

ANALYTIC TRANSIT SURVEYS

LGB, KM, JNW

Memo for internal use

ABSTRACT

We describe a few different toy transit surveys.

First, we derive the signal to noise distribution of planets detected in a survey where all the stars and planets are the same (except some stars are twin binary systems). We give expressions for the number of detected planets, and for possible errors one can make when deriving occurrence rates.

Then, we allow the properties of the secondary in binary systems to vary. A distribution of light ratios introduces additional biases, which we discuss. We reassess the errors related to occurrence rates.

Subject headings: planets and satellites: detection

1. MODEL #1: FIXED STARS, FIXED PLANETS, FIXED LIGHT-RATIO BINARIES

Ah, transiting planets! We learn so much by studying them. But how much can we actually learn, and how much is messed up by binarity?

Imagine the following transit survey:

1. You are going to observe the entire sky for a duration T_{obs} , with a detector of area A , and known bandpass. Your detector is photon-noise limited.
2. You are interested only in detecting planets of radius R_p , and orbital period P . For instance, $R_p = R_{\oplus}$, $P = 1$ year.
3. You are only interested in detecting them around stars of radius R_1 , and luminosity L_1 . For instance, G2V dwarfs.
4. You can only detect your ‘nominal planet’ when you observe signals with $S/N > (S/N)_{\text{min}}$. For a photon-noise limited survey, this minimum signal to noise ratio is equivalent to a minimum flux, F_{min} .
5. However, you opt to observe all the “points” on-sky with apparent magnitude $m < m_{\text{lim}}$, or equivalently with energy flux in your bandpass greater than some limit $F > F_{\text{lim}}$. (You will not be able to detect planets around the faintest of these, and will need to perform a completeness correction to later derive occurrence rates.) Generally $F_{\text{lim}} < F_{\text{min}}$, *e.g.*, whenever the original magnitude cut defines a “source catalog” from which “searchable stars” are selected, as in the KIC [Batalha et al, 2010, Brown et al 2011] and the TIC [Stassun et al 2017].

You get funding, and perform the S/N -limited survey. All planets with $S/N > (S/N)_{\text{min}}$ are detected. The rest are not. You now wish to derive an occurrence rate for planets of radius R_p and orbital period P . Assume your universe is a universe in which:

- The true population of “points” (stellar systems, all unresolved) comprises both single and double star systems. Single star systems have luminosity in the observed bandpass L_1 , radii R_1 , and effective

temperature $T_{\text{eff},1}$. Double star systems have luminosity in the observed bandpass $L_d = (1 + \gamma_R)L_1$, for $\gamma_R = L_2/L_1$ the ratio of the luminosity of the secondary to the primary. In this section, γ_R is a constant across the population of star systems. The ratio of the two number densities in a volume-limited sample is the binary fraction¹.

- The true population of planets around these stars is as follows:

A fraction $\Gamma_{t,s}$ of stars in single star systems have a planet of radius R_p , with orbital period P .

A fraction $\Gamma_{t,d}$ of each star in a double star system has a planet of radius R_p , with orbital period P . For instance, if $\Gamma_{t,s} = \Gamma_{t,d} = 0.1$, on average each double system contributes 0.2 planets, and each single system 0.1 planets. Any astrophysical difference in planet formation between singles and binaries is captured by these two terms. Note that we are taking a limit by assuming the primary and secondary of a binary system host planets at equal rates. If the secondary is non-identical to the primary, it’s clearly wrong.

For simplicity, one might wish to imagine the observer correctly pre-selects all of the “searchable stars”, in the style of [Pepper et al, 2003]. In other words, they would begin by knowing $(S/N)_{\text{min}}$, and then only observe stars for which they would have adequate flux to achieve it – setting $F_{\text{lim}} = F_{\text{min}}$. This simplification would naively allow the observer to ignore completeness effects when deriving occurrence rates, since their completeness is 100%.

We do not begin this way because in a stellar population of both singles and binaries, the above approach ignores necessary incompleteness for binary systems. If we set the limiting flux to give 100% completeness for single stars, we will unwittingly select binaries near that flux limit, for which dilution will push transit signals below the detection threshold. This is because a flux limit

¹ The binary fraction is equivalent to the multiplicity fraction if there are no triple, quadruple, ... systems.

maps onto three different maximum detection distances for a population of single stars and twin binaries:

1. $d_{\max,s}$: the maximum distance out to which single stars are selected. This can be chosen to coincide with the maximum distance out to which planets are detectable around single stars.
2. $d_{\max,d}^*$: the maximum distance out to which double stars are selected. This is different from $d_{\max,d}^p$, the maximum distance out to which planets are detectable around double stars. For a population with fixed γ_R , since $S/N \propto \mathcal{D}L^{-1/2}d$,

$$d_{\max,d}^p = (1 + \gamma_R)^{-1/2} d_{\max,s} = (1 + \gamma_R)^{-1} d_{\max,d}^* \quad (1)$$

For $\gamma_R = 1$, this means only 1 in 8 selected binary stars can yield detectable planets.

We develop tools to address this and related completeness effects from the outset.

Consider the following questions:

1. How many single and double star systems, respectively, are in the sample? Correspondingly, how many stars are in the sample?
2. How many planets are in the sample? (Orbiting single stars, and orbiting double stars respectively).
3. What is the true occurrence rate?
4. How many planets are detected?
5. What occurrence rate does astronomer A, who has never heard of binary star systems, or completeness, derive for planets of radius R_p and period P ?
6. What occurrence rate does astronomer B, who accounts for the “2 for 1” effect of binarity (*i.e.* that the sample actually has more stars than astronomer A thought) derive? (Astronomer B neglects completeness).
7. What about astronomer C, who accounts for “2 for 1” *and* misclassification due to diluted radii? In other words, astronomer C did a combination of high resolution imaging and RV followup on every candidate, and correctly classifies the planetary radii in every case.
8. What about astronomer D, who additionally notes the importance of completeness?

1.1. How many stars are in the sample?

Let N_s be the number of single star systems, and N_d the number of double star systems. Then the total number of stars in the sample is

$$N_{\text{stars}} = N_s + 2N_d. \quad (2)$$

In a magnitude-limited sample in which stars are uniformly distributed in volume, the number of stars will be the number density times the volume. If the volume is taken to be a sphere over which the number density is uniform,

$$N_i = n_i \frac{4\pi}{3} d_{\max,i}^3, \quad (3)$$

for $i \in \{\text{single}, \text{double}\} \equiv \{s, d\}$, and

$$\frac{n_d}{n_s} = \text{binary fraction} \equiv \text{BF} \quad (4)$$

by definition. The absolute normalization of the number density is a measured quantity, as is the binary fraction. For G2V dwarfs, the latter is ≈ 0.45 [Duchene & Kraus, 2013]. The former is given by [Bovy 2017].

$d_{\max,i}$ in Eq. 3 is the maximum distance corresponding to the given magnitude limit:

$$d_{\max,i} = \left(\frac{L_i}{4\pi F_{\min}} \right)^{1/2}, \quad (5)$$

where the limiting flux in the bandpass F_{\min} can also be stated in terms of the limiting magnitude m_{\min} ,

$$m_{\min} = m_0 - \frac{5}{2} \log_{10} \left(\frac{F_{\min}}{F_0} \right), \quad (6)$$

for m_0 a zero-point magnitude and F_0 its corresponding flux (as always, everything is implicitly written in a defined bandpass).

In Eq. 5, again $i \in \{\text{single}, \text{double}\}$, and as a consequence the maximum distance to which binary stars will be selected is greater than that of single stars, simply as a consequence of imposing a magnitude cut. The ratio of double to single systems is

$$\frac{N_d}{N_s} = \frac{n_d}{n_s} \left(\frac{d_{\max,d}}{d_{\max,s}} \right)^3 \quad (7)$$

$$= \text{BF} \times (1 + \gamma_R)^{3/2}. \quad (8)$$

In the nominal case of twin binaries ($\gamma_R = 1$), with a binary fraction $\text{BF} = 0.5$, there are $\sqrt{2}$ more binary systems than single systems in the sample. Correspondingly, there are $2\sqrt{2}$ more stars in binary systems than stars in single systems.

As a comment on Eq. 3, if we wished to write a stellar number density profile that accounted for the vertical structure of the Milky Way, we might choose a profile either $\propto \exp(-z/H)$, or $\propto \text{sech}^2(z/H)$ for z the distance from the galactic midplane and H a scale-height. Both density profiles would lead closed form analytic solutions.

1.2. How many planets are in the sample?

The number of planets in the sample is

$$\begin{aligned} N_{\text{planets}} &= N_{\text{planets in single star systems}} + \\ &\quad N_{\text{planets in double star systems}} \\ &= \Gamma_{t,s} N_s + 2\Gamma_{t,d} N_d. \end{aligned} \quad (9)$$

The factor of 2 accounts for the fact that there are twice as many stars in double star systems.

1.3. What is the true occurrence rate?

The “true occurrence rate” is the average number of planets per star. Thus

$$\Gamma_t = \frac{N_{\text{planets}}}{N_{\text{stars}}} \quad (11)$$

$$\Gamma_t = \frac{\Gamma_{t,s} N_s + 2\Gamma_{t,d} N_d}{N_s + 2N_d}. \quad (12)$$

1.4. How many planets are detected?

The total number of planet detections is the sum of the number of planets detected in single star systems $N_{\text{det},s}$ and the number of planets detected in double star systems $N_{\text{det},d}$. These can be expressed individually. The former is

$$N_{\text{det},s} = N_s \Gamma_{t,s} f_{s,\text{geom}} f_{s,S/N > (S/N)_{\min}}, \quad (13)$$

where the product $N_s \Gamma_{t,s}$ is the number of planets in the single star systems of the sample, $f_{s,\text{geom}} \equiv f_{s,g}$ is the geometric transit probability, and $f_{s,(S/N > (S/N)_{\min})} \equiv f_{s,c}$ is the fraction of these transiting planets that are observed with signal to noise greater than the minimum detection threshold (the completeness). Analogously,

$$N_{\text{det},d} = 2N_d \Gamma_{t,d} f_{d,g} f_{d,c}, \quad (14)$$

where now $2N_d \Gamma_{t,d}$ is the number of planets in the double star systems of the sample, the geometric transit probability is the same (if we have twin binaries – not if we consider a distribution of secondaries) and the completeness term must account for any differences in the signal to noise distribution that come from stellar binarity.

1.4.1. Analytic completeness

Since the geometric transit probability is known, the only terms we have yet to compute are the completeness terms, $f_{i,c}$ for $i \in \{\text{single}, \text{double}\}$. We proceed as follows.

The signal S for a box-car train transiting planet is

$$S = \delta \mathcal{D} \quad (15)$$

$$= \left(\frac{R_p}{R_\star} \right)^2 \mathcal{D}, \quad (16)$$

for R_p the planet's radius, R_\star that of its host star, and \mathcal{D} the dilution parameter defined as

$$\mathcal{D} = \begin{cases} L_1/L_d, & \text{if binary and target primary} \\ \gamma_R L_1/L_d, & \text{if binary and target secondary} \\ 1, & \text{if single,} \end{cases}$$

$$= \begin{cases} (1 + \gamma_R)^{-1}, & \text{if binary and target primary} \\ (1 + \gamma_R^{-1})^{-1}, & \text{if binary and target secondary} \\ 1, & \text{if single,} \end{cases} \quad (17)$$

where L_1, L_d , and γ_R were defined in the opening monograph.

Assuming the only source of noise is Poissonian counting noise, the noise N can be written

$$N = \frac{1}{\sqrt{N_\gamma}}, \quad (18)$$

for N_γ the number of photons received by the detector. This noise model is a useful simplification – see [Howell 2006, pg 75] for the full CCD equation. The number of received photons can be written

$$N_\gamma = F_\gamma^N A N_{\text{tra}} T_{\text{dur}}, \quad (19)$$

for F_γ^N the photon number flux from the system [$\text{ph cm}^{-2} \text{s}^{-1}$], A the detector area, T_{dur} the transit duration, and N_{tra} the number of transits observed, which is multiplied in assuming the transits are “phase-folded”.

Thus the signal to noise ratio can be written

$$S/N = \delta \mathcal{D} \sqrt{F_\gamma^N A N_{\text{tra}} T_{\text{dur}}}. \quad (20)$$

This means we can write the minimum number flux of photons required for a detection at threshold as

$$F_{\text{lim}}^N = \left[\left(\frac{S}{N} \right)_{\min} \frac{1}{\delta \mathcal{D}} \right]^2 \frac{1}{A N_{\text{tra}} T_{\text{dur}}}. \quad (21)$$

To convert this to F_{lim} , multiply by the average photon energy in the bandpass.

In passing, given the parameters that define a survey and planet type, Eq. 20 would need to be re-expressed with N_{tra} roughly the ratio of the observing baseline to the planet period, and T_{dur} a function of R_\star, P, a , and impact parameter b , and then perhaps averaged over b . We leave them as-is for subsequent development.

The interesting term in Eq. 20 that changes between stars of the same binarity class in our idealized sample is the square root of F_γ^N . This is the term that leads to a distribution of signal to noises for different stars. The completenesses $f_{i,c}$ can be directly expressed in terms of those probability density functions:

$$f_{i,c} = \int_{(S/N)_{\min}}^{\infty} d \left(\frac{S}{N} \right)_i \text{prob} \left(\frac{S}{N} \right)_i. \quad (22)$$

We keep the subscript i because the signal to noise distributions are different for the cases of single star systems ($i = s$) and double star systems ($i = d$). Notably:

- The dilution differs (Eq. 17).
- The photon number flux from the system differs.

To simplify notation, we let $x_i \equiv (S/N)_i$, and rewrite Eq. 22 as

$$f_{i,c} = \int_{x_{\min}}^{\infty} dx_i \text{prob}(x_i). \quad (23)$$

1.4.2. Deriving $\text{prob}(x_i)$

We want expressions for the probability density function of the observed signal to noise ratio, $\text{prob}(x_i)$, for both the single and binary system case.

First, note that a star placed uniformly in the volume of the search space will have a probability density function for its distance r from the origin of

$$\text{prob}(r) = \frac{3r^2}{d_{\text{max}}^3}, \quad (24)$$

where the appropriate maximum distances should be substituted per Eq. 5. Noting the transformation rule for probability density functions, we can evaluate the probability of a star having a observed number flux $F_{i,\gamma}^N$ in the bandpass,

$$\text{prob}(F_{i,\gamma}^N) = \text{prob}(r(F_{i,\gamma}^N)) \left| \frac{dr}{dF_{i,\gamma}^N} \right| \quad (25)$$

$$= \frac{3}{2d_{\text{max}}^3} c_i^{3/2} (F_{i,\gamma}^N)^{-5/2}, \quad (26)$$

where in the latter equality we have written a “number luminosity” c_i (units of inverse time) defined for $i \in \{\text{single}, \text{double}\}$ as

$$c_i = \begin{cases} R_1^2 F_{s1,\gamma}^N, & \text{if single} \\ R_1^2 F_{s1,\gamma}^N + R_2^2 F_{s2,\gamma}^N & \text{if double.} \end{cases} \quad (27)$$

In Eq. 27, $F_{s1,\gamma}^N$ and $F_{s2,\gamma}^N$ are the photon number fluxes at the surfaces of the stars. To derive Eq. 26, we simply scaled these by the distance:

$$F_{i,\gamma}^N = \frac{c_i}{r^2}. \quad (28)$$

The surface photon number fluxes $F_{si,\gamma}^N$ in Eq. 27 are usually evaluated numerically, by convolving the wavelength-specific photon flux density of a star with the dimensionless spectral response function of the instrument. In other words,

$$F_{s,\gamma}^N = \int F_\lambda T_\lambda d\lambda. \quad (29)$$

The wavelength-specific photon flux density F_λ [$\text{ph cm}^{-2} \text{s}^{-1} \text{\AA}^{-1}$] could be from Pickles’ library, or could be a blackbody function. The transmission function is, up to factor of order unity, a step function between two wavelengths λ_{\min} and λ_{\max} . If we assume a blackbody source, Eq. 29 becomes

$$F_{s,\gamma}^N = 8\pi c \left(\frac{kT}{hc} \right)^3 \int_{hc/(\lambda_{\max} kT)}^{hc/(\lambda_{\min} kT)} \frac{u^2}{e^u - 1} du, \quad (30)$$

which can be evaluated numerically².

The importance of the functional form of $F_{s,\gamma}^N$ is that it is to first order only a function of the blackbody temperature and the bandpass wavelength bounds. Thus in the most general case $c_i(R_1, R_2, T_1, T_2, \lambda_{\min}, \lambda_{\max})$ and nothing else. The only random variable involved in the flux being received at the detector is r , so we can indeed write the flux received at the detector as in Eq. 28.

We can finally write out the probability density functions for the signal to noise ratios in the single and double-star cases by using the transformation rule for pdfs, and applying Eq. 20. For single stars,

$$\text{prob}(x_s) = \frac{3}{d_{\max,s}^3} c_s^{3/2} \delta^3 (AT_{\text{dur}} N_{\text{tra}})^{3/2} x_s^{-4}. \quad (31)$$

Analogously for double stars,

$$\text{prob}(x_d) = \frac{3}{d_{\max,d}^3} c_d^{3/2} (\mathcal{D}\delta)^3 (AT_{\text{dur}} N_{\text{tra}})^{3/2} x_d^{-4}. \quad (32)$$

1.4.3. Number of detected planets

Performing the integrals of Eq. 23, we get:

² It may help in the numerics to note that infinite series representations of this type of dimensionless integral exist and converge quickly. For instance, one can show that

$$\int_0^a \frac{u^3}{e^u - 1} du = \sum_{n=1}^{\infty} \frac{6}{n^4} - \frac{e^{-an}}{n^4} (6 + 6an + 3(an)^2 + (an)^3).$$

A similar expression exists for the similar integral in the text. [Michels 1968] explains an analogous derivation.

$$N_{\text{det},s} = N_s \Gamma_{t,s} f_{s,g} f_{s,c} \quad (33)$$

$$= N_s \Gamma_{t,s} \frac{R_\star}{a} \frac{1}{d_{\max,s}^3} c_s^{3/2} \delta^3 (AT_{\text{dur}} N_{\text{tra}})^{3/2} x_{\min}^{-3}, \quad (34)$$

and

$$N_{\text{det},d} = 2N_d \Gamma_{t,d} f_{d,g} f_{d,c} \quad (35)$$

$$= 2N_d \Gamma_{t,d} \frac{R_\star}{a} \frac{1}{d_{\max,d}^3} c_d^{3/2} (\mathcal{D}\delta)^3 (AT_{\text{dur}} N_{\text{tra}})^{3/2} x_{\min}^{-3}. \quad (36)$$

Formally, the $f_{i,c}$ terms should be written as $\min(1, \dots)$, where (\dots) is the given expression. This ensures that the fraction of planets above the signal to noise threshold is less than 1. Note that we assumed twin binaries in the geometric transit probabilities. Interestingly, the ratio of the two completeness fractions is

$$\frac{f_{d,c}}{f_{s,c}} = \left(\frac{d_{\max,s}}{d_{\max,d}} \right)^3 \left(\frac{c_d}{c_s} \right)^{3/2} \mathcal{D}^3 = (1 + \gamma_R)^{-3}, \quad (37)$$

which, for the $\gamma_R = 1$ case, gives $1/8$. (We got the same number from simple scaling laws at the outset, but the above machinery will be useful when generalizing in Sec. 2.)

The number of detected planets N_{det} is the sum of the two appropriate equations above, and can be written

$$N_{\text{det}} = \left(\frac{\delta}{x_{\min}} \right)^3 (AT_{\text{dur}} N_{\text{tra}})^{3/2} \frac{R_\star}{a} \times \left[N_s \Gamma_{t,s} \frac{c_s^{3/2}}{d_{\max,s}^3} + 2N_d \Gamma_{t,d} \frac{c_d^{3/2}}{d_{\max,d}^3} \right]. \quad (38)$$

All terms can be input to a computer, and checked against a Monte Carlo simulation if desired.

1.5. Astronomer A ignores binarity

Astronomer A has never heard of binary star systems. Nor has he heard of completeness corrections. He knows about geometric transit probabilities. What occurrence rate does he derive for planets of radius R_p and period P ?

The *total* occurrence rate (number of planets divided by number of “stars”) for Astronomer A would be $N_{\text{det}}/(N_s + N_d)$, up to the geometric correction. However, even though Astronomer A does not know about binaries, the radii he derives for any planets in binary systems are too small, by a factor $\sqrt{\mathcal{D}}$ (for twin binaries, they are all $R_p/\sqrt{2}$). Astronomer A wants an occurrence rate for planets of radius R_p and period P . The answer is thus

$$\Gamma_{A,R_p} = \frac{N_{\text{det},s}/f_{s,g}}{N_s + N_d}. \quad (39)$$

This astronomer will also think there is a second population of planets, with radius $R_p\sqrt{\mathcal{D}}$, and will thus rush to Nature claiming to also have derived a second occurrence rate,

$$\Gamma_{A,R_p\sqrt{\mathcal{D}}} = \frac{N_{\text{det},d}/f_{s,g}}{N_s + N_d}. \quad (40)$$

Note in the second case that this astronomer thinks these are single stars, and generally will miscompute the geometric transit probability.

1.6. Astronomer B counts host stars correctly

Astronomer B can somehow account correctly for the “2 for 1” effect of binarity, *i.e.* that the sample actually has more stars than astronomer A thought.

By the same token as above,

$$\Gamma_{B,R_p} = \frac{N_{\text{det},s}/f_{s,g}}{N_s + 2N_d}, \quad (41)$$

and

$$\Gamma_{B,R_p\sqrt{D}} = \frac{N_{\text{det},d}/f_{d,g}}{N_s + 2N_d}. \quad (42)$$

1.7. Astronomer C counts host stars correctly and figures out diluted radii

Astronomer C did high resolution imaging followup on every candidate, and correctly classifies the planetary radii. Thus, she also knows which planets are in binary systems, and which are in single star systems.

She knows that the purported population of planets with radii $R_p\sqrt{D}$ does not exist. All detected planets from this survey have radii R_p . She computes an occurrence rate

$$\Gamma_{C,R_p} = \frac{N_{\text{det},s}/f_{s,g} + N_{\text{det},d}/f_{d,g}}{N_s + 2N_d} \quad (43)$$

the closest yet to the true rate (Sec. 1.3).

1.8. Astronomer D counts host stars correctly, figures out diluted radii, and accounts for completeness

Astronomer D, knows which detections were around binaries and the associated radius correction. They do injection recovery, and derive correct estimates for their completeness functions about single stars $f_{s,c}$, and about double star systems, $f_{d,c}$. With this knowledge in hand, they compute the respective single and binary occurrence rates

$$\Gamma_{t,s} = \frac{N_{\text{det},s}}{N_s f_{s,g} f_{s,c}}, \quad (44)$$

$$\Gamma_{t,d} = \frac{N_{\text{det},d}}{2N_d f_{d,g} f_{d,c}}. \quad (45)$$

With these in hand, they derive the overall occurrence rate

$$\Gamma_{D,R_p} = \frac{N_{\text{det},s}/(f_{s,g} f_{s,c}) + N_{\text{det},d}/(f_{d,g} f_{d,c})}{N_s + 2N_d} \quad (46)$$

$$= \frac{\Gamma_{t,s} N_s + 2\Gamma_{t,d} N_d}{N_s + 2N_d}. \quad (47)$$

Although we admit it’s been a bit of a slog, it happens that Astronomer D’s occurrence rate is the true occurrence rate (cf. Eq. 12).

All it takes is \approx a full semester at Keck, a system-by-system analysis, and perfect understanding of the completeness of the detection efficiency for single and double star systems.

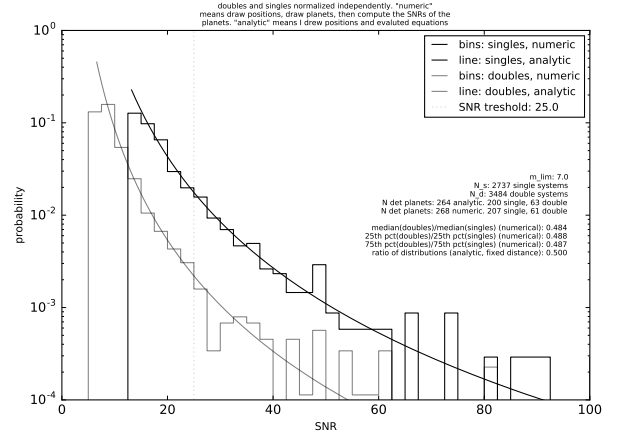


FIG. 1.— Comparison of analytic and numeric probability density functions of the SNR in an idealized transit survey. The analytic lines are Eqs. 31 and 32 for the planet populations orbiting single and binary stars. The underlying stepped histogram is output from Monte Carlo simulations. Poisson noise leads to a small deviation at the faint and bright limits, but the numerics and analytics otherwise agree.

1.9. Numerical verification

To check the preceding analytic development, we implemented a Monte Carlo simulation of this idealized transit survey. To run the survey, we defined the instrument specifications (detector area and transmission function), the stellar population (binary fraction, total number density of a given stellar class, fixed stellar properties), the planet population (fixed planet radius, period, and occurrence rate about single and binary stars), and finally the survey parameters (observing baseline, minimum SNR for “detection”). We then randomly drew star positions, randomly assigned planets to stars in single and binary systems, and computed the resulting signal to noise (Eq. 20) with which the transits would be observed. As in the preceding analytics, we assumed “twin” binaries (same stellar radii, same effective temperature, and dilution does not depend on which stellar binary is the “target”).

The results are shown in Fig. 1, and indicate that the analytic probability distribution functions Eqs. 31, 32, and the number of detections (Eq. 38) are correct.

A point evident in Fig. 1 is that, for fixed planet parameters, and fixed stellar parameters (R_\star , L_\star , and distance r) the SNR distribution for planets in binaries is poorer than that of planets in single star systems. We can see analytically that this simply due to dilution:

$$\frac{\text{prob}(x_d)}{\text{prob}(x_s)} = (1 + \gamma_R)^{-1} = D. \quad (48)$$

Deriving this simple form requires noting that the ratios of the bandpass-specific number luminosities (Eq. 27) is equal to the ratio of the bandpass-specific energy luminosities (otherwise a term with c_s/c_d must be included).

1.10. Representative numbers for a few cases

1.10.1. Twin binaries: if we ignore binarity, for what fraction of detections do we misclassify the radii?

Ignoring binarity, we will detect $N_{\text{det},s}$ planets around single stars, and $N_{\text{det},d}$ planets around double stars. The

latter set will be assumed to have radii $R_p\sqrt{\mathcal{D}}$. The fraction of detections with misclassified radii can then be written

$$\frac{N_{\text{det},d}}{N_{\text{det},s} + N_{\text{det},d}} = \frac{1}{1 + \alpha}, \quad (49)$$

for

$$\alpha \equiv \frac{1}{2(\text{BF})} (1 + \gamma_R)^{3/2} \frac{\Gamma_{t,s}}{\Gamma_{t,d}}. \quad (50)$$

For the nominal G2V dwarf case of $\text{BF} = 0.45$, twin binaries with equal occurrence rates this produces a misclassification rate of 24%, in agreement with Fig. 1.

1.10.2. *Twin binaries: if we ignore binarity, how wrong is our occurrence rate for planets of radius R_p ?*

This is almost simply asking “what is the relative difference between the occurrence rates derived by Astronomers D and A for planets of radius R_p ?” However, in the more realistic case, Astronomer A also has derived a completeness, which we assume is the same as for Astronomer D in the single star case. So Astronomer A now misclassifies planetary radii, and miscounts the total number of stars, but knows his completeness for single stars. Astronomer D corrects all these errors. Then

$$\Gamma_{A,R_p} = \frac{N_{\text{det},s}/(f_{s,g}f_{s,c})}{N_s + N_d}, \quad (51)$$

and the relative difference between the two occurrence rates is

$$\left| \frac{\Gamma_{A,R_p} - \Gamma_{D,R_p}}{\Gamma_{A,R_p}} \right| = \left| 1 - \frac{\Gamma_{D,R_p}}{\Gamma_{A,R_p}} \right| \quad (52)$$

$$= \left| 1 - \left(\frac{\Gamma_{t,s}N_s + 2\Gamma_{t,d}N_d}{N_s + 2N_d} \cdot \frac{N_s + N_d}{\Gamma_{t,s}N_s} \right) \right| \quad (53)$$

$$= \left| 1 - \frac{(1 + 2\beta\Gamma_{t,d}/\Gamma_{t,s})(1 + \beta)}{(1 + 2\beta)} \right|, \quad (54)$$

for

$$\beta \equiv N_d/N_s = \text{BF} \times (1 + \gamma_R)^{3/2}. \quad (55)$$

For the nominal G2V dwarf case of $\text{BF} = 0.45$ with twin binaries ($\gamma_R = 1$) and $\Gamma_{t,d} = \Gamma_{t,s}$ this gives a relative error of 127%. For instance, in the numerical simulation corresponding to Fig. 1, Astronomer A finds $\Gamma_{A,R_p} = 0.22$, while Astronomer D derives the true (input) occurrence rate of $\Gamma_{D,R_p} = 0.5$.

2. MODEL #2: FIXED STARS, FIXED PLANETS, VARYING LIGHT-RATIO BINARIES

Same story as Model #1, except now γ_R varies across the population of star systems. It does so because the underlying mass ratio varies. Since we are interested in solar type binaries, we take the distribution of the mass ratio $q = M_2/M_1$ by approximating [Rhagavan et al 2010, Fig 16]:

$$\text{prob}(q) = \begin{cases} c_q & 0.1 < q \leq 1 \\ 0 & \text{otherwise,} \end{cases} \quad (56)$$

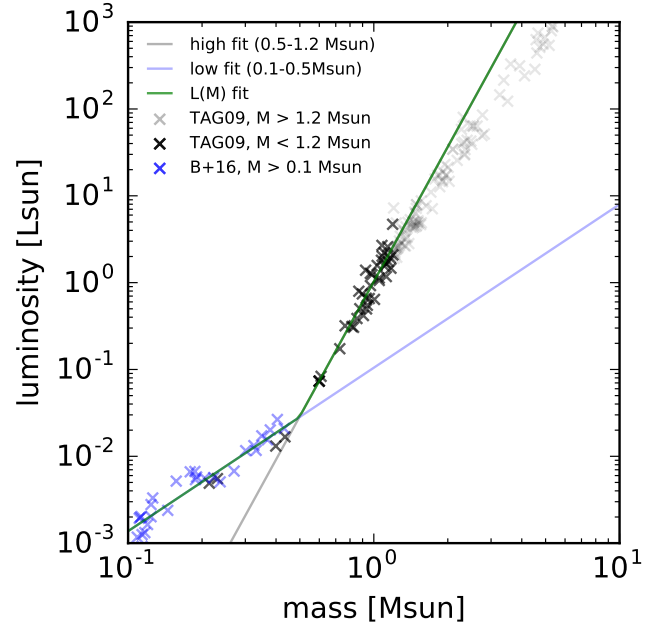


FIG. 2.— Empirical fit to main sequence dwarf mass luminosity data from Torres et al. [2009] and Benedict et al. [2016]. The “low fit” is a least squares fit to data from $0.1 - 0.5 M_\odot$, and the “high fit” is to data above that, and below the Kraft break ($1.2 M_\odot$). The $L(M)$ relation taken for subsequent numerics is the maximum of the two fits.

for $c_q = 1/9$ to normalize the distribution to 1. The mass-luminosity relation can, for analytic convenience, be approximated as $L = M^\alpha$, with the lore-value of α being 3.5. While we use this in subsequent analytic development, for numerics we fit a lines to mass-luminosity data collected by Torres et al. [2009] for dwarfs above M , and Benedict et al. [2016] for dwarfs below M . To convert Benedict et al. [2016]’s reported M_V values to absolute luminosities, we interpolated over E. Mamajek’s table³. In log-log space, we let the intersection point of the lines float, and make various cuts on the data as indicated in Fig 2. The fit parameters are available in a footnote⁴.

The other changes between Model #2 and Model #1 are:

1. We allow a third occurrence rate, splitting $\Gamma_{t,d} \rightarrow (\Gamma_{t,d1}, \Gamma_{t,d2})$ where the latter terms represent the occurrence rate around the primary of a double star system, and the occurrence rate around the secondary of a double star system. For the case of *e.g.*, $q = 0.1$, this seems more relevant. We can then take the limit that says “any planet around the secondary is as likely as around the primary”, or go to the opposite extreme of “there are only planets around the primaries, not the secondaries”.
2. Our binary systems have secondaries with varying masses. Thus they have varying radii. Taking [Demircan & Kahraman 1991]’s empirical fit to

³ pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt, downloaded 2017.08.02

⁴ m lo: 1.8818719873988132 – c lo: -0.9799647314108376 – m hi: 5.1540712426599882 – c hi: 0.0127626185389781 – M at merge: 0.4972991257826812 – L at merge: 0.0281260412126928.

eclipsing binary data,

$$R = 1.06M^{0.945} \quad (57)$$

for all the stars in our desired mass range ($M < 1.66M_\odot$, D&K 1991's stated bounds), and for each in solar units. This only affects the transit probabilities about the secondaries of binary star systems.

We then ask the same questions as for Model #1.

2.1. How many stars are in the sample?

Let N_s be the number of single star systems, and N_d the number of double star systems. N_s is the same as in Sec. 1.

Now

$$N_d(\gamma_R) = n_d \frac{4\pi}{3} d_{\max,d}^3(L_1^N, \gamma_R, F_{\lim}^N) \quad (58)$$

for

$$d_{\max,d}(L_1^N, \gamma_R, F_{\lim}^N) = d_{\max,s}(L_1^N, F_{\lim}^N) \times (1 + \gamma_R)^{1/2}. \quad (59)$$

Since q is a random variable, γ_R is a random variable, and $d_{\max,d}$ is a random variable. Since N_d is a function of $d_{\max,d}$, the number of double star systems becomes a random variable. The expected (mean) number of double star systems in the sample is

$$\langle N_d \rangle = \int_0^\infty N_d \text{prob}(N_d) dN_d. \quad (60)$$

By applying the chain rule for probability density functions, the distribution $\text{prob}(N_d)$ can be written

$$\text{prob}(N_d) = \text{prob}(q(\gamma_R)) \left| \frac{dq}{d\gamma_R} \right| \left| \frac{d\gamma_R}{dd_{\max,d}} \right| \left| \frac{dd_{\max,d}}{dN_d} \right|. \quad (61)$$

Doing some algebra, and assuming $\gamma_R = q^3$, this can be shown to be

$$\text{prob}(N_d) = \frac{2}{81} N_d^{-1/3} \left(\frac{3}{4\pi n_d} \right)^{2/3} d_{\max,s}^{-2/3} \times \left[\left(\frac{3N_d}{4\pi n_d} \right)^{2/3} - d_{\max,s}^2 \right]^{-2/3} \quad (62)$$

over the interval

$$\text{lower bound} = \frac{4\pi n_d}{3} (\sqrt{0.1^3 + 1} d_{\max,s})^3 \quad (63)$$

$$\text{upper bound} = \frac{4\pi n_d}{3} (\sqrt{2} d_{\max,s})^3, \quad (64)$$

and outside the stated interval $\text{prob}(N_d) = 0$.

While a closed analytic expression for $\langle N_d \rangle$ does exist, it is messy, and it does not yield much intuition. Instead, summarizing the important points and supporting them numerically:

- In a volume-limited sample of binary star systems in which the primary mass is fixed, and the mass ratio is drawn from a bounded uniform distribution, the distribution of γ_R will be biased towards low values ($\gamma_R \approx 0.1$). This is shown in Fig. 3.

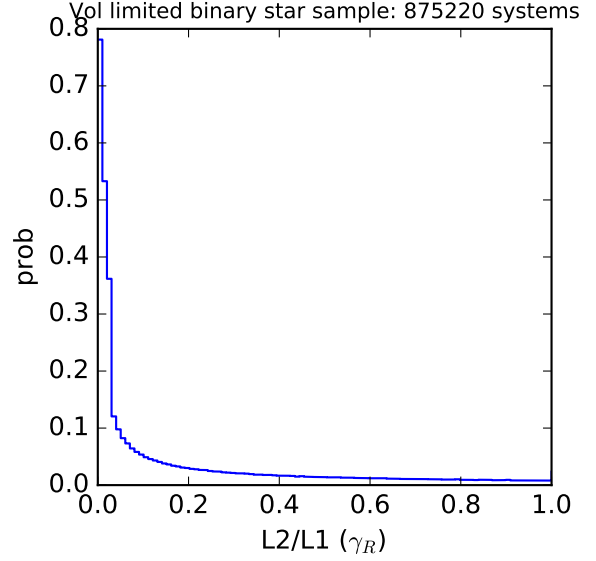


FIG. 3.— The distribution of the luminosity ratio for a volume limited sample of binary stars. $\text{BF} = 0.45$ [Duchene and Kraus 2013]; total number density from Bovy 2017; $M(L)$ relation from Fig. 2; mass ratios are drawn from Eq. 56.

- In a magnitude-limited sample of binary star systems in which the primary mass is fixed, and the mass ratio of the *population* is drawn from a bounded uniform distribution, the observed distribution of mass ratios will be biased towards high values. You will see more twins, because they are detectable out to a greater distance⁵. This is shown in Fig. 4.
- In a magnitude-limited sample of binary star systems in which the primary mass is fixed, the distribution of γ_R will be biased towards low values (≈ 0.1), but less so than in a volume-limited sample. This is shown in Fig. 5.

In passing, the bias in $\text{prob}(\gamma_R)$ towards low luminosity ratios can be seen analytically. If we assume $\gamma_R = q^\alpha$, then

$$\text{prob}(\gamma_R) = \frac{1}{9\alpha} \gamma_R^{\frac{1-\alpha}{\alpha}} \quad \text{for } (0.1)^\alpha < \gamma_R < 1, \quad (65)$$

and otherwise zero. For instance if $\alpha = 3$, $\text{prob}(\gamma_R) \propto \gamma_R^{-2/3}$ and the domain extends from 1 to 10^{-3} , where the probability distribution peaks.

How many stars are in the sample?— We return to the original question: how many stars?

$$\langle N_{\text{stars}} \rangle = N_s + 2\langle N_d \rangle, \quad (66)$$

as before. Since we are not giving analytic expressions for either of the $\langle \dots \rangle$ terms, we just note that in the case of a population with a single γ_R value, $N_d/N_s = \text{BF} \times (1 + \gamma_R)^{3/2}$. A lower bound for this ratio will always be the binary fraction. For a single run of a Monte Carlo code, where $\text{mean}(\gamma_R) = 0.28$, $\text{median}(\gamma_R) = 0.13$, and

⁵ This is a scarcity systematic w.r.t. the claimed intrinsic excess of twin binaries.

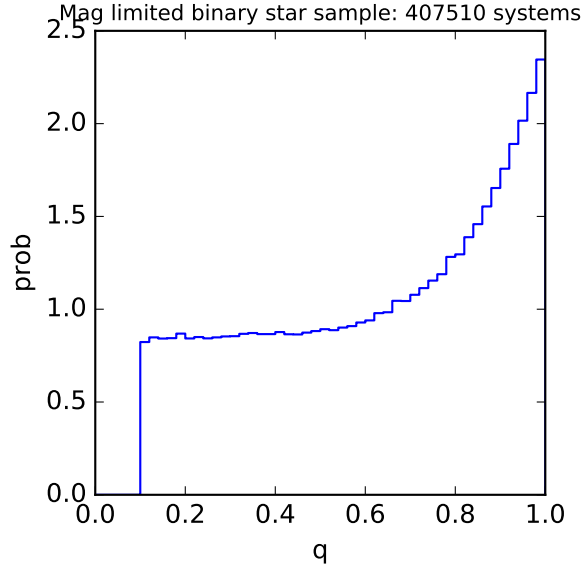


FIG. 4.— The distribution of the mass ratio for a magnitude limited sample of binary stars. $BF = 0.45$ [Duchene and Kraus 2013]; total number density from Bovy 2017; $M(L)$ relation from Fig. 2; mass ratios are drawn from Eq. 56 – a *uniform* distribution in a volume-limited sample!

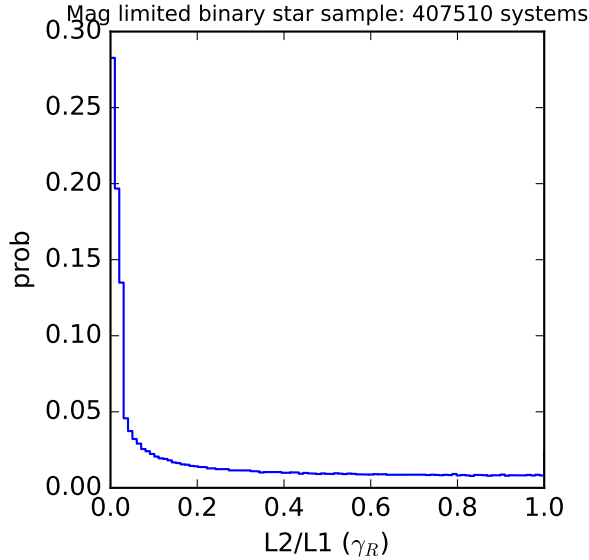


FIG. 5.— The distribution of the luminosity ratio for a magnitude limited sample of binary stars. $BF = 0.45$ [Duchene and Kraus 2013]; total number density from Bovy 2017; $M(L)$ relation from Fig. 2; mass ratios are drawn from Eq. 56.

the distribution was that given in Fig. 5, the observed ratio was $N_d/N_s = 0.59$. This corresponds to a single-valued population with $\gamma_R \sim 0.20$, a number between the mean and median of the true distribution. It means ~ 1.2 stars in binary star systems for every star in a single star system from this sample.

We also note in passing that $N_d/N_s = 0.59$ is a smaller ratio than for the case of a fixed γ_R population with $\gamma_R = 1$, which gave $N_d/N_s = 1.27$. The difference is that in the latter population, the volume of searchable

stars is greater. Based on the fixed- γ_R population scaling law $N_d/N_s \propto (1 + \gamma_R)^{3/2}$, we would expect the ratio of single stars to be $\sim (2/1.1)^{3/2} = 2.4$ times smaller, which is roughly (though not exactly) what we observe.

2.2. How many planets are in the sample?

The mean number of planets in the sample is

$$\langle N_{\text{planets}} \rangle = N_{\text{planets in single star systems}} + \quad (67)$$

$$N_{\text{planets in double star systems}} \\ = \Gamma_{t,s} N_s + (\Gamma_{t,d1} + \Gamma_{t,d2}) \langle N_d \rangle. \quad (68)$$

What was formerly a factor of 2 has now been split into the true occurrence rates about the primaries and secondaries of double star systems.

2.3. What is the true occurrence rate?

The “true occurrence rate” is the average number of planets per star. Thus

$$\Gamma_t = \frac{\langle N_{\text{planets}} \rangle}{\langle N_{\text{stars}} \rangle} \quad (69)$$

$$\Gamma_t = \frac{\Gamma_{t,s} N_s + (\Gamma_{t,d1} + \Gamma_{t,d2}) \langle N_d \rangle}{N_s + 2 \langle N_d \rangle}. \quad (70)$$

2.4. How many planets are detected?

The total number of planet detections is the sum of the number of planets detected in single star systems $N_{\text{det},s}$ and the number of planets detected in double star systems $N_{\text{det},d}$. The latter of these is a random variable, and can be further split into the primary and secondary contributions $N_{\text{det},d1}$, $N_{\text{det},d2}$.

The number of planets detected in single star systems is

$$N_{\text{det},s} = N_s \Gamma_{t,s} f_{s,g} f_{s,c}, \quad (71)$$

where the product $N_s \Gamma_{t,s}$ is the number of planets in the single star systems of the sample, $f_{s,g}$ is the geometric probability of the planets transiting, and $f_{s,c}$ is the fraction of transiting planets around single star systems that are detected (the completeness).

The mean number of planets detected in double star systems is

$$\langle N_{\text{det},d} \rangle = \langle N_d \rangle (\Gamma_{t,d1} f_{d1,g} f_{d1,c} + \Gamma_{t,d2} f_{d2,g} f_{d2,c}). \quad (72)$$

Now $\langle N_d \rangle \Gamma_{t,d1}$ is the mean number of planets orbiting the primaries of double star systems, ditto the corresponding expression for the secondaries. For this population, $f_{d1,g} = f_{s,g}$, but $f_{d1,g} \neq f_{d2,g}$. The completeness fractions differ between the primary and secondary because the dilution as a function of the light ratio, $\mathcal{D}(\gamma_R)$, has different behavior for the two cases. We show this in Fig. 6. Specifically, a planet orbiting the secondary will have worse dilution than a planet orbiting the primary, and consequently lower completeness.

We could probably derive analytic expressions for all the terms between brackets in Eq. 72, completeness included. I’m not yet convinced we need to – we should get cancellations for most of the questions we want to ask.

2.4.1. Analytic completeness

todo, if necessary

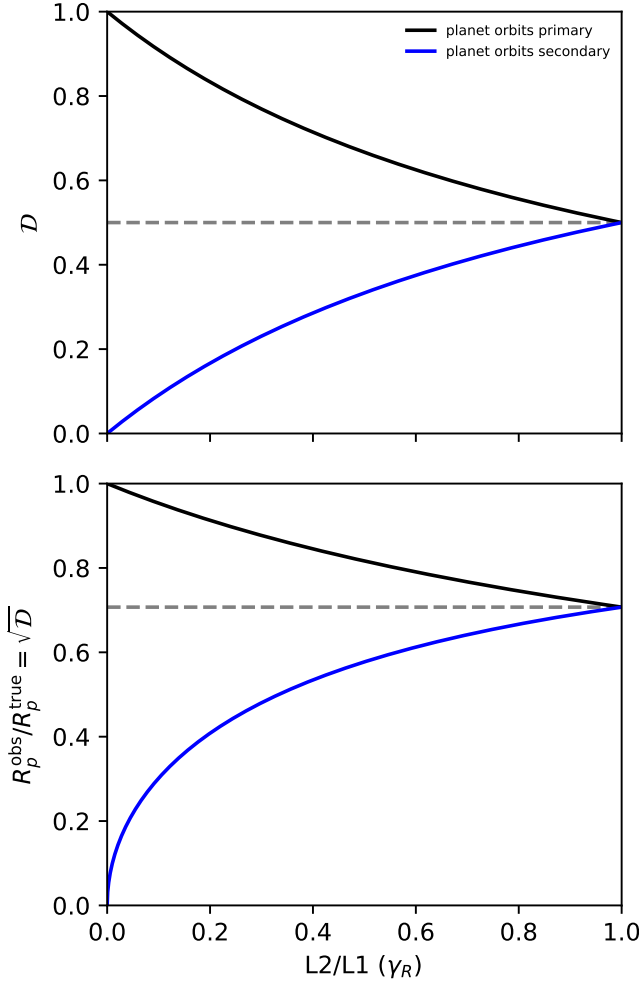


FIG. 6.— Dilution \mathcal{D} and observed to true radius ratio $\sqrt{\mathcal{D}}$ plotted against binary light ratio γ_R . This illustrates Eq. 17. The lower panel’s horizontal line is at $1/\sqrt{2}$. The implication is that if a detected planet orbits the secondary and you do not know it, you cannot measure R_p to better than $\approx 29\%$ accuracy. Conversely, if it orbits the primary and you do not know it, you will get the radius wrong by at most $\approx 29\%$.

2.4.2. Deriving $\text{prob}(x_i)$

todo, if necessary

2.4.3. Number of detected planets

todo, if necessary

2.5. Astronomer A ignores binarity

Just as in Sec. 1, Astronomer A

- has never heard about binary star systems.
- has never heard about completeness corrections.
- knows about geometric transit probabilities.

What occurrence rate does he derive for planets of radius R_p and period P ?

The answer is the same as in Sec. 1,

$$\Gamma_{A,R_p} = \frac{N_{\text{det},s}/f_{s,g}}{N_s + N_d}. \quad (73)$$

However, now rather than planets detected in binary systems being perceived as a second population of planets with fixed radius $R_p\sqrt{\mathcal{D}}$, there will be an apparent spectrum of diluted radii, since \mathcal{D} varies by system. Writing the number of detections in double systems as a function of observed radius, $N_{\text{det},d}(R_p^{\text{obs}})$, Astronomer A would derive an occurrence rate for this separate population of

$$\Gamma_A(R_p^{\text{obs}}) = \frac{N_{\text{det},d}(R_p^{\text{obs}})/f_{s,g}}{N_s + N_d}. \quad (74)$$

Note that this astronomer thinks these planets orbit single stars, and so miscomputes the geometric transit probability for the secondaries.

A more realistic question is “what fraction of multiples have γ_R smaller enough so that you’d still observe a radius close to the true one?”. We discuss this in Sec. 2.10.3.

2.6. Astronomer B counts host stars correctly

Same as astronomer A above, except the denominator becomes $N_s + 2N_d$.

2.7. Astronomer C counts host stars correctly and figures out diluted radii

In Sec. 1, Astronomer C only had to do high resolution imaging to measure γ_R for each system. Since for $\gamma_R = 1$, the dilution is single-valued (Fig. 6), this immediately told her the true planet radius (since the primary/secondary distinction was meaningless).

In Model #2’s universe, the dilution is double-valued for a given γ_R . To correctly work out the diluted radii, Astronomer C now needs more than high resolution imaging – she needs to resolve the stars during transit to discover which star the planet orbits. She also needs to derive masses for the primaries and secondaries (to make the geometric transit probability correction).

Only after doing this does she find that all detected planets from this survey have radii R_p . She computes an occurrence rate

$$\Gamma_{C,R_p} = \frac{N_{\text{det},s}/f_{s,g} + N_{\text{det},d1}/f_{d1,g} + N_{\text{det},d2}/f_{d2,g}}{N_s + 2N_d} \quad (75)$$

the closest yet to the true rate (Sec. 1.3).

2.8. Astronomer D counts host stars correctly, figures out diluted radii, and accounts for completeness

Same story as for Model #1. Now, they do injection recovery and derive correct estimates for their completeness functions about single stars $f_{s,c}$, the primaries of double star systems, $f_{d1,c}$, and the secondaries of double star systems $f_{d2,c}$. They compute the respective single and binary occurrence rates

$$\Gamma_{t,s} = \frac{N_{\text{det},s}}{N_s f_{s,g} f_{s,c}}, \quad (76)$$

$$\Gamma_{t,d1} = \frac{N_{\text{det},d1}}{N_d f_{d1,g} f_{d1,c}}, \quad (77)$$

$$\Gamma_{t,d2} = \frac{N_{\text{det},d2}}{N_d f_{d2,g} f_{d2,c}}. \quad (78)$$

With these in hand, they derive the overall occurrence rate

$$\Gamma_{D,R_p} = \frac{N_{\text{det},s}/f_s + N_{\text{det},d1}/f_{d1} + N_{\text{det},d2}/f_{d2}}{N_s + 2N_d} \quad (79)$$

$$= \frac{\Gamma_{t,s}N_s + (\Gamma_{t,d1} + \Gamma_{t,d2})N_d}{N_s + 2N_d}, \quad (80)$$

where $f_s \equiv f_{s,g}f_{s,c}$, $f_{d1} \equiv f_{d1,g}f_{d1,c}$, $f_{d2} \equiv f_{d2,g}f_{d2,c}$ for brevity. Astronomer D's occurrence rate is the true occurrence rate (cf. Eq. 70).

2.9. Numerical verification

todo, if needed

2.10. Representative numbers for a few cases

We have not given analytic expressions for the number of double star systems N_d because even with simplified $M(L)$ relations, they are unwieldy. However, we can ask similar numerical questions as from Sec. 1, and compare the answers.

2.10.1. If we ignore binarity, for what fraction of detections do we misclassify the radii?

Ignoring binarity, we will detect $N_{\text{det},s}$ planets around single stars, and $N_{\text{det},d}$ planets around double stars. The latter set will be assumed to have radii $R_p\sqrt{\mathcal{D}}$, where \mathcal{D} is the vector of dilutions appropriate for each binary star system. The fraction of detections with misclassified radii can then be written

$$\frac{N_{\text{det},d}}{N_{\text{det},s} + N_{\text{det},d}} = \frac{1}{1 + \alpha}, \quad (81)$$

for

$$\alpha \equiv \frac{N_d(\Gamma_{t,d1}f_{d1} + \Gamma_{t,d2}f_{d2})}{N_s\Gamma_{t,s}f_s}, \quad (82)$$

where $f_s \equiv f_{s,g}f_{s,c}$, $f_{d1} \equiv f_{d1,g}f_{d1,c}$, $f_{d2} \equiv f_{d2,g}f_{d2,c}$ for brevity.

The different secondary completeness matters here if proceeding analytically. Would be easier to do numerically. Regardless, this is not the most interesting question right now.

2.10.2. If we ignore binarity, how wrong is our occurrence rate for planets of radius R_p ?

Write $\Gamma_{A, \text{planets of } R_p} = \Gamma_{A,R_p}$, and similarly for D. Just as in Model #1, the answer to “what is the relative difference between the occurrence rates derived by Astronomers D and A for planets of radius R_p ?” is simpler when we assume that Astronomer A has also derived a completeness, which we assume is the same as for Astronomer D in the single star case. So Astronomer A now misclassifies planetary radii, and miscounts the total number of stars, but knows his completeness for single stars. Astronomer D corrects all these errors. Then

$$\Gamma_{A,R_p} = \frac{N_{\text{det},s}/(f_{s,g}f_{s,c})}{N_s + N_d}. \quad (83)$$

The relative difference between the two occurrence

rates, $\Delta \equiv |(\Gamma_{A,R_p} - \Gamma_{D,R_p})/\Gamma_{A,R_p}|$, is

$$\Delta = \left| 1 - \frac{\Gamma_{D,R_p}}{\Gamma_{A,R_p}} \right| \quad (84)$$

$$= \left| 1 - \left(\frac{\Gamma_{t,s}N_s + (\Gamma_{t,d1} + \Gamma_{t,d2})N_d}{N_s + 2N_d} \cdot \frac{N_s + N_d}{\Gamma_{t,s}N_s} \right) \right| \quad (85)$$

$$= \left| 1 - \frac{(1 + \beta)(\Gamma_{t,d1} + \Gamma_{t,d2})/\Gamma_{t,s}(1 + \beta)}{(1 + 2\beta)} \right| \quad (86)$$

$$= \left| 1 - \left([1 + \chi\beta] \cdot \frac{1 + \beta}{1 + 2\beta} \right) \right| \quad (87)$$

for

$$\beta \equiv N_d/N_s, \quad \chi \equiv (\Gamma_{t,d1} + \Gamma_{t,d2})/\Gamma_{t,s}. \quad (88)$$

In the limit where the R_p planet has the same occurrence rate about every host type regardless of its mass, $\chi = 2$, and this simplifies to $\Delta = \beta$, for a relative error of 59%. This is less than for Model #1 because of the smaller volume of searchable binary stars in Model #2.

If we consider the opposite limit of the R_p planet *only* existing around single stars and the primaries of double star systems with equal occurrence, $\chi = 1$, and $\Delta = \beta^2/(1 + 2\beta)$. For $\beta = 0.59$, this gives $\Delta = 0.16$.

2.10.3. What if we ignore binarity, but count derived planet radii that are “close enough”?

Let x be the acceptable margin of error on the observed radius, so that we are interested in counting the number of planets detected with $R_p < R_p^{\text{obs}} < (1 \pm x)R_p$. For Model #2, since all planets have radius R_p , we only care about the smaller radius case, *i.e.*, we want to count $R_p < R_p^{\text{obs}} < (1 - x)R_p$, for instance with $x = 0.1$.

All of these diluted detections will be from binary systems. The number of planets detected in double star systems with observed radius R_p^{obs} is

$$N_{\text{det},d}(R_p^{\text{obs}}) = N_{\text{det},d1}(R_p^{\text{obs}}) + N_{\text{det},d2}(R_p^{\text{obs}}) \quad (89)$$

$$= \Gamma_{t,d1}f_{d1}(R_p^{\text{obs}})N_{d1}(R_p^{\text{obs}}) + \Gamma_{t,d2}f_{d2}(R_p^{\text{obs}})N_{d2}(R_p^{\text{obs}}). \quad (90)$$

As always, the fractional terms f_{di} for $i \in 1, 2$ encompass geometric and completeness corrections. We're writing both them and the number of stars for which a given observed radius would be seen as functions of R_p^{obs} .

To save on notational horribleness, write $R'_p \equiv \{(1 - x)R_p < R_p^{\text{obs}} < R_p\}$, in other words let R'_p denote the set of planets that are observed in the desired radius range.

Then writing $N_{d1}(R'_p)$ is relatively straight-forward:

$$N_{d1}(R'_p) = \int_0^{\min(1, \gamma_{R,u})} N_d(\gamma_R) \text{prob}(\gamma_R) d(\gamma_R), \quad (91)$$

where the upper limit is set by equating the limiting radius $(1 - x)R_p$ to the observed diluted radius $(1 + \gamma_R)^{-1/2}R_p$, giving

$$\gamma_{R,u} \equiv \frac{x(2 - x)}{(1 - x)^2}. \quad (92)$$

If $x > (1 - 2^{-1/2})$, bad things happen (cf. Fig. 6), hence the need for the minimum.

The analogous equation for secondaries is

$$N_{d2}(R'_p) = \begin{cases} \int_{\gamma_{R,l}}^1 N_d(\gamma_R) \text{prob}(\gamma_R) d(\gamma_R), & x > 1 - 2^{-1/2} \\ 0, & x \leq 1 - 2^{-1/2} \end{cases} \quad (93)$$

where

$$\gamma_{R,l} \equiv \frac{(1-x)^2}{x(2-x)}. \quad (94)$$

Eq. 93 accounts for the fact that you cannot measure R_p to better than $\approx 29\%$ if the planet orbits the secondary of an unidentified binary. This was originally noted in Fig. 6.

If we were forward-modelling (*i.e.* we actually wanted to compute $N_{\text{det},d}(R'_p)$) we would need to evaluate the completeness term and this “fraction of slightly diluted stars” terms together, and integrate jointly to find $N_{\text{det},d1}(R'_p)$ and $N_{\text{det},d2}(R'_p)$.

If we ask the right question, this will not be necessary.

Astronomer A misclassifies planetary radii, miscounts the total number of stars, but knows his completeness and geometric correction for single stars as a function of radius $f_s(R'_p)$. His occurrence rate for planets of radius near R_p , *i.e.* the set of planets with $\{(1-x)R_p < R_p^{\text{obs}} < R_p\}$, is

$$\Gamma_{A,R'_p} = \frac{(N_{\text{det},s} + N_{\text{det},d}(R'_p))/f_s(R'_p)}{N_s + N_d}. \quad (95)$$

For him,

$$f_s(R'_p) = f_{s,g}(R'_p)f_{s,c}(R'_p). \quad (96)$$

Astronomer D, who had everything right, doesn't change anything.

The new difference between the two rates is

$$\Delta' \equiv \left| \frac{\Gamma_{A,R'_p} - \Gamma_{D,R_p}}{\Gamma_{A,R'_p}} \right| \quad (97)$$

$$= \left| 1 - \frac{\Gamma_{D,R_p}}{\Gamma_{A,R'_p}} \right| \quad (98)$$

$$= \left| 1 - \left(\frac{1 + \chi\beta}{1 + \xi} \cdot \frac{1 + \beta}{1 + 2\beta} \right) \right|, \quad (99)$$

for

$$\beta \equiv N_d/N_s, \quad \chi \equiv (\Gamma_{t,d1} + \Gamma_{t,d2})/\Gamma_{t,s}, \quad (100)$$

and

$$\xi \equiv \frac{N_{\text{det},d}(R'_p)}{f_s(R'_p)} \cdot \frac{1}{N_s \Gamma_{t,s}}. \quad (101)$$

If we let $x < 1 - 2^{-1/2}$, for instance $x = 0.1$, then $N_{d2}(R'_p) = 0$. The expression for ξ simplifies to

$$\chi = \frac{N_{d1}(R'_p)}{N_s} \cdot \frac{\Gamma_{t,d1}}{\Gamma_{t,s}} \cdot \frac{f_{d1}(R'_p)}{f_s(R'_p)}, \quad (\text{if } x < 1 - 2^{-1/2}). \quad (102)$$

If we assume $\Gamma_{t,d1} = \Gamma_{t,s}$, as we were doing anyway, Eq. 102 is bounded above by β . This is because $N_{d1}(R'_p) \leq N_{d1} = N_d$, and $f_{d1}(R'_p)/f_s(R'_p) \leq 1$, since

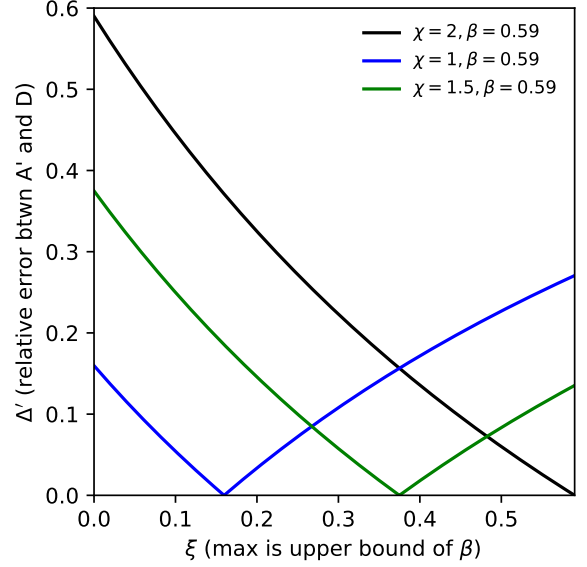


FIG. 7.— *Y axis*: relative error between the Astronomer who ignores binarity but counts planetary radii that are “close enough” (A') and the Astronomer who knows everything (D). *X axis*: χ as defined in Eq. 102. Taking a guess at the true value, if $N_{d1}(R'_p) = N_d/2$, and the occurrence rates between ‘d1’ and ‘s’ are the same, and we put $\gamma_R = 0.1$ into the completeness ratio, then $\xi = (\beta/2) \times (1 + 0.1)^{-3} = 0.22$. The relative error then varies between $\approx 5\% - 35\%$, depending on what fraction of secondaries you assume host planets.

over a fixed radius interval dilution will always yield a lower completeness fraction for double star systems than singles. Recall for the fixed γ_R population we showed that $f_{d1}/f_s = (1 + \gamma_R)^{-3}$.

Regardless, the cheap way of making the estimate without deriving the actual completeness fractions, or even running numerics, is just to plot the error as a function of ξ . We do this in Fig. 7, which shows that the relative error varies between $\approx 5\% - 35\%$, depending on what fraction of secondaries you assume host planets.