

## A TOY ANALYTIC TRANSIT SURVEY: FIXED LIGHT RATIO BINARIES

LGB, KM, JNW

*Memo for internal use*

### ABSTRACT

We discuss various errors one can make by ignoring binarity when deriving occurrence rates. In this memo, we assume the stars in any binary system have fixed light ratio.

*Subject headings:* planets and satellites: detection

### STATEMENT OF PROBLEM

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 survey, similar to that outlined by [Pepper et al, 2003]:

1. You are going to observe the entire sky for a duration  $T_{\text{obs}}$ , 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 only want to detect them around stars of radius  $R_1$ , and luminosity  $L_1$ . For instance, G2V dwarfs.
4. Since your detections will be S/N limited, you want only to observe stars for which your target can be detected with  $S/N > (S/N)_{\text{min}}$ .
5. For a photon-noise limited survey, the signal to noise limit is equivalent to a magnitude (flux) limit. So you select all the points on the sky above a flux limit, *i.e.* with flux  $F_{\text{pt}} > F_{\text{min}}$ , for  $F_{\text{min}}$  defined by your target planet and stellar types, and your survey specifications.
6. You carry out a transit survey, and detect many transiting planets.

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 memo,  $\gamma_R$  is assumed to be a constant across the population of star systems. The ratio of the two number densities in a volume-limited sample is the binary fraction<sup>1</sup>.

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

- 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.<sup>2</sup>

Consider the following set of 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, 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?
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.

<sup>2</sup> We’ll see that it’s an extreme limit to assume the primary and secondary of binary systems host planets at equal rates if the secondary is non-identical to the primary.

### 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 can be written

$$N_{\text{stars}} = N_s + 2N_d. \quad (1)$$

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,

$$N_i = n_i \frac{4\pi}{3} d_{\text{max},i}^3, \quad (2)$$

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

$$\frac{n_d}{n_s} = \text{binary fraction} \equiv \text{BF} \quad (3)$$

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

$d_{\text{max},i}$  in Eq. 2 is the maximum distance corresponding to the given magnitude limit:

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

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

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

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. 4, 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, assuming  $\gamma_R$  is a constant for the entire population, is

$$\frac{N_d}{N_s} = \frac{n_d}{n_s} \left( \frac{d_{\text{max},d}}{d_{\text{max},s}} \right)^3 \quad (6)$$

$$= \text{BF} \times (1 + \gamma_R)^{3/2}. \quad (7)$$

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. 2, 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.

#### 1.1. Actually finding $F_{\text{lim}}$

To express  $F_{\text{lim}}$  in terms of the survey's defining parameters, we need to convert between  $(S/N)_{\text{lim}}$  and  $F_{\text{lim}}$ . We proceed as follows.

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

$$S = \delta \mathcal{D} \quad (8)$$

$$= \left( \frac{R_p}{R_\star} \right)^2 \mathcal{D}, \quad (9)$$

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

$$\mathcal{D} = \begin{cases} L_1/L_d, & \text{if binary and target primary} \\ \gamma_R L_1/L_d, & \text{if binary and target secondary} \\ 1, & \text{if single,} \end{cases} \quad (10)$$

where  $L_1$ ,  $L_d$ , and  $\gamma_R$  were defined in the opening monograph.

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

$$N = \frac{1}{\sqrt{N_\gamma}}, \quad (11)$$

for  $N_\gamma$  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

$$N_\gamma = F_\gamma^N A N_{\text{tra}} T_{\text{dur}}, \quad (12)$$

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

Thus the signal to noise ratio can be written

$$S/N = \delta \mathcal{D} \sqrt{F_\gamma^N A N_{\text{tra}} T_{\text{dur}}}. \quad (13)$$

In passing, given the parameters that define a survey and planet type, Eq. 13 would need to be re-expressed with  $N_{\text{tra}}$  roughly the ratio of the observing baseline to the planet period, and  $T_{\text{dur}}$  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.

Using Eq. 13, we can write the minimum number flux of photons required for a detection at threshold as

$$F_{\text{lim}}^N = \left[ \left( \frac{S}{N} \right)_{\text{min}} \frac{1}{\delta \mathcal{D}} \right]^2 \frac{1}{A N_{\text{tra}} T_{\text{dur}}}. \quad (14)$$

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

### 2. HOW MANY PLANETS ARE IN THE SAMPLE?

The number of planets in the sample is

$$N_{\text{planets}} = N_{\text{planets in single star systems}} + \quad (15)$$

$$N_{\text{planets in double star systems}} \\ = \Gamma_{t,s} N_s + 2 \Gamma_{t,d} N_d. \quad (16)$$

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

### 3. WHAT IS THE TRUE OCCURRENCE RATE?

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

$$\Gamma_t = \frac{N_{\text{planets}}}{N_{\text{stars}}} \quad (17)$$

$$\Gamma_t = \frac{\Gamma_{t,s}N_s + 2\Gamma_{t,d}N_d}{N_s + 2N_d}. \quad (18)$$

#### 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. Since we selected stars for which there was enough light to make detections, there is no need to compute the S/N distribution of “threshold-crossing events”.

The number of planets detected in single star systems is

$$N_{\text{det},s} = N_s \Gamma_{t,s} f_{s,\text{geom}}, \quad (19)$$

where the product  $N_s \Gamma_{t,s}$  is the number of planets in the single star systems of the sample, and  $f_{s,\text{geom}}$  is the geometric transit probability. The number of planets detected in double star systems is

$$N_{\text{det},d} = 2N_d \Gamma_{t,d} f_{d,\text{geom}}, \quad (20)$$

where now  $2N_d \Gamma_{t,d}$  is the number of planets in the double star systems of the sample. The number of detected planets  $N_{\text{det}}$  is the sum of the two previously written equations.

#### 5. ASTRONOMER A IGNORES BINARITY

Astronomer A has never heard of binary star systems. What occurrence rate does he derive for planets of radius  $R_p$  and period  $P$ ?

The *total* occurrence rate (number of planets divided number of “stars”) for Astronomer A would be  $(N_{\text{det}}/f_{s,\text{geom}})/(N_s + N_d)$ . 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  $\sqrt{D}$ . The question asks what occurrence rate is derived for planets of radius  $R_p$  and period  $P$ . The answer is

$$\Gamma_{A, \text{ planets of } R_p} = \frac{N_{\text{det},s}/f_{s,\text{geom}}}{N_s + N_d} \quad (21)$$

$$= \frac{\Gamma_{t,s}N_s}{N_s + N_d} \quad (22)$$

This astronomer will also think there is a second population of planets, with radius  $R_p\sqrt{D}$  (a constant number for  $\gamma_R$  = constant). He will then also claim have derived a second occurrence rate,

$$\Gamma_{A, \text{ planets of } R_p\sqrt{D}} = \frac{N_{\text{det},d}/f_{d,\text{geom}}}{N_s + N_d}, \quad (23)$$

where at least for the twin binary case the geometric completeness term is the same as for Eq. 21.

#### 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,

$$\Gamma_{B, \text{ planets of } R_p} = \frac{N_{\text{det},s}/f_{s,\text{geom}}}{N_s + 2N_d}, \quad (24)$$

and

$$\Gamma_{B, \text{ planets of } R_p\sqrt{D}} = \frac{N_{\text{det},d}/f_{d,\text{geom}}}{N_s + 2N_d}. \quad (25)$$

#### 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. Since there are no completeness corrections, she has everything she needs.

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

$$\Gamma_{C, \text{ planets of } R_p} = \frac{N_{\text{det},s}/f_{s,\text{geom}} + N_{\text{det},d}/f_{d,\text{geom}}}{N_s + 2N_d} \quad (26)$$

$$= \frac{\Gamma_{t,s}N_s + 2\Gamma_{t,d}N_d}{N_s + 2N_d}. \quad (27)$$

With  $\approx$  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, Astronomer C has found the true occurrence rate (cf. Eq. 18).

#### 8. NUMERICAL VERIFICATION

As a check on the preceding analytic discussion, 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, binary light ratio, 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. 13) 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 derived in an earlier version of this memo are correct.

A point evident in Fig. 1 is that, for fixed planet parameters, and fixed stellar parameters ( $R_*$ ,  $L_*$ , and distance  $r$ ), and a constant  $\gamma_R$  across the population, the SNR distribution for planets in binaries is poorer than that of planets in single star systems. We can see analytically that this can be written only as function of the binary light ratio:

$$\frac{\text{prob}(x_d)}{\text{prob}(x_s)} = (1 + \gamma_R)^{-1}. \quad (28)$$

Deriving this simple form requires noting that the ratios of the bandpass-specific number luminosities is equal

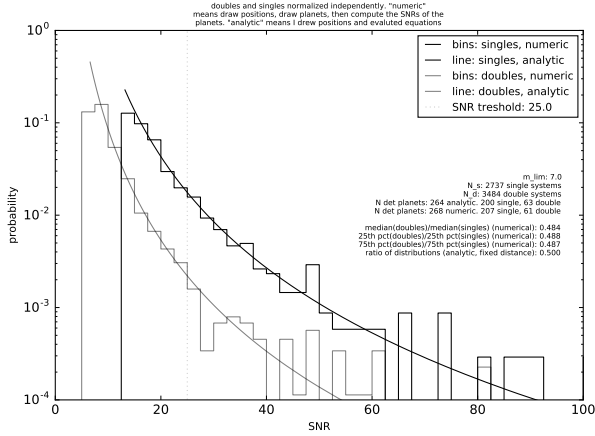


FIG. 1.— Comparison of analytic and numeric probability density functions of the SNR in an idealized transit survey. The analytic lines are from a previous version of this memo (20170728) for the planet populations orbiting single and binary stars. The underlying stepped histogram is output from Monte Carlo simulations. Poisson noise leads to a small deviation at the faint and bright limits, but the numerics and analytics otherwise agree.

to the ratio of the bandpass-specific energy luminosities (otherwise a term with  $c_s/c_d$  must be included).

## 9. REPRESENTATIVE NUMBERS FOR A FEW CASES

### 9.1. 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  $R_p\sqrt{D}$ . The fraction of detections with misclassified radii can then be written

$$\frac{N_{\text{det},d}}{N_{\text{det},s} + N_{\text{det},d}} = \frac{1}{1 + \alpha}, \quad (29)$$

for

$$\alpha \equiv \frac{1}{2(\text{BF})} (1 + \gamma_R)^{3/2} \frac{\Gamma_{t,s}}{\Gamma_{t,d}}. \quad (30)$$

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.

### 9.2. 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.

For brevity, write  $\Gamma_{A, \text{planets of } R_p} = \Gamma_{A,R_p}$ , and similarly for D. Then the relative difference between the two occurrence rates is

$$\left| \frac{\Gamma_{A,R_p} - \Gamma_{D,R_p}}{\Gamma_{A,R_p}} \right| = \left| 1 - \frac{\Gamma_{D,R_p}}{\Gamma_{A,R_p}} \right| \quad (31)$$

$$= \left| 1 - \left( \frac{\Gamma_{t,s}N_s + 2\Gamma_{t,d}N_d}{N_s + 2N_d} \cdot \frac{N_s + N_d}{\Gamma_{t,s}N_s} \right) \right| \quad (32)$$

$$= \left| 1 - \frac{(1 + 2\beta\Gamma_{t,d}/\Gamma_{t,s})(1 + \beta)}{(1 + 2\beta)} \right|, \quad (33)$$

for

$$\beta \equiv N_d/N_s = \text{BF} \times (1 + \gamma_R)^{3/2}. \quad (34)$$

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  $\Gamma_{D,R_p} = 0.5$ . For the more “realistic” case of  $\gamma_R = 0.1$ ,  $\Gamma_{t,d} = \Gamma_{t,s}$ , the relative error is 52%. Not as bad, and this latter number should be looked at skeptically because the “equal occurrence rates for primary and secondary” assumption may be unlikely with this light ratio as a median.