Draft version
Preprint typeset using LaTeX style emulateapj v. 12/16/11

THE EFFECTS OF STELLAR BINARITY ON TRANSIT SURVEY OCCURRENCE RATES

LGB, KM, JNW

Draft version

ABSTRACT

What errors does ignoring binarity introduce on the occurrence rates derived from transit surveys?

Subject headings: planets and satellites: detection

1. INTRODUCTION

An astronomer who does not believe in binaries wants to measure the occurrence of planets of a certain type around stars of a certain type. In other words, they wish to find the mean number of planets within specific planetary (R_p, P) bounds orbiting the stars in a given volume of stellar $(M_\star, R_\star, L_\star)$ phase-space. They perform a signal-to-noise limited transit survey and detect $N_{\rm det}$ transit signals that appear to be from planets of the desired type. They calculate the number of stars $N_\star$ that appear to be searchable for the desired type of planet, and also of the desired stellar type. These "searchable stars" are the points on the sky for which they think planets are observed with 100% detection efficiency. Accounting for the geometric transit probability $f_{\rm g}$, they compute an apparent occurrence rate $\Gamma_{\rm a}$:

$$\Gamma_{\rm a} = \frac{N_{\rm det}}{N_\star} \times \frac{1}{f_g}. \quad (1)$$

There are many potential pitfalls. Some genuine transit signals can be missed by the detection pipeline. Some apparent transit signals are spurious, from noise fluctuations, failures of 'detrending', or instrumental effects. Stars and planets can be misclassified due to statistical and systematic errors in the measurements of their properties. Poor angular resolution causes false positives due to blends with background eclipsing binaries. *Et cetera.*

Here we focus on problems that arise from the fact that stars of the desired type often exist in binaries. We assume for simplicity that all binaries in the transit survey are spatially unresolved.

One immediate complication is that – due to dynamical stability or some aspect of planet formation – the occurrence rate of planets of the desired type may differ between binary and single-star systems. If "occurrence rate" is defined purely as the mean number of planets within set radius and period bounds per star in a given interval $\Delta M_\star, \Delta R_\star, \Delta L_\star$, it must include an implicit marginalization over stellar multiplicity. Otherwise, one must discuss "occurrence rates in single star systems", "occurrence rates about primaries of double systems", and "occurrence rates about secondaries of double star systems".

Outside of bonafide astrophysical differences, there are observational biases. A given apparently-searchable star may truly be a single star of the desired type. If not, and ignoring higher order multiples, one of the following must be true:

1. The apparently-searchable star is a binary system, with two searchable stars of the desired type. $N_\star$ is under-counted by one for every such system.
2. The apparently-searchable star is a binary system, with one searchable star of the desired type. The searchable star of the desired type could be the primary or the secondary. $N_\star$ is correctly counted for every such system, though there may be systematic errors in estimates of stellar parameters.
3. The apparently-searchable star is a binary system, in which neither of the components is a searchable star, but at least one is of the desired type. $N_\star$ is over-counted by one for every such system.

One might also imagine an apparently-searchable star which in no stellar component is of the desired type. We will not consider errors of this type. We will assume that all the apparently-searchable stars are either single stars of the desired type, or binaries in which either the primary or else both components are of the desired type.

The above enumerates possible errors when counting $N_\star$. When tallying $N_{\rm det}$, the detected signals from planets of the desired type, neglecting binarity introduces the following error cases:

1. A detected signal is from a star of undesired type (by the above assumptions, the secondary of a binary system), from a planet of undesired type, and is incorrectly counted as a planet of desired type.
2. A detected signal is from a star of desired type (the primary or secondary of a binary systems), from a planet of undesired type, and is incorrectly counted as a planet of desired type. This and the case above occur when the host star parameters are miscalculated (*e.g.,* the host star is the faint secondary, but is assumed to be primary), or when the constant diluting light from the binary companion is not corrected. $N_{\rm det}$ should be lowered by 1 for every such case.
3. A detected signal is from a star of desired type, from a planet of desired type, but incorrectly counted as planet of undesired type. $N_{\rm det}$ should be raised by 1 for every such case.
4. An *undetected* signal from a planet of the desired type, around a star of the desired type, was not detected because of dilution from the companion (because the SNR floor of the survey is set for single stars). This happens during case #3 in $N_\star$-counting errors (the apparently-searchable star is

a binary system, in which neither component is in fact searchable). If $N_\star$ is corrected for this, *i.e.,* the system is correctly counted as "not searchable", no correction to $N_{\rm det}$ is needed.

Errors (1)-(3) in calculating $N_{\rm det}$ above broadly fall under "radius misclassification", while error (4) falls under "completeness miscalculation". Note that in the above list, a planet cannot be both "of desired type" and "orbiting a star of undesired type". We ignore this type of error for analytic convenience, but `reintroduce it later during numerics`.

Our approach is step-by-step starting from a very simple scenario.

- We consider a universe consisting of only the desired types of stars and planets. We give analytic expressions for eta(apparent)/eta(true) taking into account all applicable errors.
- We allow for a distribution of the mass ratio $dn/dq \propto q^\gamma$, and a concurrent power-law distribution in the luminosity ratio, $dn/dL \propto L^\beta$. We give analytic expressions for eta(apparent)/eta(true) taking into account errors of type 2.
- We allow for a nonzero tolerance ΔR_p in the planet radius, and tolerance $\Delta L_\star$ in stellar luminosity, to qualify as "desirable". We recalculate eta(apparent)/eta(true). (LB: I think this has to become numerical to actually evaluate the occurrence rate correction factor, X_Γ.)
- We assume the planets have a power-law radius distribution, $d\eta/dR_p \propto r^\alpha$. We calculate the apparent $d\eta/dR_p$ taking into account errors of type 1-3. We do so for stars of the desired type within some tolerance dL.
- Everything up to now is more-or-less analytic, perhaps supported by numerical checking or numerical integration. At the end we summarize the radius-dependence of the occurrence rate shifting. "Accounting only for dilution & its numerator effect, big radius planets have underestimated inferred rates, the smallest radius planets have overestimated inferred rates, and the in-between radius planets are in-between"

We then dig in, w/ numerics, to what happens with different assumed completeness vs Rp functions. We note that the the claimed HJ occ rate discrepency btwn RV & transit surveys gets much weaker when accounting for binarity. This is b/c the more believable occ rate papers do not heavily rely on their completeness calibration – they know they're complete. This is a much more understandable regime than the wonky completeness estimate dependent zone of eta-Earth for Kepler.

Finally, for numerics, we simulate a sample of 20,000 apparently-searchable-stars (or however many "stars" appear to be searchable for Earthlike planets in the Kepler field). Basically this means a magnitude limited survey of 20,000 "stars" that are either single or are binaries with the primary of the desired type. The sample has a realistic distribution of binaries (occurrence and luminosity ratio). We calculate eta(apparent)/eta(true). We can also play with errors of type 4 in this numerical model (i.e. we let the relative occurrence rates between primaries and secondaries vary). (It may not worth doing so for the analytic work– though it's pretty simple. The way to play with this is just present both limits, and then also present a few weight factors (Furlan+17 do a good job).

We discuss "how do we interpret Fulton+ 2017?" That paper did (or claims to have done) a good job at weeding out numerator errors. If true, means $\Gamma(R_{\rm p})$ has the right shape, and just the wrong normalization. Important if we ever manage to convince ourselves "brightest transit hosts" could work.

In passing, at the end, we mention how this changes attempts at eta-Earth measurements.

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

Stellar binarity introduces convoluted effects on transit surveys. To disentangle them, imagine the following idealized survey.

You observe the entire sky for some duration with a detector of known area and bandpass. Your detector is photon-noise limited. You are interested only in detecting planets of radius R_p, and orbital period P. For instance, $R_p = R_\oplus$, $P = 1$ year. You are only interested in detecting them around stars of radius R_1, and luminosity L_1. For instance, G2V dwarfs. You can only detect your desired type of planet when you observe signals with $\mathrm{S/N} > (\mathrm{S/N})_{\rm min}$. For a photon-noise limited survey, this minimum signal to noise ratio is equivalent to a minimum flux required for detection, $F_{\rm min}$.

You get funding, and perform the S/N-limited survey. You observe all the "points" on-sky with apparent magnitude $m < m_{\rm min}$, or equivalently with energy flux in your bandpass greater than some limit $F > F_{\rm min}$. All planets with $\mathrm{S/N} > (\mathrm{S/N})_{\rm min}$ 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 only single and double star systems. Single star systems have luminosity in the observed bandpass L_1, radii R_1, and effective temperature $T_{\rm eff,1}$. 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, $\gamma_R = 1$ across the population of star systems – all binaries are assumed to be twin systems. The ratio of the number density of binary systems and the number density of all systems in a volume-limited sample is the binary fraction[1].
- The true population of planets around these stars is as follows:
  - A fraction $\Gamma_{t,s}$ of stars in single star systems have planets of radius R_p, with orbital period P.

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

– A fraction $\Gamma_{t,d}$ of primary stars in double star systems have planets of radius R_p, with orbital period P. A fraction $w_{d2}\Gamma_{t,d}$ of secondary stars in a double star systems have the same type of planet, where the weight factor w_{d2} specifies the secondary star occurrence rate as a fraction of the primary star occurrence rate. Any astrophysical difference in planet formation between singles and binaries is captured by $\Gamma_{t,s}, \Gamma_{t,d}$, and w_{d2}. If one assumes the primary and secondary of a binary system host planets at equal rates, this is the $w_{d2} = 1$ limit. Similarly, if one assumes secondaries cannot host planets, this is the $w_{d2} = 0$ limit.

The assumption of [Pepper et al, 2003] in a very similar context was that the observer correctly pre-selects all of the "searchable stars". This simplification would allow the observer to ignore completeness effects when deriving occurrence rates, since their completeness is 100%.

The above approach ignores 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 maps onto three distinct maximum detection distances for a population of single stars and twin binaries:

1. $d_{\max,s}$: the maximum distance out to which the desired type of planet is detectable around single stars. These are the selected "searchable, single stars".

2. $d_{\max,d}$: the maximum distance out to which double stars are selected (down to the same minimum flux as that defining $d_{\max,s}$). This is different from $d^{\rm p}_{\max,d}$, the maximum distance out to which planets are detectable around double stars. For a population with fixed γ_R, since S/N $\propto \mathcal{D}L^{1/2}d^{-1}$,

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

For $\gamma_R = 1$, this means only 1 in 8 apparently-searchable stars which are in fact binary systems can yield detectable planets (the ratio of volume in which planets can be detected to the actual selected volume). For a more realistic sample, $\gamma_R \approx 0.1$, and there is an interesting asymmetry. One finds the completeness for planets which orbit the primary star, $f_{d1,c}$ to be

$$f_{d1,c} = (1+\gamma_R)^{-3} \approx 3/4. \quad (3)$$

However, for planets which orbit the secondary star, the completeness is

$$f_{d2,c} = (1+\gamma_R^{-1})^{-3} \approx 3/4000. \quad (4)$$

We note in passing that we are considering a SNR-limited survey, which in this case is the same as a magnitude-limited one. The observer might prefer to perform a completeness-limited survey, which in this case would be the same as a volume-limited one with limiting distance $d^{\rm p}_{\max,d}$. This change would likely improve interpretability.

With the stage set, we ask:

2.1. *What is the true occurrence rate?*

The "true occurrence rate" Γ_t is the average number of planets of the desired type per star of the desired type:

$$\Gamma_t = \frac{N_{\rm planets}}{N_{\rm stars}} \quad (5)$$

$$\Gamma_t = \frac{\Gamma_{t,s}N_s + (1+w_{d2})\Gamma_{t,d}N_d}{N_s + 2N_d}. \quad (6)$$

where N_s is the number of single star systems, and N_d the number of double star systems. The factor of 2 in the total number of stars accounts for the fact that there are twice as many stars in double star systems.

To develop more analytic expressions in terms of observables, assume that the stars are homogeneously distributed[2]. Counting the number of single and double systems in their respective selected volumes,

$$N_i = n_i \frac{4\pi}{3} d^3_{\max,i}, \quad (7)$$

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

$$\frac{n_d}{n_s + n_d} = \text{binary fraction} \equiv \text{BF} \quad (8)$$

The absolute normalization of the number density $n_s + n_d$ is an observed quantity [Bovy 2017], as is the binary fraction. For "sun-like" dwarfs with $0.7 < M_\star/M_\odot < 1.3$, BF $= 0.44 \pm 0.02$ [Duchene & Kraus, 2013, Raghavan et al 2010]. Note that these latter authors really quote a *multiplicity* fraction of 0.44 for sun-like dwarfs – we count the $\approx$ 25% of multiple systems that are higher order multiples as doubles.

$d_{\max,i}$ in Eq. 7 is the maximum distance corresponding to the given magnitude limit:

$$d_{\max,i} = \left(\frac{L_i}{4\pi F_{\min}}\right)^{1/2}, \quad (9)$$

where the limiting flux in the bandpass $F_{\min}$ could equivalently be stated in terms of a limiting magnitude $m_{\min}$. In Eq. 9, again $i \in \{\text{single}, \text{double}\}$. Simply as a consequence of imposing a magnitude cut, the maximum distance to which binary stars will be selected is greater than that of single stars. The ratio of double to single systems in the observed SNR-limited sample is then

$$\frac{N_d}{N_s} = \frac{n_d}{n_s}\left(\frac{d_{\max,d}}{d_{\max,s}}\right)^3 \quad (10)$$

$$= \frac{\text{BF}}{1-\text{BF}} \times (1+\gamma_R)^{3/2}. \quad (11)$$

In the case of twin binaries ($\gamma_R = 1$), with a binary fraction BF $= 0.5$, there are $2^{3/2} \approx 2.8$ times more binary

[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.

systems than single systems in the SNR-limited sample. Since the completeness fraction for binary systems is 1/8, there are $2^{-3/2} \approx 0.35$ times fewer binary systems in the sample that can yield planets as single star systems in the system that can yield planets.

2.2. *What is the measured occurrence rate?*

The observer who has never heard of binary star systems wants to find the occurrence rate of planets with radius R_p and period P, orbiting stars of radius R_1 and luminosity L_1. They detected $N_{\rm det}$ transit signals that appear to be from planets of this desired type. The occurrence rate they then find, *i.e.* the number of planets divided by number of "stars" (really, stellar systems) is

$$\Gamma_{\rm a} = \frac{N_{\rm det}}{N_s + N_d} \times \frac{1}{f_g}. \quad (12)$$

Since only the transit depth is observed, the planetary radii apparent to this astronomer, $R_{\rm p,a}$, are different from the true planetary radii $R_{\rm p,t}$. This is because the assumed stellar radii $R_{\star,\rm a}$ may differ from the true stellar radii $R_{\star,\rm t}$, and the flux will be diluted:

$$R_{p,\rm a} = R_{p,\rm t} \frac{R_{\star,\rm a}}{R_{\star,\rm t}} \mathcal{D}^{1/2}, \quad (13)$$

where the dilution parameter $\mathcal{D}$ is defined for binary systems as

$$\mathcal{D} = \begin{cases} L_1/L_d, & \text{if planet orbits primary} \\ \gamma_R L_1/L_d, & \text{if planet orbits secondary} \end{cases}$$
$$= \begin{cases} (1+\gamma_R)^{-1}, & \text{if planet orbits primary} \\ (1+\gamma_R^{-1})^{-1}, & \text{if planet orbits secondary} \end{cases} \quad (14)$$

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

In our twin binary thought-experiment, dilution means that all planets detected in binary systems will have apparent radii of $R_p/\sqrt{2}$. Since for twin binaries the stellar radii are the same, the transit probability will be correctly calculated.

The total number of planet detections, irrespective of whether they are "of the desired type", is the sum of the number of planets detected in single star systems $N_{{\rm det},s}$ and the number of planets detected in double star systems $N_{{\rm det},d}$. Expressing each individually,

$$N_{{\rm det},s} = N_s \Gamma_{t,s} f_{s,g} f_{s,c}, \quad (15)$$

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 transit probability, and $f_{s,c}$ is the fraction of these transiting planets that are observed with signal to noise greater than the minimum detection threshold (the completeness). For our SNR-limited, twin-binary survey, $f_{s,g} = R_1/a \equiv f_g$, and $f_{s,c} = 1$.

Analogously,

$$N_{{\rm det},d} = (1 + w_{d2}) N_d \Gamma_{t,d} f_{d,g} f_{d,c}, \quad (16)$$

where now $(1+w_{d2})N_d\Gamma_{t,d}$ is the number of planets in the double star systems of the sample, the geometric transit probability $f_{d,g} = R_1/a$ for our twin-binary survey, and $f_{d,c} = 1/8$ by the geometric argument presented in the initial discussion of limiting distances.

The occurrence rate this astronomer derives for planets of assumed radius R_p orbiting stars thought to be of the desired type is then

$$\Gamma_{\rm a} = \frac{\Gamma_{t,s} N_s}{N_s + N_d}. \quad (17)$$

2.3. *Errors on derived occurrence rates for twin-binary thought experiment*

The ratio of the true to calculated occurrence rates for planets of radius R_p, X_Γ, is

$$X_\Gamma = \frac{\Gamma_t}{\Gamma_{\rm a}} \quad (18)$$
$$= \frac{\Gamma_{t,s} N_s + (1+w_{d2})\Gamma_{t,d} N_d}{\Gamma_{t,s} N_s} \cdot \frac{N_s + N_d}{N_s + 2N_d} \quad (19)$$
$$= \left(1 + \frac{(1+w_{d2})\Gamma_{t,d}}{\Gamma_{t,s}}\beta\right) \cdot \frac{1+\beta}{1+2\beta}, \quad (20)$$

for

$$\beta \equiv N_d/N_s = \frac{\rm BF}{1-{\rm BF}} \times (1+\gamma_R)^{3/2}. \quad (21)$$

For the nominal G2V dwarf case of BF = 0.44, if we assume that all stars have equal occurrence rates, $\Gamma_{t,s} = \Gamma_{t,d}$ and $w_{d2} = 1$, then we find $X_\Gamma = 1 + \beta = 1 + 2^{3/2} \approx 4 -$ the observed occurrence rate is underestimated by a factor of 4.

For the planets around double stars with observed radii $R_p\sqrt{\mathcal{D}}$, there is also the error of their misclassification. The fraction of detections with misclassified radii can be written

$$\frac{N_{\rm det,d}}{N_{\rm det,s} + N_{\rm det,d}} = \frac{1}{1+\zeta}, \quad (22)$$

for

$$\zeta \equiv \frac{N_s \Gamma_{t,s} f_{s,g} f_{s,c}}{(1+w_{d2}) N_d \Gamma_{t,d} f_{d,g} f_{d,c}}. \quad (23)$$

For our nominal G2V dwarf case, the twin binaries with equal occurrence rates gives $\zeta = 4/\beta = \sqrt{2}$, so 41% of detections have misclassified radii, all of $R_p/\sqrt{2}$.

Of course more generally, the observer might define their desired "type of planet" to be anything within some margin of R_p. Considering this for the idealized twin-binary case, if the margin includes the planets in binary systems, the number of detections "of the desired type" would change. The correction factor X_Γ would then become

$$X_\Gamma = \frac{\Gamma_{t,s} + (1+w_{d2})\Gamma_{t,d}\beta}{\Gamma_{t,s} + (1+w_{d2})\Gamma_{t,d} f_{d,c}} \cdot \frac{1+\beta}{1+2\beta}, \quad (24)$$

which simplifies to $(1+2f_{d,c})^{-1}(1+\beta) = 4(1+\beta)/5 \approx 3$ for our twin-binary thought experiment. This makes sense: increasing the radius interval of "desired planets" decreases the error in the numerator, but the error of mis-counted stars in the denominator remains identical.

3. MODEL #2: FIXED PRIMARIES, FIXED PLANETS, VARYING LIGHT-RATIO SECONDARIES

The model of Sec. 2 is idealized, but it helps develop intuition. Now, we let the light ratio $\gamma_R = L_2/L_1$ vary across the population of star systems. 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.

One parametrization for the distribution of binary mass ratios in a volume-limited sample is $f(q) \propto q^\gamma$. [Duchene and Kraus 2013], fitting all the multiple systems of [Raghavan et al (2010)]'s Fig 16, quote $\gamma = 0.28 \pm 0.05$ for $0.7 < M_\star/M_\odot < 1.3$. Examining only the binary systems of Rhagavan et al 2010, Fig 16, the distribution is roughly uniform,

$$\mathrm{prob}(q) = \begin{cases} c_q & 0.1 < q \leq 1 \\ 0 & \text{otherwise,} \end{cases} \tag{25}$$

provided we ignore the peak at high mass ratios for analytic convenience. $c_q = 1/9$ is the normalization. We choose this latter mass ratio distribution.

For the mass-luminosity relation, for analytic convenience we assume $L \propto M^\alpha$, with the lore-value of α being 3.5. As a simple check, we fit lines to mass-luminosity data[3]. In log-log space, we let the intersection point of the lines float. The resulting fit parameters are available in a footnote[4]. Though the data show a break at $M_\star \approx 0.5 M_\odot$, the $L = M^{3.5}$ relation is a reasonable fit for $M \lesssim 2M_\odot$.

The final addition to our model is that we must assume a stellar mass-radius relation. This affects the transit probabilities and signal sizes about the secondaries of binary star systems. Since again, our main purpose is to develop intuition about observational biases, we approximate this as

$$R \propto M. \tag{26}$$

In passing, we note that while $w_{d2} = 1$ was a reasonable assumption in Sec. 2 (since secondaries were identical to primaries), in binary systems we now have secondaries with different stellar properties from primaries. Recall that our occurrence rate is defined for planets of a given type orbiting stars of a given type. Assume our universe is one in which the primary mass of all stars is $M_1 = M_\odot$. If this is the case, we might define our "stars of interest" to be those with $0.7 < M_\star/M_\odot < 1.3$. Then, in a volume-limited sample, 1 in 3 secondaries would be "of interest", and could contribute to the true occurrence rate. The general point is that w_{d2} will be less than 1, unless planets of the desired type are more common around secondaries of lower mass (as may actually be true for small planets around M dwarfs). Keeping this in mind, we ask the same questions as for Model #1.

3.1. *What is the true occurrence rate?*

The true occurrence rate is the average number of planets of the desired type per star of the desired type. The number of planets of the desired type is

$$N_{\mathrm{planets}} = \Gamma_{t,s} N_s + \Gamma_{t,d} N_d + \Gamma_{t,d} w_{d2} f_{\mathrm{desired}} N_d, \tag{27}$$

where as before N_s and N_d are the number of single and double systems, and the fraction f_{desired} is defined so that $f_{\mathrm{desired}} N_d$ is the number of secondaries that are of the desired type, and as before $\Gamma_{t,d} w_{d2}$ is the average number of planets each of these desired stars hosts.

Counting the number of desired stars, N_s is the same as in Sec. 2. However

$$N_d(\gamma_R) = n_d \frac{4\pi}{3} d^3_{\mathrm{max,d}} \tag{28}$$

for

$$d_{\mathrm{max,d}} = d_{\mathrm{max,s}} \times (1 + \gamma_R)^{1/2}, \tag{29}$$

where $d_{\mathrm{max,s}}$ is given by Eq. 9.

For a given binary system, q, γ_R, and thus $d_{\mathrm{max,d}}(\gamma_R)$ are all random variables. Since N_d is a function of $d_{\mathrm{max,d}}$, the number of double star systems becomes a random variable, drawn from the ensemble of all possible transit surveys. Applying the chain rule for probability density functions, the distribution function for the number of binary systems, prob(N_d), can be written

$$\mathrm{prob}(N_d) = \mathrm{prob}(q(\gamma_R)) \left|\frac{\mathrm{d}q}{\mathrm{d}\gamma_R}\right| \left|\frac{\mathrm{d}\gamma_R}{\mathrm{d}d_{\mathrm{max,d}}}\right| \left|\frac{\mathrm{d}d_{\mathrm{max,d}}}{\mathrm{d}N_d}\right|. \tag{30}$$

Necessary aside on biases in mass ratio distributions— Before blindly plugging in Eq. 25's assumed mass ratio distribution to find the expected number of double systems, we must emphasize that Eq. 25's distribution applies only for volume-limited samples. If one is interested in a SNR-limited transit survey (or really any magnitude-limited survey), the Malmquist-like bias associated with binarity must be included.

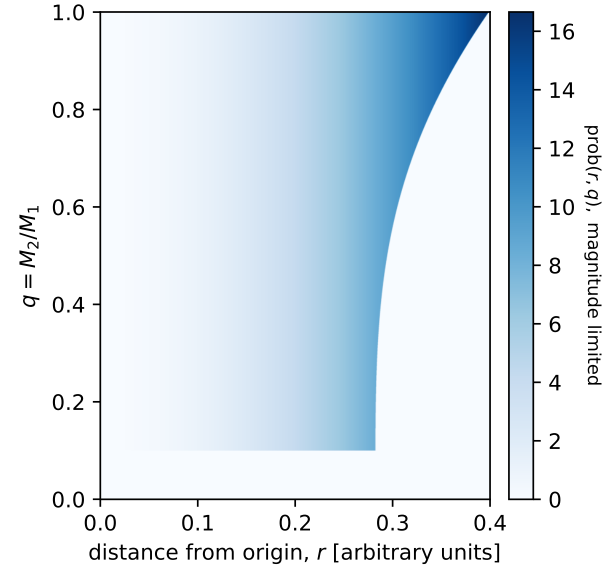

FIG. 1.— The joint probability distribution of a binary star's position and mass ratio in a magnitude-limited sample. This plot assumes $L \propto M^{3.5}$. Note that the marginalized mass ratio distribution is uniform at any given distance.

[3] Data collected by Torres et al. [2009] for dwarfs earlier than spectral class M, and Benedict et al. [2016] for dwarfs later than spectral class M. To convert Benedict et al. [2016]'s reported M_V values to absolute luminosities, we interpolated over E. Mamajek's table `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.

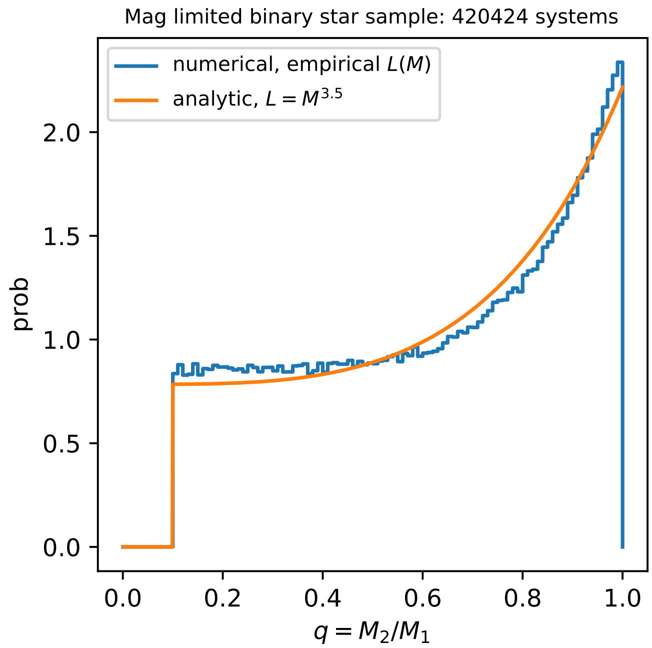

FIG. 2.— The distribution of the mass ratio for a magnitude limited sample of binary stars. The underlying mass ratios are drawn from Eq. 25 – a *uniform* distribution in a volume-limited sample. The entire bias can be understood analytically (Eq. 33) to good precision. The numerical comparison uses the empirical mass-luminosity relation shown in Fig. **??**.

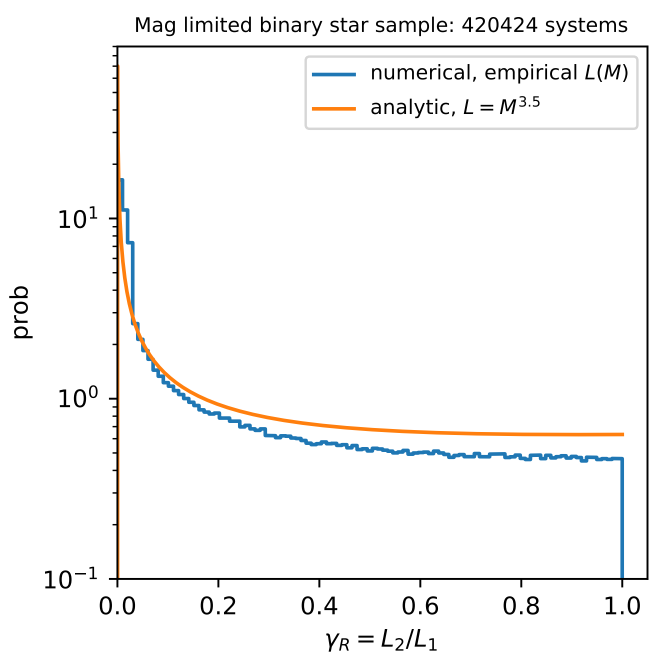

FIG. 3.— The distribution of the luminosity ratio for a magnitude limited sample of binary stars. Despite the preference in the sample towards large q binaries (Fig. 2), the bias is towards small γ_R binaries because of the steepness of the mass-luminosity relation. The underlying mass ratios are uniform in volume. The numerical comparison uses the empirical mass-luminosity relation shown in Fig. **??**.

The distribution function for the mass ratio of binaries in a magnitude limited sample can be found by marginalizing over the joint distribution for a binary star's position r and mass ratio q. Explicitly, using 'ml' as a subscript for 'magnitude-limited',

$$\mathrm{prob}_{\mathrm{ml}}(r, q) = \mathrm{prob}(q|r)\mathrm{prob}_{\mathrm{ml}}(r), \tag{31}$$

where $\mathrm{prob}(q|r)$ is given by Eq. 25. One can show

$$\mathrm{prob}_{\mathrm{ml}}(r, q) \propto r^2 \quad \text{if } r < d_{\mathrm{max,d}}(q) \text{ and } 0.1 < q < 1, \tag{32}$$

and otherwise zero. The resulting joint probability distribution is plotted in Fig. 1. Marginalizing the joint distribution over distance, and assuming $L \propto M^\alpha$, one finds

$$\mathrm{prob}_{\mathrm{ml}}(q) \propto (1 + q^\alpha)^{3/2} \quad \text{if } 0.1 < q \leq 1, \tag{33}$$

and otherwise 0. This marginalized distribution – the distribution function of mass ratios for binaries in a magnitude-limited sample – is plotted in Fig. 2. Reworded, the probability of drawing a binary of mass ratio q in a magnitude-limited sample scales as $d^3_{\mathrm{max,d}}(q)$.

Applying Eq. 30 with the correct magnitude-limited distributions gives

$$\mathrm{prob}(N_d) \propto N_d^{2/3} \left(\left(\frac{N_d n_s}{N_s n_d} \right)^{2/3} - 1 \right)^{\frac{1-\alpha}{\alpha}} \tag{34}$$

over the interval

$$N_d^{\mathrm{lower}} = \frac{4\pi n_d}{3} (\sqrt{(0.1)^\alpha + 1} d_{\mathrm{max,s}})^3, \tag{35}$$

$$N_d^{\mathrm{upper}} = \frac{4\pi n_d}{3} (\sqrt{2} d_{\mathrm{max,s}})^3, \tag{36}$$

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

An interesting intermediate result is the magnitude-limited luminosity ratio distribution,

$$\mathrm{prob}_{\mathrm{ml}}(\gamma_R) \propto (1 + \gamma_R)^{3/2} \gamma_R^{\frac{1-\alpha}{\alpha}}, \tag{37}$$

for $(0.1)^\alpha < \gamma_R < 1$, and otherwise zero. For $\alpha = 3.5$, $\mathrm{prob}_{\mathrm{ml}}(\gamma_R) \propto (1 + \gamma_R)^{3/2} \gamma_R^{-5/7}$, which tells us the distribution will be highly biased towards low-light ratio binaries, as shown in Fig. 3.

The reason we derived all this is that we wanted to know the number of searchable binary stars. For any given realization of the survey, this number is a random draw. Computing the *expected value* for the number of double star systems in the sample,

$$\langle N_d \rangle = \int_0^\infty N_d \, \mathrm{prob}(N_d) \, \mathrm{d}N_d. \tag{38}$$

Letting $\alpha = 3.5$, this can be expressed as

$$\langle N_d \rangle \propto \mathcal{R}^{-3/2} \int_{0.1^{7/2}+1}^{2} u^3 (u - 1)^{-5/7} \, \mathrm{d}u, \tag{39}$$

where non-dimensional constants are not shown, $u = \mathcal{R} N_d^{2/3}$, and $\mathcal{R}$ is defined as

$$\mathcal{R} \equiv \left(\frac{3}{4\pi n_d} \right)^{2/3} d^{-2}_{\mathrm{max,s}} = \left(\frac{n_s}{N_s n_d} \right)^{2/3}. \tag{40}$$

The integral in Eq. 39 has a closed-form solution in terms of polynomials. Once evaluated, and after normalizing prob(N_d), to 4 significant digits it gives

$$\langle N_d \rangle = 1.590 \left(\frac{4\pi n_d}{3} \right) d^3_{\mathrm{max,s}}. \tag{41}$$

This gives the useful expression

$$\frac{\langle N_d \rangle}{N_s} = 1.590 \, \frac{n_d}{n_s} \tag{42}$$

Taking the sun-like star binary fraction of BF = 0.44, this gives $\langle N_d \rangle / N_s = 1.25$. Compared to the binary-twin thought experiment of Sec. 2, in which the SNR-limited sample had ≈ 2.8 times more binary systems than single systems, there are now ≈ 1.25 times as many. This is sensible, because the mean light ratio has decreased; it was previously 1, it is now $\langle \gamma_R \rangle \approx 0.27$.

What is the true occurrence rate?— Returning to the original question, the total number of stars for any given realization of the survey is

$$N_{\mathrm{stars,tot}} = N_s + 2N_d. \tag{43}$$

Since N_d is a random number, so is the total number of stars. By assumption all primaries are of the desired type, but for secondaries only a fraction f_{desired} of them are of the desired type:

$$N_{\mathrm{stars}} = N_s + (1 + f_{\mathrm{desired}})N_d. \tag{44}$$

Explicitly,

$$f_{\mathrm{desired}} = \frac{\int_{M_{\mathrm{min}}}^{M_{\mathrm{max}}} \mathrm{prob_{ml}}(q)\,\mathrm{d}q}{\int_{0.1M_\odot}^{M_\odot} \mathrm{prob_{ml}}(q)\,\mathrm{d}q}. \tag{45}$$

For instance, if we define the "desired type" to be $0.7 < M_\star/M_\odot < 1.3$, we find $f_{\mathrm{desired}} \approx 0.475$.

For a given survey draw, the number of planets of the desired type is then

$$N_{\mathrm{planets}} = \Gamma_{t,s}N_s + \Gamma_{t,d}N_d + f_{\mathrm{des}}\Gamma_{t,d}w_{d2}N_d, \tag{46}$$

and the drawn "true occurrence rate" is

$$\Gamma_t = \frac{N_{\mathrm{planets}}}{N_{\mathrm{stars}}} \tag{47}$$

$$\Gamma_t = \frac{\Gamma_{t,s}N_s + (1 + f_{\mathrm{desired}}w_{d2})\Gamma_{t,d}N_d}{N_s + (1 + f_{\mathrm{desired}})N_d}. \tag{48}$$

3.2. *What is the measured occurrence rate?*

The observer who has never heard of binary star systems wants to find the occurrence rate of planets with radius R_p and period P, orbiting stars with mass $0.7 < M_\star/M_\odot < 1.3$ (with corresponding stellar radii and luminosities). They detected N_{det} transit signals that appear to be from planets of this desired type. The occurrence rate they then find, *i.e.* the apparent number of planets divided by the apparent number of "stars" (really, stellar systems) is

$$\Gamma_{\mathrm{a}} = \frac{N_{\mathrm{det}}}{N_s + N_d} \times \frac{1}{f_{s,g}}. \tag{49}$$

Unlike in Sec. 2, where the observer saw signals of only two apparent radii (undiluted and diluted), they will now find signals with a spectrum of apparent radii. The dilutions $\mathcal{D}$ in binary systems, given by Eq. 14, vary depending on which star the planet orbits, and depending on the system's luminosity ratio γ_R. Moreover, the apparent planetary radii will be affected by the assumptions the observer makes about the apparently-single star's radius, $R_{\star,\mathrm{a}}$. We posit that the observer assumes $R_{\star,\mathrm{a}} = R_1$; in other words they assume that the primary star hosts any signal they observe. The apparent planetary radii, $R_{\mathrm{p,a}}$, are then

$$R_{p,\mathrm{a}} = R_p \frac{R_1}{R_{\star,\mathrm{t}}} \mathcal{D}^{1/2}. \tag{50}$$

The drawn number of planet detections of the desired type is

$$N_{\mathrm{det}} = N_{\mathrm{det},s} + N_{\mathrm{det},d}. \tag{51}$$

As in Sec. 2,

$$N_{\mathrm{det},s} = N_s\Gamma_{t,s}f_{s,g}f_{s,c}, \tag{52}$$

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 transit probability, and $f_{s,c}$ is the fraction of these transiting planets that are observed with signal to noise greater than the minimum detection threshold (the completeness). For our SNR-limited, varying-γ_R survey, $f_{s,g} = R_1/a \equiv f_g$, and $f_{s,c} = 1$.

For planets detected in binary systems, only a fraction will be "of the desired type", since only a fraction of the stars are of the desired type. This is the same factor f_{desired} as previously discussed. Accounting for this and the independent geometric and completeness probabilities,

$$N_{\mathrm{det},d} = N_d\Gamma_{t,d}(f_{d1,g}f_{d1,c} + f_{\mathrm{desired}}w_{d2}f_{d2,g}f_{d2,c}). \tag{53}$$

For any given system,

$$f_{d1,c} = (1 + \gamma_R)^{-3} \tag{54}$$

$$f_{d1,g} = f_{s,g} \tag{55}$$

$$f_{d2,c} = (1 + \gamma_R^{-1})^{-3} \tag{56}$$

$$f_{d2,g} = f_{s,g}q^{2/3} \tag{57}$$

3.3. *Errors on derived occurrence rates*

Recall $X_\Gamma = \Gamma_t/\Gamma_a$. Let $f_{\mathrm{desired}} = f_{\mathrm{d}}$ for brevity. Combining the above results, for any draw from the ensemble of transit surveys the necessary correction to the derived occurrence rate is

$$X_\Gamma = \frac{\Gamma_{t,s}N_s + (1 + f_{\mathrm{d}}w_{d2})\Gamma_{t,d}N_d}{N_s + (1 + f_{\mathrm{d}})N_d} \times \frac{(N_s + N_d)f_{s,g}}{N_s\Gamma_{t,s}f_{s,g} + N_d\Gamma_{t,d}(f_{d1,g}f_{d1,c} + f_{\mathrm{d}}w_{d2}f_{d2,g}f_{d2,c})}. \tag{58}$$

Letting $\beta = N_d/N_s$,

$$X_\Gamma = \frac{1 + \beta}{1 + \beta(1 + f_{\mathrm{d}})} \times \frac{1 + \beta(\Gamma_{t,d}/\Gamma_{t,s})(1 + f_{\mathrm{d}}w_{d2})}{1 + \beta(\Gamma_{t,d}/\Gamma_{t,s})(f_{d1,c} + f_{\mathrm{d}}w_{d2}f_{d2,g}f_{d2,c}/f_{s,g})}. \tag{59}$$

One typically assumes that $\Gamma_{t,s} = \Gamma_{t,d}$, and that the fraction of secondaries of the desired type with planets, w_{d2} can be set to 1. If so,

$$X_\Gamma = \frac{1+\beta}{1+\beta(f_{d1,c} + f_{\rm d} f_{d2,g} f_{d2,c}/f_{s,g})}. \tag{60}$$

If $\gamma_R \ll 1$, $f_{d1,c} \approx 1$, $f_{d2,c} \ll f_{d1,c}$, and the number of detections from secondaries of binary systems is negligible. In this limit, binarity does not affect the derived occurrence rate:

$$X_\Gamma = 1 \quad \text{(limit of } \gamma_R \ll 1). \tag{61}$$

For the stellar population described in this model, with a mean light ratio $\langle \gamma_R \rangle \approx 0.27$, this limit is not strictly applicable. To assess the importance of the error, we want the mean value of X_Γ, or the median, or else some other summary statistic. Recall that all the terms in Eq. 60 are random variables, and note that they are interdependent. This latter fact impedes us from analytically expressing these descriptive statistics.

Running a Monte Carlo simulation in which we fix the planet and primary star properties, draw the binary mass ratio from Eq. 33, and take the luminosity as $L \propto M^{3.5}$, we find a mean of

$$\langle X_\Gamma \rangle = 1.291 \quad \text{(Monte Carlo result)}, \tag{62}$$

and a median that is essentially the same.