Draft version October 20, 2017
Typeset using LaTeX **modern** style in AASTeX61

THE EFFECTS OF BINARITY ON PLANET OCCURRENCE RATES MEASURED BY TRANSIT SURVEYS

L. G. Bouma, K. Masuda, J. N. Winn

ABSTRACT

The aim of this work is to clarify what biases, if any, stellar binarity introduces to occurrence rates inferred from transit surveys. Ignoring binarity leads to diluted planetary radii, overestimated detection efficiencies, and an undercounted number of selected stars. These effects skew occurrence rate measurements in different directions, and we develop simplified analytic and numerical models to clarify which are the most important. For models in which all planets have the same radii, we find that ignoring binarity leads to an inferred occurrence rate at the true radius two to three times smaller than the actual rate (an upper limit to the possible error). More realistically, using Howard et al. (2012)'s radius distribution and a simple stellar model, we find that ignoring binarity leads to a $\sim 10 - 30\%$ underestimate of the number of planets per star over most planetary radii. In particular, the 30% underestimate applies to hot Jupiter occurrence rates, which suggests that the discrepancy between rates measured by *Kepler* and RV surveys might be attributable to the *Kepler* sample containing more binaries.

1. INTRODUCTION

TODO

2. METHOD

2.1. *Numerical Approach*

To progressively gain intuition, we study the following idealized transit surveys:

- Model #1: fixed stars, fixed planets, twin binaries;
- Model #2: fixed planets and primaries, varying secondaries;
- Model #3: fixed primaries, varying planets and secondaries.

Assuming a signal-to-noise limited survey, we would like to find the true occurrence rate density, and also that inferred by an observer ignoring binarity. Though analytic or semi-analytic solutions exist for each model, beyond Model #2 the equations become burdensome. To preserve simplicity, we develop a Monte Carlo approach, which works as follows.

First, the user specifies the model class (#1, 2, or 3) and various free parameters describing the stellar and planetary population. Most importantly, these parameters include the binary fraction, and the true planet occurrence rates around single stars, primaries, and secondaries.

We then generate the population of selected stars. Each selected star has a type (single, primary, secondary), a binary mass ratio (if it is not single), and the property of whether it is "searchable". The absolute number of stars is arbitrary. The relative number of binaries to singles is calculated according to analytic formulae. The binary mass ratios are sampled from the appropriate magnitude-limited distributions.

Whether a star is "searchable" depends entirely on its "completeness" fraction. By "completeness", we mean the ratio of the actual number of searchable stars to the assumed number of searchable stars (for a given planet size, period, etc.). Assuming homogeneously distributed stars, we will show (Sec. 3.1) that this is equivalent to the ratio of the searchable to selected volumes. For single stars, we assume these volumes are identical – exactly the case discussed by Pepper et al. (2003). For binaries, this volume ratio is a function of only the binary mass ratio.

To assign planets, each selected star receives a planet at the initially specified rate, according to its type. The radii of planets are assigned independently of any host system property, as sampled from $p_r(r) \sim r^\delta$ for Model #3 or a δ–function for Models #1 and #2. A planet is detected when a) it transits, and b) its host star is searchable.

The probability of transiting single stars in our model is assumed to be known, and so it is mostly corrected by the observer attempting to infer an occurrence rate. The only bias is for secondaries, which can be smaller than primaries in Models #2 and #3. This effect is included when computing the transit probability.

For detected planets, apparent radii are computed according to analytic formulae that account for both dilution and the misclassification of stellar radii. We assume that the observer thinks all transits are around primaries.

The rates are then computed in bins of true planet radius and apparent planet radius. In a given radius bin, the true rate is found by counting the number of planets that exist around selected stars of all types (singles, primaries, secondaries), and dividing by the total number of these stars. The apparent rates are found by counting the number of detected planets that were found in an apparent radius bin, dividing by the geometric transit probability for single stars, and dividing by the apparent total number of stars.

The simplest realization of this scheme is described analytically in Sec. 3.1, but we first clarify our terminology.

2.2. *Analytic Preliminaries*

Define the occurrence rate density, Γ, as the expected number of planets per star per natural logarithmic bin of planetary and stellar phase space:

$$\Gamma(\vec{x}) = \frac{d^n\Lambda}{\prod_{i=1}^{n} \mathrm{d}\ln x_i}. \tag{1}$$

$\vec{x}$ is an n-dimensional list of the continuous physical properties that might affect the occurrence rate density. For example, $\vec{x} = (r, P, R)$ where r is the planet radius, P is its orbital period, and R is the host star radius. The occurrence rate Λ is found by integrating the rate density over a specified volume of phase space.

The previous definition implicitly marginalizes the rate density over stellar multiplicity. For simplicity, this work only considers single and binary star systems. Then for a selected population of stars and planets the rate density can be written

$$\Gamma(\vec{x}) = \sum_{i=0}^{2} w_i \Gamma_i(\vec{x}) = \sum_i w_i \Lambda_i p_i(\vec{x}) \tag{2}$$

where $i = 0$ corresponds to single star systems, $i = 1$ primaries of binaries, and $i = 2$ secondaries of binaries. Λ_i is the occurrence rate integrated over all possible phase space for the i^{th} system type, $p_i(\vec{x})$ is the joint probability density function so that $\Gamma_i(\vec{x}) = \Lambda_i p_i(\vec{x})$, and the weights are given by

$$w_i = N_i/N_{\text{tot}}, \tag{3}$$

for $N_{\text{tot}} = \sum_i N_i$ the total number of selected stars, and N_0, N_1, N_2 the number of selected single stars, primaries, and secondaries respectively[1]. The relationship between the rate Λ over a desired volume of phase space Ω_{desired} and Λ_i is

$$\Lambda = \sum_i \left[\left(w_i \Lambda_i \int_{\Omega_{\text{desired}}} p_i(\vec{x})\,\mathrm{d}\Omega \right) \cdot \left(w_i \Lambda_i \int_{\Omega_{\text{total}}} p_i(\vec{x})\,\mathrm{d}\Omega \right)^{-1} \right], \tag{4}$$

[1] Eq. 2 follows by writing the i^{th} system type's rate density as some normalization multiplied by a probability density: $\Gamma_i(\vec{x}) = \mathcal{Z}_i p_i(\vec{x})$. For Eq. 1 to hold, we must have $\mathcal{Z}_i = \Lambda_i$.

where the inverse term is unity if $p_i(\vec{x})$ is appropriately normalized.

A transit survey will have a rate density of detected planets $\hat{\Gamma}$, which for each system type will be the product of the rate density and the detection efficiency $Q_i(\vec{x})$:

$$\hat{\Gamma}(\vec{x}) = \sum_i Q_i(\vec{x})\Gamma_i(\vec{x}) \equiv \sum_i \hat{\Gamma}_i(\vec{x}), \tag{5}$$

where again the index i is over each type of system (singles, primaries, and secondaries). The detection efficiency includes the geometric transit probability, as well as any incompleteness effects. Foreman-Mackey et al. (2014) discuss how this is calculated in practice.

3. ANALYTIC AND NUMERICAL RESULTS

3.1. *Model #1: fixed stars, fixed planets, twin binaries*

Consider a universe in which all planets are identical, and all stars are either single or twin stars with otherwise identical physical properties. The occurrence rate density at a planet radius r, stellar radius R, and semimajor axis a for this model is

$$\Gamma(r, R, a) = \sum_i w_i \Lambda_i \delta^3(r_p, R_\star, a_p), \tag{6}$$

for r_p and $R_\star$ some fixed planet and stellar radii, a_p a fixed semi-major axis, and δ the Dirac delta function, whose compact form will be used for brevity. The occurrence rate over any interval that includes r_p, $R_\star$, and a_p is

$$\Lambda = \sum_i w_i \Lambda_i = \frac{\sum_i N_i \Lambda_i}{N_{\rm tot}}. \tag{7}$$

The rate is zero over intervals that do not.

To express the rate density of detected planets, $\hat{\Gamma} = \sum Q_i \Gamma_i$, we need the detection efficiencies for each system type, which are products of the geometric and selection probabilities:

$$Q_i(\vec{x}) = Q_{g,i}(\vec{x}) Q_{c,i}(\vec{x}), \quad \text{where } \vec{x} = (r, R, a). \tag{8}$$

Similar to Pepper et al. (2003), but in a new context, we take Q_c as the ratio of the number of stars that were searchable to the number of stars that were selected. Assuming a homogeneous distribution of stars, this gives

$$Q_{c,i}(\vec{x}) = \left(\frac{d_{{\rm det},i}(\vec{x})}{d_{\rm sel}(\vec{x})}\right)^3, \tag{9}$$

for $d_{\rm sel}$ the maximum distance to which surveyed stars are selected, and $d_{{\rm det},i}$ the maximum distance to which planets can actually be detected about the $i^{\rm th}$ system type. Note that $d_{\rm sel} \geq d_{{\rm det},i}$. In a signal-to-noise limited transit survey in which the observer does not know which stars are binaries,

$$d_{\rm sel} \propto (r/R)^2 (L_{\rm sys} T_{\rm dur} A N_{\rm tra})^{1/2}, \tag{10}$$

for $L_{\rm sys} = L_1(1+\gamma_R)$ the system luminosity, $T_{\rm dur}$ the transit duration, A the detector area, and $N_{\rm tra}$ the number of observed transits. However,

$$d_{{\rm det},i} \propto \mathcal{D}_i (r/R)^2 (L_{\rm sys} T_{\rm dur} A N_{\rm tra})^{1/2}, \tag{11}$$

for the dilution $\mathcal{D}_i$ given by

$$\mathcal{D}_i = \begin{cases} 1 & \text{for } i=0, \text{ single} \\ L_1/L_{\rm sys} = (1+\gamma_R)^{-1}, & \text{for } i=1, \text{ primary} \\ \gamma_R L_1/L_{\rm sys} = (1+\gamma_R^{-1})^{-1}, & \text{for } i=2, \text{ secondary}, \end{cases} \tag{12}$$

where the light ratio γ_R of a given binary is defined as the ratio of the luminosity of the secondary to the primary.

The maximum detectable distance to single stars is assumed to be known, and so $d_{\rm sel,0} = d_{\rm det,0}$. For binary systems there is a necessary incompleteness, and combining Eqs. 8 through 12 yields

$$Q_0 = Q_{g,0} Q_{c,0} = Q_{g,0} \tag{13}$$
$$Q_1 = Q_{g,1} Q_{c,1} = Q_{g,0}(1+q^{\alpha})^{-3} \tag{14}$$
$$Q_2 = Q_{g,2} Q_{c,2} = Q_{g,0} q^{2/3} (1+q^{-\alpha})^{-3} q^{-5}, \tag{15}$$

for $Q_{g,0} = R/a$, the transit probability in single star systems. The factors of $q^{2/3}$ and q^{-5} in Eq. 15 come from the assumed stellar mass-luminosity-radius relation: $R \propto M \propto L^{1/\alpha}$. For $q=1$ both terms evaluate to unity, but they will later become relevant.

Summarizing, we have written the rate density for each system type (Eq. 6) and the detection efficiency for each system type (Eq. 13-15), and so have fully specified the rate density of detected planets, in addition to the true rate density.

What does an observer ignoring binarity infer?—An observer who ignores binarity assumes a detection efficiency $\tilde{Q} = Q_0$, measures a detected planet rate density $\tilde{\Gamma}$, and infers an apparent rate density Γ_a. Analogous to Eq. 5,

$$\tilde{\Gamma} = \Gamma_a \tilde{Q}. \tag{16}$$

Accounting for dilution, one can show

$$\Gamma_a = w_a \Lambda_0 \delta^3(r_p, R_\star, a_p) + w_b(\Lambda_1 Q_{c,1} + \Lambda_2 Q_{c,2}) \delta^3(r_p/\sqrt{2}, R_\star, a_p), \tag{17}$$

for $w_a = N_0/(N_0+N_1)$, and $w_b = N_1/(N_0+N_1)$. This observer miscounts the number of total searched stars, does not correct for incompleteness, and misclassifies the planetary radii because of dilution.

Correction to inferred rate density and inferred rate—Define a rate density correction factor, X_Γ, as the ratio of the apparent to true rate densities:

$$X_\Gamma \equiv \frac{\Gamma_a}{\Gamma}. \tag{18}$$

This factor can be a function of whatever parameters Γ_a and Γ depend on; in this study, the planet radius is most relevant. For the twin-binaries model,

$$X_\Gamma(r) = \frac{w_a \Lambda_0 \delta^3(r_p) + w_b(\Lambda_1 Q_{c,1} + \Lambda_2 Q_{c,2})\delta^3(r_p/\sqrt{2})}{(w_0\Lambda_0 + w_1\Lambda_1 + w_2\Lambda_2)\delta^3(r_p)} \tag{19}$$

where $\delta^3(r_p)$ is shorthand for $\delta^3(r - r_p, R - R_\star, a - a_p)$.

If we take the rates Λ_i to be equal, applying the definitions of the weights gives a rate density correction factor at $r = r_p$ of $X_\Gamma(r_p) = (1+\mu)^{-1}$, where

$$\mu \equiv \frac{N_1}{N_0} = \frac{n_b}{n_s}\left(\frac{d_{\mathrm{sel,b}}}{d_{\mathrm{sel,s}}}\right)^3 = \frac{\mathrm{BF}}{1-\mathrm{BF}}(1+\gamma_R)^{3/2}, \tag{20}$$

for n_b and n_s the number density of binaries and singles in a volume limited sample. Using Raghavan et al. (2010)'s $0.7 - 1.3 M_\odot$ multiplicity fraction as our binary fraction[2], we set $\mathrm{BF} = 0.44$. The resulting correction to the rate density is $X_\Gamma(r_p) \approx 0.31$. The correction at $r_p/\sqrt{2}$ is infinite. The numerical realization of this model agrees with these analytic values, and its output is shown in Fig 1. If instead we assume that $\Lambda_0 = \Lambda_1$, but that $\Lambda_2 = 0$, we find $X_\Gamma(r_p) = (1+2\mu)/(1+\mu)^2$. Taking the same binary fraction, this evaluates to $X_\Gamma(r_p) \approx 0.53$. Since the correction to the rate is equal to that of rate density, at $r = r_p$, the occurrence rate is underestimated by a factor of roughly 2 to 3.

3.2. *Model #2: fixed planets and primaries, varying secondaries*

The main use of our binary-twin model is to help develop intuition. We now let the light ratio $\gamma_R = L_2/L_1$ vary across the binary population. It does so because the underlying mass ratio $q = M_2/M_1$ varies. We keep the primary mass fixed as M_1, which is also the mass of all single stars.

We parametrize the distribution of binary mass ratios in a volume-limited sample as a power law: $p(q) \propto q^\beta$. For binaries with solar-type primaries[3], β is probably between 0 and 0.3. Since we assume stars are a one-parameter family, $R \propto M \propto L^{1/\alpha}$, a drawn value of q determines everything about a secondary.

The rate density in this model, $\Gamma(\vec{x})$, is the sum over system types of $w_i \Lambda_i p_i(\vec{x})$:

$$\Gamma(\vec{x}) = \delta^4(r_p, R_\star, a_p, P_p)(w_0\Lambda_0 + w_1\Lambda_1) + w_2\Lambda_2\delta^3(r_p, P_p, a_p)p_2(q), \tag{21}$$

[2] The binary fraction is the fraction of systems in a volume-limited sample that are binary. It is equivalent to the multiplicity fraction if there are no triple, quadruple, or higher order multiples.

[3] Duchene and Kraus (2013), fitting all the multiple systems of Raghavan et al. (2010)'s Fig 16, find $\beta = 0.28 \pm 0.05$ for $0.7 < M_\star/M_\odot < 1.3$. Examining only the binary systems of Rhagavan et al 2010, Fig 16, the distribution seems roughly uniform, $\beta \approx 0$, except for a claimed excess of twin binaries with $q \approx 1$, and an obvious lack of $q < 0.1$ stellar companions.

where the semimajor axis of the planet must be such that its period is P_p, and $p_2(q)$ is expressed in terms of the mass ratio instead of the secondary star's radius for convenience (q and R_2 are interchangeable). The probability that a secondary hosts a planet, as a function of the mass ratio, is

$$p_2(q) = p(\text{has planet}\,|\,\text{secondary with } q) \times p(\text{secondary with } q) \tag{22}$$

$$p_2(q) \propto q^{\gamma+\beta}(1+q^{\alpha})^{3/2}. \tag{23}$$

We take first term, $p(\text{has planet}\,|\,\text{secondary with } q)$, as a power law of q with exponent γ. For the second term, since the selected sample at a given (r, P, a) is magnitude-limited, $p(\text{secondary with } q)$ is the product of the volume limited probability and a Malmquist-like bias $(1+q^{\alpha})^{3/2}$. In passing, we note that various authors (ourselves included) have incorrectly used volume-limited binary distributions in Monte Carlo simulations of transit surveys (Sullivan et al 2015, Hartman and Bakos XX, Guenther et al 2017, Bouma et al 2017).

The occurrence rate corresponding to Eq. 21's rate density for a desired volume of phase space Ω_{desired} is given by Eq. 4. Specifying the desired mass ratios of interest as $q_{\min} < q < q_{\max}$, this simplifies to

$$\Lambda = \frac{N_0\Lambda_0 + N_1\Lambda_1 + N_2\Lambda_2 f_2}{N_{\text{tot}}}, \tag{24}$$

for

$$f_2 \equiv \left(\int_{q_{\min}}^{q_{\max}} p_2(q)\,\mathrm{d}q\right)\cdot\left(\int_0^1 p_2(q)\,\mathrm{d}q\right)^{-1}. \tag{25}$$

The detected rate density, $\hat{\Gamma} = \sum_i Q_i\Gamma_i$, will be specified by the detection efficiencies for each type of system. These are identical to Eqs. 13-15. The detection efficiency for secondaries (Eq. 15) includes the transit probability from the smaller stellar radius, and combines dilution, the transit duration, and stellar radius for the completeness probability.

What does an observer ignoring binarity infer?—As a reminder, the apparent rate density is found by correcting the detected apparent rate density for the transit probability: $\Gamma_a = \tilde{\Gamma} Q_{g,0}^{-1}$. The observer's errors are as follows:

1. The true planetary radii r are interpreted as apparent radii r_a. The apparent radii depend on the system type:

$$r_a = \begin{cases} r_p(1+q^{\alpha})^{-1/2} & \text{for } i=1, \text{ primary} \\ r_p(1+q^{-\alpha})^{-1/2}q^{-1}, & \text{for } i=2, \text{ secondary.} \end{cases} \tag{26}$$

   The factor of q^{-1} for the secondary case accounts for the observer assuming that the host star is the primary.

2. The true stellar radii R are all thought to be $R_\star$. In “reality”, this only holds for single stars and primaries.

3. The selected and searchable stars are miscounted.

To write the apparent rate density as a function of the apparent radius r_a, we marginalize out the planet period, semimajor axis, and stellar radius (or equivalently the mass ratio, for binaries):

$$\Gamma_a(r_a) = w_a \Lambda_0 \delta(r_p) + w_b \Lambda_1 I_1(r_a) + w_b \Lambda_2 I_2(r_a), \tag{27}$$

where the detection efficiencies are given in Eqs. 13-15, and as in the first model, $w_a = N_0/(N_0+N_1)$, $w_b = N_1/(N_0+N_1)$. The ratio of primaries to singles, μ, is now given by a variant of Eq. 20:

$$\mu \equiv \frac{N_1}{N_0} = \frac{\mathrm{BF}}{1-\mathrm{BF}} \left(2^{3/2} - \int_1^{\sqrt{2}} u^2 (u^2-1)^{1/\alpha} \, \mathrm{d}u \right), \tag{28}$$

where the latter dimensionless integral is easily found numerically. The $I_1(r_a)$ and $I_2(r_a)$ terms marginalize over the joint distribution of apparent radius and mass ratio:

$$\begin{aligned} I_i(r_a) &= \int_0^1 p(\text{has detected planet}, r_a, q | \text{star is type } i) \, \mathrm{d}q, \quad \text{for } i \in \{1,2\}, \qquad (29) \\ &= \int_0^1 p(\text{has detected planet} | r_a, q, \text{star is type } i) \\ &\qquad \times p(r_a | q, \text{star is type } i) \, p(q | \text{star is type } i) \, \mathrm{d}q. \qquad (30) \end{aligned}$$

The first term is the detection efficiency; the second is a δ-function of the apparent radius; the latter is the mass ratio distribution given by Eq. 23. The analytic solution for $i=1$ is

$$I_1(r_a) = \frac{1}{\mathcal{N}_1} \left(\frac{r_p}{r_a}\right)^{-3} \left(\left(\frac{r_p}{r_a}\right)^2 - 1 \right)^{\frac{\gamma+\beta}{\alpha}}, \quad \text{for } r_p/\sqrt{2} < r_a < r_p, \tag{31}$$

where $\mathcal{N}_1$ is the normalization term of the binary mass ratio distribution (Eq. 23): $\mathcal{N}_1 = \int_0^1 q^{\gamma+\beta}(1+q^\alpha)^{3/2} \mathrm{d}q$.

For $i=2$, there is no analytic solution, because evaluating the integral requires imposing the constraint that $r_a = r_p(1+q^{-\alpha})^{-1/2} q^{-1}$. This equation can be re-written

$$\left(\frac{r_p}{r_a}\right)^2 = q^2 + q^{-\alpha+2}, \tag{32}$$

which has no analytic solution except for special values of α, the mass-luminosity exponent. For $\alpha = 3.5$, our nominal case, semianalytic solutions can be found.

However, since our main interest is in understanding the qualitative behavior of the solutions, we focus instead on a few limiting cases.

Correction to inferred rate density—Recall that the rate density correction factor, X_Γ, is the ratio of the apparent to true rate densities. We consider a "nominal model" in which the stellar population is similar to Sun-like stars in the local neighborhood: BF $= 0.44$, $\alpha = 3.5$, $\beta = 0$. Our default assumption is also that the occurrence of planets is independent of stellar mass ($\gamma = 0$), so secondaries have the same occurrence rate as primaries and single stars. Under these assumptions, the planetary rate density is

$$\Gamma(r) \approx \delta(r_p)\left(\Lambda_0 + \Lambda_1 + \Lambda_2\right)/3, \tag{33}$$

where the coefficients of 1/3 are accurate to within one percent of the true coefficients. Ignoring binarity, the observer finds an apparent rate density

$$\Gamma_a(r) = c_0\Lambda_0\delta(r_p) + c_1\Lambda_1 I_1(r_a) + c_2\Lambda_2 I_2(r_a), \tag{34}$$

for $c_0 \approx 0.49$, $c_1 \approx 0.32$, $c_2 \approx 0.03$. Evaluating the correction term at $r = r_p$, since $\lim_{q\to 0} p_i(r_a) = 0$ for $i \in \{1, 2\}$,

$$X_\Gamma(r = r_p) \approx \frac{3c_0\Lambda_0}{\Lambda_0 + \Lambda_1 + \Lambda_2}. \tag{35}$$

If all the rates are equal, $X_\Gamma(r = r_p) \approx 0.49$. If there are no planets around the secondaries, $X_\Gamma(r = r_p) \approx 0.74$.

We also run this model in our numerical simulation. The results are shown in Figs 2 and 3. They produce the same correction factors as predicted analytically. They also show a peak in the inferred rate at $r_p/\sqrt{2}$ from secondaries. This effect requires the observer to incorrectly estimate the host star's radius; if they somehow knew the host star's radius, but did not correct for binary dilution, the peak would instead be at an apparent radius of 0.

To summarize, dilution produces a spectrum of apparent planetary radii. In this model, this produces overestimated rates everywhere except where there are actually planets, where the rate is underestimated by a factor of 2.

3.3. *Model #3: Fixed primaries, varying planets and secondaries*

See Fig. 4.

4. DISCUSSION

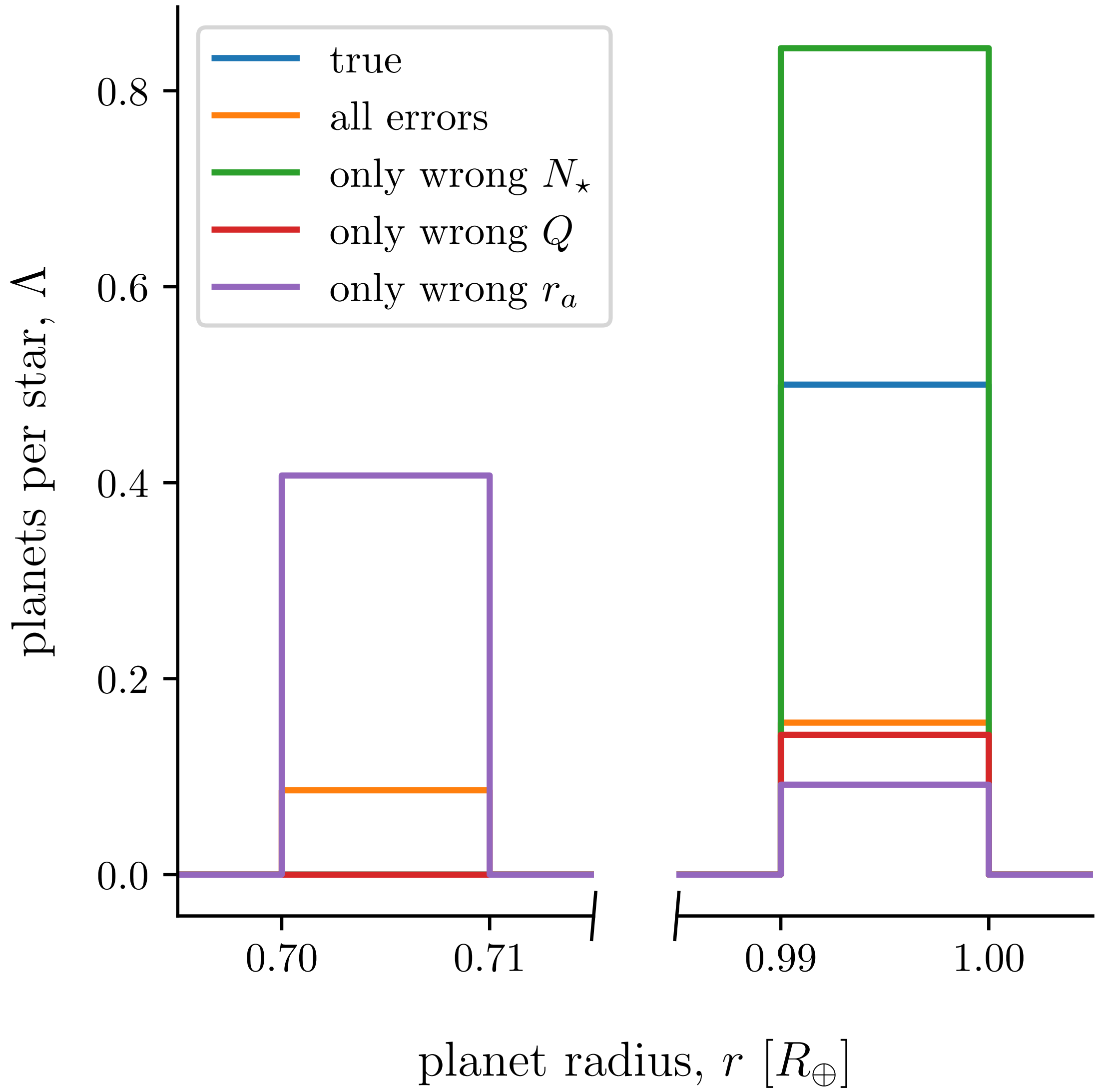

Figure 1. Inferred planet occurrence rates as a function of planet radius in Model #1. This model has fixed stars, fixed planets, twin binaries; Note the break in the x axis. If the true planet radius is r_p, all planets detected in binaries will have apparent radii $r_a = r_p/\sqrt{2}$. Only 1 in 8 selected binaries is actually searchable (see Sec. 3.1).

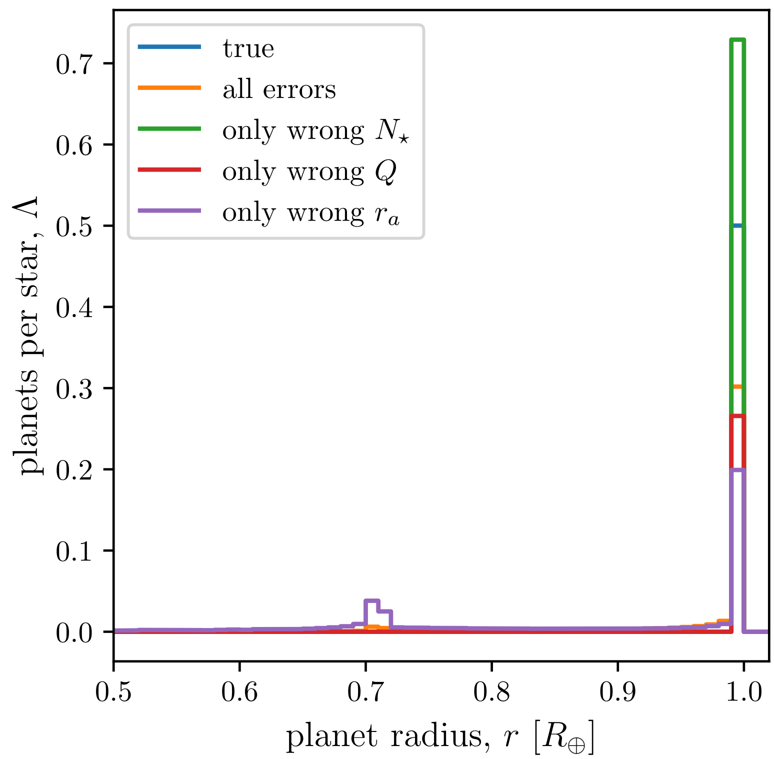

Figure 2. Inferred planet occurrence rates as a function of planet radius in Model #2. This model has fixed planets and primaries, but varying secondaries.

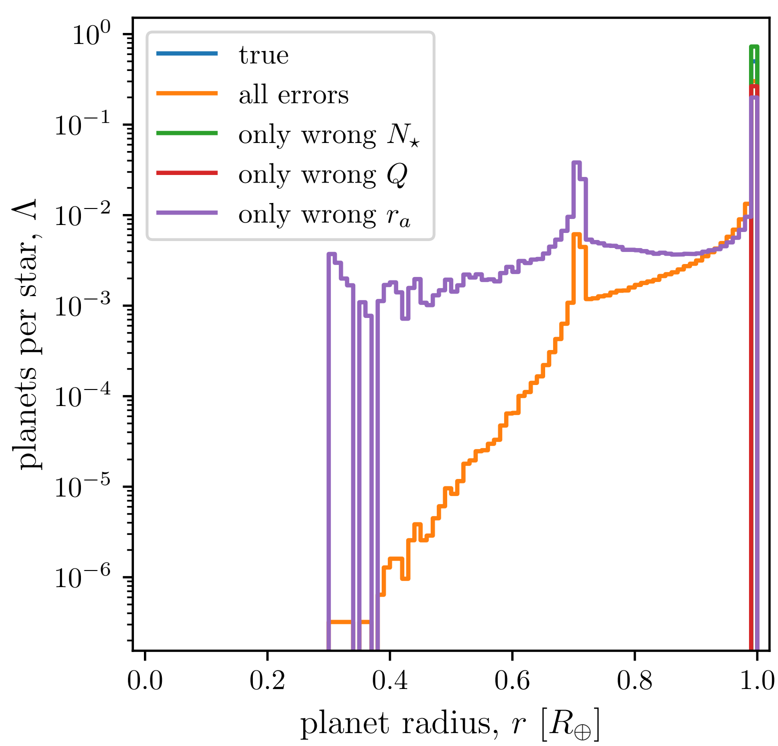

Figure 3. Same as Fig. 2, but with logarithmic y-axis, and different x scale.

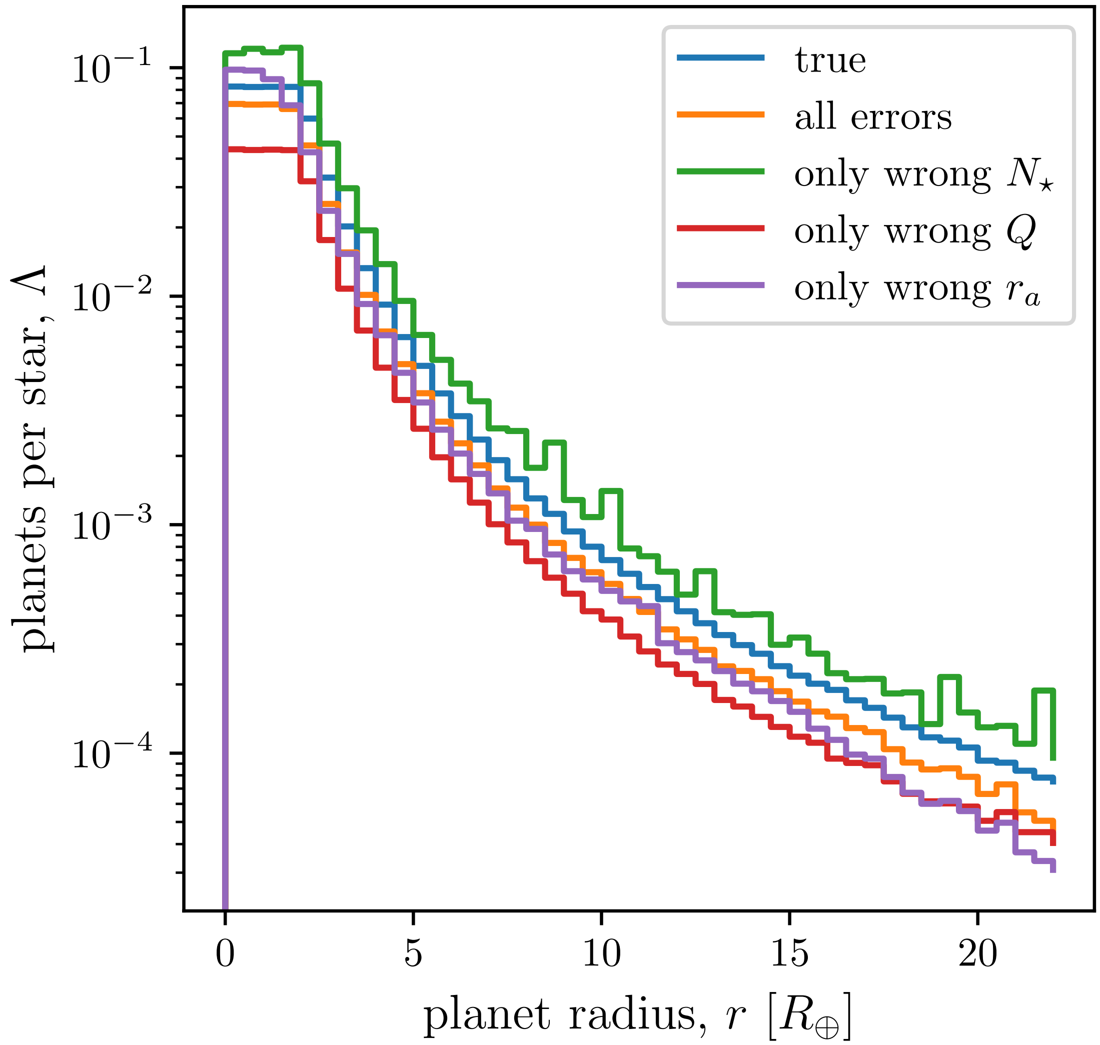

Figure 4. Inferred planet occurrence rates as a function of planet radius in Model #3. This model has fixed primaries and single stars, but varying secondaries. The true planet radius distribution is a power law with exponent -2.92 above $2R_\oplus$, below which it is uniform (e.g., Howard et al., 2012).