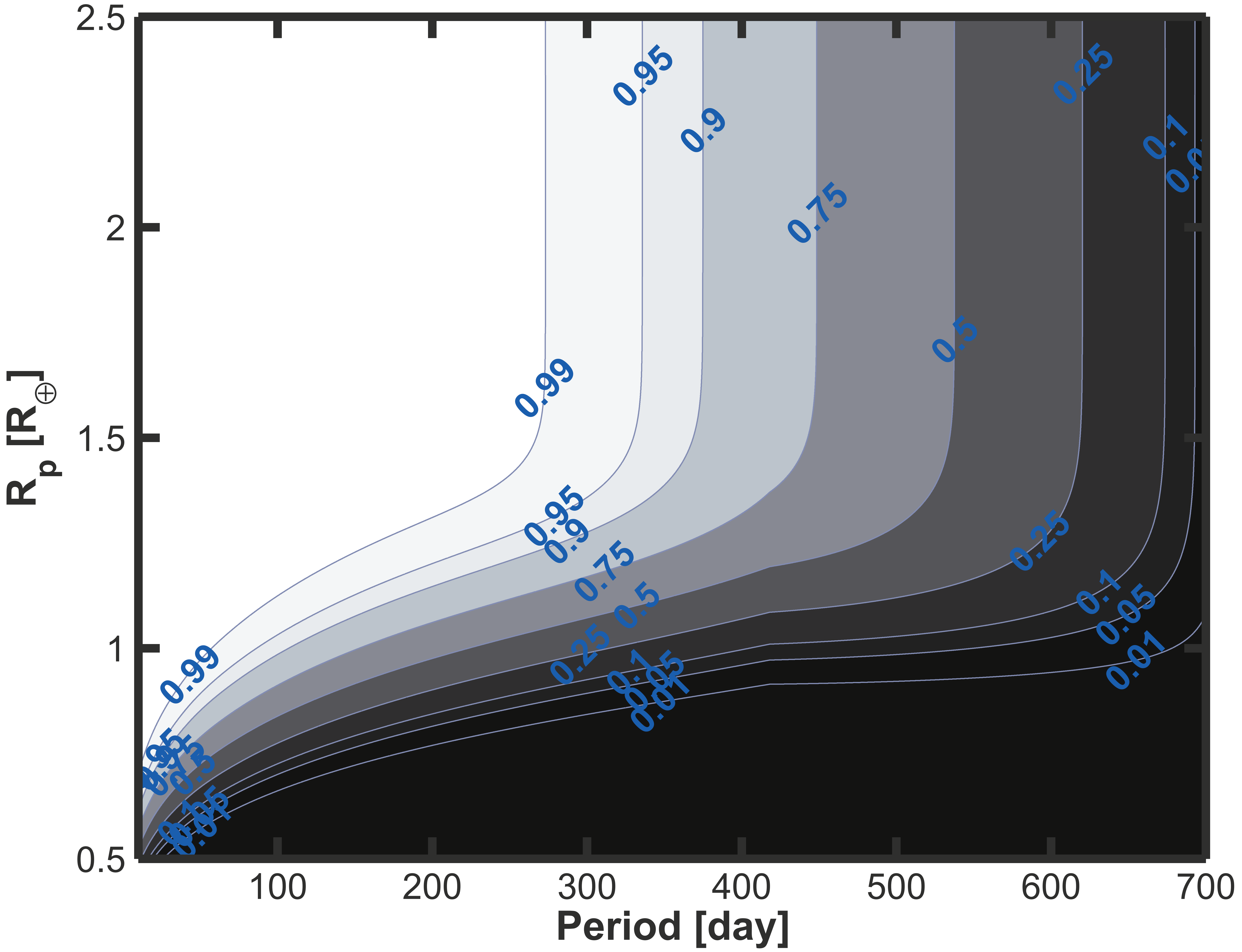
see: dfm.io/posts/histogram1

Inverse detection efficiency is not the right estimator.

Instead, take the fraction of detections and divide by the average detection efficiency*.

* averaged over *the correct distribution* for all planet and star properties

The key ingredient is the detection efficiency model.



Burke, Christiansen et al. (2015)

Remember: an occurrence rate depends on a lot of decisions!

1. **Stellar sample**
2. **Range of planet parameters**
3. **Units**
4. **Planet multiplicity**

4

Complications

1. **Multiplicity**
   (planetary and stellar)
2. **Uncertainties**
3. **False positives**
4. **Heterogeneous catalogs**

You end up needing to do an integral over all the properties of all the planets and false positives that you didn't observe.

1

Mathematica™ can't do that integral.

2

Eric Agol can't do that integral.

3

MCMC can't do that integral*.

* in finite time.

This is where you use approximate Bayesian computation (ABC).

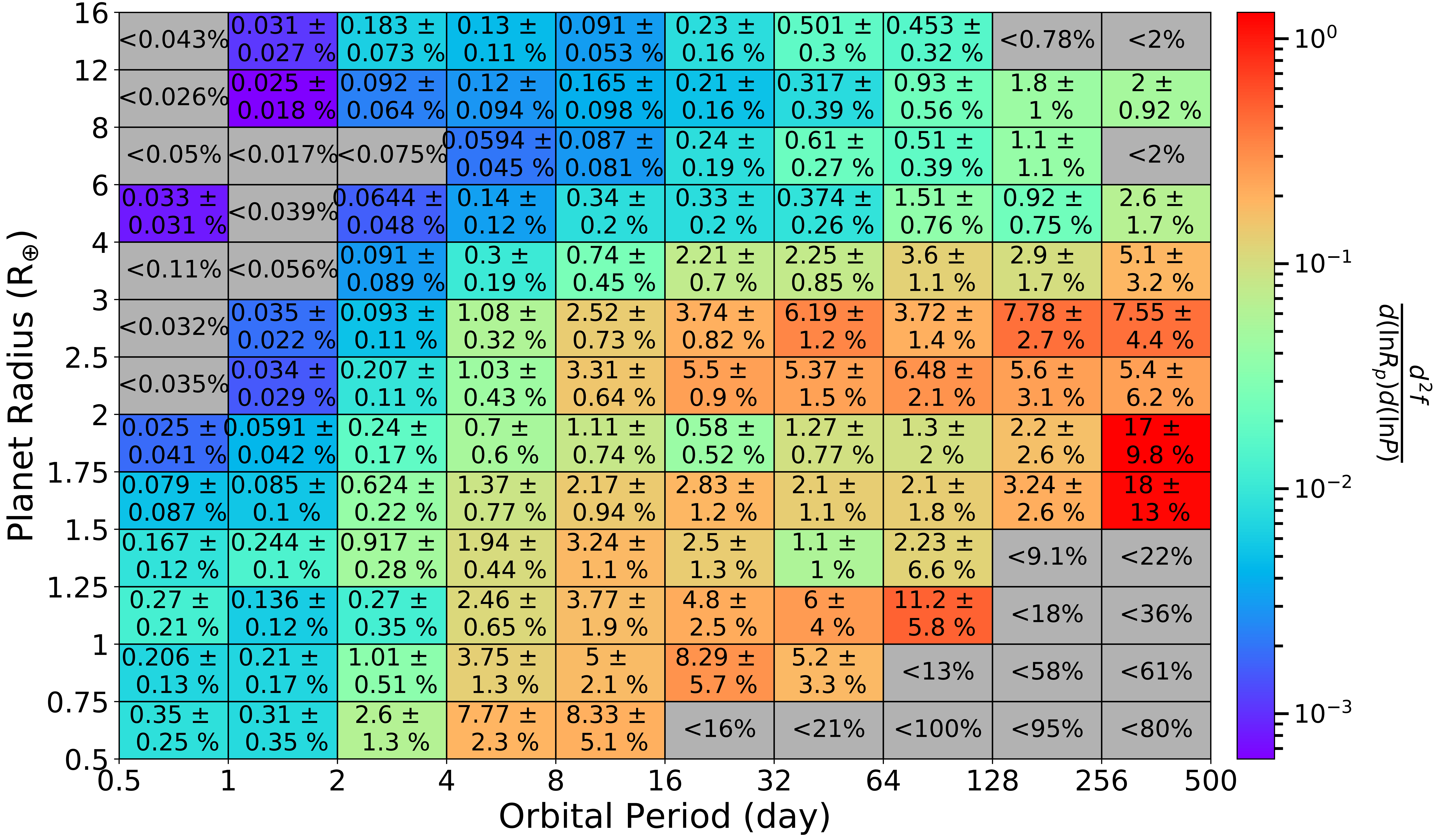
This is where you use ~~approximate Bayesian computation (ABC).~~ likelihood-free inference.

Likelihood-free inference is a method for doing rigorous inference with stochastic models.

a realistic catalog

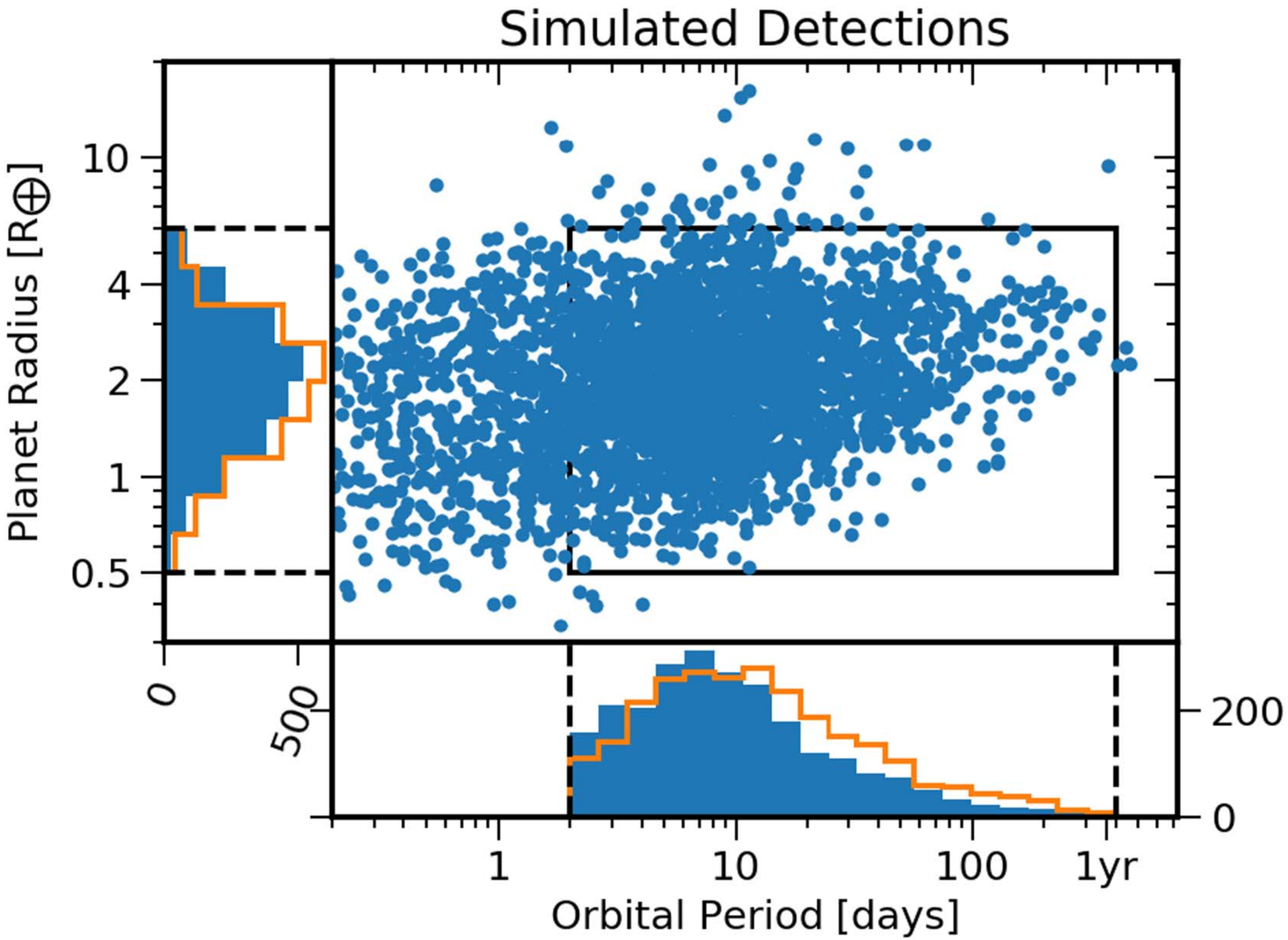
If you can simulate it then you can do inference.

The *promise* of "likelihood-free inference".

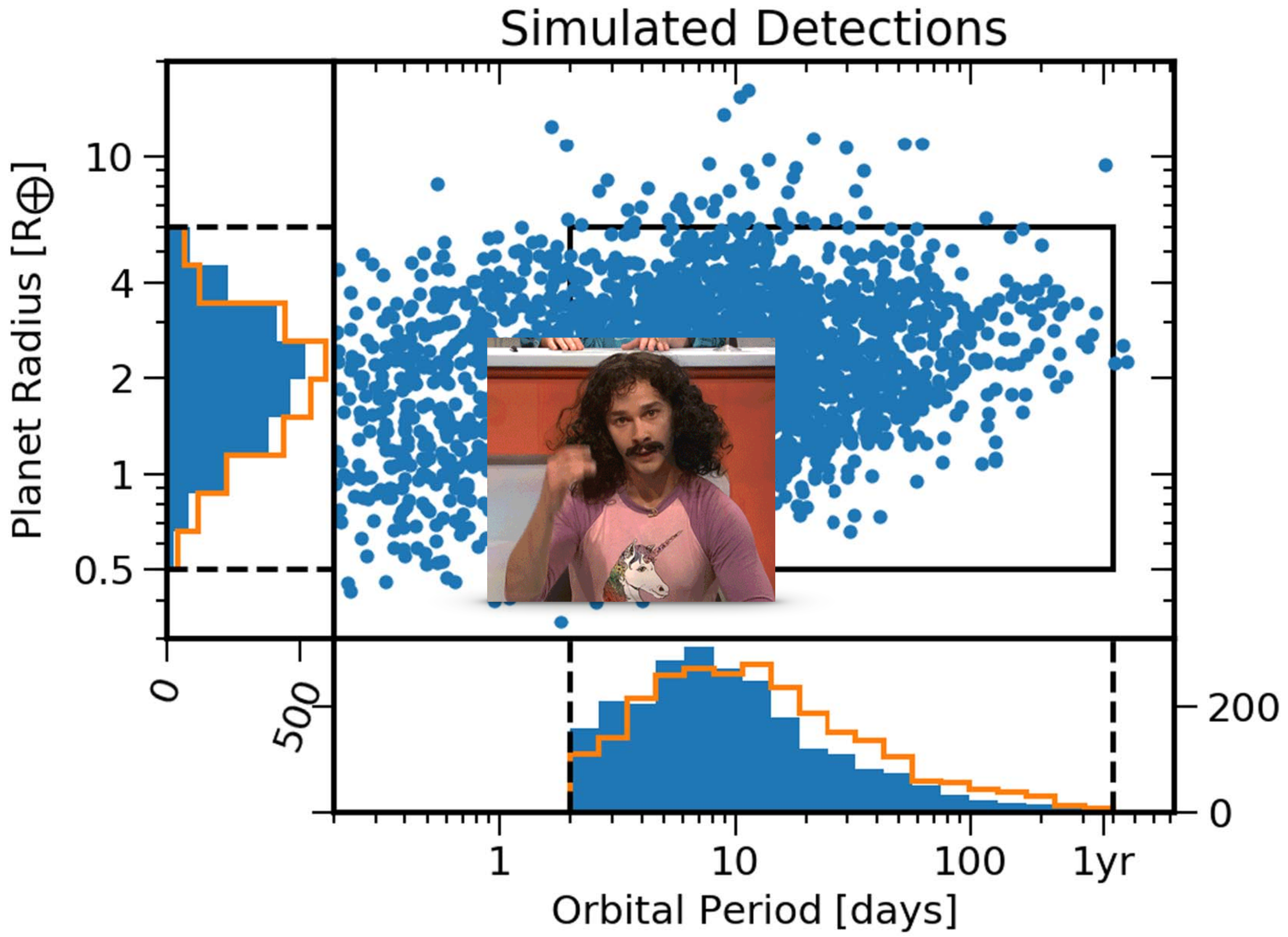


Hsu et al. (2019)

There's still lots to do!



EPOS; Mulders et al. (2018)



EPOS; Mulders et al. (2018)

5

Take homes

An occurrence rate needs to come with a lot of metadata.

Comparing occurrence rates:

Check the units.

Check the parameter ranges.

Don't sum the inverse detection probabilities for your planets!

* a more reliable estimator *is just as easy* to compute!

If you're using a method that seems intuitive, make sure the math checks out!

Likelihood-free inference seems like a promising way forward.

* a.k.a. Approximate Bayesian Computation (ABC)

It's over.

Extras.

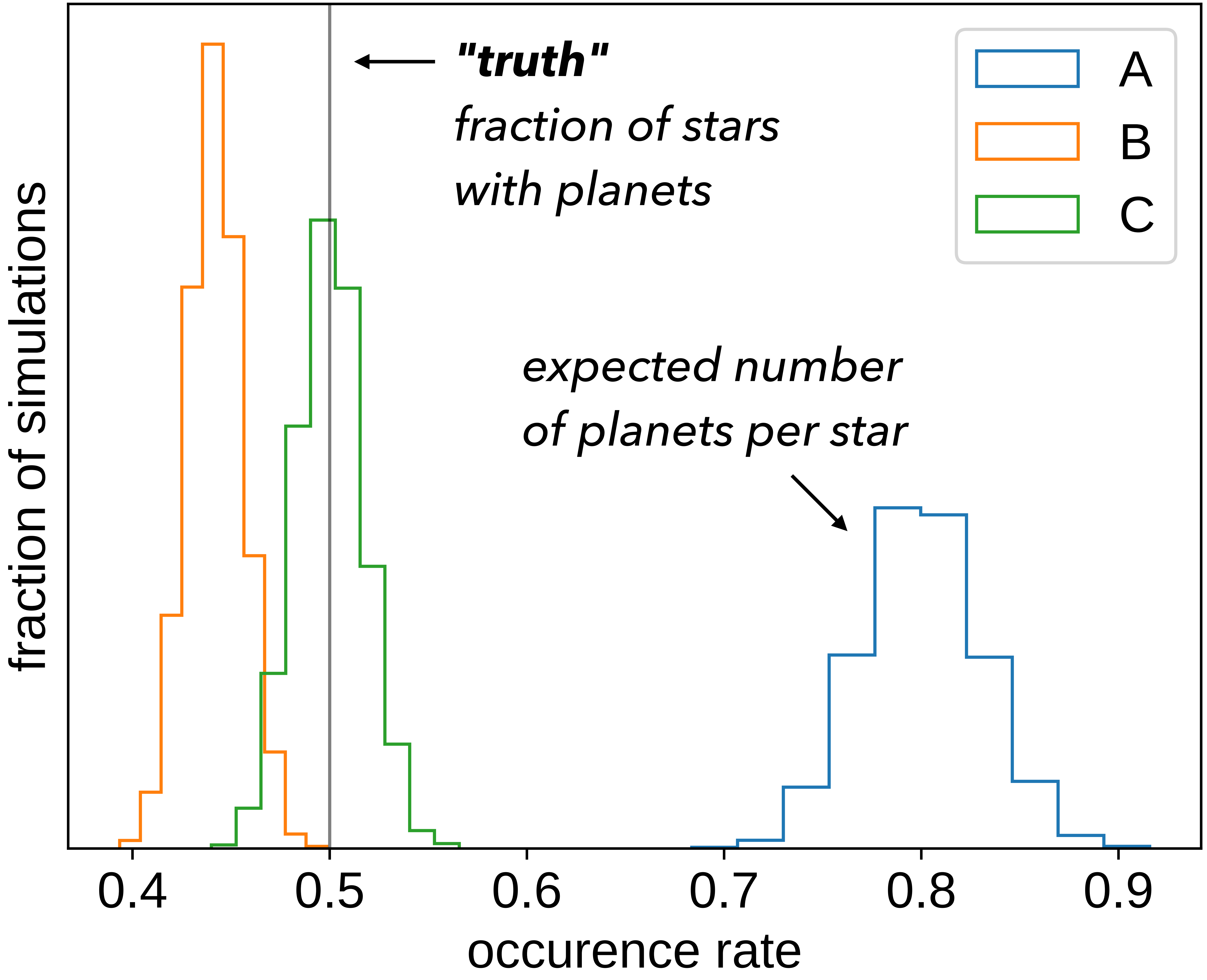
$$p(\{n_j\}, \{x_j\} \mid Q) = [1 - Q\, P_0]^{N_0} \left[\prod_{j=1}^{N_1} Q\, p(x_j)\, P(n_j = 1 \mid x_j,\, q_j = 1) \right]$$

$$\log p(\{n_j\}, \{x_j\} \,|\, Q) = N_0 \log\left(1 - Q\,P_0\right) + N_1 \log Q + \text{constant}$$

$$\log p(\{n_j\}, \{x_j\} \mid Q) = N_0 \log(1 - Q\, P_0) + N_1 \log Q + \text{constant}$$

$$\downarrow$$

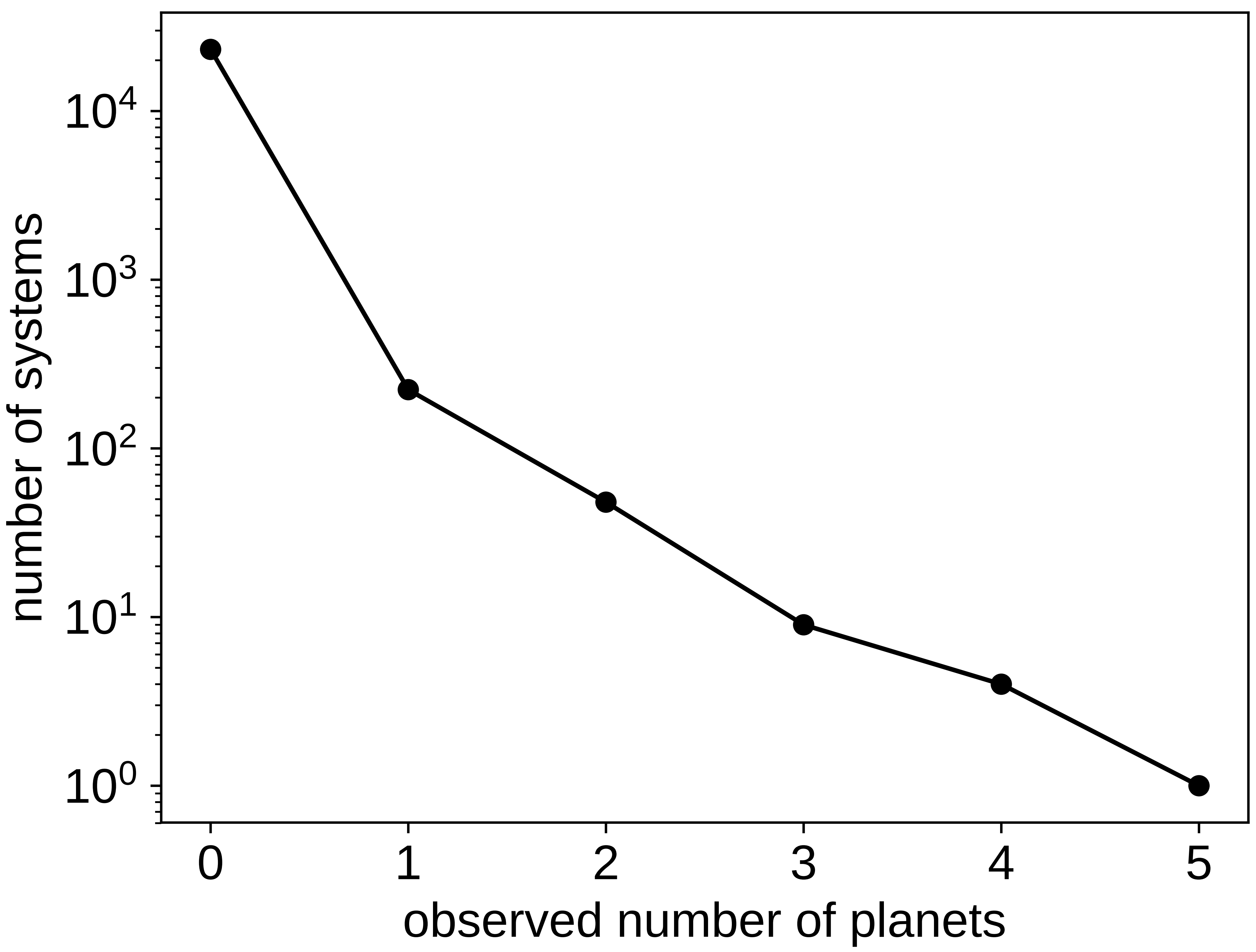
$$Q = \frac{N_1}{N_0 + N_1} \frac{1}{P_0}$$

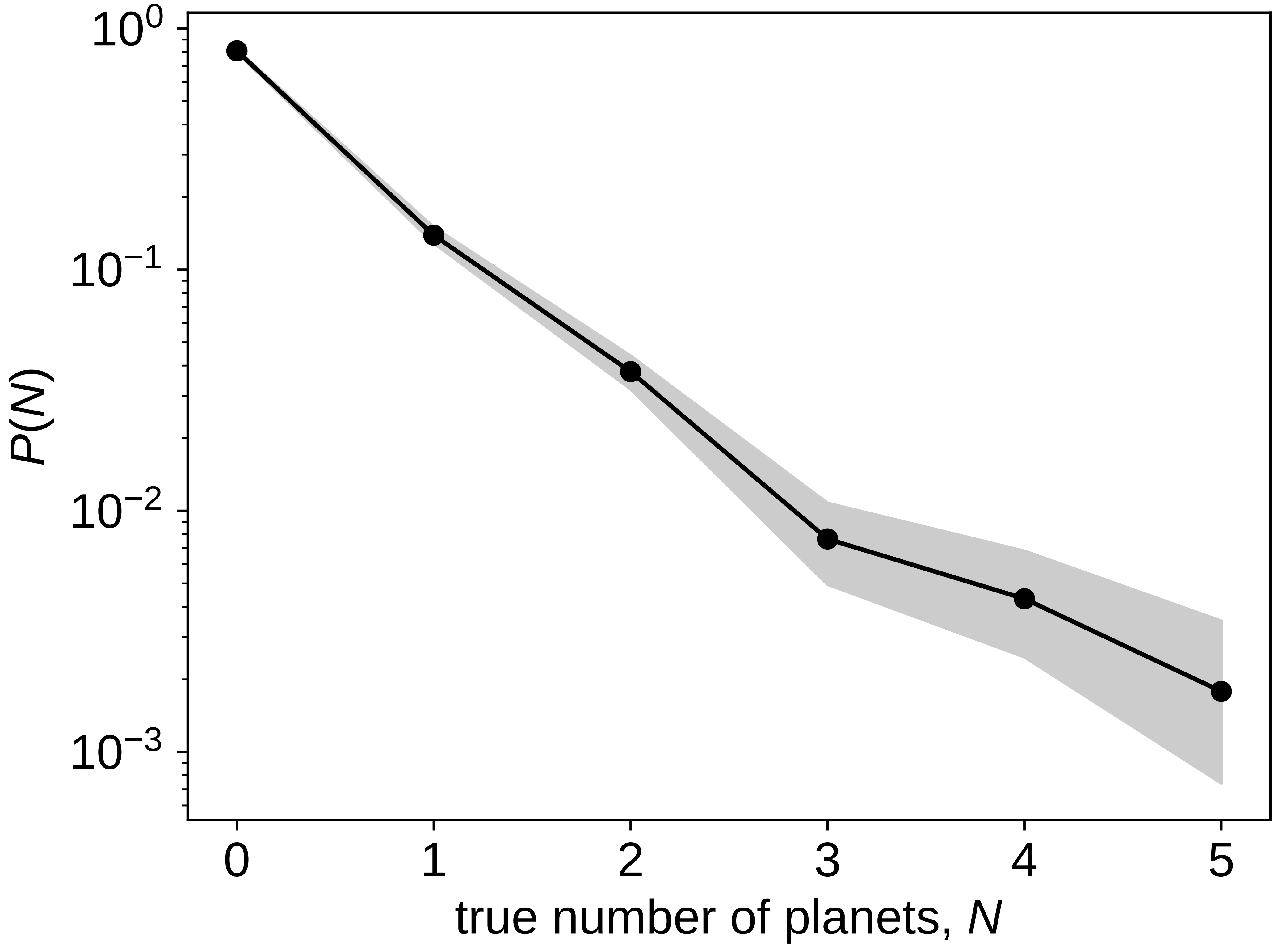


Simulations

github.com/dfm/exostar19

Note: this is preliminary & really just a toy...

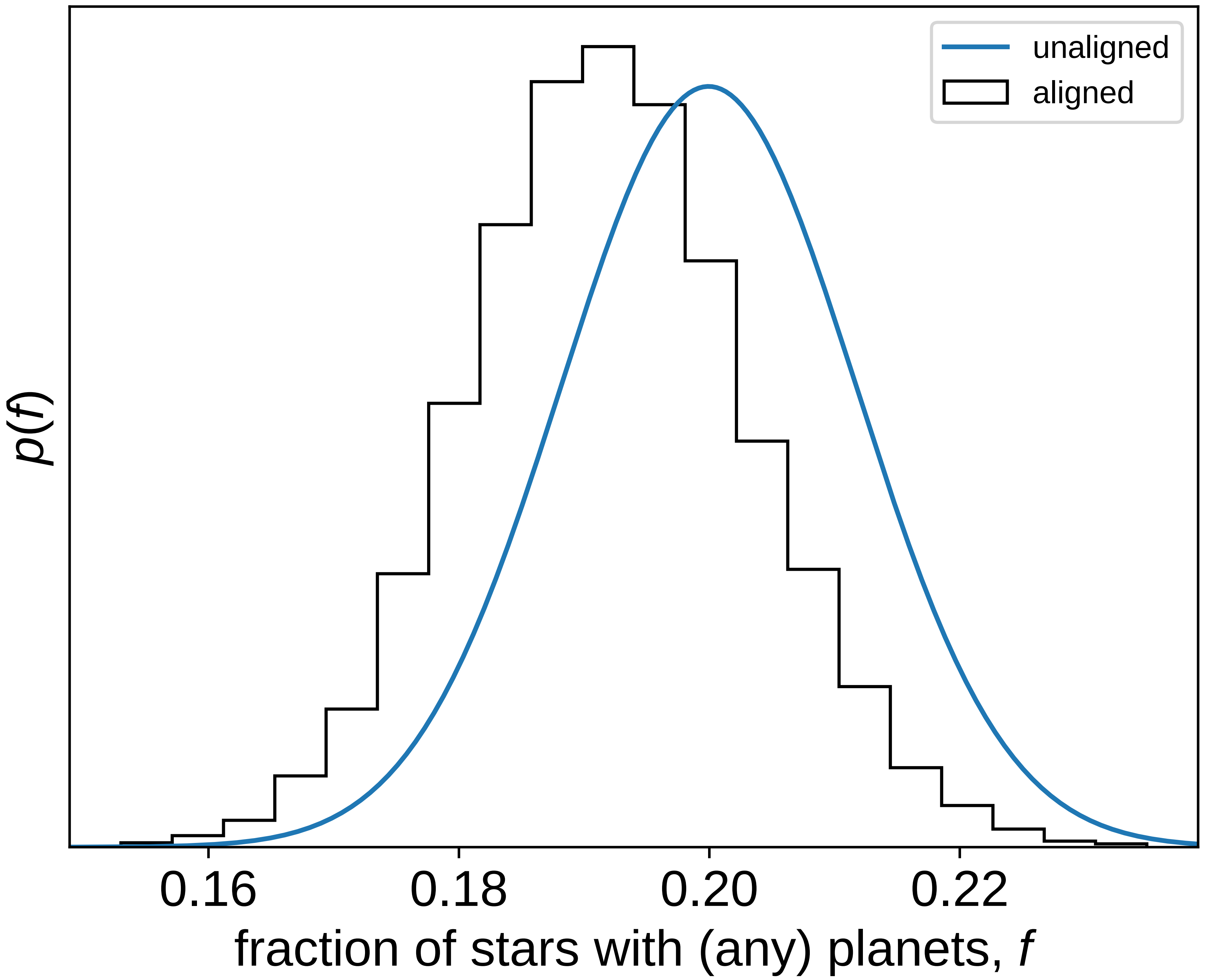


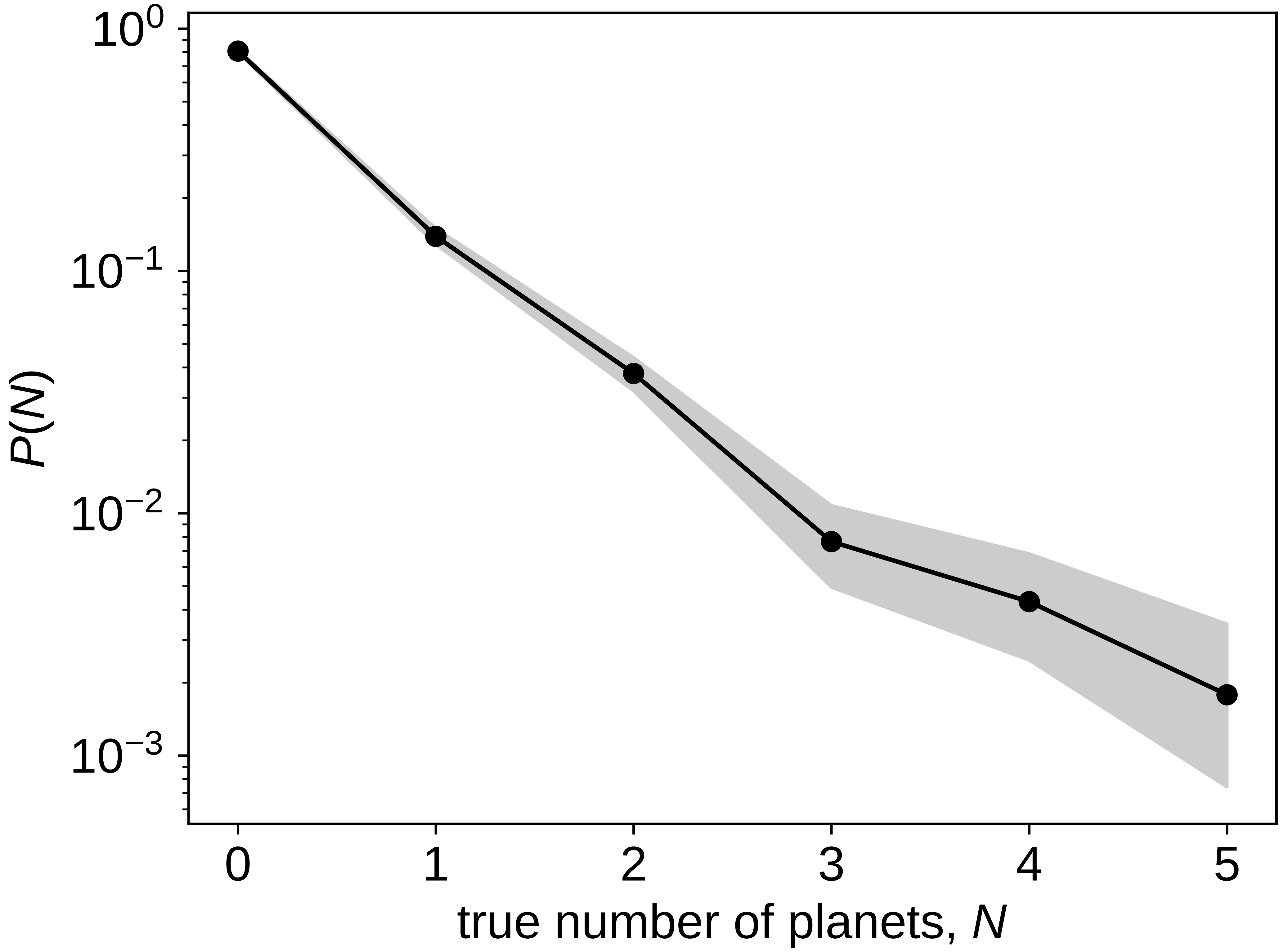


assuming:
no mutual inclination
only geometric transit probability

$0.5 < R_P/R_{Earth} < 8$; $10 < a/R_{star} < 30$
Kepler data:
github.com/dfm/exostar19

Note: this is preliminary & really just a toy...





assuming:
no mutual inclination
only geometric transit probability

$0.5 < R_P/R_{Earth} < 8$; $10 < a/R_{star} < 30$

Kepler data:

github.com/dfm/exostar19