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. For single stars, this may simply be from erroneous stellar parameters. In binaries, one might imagine the combined light of two stars somehow making them resemble a single star 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 S/N > $({\rm S/N})_{\rm min}$. 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 S/N > $({\rm S/N})_{\rm min}$ 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:

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

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

- 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.

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:

$$\Gamma_t = \frac{N_{\rm planets}}{N_{\rm stars}} \quad (3)$$

$$\Gamma_t = \frac{\Gamma_{t,s}N_s + (1+w_{d2})\Gamma_{t,d}N_d}{N_s + 2N_d}. \quad (4)$$

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,

$$N_i = n_i \frac{4\pi}{3} d^3_{\max,i}, \quad (5)$$

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

$$\frac{n_d}{n_s + n_d} = \text{binary fraction} \equiv \text{BF} \quad (6)$$

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. 5 is the maximum distance corresponding to the given magnitude limit:

$$d_{\max,i} = \left(\frac{L_i}{4\pi F_{\min}}\right)^{1/2}, \quad (7)$$

where the limiting flux in the bandpass $F_{\min}$ could equivalently be stated in terms of a limiting magnitude $m_{\min}$. In Eq. 7, 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

$$\frac{N_d}{N_s} = \frac{n_d}{n_s}\left(\frac{d_{\max,d}}{d_{\max,s}}\right)^3 \quad (8)$$

$$= \frac{\text{BF}}{1-\text{BF}} \times (1+\gamma_R)^{3/2}. \quad (9)$$

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 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] 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.

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

$$\Gamma_{\rm a} = \frac{N_{\rm det}}{N_s + N_d} \times \frac{1}{f_g}. \tag{10}$$

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:

$$R_{p,\rm a} = R_{p,\rm t} \frac{R_{\star,\rm a}}{R_{\star,\rm t}} \mathcal{D}^{1/2}, \tag{11}$$

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

$$\begin{aligned}\mathcal{D} &= \begin{cases} L_1/L_d, & \text{if planet orbits primary} \\ \gamma_R L_1/L_d, & \text{if planet orbits secondary}\end{cases} \\ &= \begin{cases} (1+\gamma_R)^{-1}, & \text{if planet orbits primary} \\ (1+\gamma_R^{-1})^{-1}, & \text{if planet orbits secondary}\end{cases}\end{aligned} \tag{12}$$

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,

$$N_{{\rm det},s} = N_s \Gamma_{t,s} f_{s,g} f_{s,c}, \tag{13}$$

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,

$$N_{{\rm det},d} = (1 + w_{d2}) N_d \Gamma_{t,d} f_{d,g} f_{d,c}, \tag{14}$$

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

$$\Gamma_{\rm a} = \frac{\Gamma_{t,s} N_s}{N_s + N_d}. \tag{15}$$

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

$$\begin{aligned} X_\Gamma &= \frac{\Gamma_t}{\Gamma_{\rm a}} && (16) \\ &= \frac{\Gamma_{t,s}N_s + (1+w_{d2})\Gamma_{t,d}N_d}{\Gamma_{t,s}N_s} \cdot \frac{N_s + N_d}{N_s + 2N_d} && (17) \\ &= \left(1 + \frac{(1+w_{d2})\Gamma_{t,d}}{\Gamma_{t,s}}\beta\right) \cdot \frac{1+\beta}{1+2\beta}, && (18) \end{aligned}$$

for

$$\beta \equiv N_d/N_s = \frac{\rm BF}{1-{\rm BF}} \times (1+\gamma_R)^{3/2}. \tag{19}$$

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

$$\frac{N_{\rm det,d}}{N_{\rm det,s} + N_{\rm det,d}} = \frac{1}{1+\zeta}, \tag{20}$$

for

$$\zeta \equiv \frac{N_s \Gamma_{t,s} f_{s,g} f_{s,c}}{(1+w_{d2}) N_d \Gamma_{t,d} f_{d,g} f_{d,c}}. \tag{21}$$

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

$$X_\Gamma = \frac{\Gamma_{t,s} + (1+w_{d2})\Gamma_{t,d}\beta}{\Gamma_{t,s} + (1+w_{d2})\Gamma_{t,d} f_{d,c}} \cdot \frac{1+\beta}{1+2\beta}, \tag{22}$$

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 of course idealized, but it helps for developing intuition. We now 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 ± 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,

$$f(q) = \begin{cases} c_q & 0.1 < q \leq 1 \\ 0 & \text{otherwise}, \end{cases} \tag{23}$$

provided we ignore the peak at high mass ratios for analytic convenience. $c_q = 1/9$ is the normalization. We use 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 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[3]. In log-log space, we let the intersection point of the lines float, and made various cuts on the data as indicated in Fig 6. The fit parameters are available in a footnote[4]. Though the data show a break at $M_\star \approx 0.5M_\odot$, 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. [Demircan & Kahraman 1991]'s empirical fit to eclipsing binary data,

$$R = 1.06M^{0.945} \tag{24}$$

applies for all the stars in our desired mass range ($M < 1.66M_\odot$, D&K 1991's stated bounds), where each quantity is given in solar units. Since again, our main purpose is to develop intuition about observational biases, we approximate this as $R \propto M$.

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

$$N_{\rm planets} = \Gamma_{t,s}N_s + \Gamma_{t,d}N_d + \Gamma_{t,d}w_{d2}f_{\rm desired}N_d, \tag{25}$$

[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.

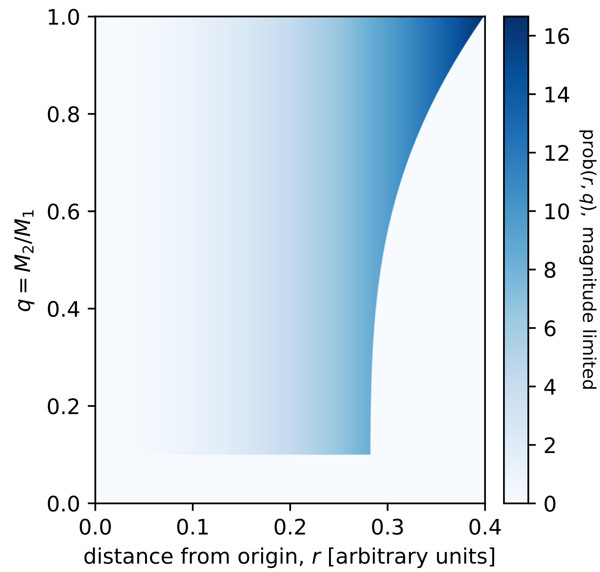

FIG. 1.— The joint probability distribution of a binary star's position and mass ratio in a magnitude-limited sample. This plot assumes a different $M(L)$ relation than previous plots: $L = M^3$. Note that the marginalized mass ratio distribution is uniform at any given distance.

where as before N_s and N_d are the number of single and double systems, and the fraction $f_{\rm desired}$ is defined so that $f_{\rm desired}N_d$ 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

$$N_d(\gamma_R) = n_d \frac{4\pi}{3} d^3_{\rm max,d}(\gamma_R) \tag{26}$$

for

$$d_{\rm max,d}(\gamma_R) = d_{\rm max,s} \times (1 + \gamma_R)^{1/2}, \tag{27}$$

where $d_{\rm max,s}$ is given by Eq. 7.

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

$$\text{prob}(N_d) = \text{prob}(q(\gamma_R)) \left|\frac{\mathrm{d}q}{\mathrm{d}\gamma_R}\right| \left|\frac{\mathrm{d}\gamma_R}{\mathrm{d}d_{\rm max,d}}\right| \left|\frac{\mathrm{d}d_{\rm max,d}}{\mathrm{d}N_d}\right|. \tag{28}$$

Necessary aside on biases in mass ratio distributions— Before blindly plugging in Eq. 23's assumed mass ratio distribution to find the expected number of double systems, we must emphasize that Eq. 23'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.

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 sub-

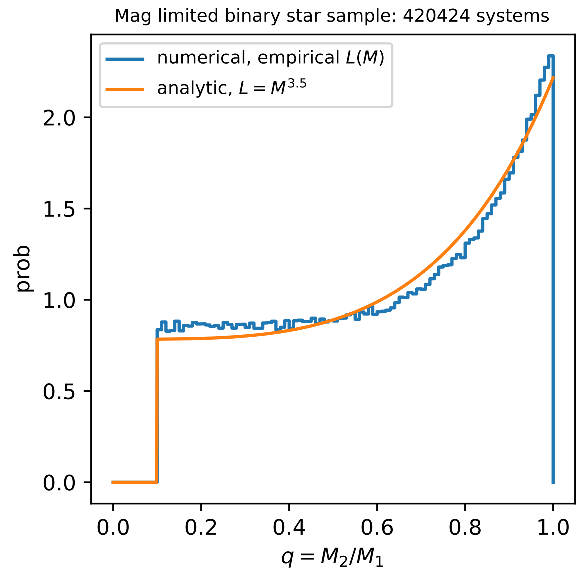

FIG. 2.— The distribution of the mass ratio for a magnitude limited sample of binary stars. The underlying mass ratios are drawn from Eq. 23 – a *uniform* distribution in a volume-limited sample. The entire bias can be understood analytically (Eq. 33) to good precision. The numerical comparison uses the empirical mass-luminosity relation shown in Fig. 6.

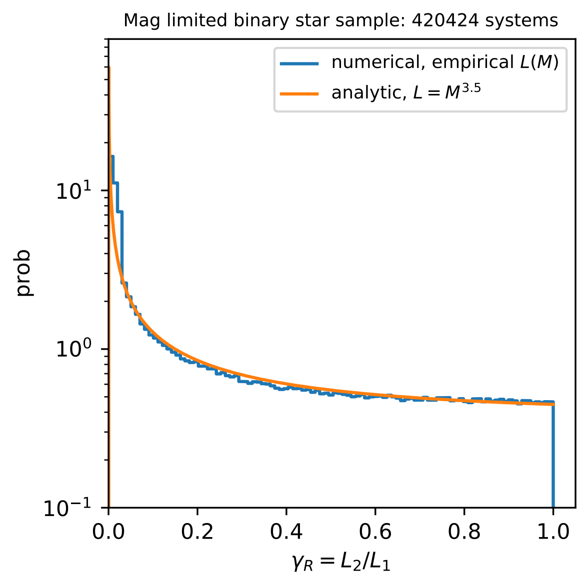

FIG. 3.— The distribution of the luminosity ratio for a magnitude limited sample of binary stars. Despite the preference in the sample towards large q binaries (Fig. 2), the bias is towards small γ_R binaries because of the steepness of the mass-luminosity relation. The underlying mass ratios are uniform in volume. The numerical comparison uses the empirical mass-luminosity relation shown in Fig. 6.

script for 'magnitude-limited',

$$\mathrm{prob_{ml}}(r,q) = \mathrm{prob}(q|r)\mathrm{prob_{ml}}(r), \tag{29}$$

where prob$(q|r)$ is given by Eq. 23, and

$$\mathrm{prob_{ml}}(r) = \int_0^1 \mathrm{prob_{ml}}(r|q)\,\mathrm{prob}(q)\,\mathrm{d}q \tag{30}$$

$$= \int_{0.1}^1 \frac{1}{3}\frac{r^2}{d^3_{\mathrm{max,d}}(q)}\,\mathrm{d}q \quad \text{if } r < d_{\max}(q), \tag{31}$$

and otherwise zero. The resulting joint probability distribution is plotted in Fig. 1.

Marginalizing the joint distribution over distance, and assuming $L \propto M^{3.5}$,

$$\mathrm{prob_{ml}}(q) = \int_0^{d_{\max}(q)} \mathrm{prob_{ml}}(r,q)\,\mathrm{d}r \tag{32}$$

$$= \mathcal{Z}\cdot(1+q^{3.5})^{3/2} \quad \text{if } 0.1 < q \le 1, \tag{33}$$

and otherwise 0, for $\mathcal{Z}$ the appropriate normalization constant. This marginalized distribution – the distribution function of mass ratios for binaries in a magnitude-limited sample – is plotted in Fig. 2.

More correctly then, we must write

$$\mathrm{prob}(N_d) = \mathrm{prob_{ml}}(q(\gamma_R))\left|\frac{\mathrm{d}q}{\mathrm{d}\gamma_R}\right|\left|\frac{\mathrm{d}\gamma_R}{\mathrm{d}d_{\mathrm{max,d}}}\right|\left|\frac{\mathrm{d}d_{\mathrm{max,d}}}{\mathrm{d}N_d}\right|. \tag{34}$$

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

$$\mathrm{prob}(N_d) = \frac{2\mathcal{Z}}{3\alpha}N_d^{1/3}\left(\frac{3}{4\pi n_d}\right)^{4/3} d^{-4}_{\mathrm{max,s}} \times\left[\left(\frac{3N_d}{4\pi n_d}\right)^{2/3} d^{-2}_{\mathrm{max,s}} - 1\right]^{\frac{1-\alpha}{\alpha}} \tag{35}$$

over the interval

$$N_d^{\mathrm{lower}} = \frac{4\pi n_d}{3}(\sqrt{0.1^3+1}\,d_{\mathrm{max,s}})^3, \tag{36}$$

$$N_d^{\mathrm{upper}} = \frac{4\pi n_d}{3}(\sqrt{2}\,d_{\mathrm{max,s}})^3, \tag{37}$$

and outside the stated interval prob$(N_d) = 0$.

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

$$\mathrm{prob_{ml}}(\gamma_R) = \frac{\mathcal{Z}}{\alpha}(1+\gamma_R)\gamma_R^{\frac{1-\alpha}{\alpha}}, \tag{38}$$

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

The reason we derived all this is that we wanted to know the number of searchable binary stars. The expected value for the number of double star systems in the sample is

$$\langle N_d\rangle = \int_0^\infty N_d\,\mathrm{prob}(N_d)\,\mathrm{d}N_d. \tag{39}$$

With some algebra, and letting $\alpha = 3.5$, this can be expressed as

$$\langle N_d\rangle = \frac{2\mathcal{Z}}{7R^{1/2}}\int_{(0.1^\alpha+1)^{1/3}}^{\sqrt{2}} u^{3/2}(u-1)^{5/7}\,\mathrm{d}u, \tag{40}$$

for

$$R \equiv \left(\frac{3}{4\pi n_d}\right)^{2/3} d_{\rm max,s}^{-2}. \quad (41)$$

The integral in Eq. 40 has a closed-form solution in terms of Gauss hypergeometric functions, ${}_2F_1$. Once evaluated, the integral takes a numerical value of 0.1829. The normalization of

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. **??**.
- 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 distance[5]. This is shown in Fig. 2, and explained analytically below.
- 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. 3.

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

$$\mathrm{prob}(\gamma_R) = \frac{1}{9\alpha}\,\gamma_R^{\frac{1-\alpha}{\alpha}} \quad \text{for } (0.1)^\alpha < \gamma_R < 1, \quad (42)$$

and otherwise zero. For instance if $\alpha = 3$, prob(γ_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?

$$\langle N_{\rm stars} \rangle = N_s + 2\langle N_d \rangle, \quad (43)$$

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

3.2. *How many planets are in the sample?*

The mean number of planets in the sample is

$$\langle N_{\rm planets} \rangle = N_{\text{planets in single star systems}} + N_{\text{planets in double star systems}} \quad (44)$$

$$= \Gamma_{t,s} N_s + (\Gamma_{t,d1} + \Gamma_{t,d2})\langle N_d \rangle. \quad (45)$$

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.

3.3. *What is the true occurrence rate?*

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

$$\Gamma_t = \frac{\langle N_{\rm planets} \rangle}{\langle N_{\rm stars} \rangle} \quad (46)$$

$$\Gamma_t = \frac{\Gamma_{t,s} N_s + (\Gamma_{t,d1} + \Gamma_{t,d2})\langle N_d \rangle}{N_s + 2\langle N_d \rangle}. \quad (47)$$

3.4. *How many planets are detected?*

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

The number of planets detected in single star systems is

$$N_{{\rm det},s} = N_s \Gamma_{t,s} f_{s,g} f_{s,c}, \quad (48)$$

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

$$\langle N_{{\rm det},d} \rangle = \langle N_d \rangle (\Gamma_{t,d1} f_{d1,g} f_{d1,c} + \Gamma_{t,d2} f_{d2,g} f_{d2,c}). \quad (49)$$

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. 4. 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. 49, completeness included. I'm not yet convinced we need to – we should get cancellations for most of the questions we want to ask.

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

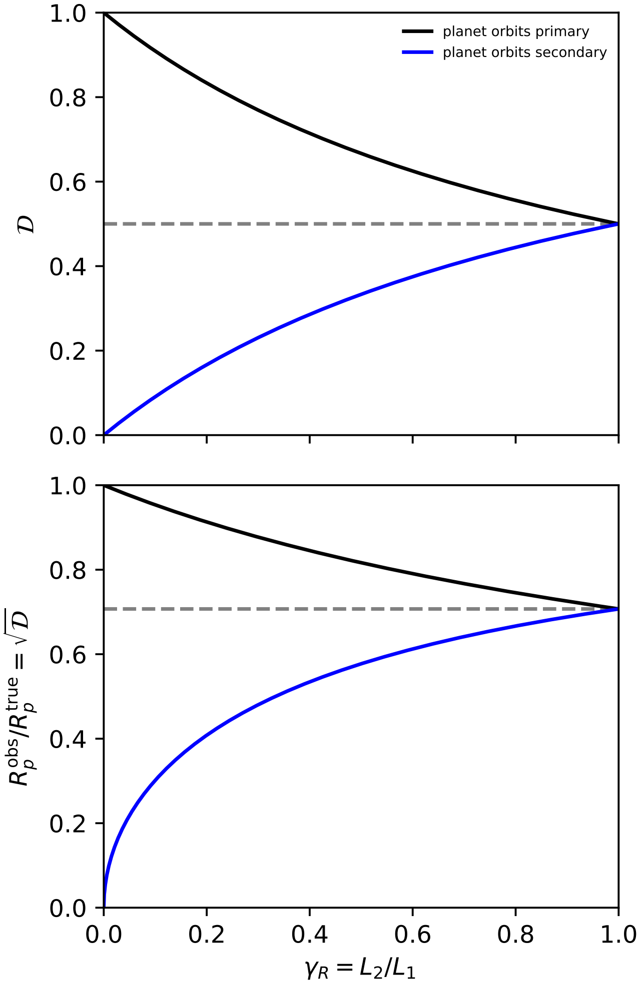

FIG. 4.— Dilution $\mathcal{D}$ and observed to true radius ratio $\sqrt{\mathcal{D}}$ plotted against binary light ratio γ_R. This illustrates Eq. 12. The lower panel's horizontal line is at $1/\sqrt{2}$. The implication is that if a detected planet orbits the secondary and you do not know it, you cannot 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 $\approx 29\%$.

3.4.1. *Analytic completeness*

`todo, if necessary`

3.4.2. *Deriving* prob(x_i)

`todo, if necessary`

3.4.3. *Number of detected planets*

`todo, if necessary`

3.5. *Astronomer A ignores binarity*

Just as in Sec. 2, 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. 2,

$$\Gamma_{\mathrm{A},R_p} = \frac{N_{\mathrm{det,s}}/f_{s,g}}{N_s + N_d}. \tag{50}$$

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, $N_{\mathrm{det,d}}(R_p^{\mathrm{obs}})$, Astronomer A would derive an occurrence rate for this separate population of

$$\Gamma_{\mathrm{A}}(R_p^{\mathrm{obs}}) = \frac{N_{\mathrm{det,d}}(R_p^{\mathrm{obs}})/f_{s,g}}{N_s + N_d}. \tag{51}$$

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. 3.10.3.

3.6. *Astronomer B counts host stars correctly*

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

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

In Sec. 2, 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. 4), 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

$$\Gamma_{\mathrm{C},R_p} = \frac{N_{\mathrm{det,s}}/f_{s,g} + N_{\mathrm{det,d1}}/f_{d1,g} + N_{\mathrm{det,d2}}/f_{d2,g}}{N_s + 2N_d} \tag{52}$$

the closest yet to the true rate (Sec. 2.1).

3.8. *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

$$\Gamma_{t,s} = \frac{N_{\mathrm{det,s}}}{N_s f_{s,g} f_{s,c}}, \tag{53}$$

$$\Gamma_{t,d1} = \frac{N_{\mathrm{det,d1}}}{N_d f_{d1,g} f_{d1,c}}, \tag{54}$$

$$\Gamma_{t,d2} = \frac{N_{\rm det,d2}}{N_d f_{d2,g} f_{d2,c}}. \tag{55}$$

With these in hand, they derive the overall occurrence rate

$$\Gamma_{{\rm D},R_p} = \frac{N_{\rm det,s}/f_s + N_{\rm det,d1}/f_{d1} + N_{\rm det,d2}/f_{d2}}{N_s + 2N_d} \tag{56}$$

$$= \frac{\Gamma_{t,s}N_s + (\Gamma_{t,d1} + \Gamma_{t,d2})N_d}{N_s + 2N_d}, \tag{57}$$

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. 47).

3.9. *Numerical verification*

`todo, if needed`

3.10. *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. 2, and compare the answers.

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

Ignoring binarity, we will detect $N_{\rm det,s}$ planets around single stars, and $N_{\rm det,d}$ 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

$$\frac{N_{\rm det,d}}{N_{\rm det,s} + N_{\rm det,d}} = \frac{1}{1+\alpha}, \tag{58}$$

for

$$\alpha \equiv \frac{N_d(\Gamma_{t,d1}f_{d1} + \Gamma_{t,d2}f_{d2})}{N_s\Gamma_{t,s}f_s}, \tag{59}$$

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.
```

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

Write $\Gamma_{\text{A, planets of } R_p} = \Gamma_{{\rm A},R_p}$, 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 somehow knows his completeness for single stars. This is possible, because Astronomer A can correctly derive a completeness estimate and geometric transit probability for single stars only. For Astronomer A' below, including the planets in binary star systems in the numerator introduces mistakes in the estimated completeness, and so the picture is more complicated. Then

$$\Gamma_{{\rm A},R_p} = \frac{N_{\rm det,s}/(f_{s,g}f_{s,c})}{N_s + N_d}. \tag{60}$$

The correction factor X_Γ is

$$X_\Gamma = \frac{\Gamma_{{\rm D},R_p}}{\Gamma_{{\rm A},R_p}} \tag{61}$$

$$= \frac{\Gamma_{t,s}N_s + (\Gamma_{t,d1} + \Gamma_{t,d2})N_d}{N_s + 2N_d} \cdot \frac{N_s + N_d}{\Gamma_{t,s}N_s} \tag{62}$$

$$= \frac{(1 + \beta(\Gamma_{t,d1} + \Gamma_{t,d2})/\Gamma_{t,s})(1+\beta)}{(1+2\beta)} \tag{63}$$

$$= (1 + \chi\beta)\cdot\frac{1+\beta}{1+2\beta} \tag{64}$$

for

$$\beta \equiv N_d/N_s, \quad \chi \equiv (\Gamma_{t,d1} + \Gamma_{t,d2})/\Gamma_{t,s}. \tag{65}$$

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 $X_\Gamma = 1 + \beta$. This is same expression as what we found for Model #1, but now the value of β will be smaller in this model because of the smaller volume of searchable binary stars. From numerics, we find $\beta = N_d/N_s = 0.59$, so the occurrence rate correction factor is $X_\Gamma(\chi = 2) = 1.59$.

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 $X_\Gamma = (1+\beta)^2/(1+2\beta)$. For $\beta = 0.59$, this gives $X_\Gamma(\chi = 1) = 1.16$. The true correction factor is somewhere between these two limits.

3.10.3. *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 $R_p < R_p^{\rm obs} < (1 \pm x)R_p$. For Model #2, since all planets have radius R_p, we only care about the smaller radius case, *i.e.*, we want to count $R_p > R_p^{\rm obs} > (1-x)R_p$, for instance with $x = 0.1$.

The diluted detections only occur in binary systems. The number of planets detected in double star systems with observed radius $R_p^{\rm obs}$ is

$$N_{\rm det,d}(R_p^{\rm obs}) = N_{\rm det,d1}(R_p^{\rm obs}) + N_{\rm det,d2}(R_p^{\rm obs}) \tag{66}$$

$$= \Gamma_{t,d1}f_{d1}(R_p^{\rm obs})N_{d1}(R_p^{\rm obs}) + \Gamma_{t,d2}f_{d2}(R_p^{\rm obs})N_{d2}(R_p^{\rm obs}). \tag{67}$$

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 $R_p^{\rm obs}$.

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

Then writing $N_{d1}(R_p')$, the number of double systems for which a (R_p, P) planet would be observed in R_p' if it orbiting the primary, is relatively straight-forward:

$$N_{d1}(R_p') = \int_0^{\min(1,\gamma_{R,u})} N_d(\gamma_R)\,{\rm prob}(\gamma_R)\,{\rm d}(\gamma_R), \tag{68}$$

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

$$\gamma_{R,u} \equiv \frac{x(2-x)}{(1-x)^2}. \tag{69}$$

If $x > (1-2^{-1/2})$, bad things happen (cf. Fig. 4), hence the need for the minimum. Note that Eq. 68 is not an expression for the number of detected planets about primaries – that quantity is expressed as $\Gamma_{t,d1} f_{d1}(R_p^{\rm obs}) N_{d1}(R_p^{\rm obs})$.

The analogous equation to Eq. 68 for $N_{d2}(R_p')$, the number of double systems for which a (R_p, P) planet would be observed in R_p' if it orbited the secondary, is

$$N_{d2}(R_p') = \begin{cases} \int_{\gamma_{R,l}}^{1} N_d(\gamma_R)\,\mathrm{prob}(\gamma_R)\,\mathrm{d}(\gamma_R), & \\ & x > 1-2^{-1/2} \\ 0, & x \leq 1-2^{-1/2} \end{cases} \tag{70}$$

where

$$\gamma_{R,l} \equiv \frac{(1-x)^2}{x(2-x)}. \tag{71}$$

Eq. 70 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. 4.

If we were forward-modelling, *i.e.* if we actually wanted to compute $N_{\rm det,d}(R_p')$, we would also need to evaluate the total detection efficiency term $f_{d1}(R_p^{\rm obs})$, which is a product of the geometric transit probability and the fraction of signals of a given depth that are detectable. We would need to integrate this product over the domain of allowable radii R_p' to find the number of detected planets.

However, the procedure of estimating the geometric transit probability, as well as the completeness, is complicated by the errors that are introduced by including binary star systems that are thought to be single in the numerator. These stars will have incorrectly estimated masses and radii, which influences both $f_c(R_p')$ and $f_g(R_p')$.

`If we ask the right question, we will not need to actually compute either term.`

Astronomer A misclassifies planetary radii, miscounts the total number of stars, and (incorrectly) estimates his completeness and geometric correction for "single" stars as a function of radius $f(R_p')$ (omitting a "s" subscript because the stars for which this incorrect occurrence rate are derived are not all single). His occurrence rate for planets of radius *near* R_p, *i.e.* the set of planets with $\{(1-x)R_p < R_p^{\rm obs} < R_p\}$, is

$$\Gamma_{{\rm A},R_p'} = \frac{(N_{\rm det,s} + N_{\rm det,d}(R_p'))/f(R_p')}{N_s + N_d}. \tag{72}$$

For him,

$$f(R_p') = f_g(R_p') f_c(R_p'). \tag{73}$$

Astronomer D, who had everything right, changes nothing.

The new correction factor is

$$X_\Gamma = \frac{\Gamma_{{\rm D},R_p}}{\Gamma_{{\rm A},R_p'}} \tag{74}$$

$$= \frac{1+\chi\beta}{1+\xi} \cdot \frac{1+\beta}{1+2\beta}, \tag{75}$$

for

$$\beta \equiv N_d/N_s, \quad \chi \equiv (\Gamma_{t,d1} + \Gamma_{t,d2})/\Gamma_{t,s}, \tag{76}$$

and

$$\xi \equiv \frac{N_{\rm det,d}(R_p')}{f(R_p')} \cdot \frac{1}{N_s \Gamma_{t,s}}. \tag{77}$$

If we let $x < 1-2^{-1/2}$, for instance $x = 0.1$, then $N_{d2}(R_p') = 0$. The expression for ξ in this case becomes

$$\xi = \frac{N_{d1}(R_p')}{N_s} \cdot \frac{\Gamma_{t,d1}}{\Gamma_{t,s}} \cdot \frac{f_{d1}(R_p')}{f(R_p')} \tag{78}$$

$$= \frac{N_{d1}(R_p')}{N_s} \cdot \frac{\Gamma_{t,d1}}{\Gamma_{t,s}} \cdot \frac{f_{g,d1}(R_p')}{f_g(R_p')} \cdot \frac{f_{c,d1}(R_p')}{f_c(R_p')}, \tag{79}$$

provided that $x < 1-2^{-1/2}$.

We have now parametrized our ignorance. The values that the latter two ratios of Eq. 79 take depend on what we assume Astronomer A' does to derive the stellar parameters for the systems that he thinks are singles but are in fact doubles. For instance, assume he somehow observed the bolometric flux from every point on the sky, and used it with a distance to obtain single-star luminosities, and thus radii and masses. His binary systems (thought to be singles) would be bluer than either component star is in reality. On the main sequence, $\rho_\star$ decreases with increasing stellar radius, so $\rho_\star$ would be underestimated for all binaries. Since $f_g \propto \rho_\star^{1/3}$, this means that $f_{g,d1}(R_p')/f_g(R_p') > 1$. However, by the same token we can place an *upper* bound on this ratio. In the worst case scenario a given star's luminosity is twice the individual star's luminosity. If $L \propto M^3 \propto R^3$, then the radius is over-estimated by at most a factor of $2^{1/3}$. Thus the density is underestimated by at most a factor of $2^{1/2}$, and $f_g(R_p')$ is underestimated by at most a factor of $2^{1/6} = 1.12$.

This begins to take us somewhere. The left-most term of Eq. 79 is bounded above by β. This is because $N_{d1}(R_p') \leq N_{d1} = N_d$. If we assume $\Gamma_{t,d1} = \Gamma_{t,s}$, as we were doing anyway, then the product of the three left-most fractions must be less than 1.12β. The only thing standing in our way of an analytic upper bound to ξ is the completeness fraction $f_{c,d1}(R_p')/f_c(R_p')$. Note that over a fixed radius interval dilution will always yield a lower completeness fraction for double star systems than singles, so $f_{c,d1}(R_p')/f_{c,s}(R_p') \leq 1$.

For the completeness term, an argument that I don't wholly believe is that

$$\frac{f_{c,d1}(R_p')}{f_c(R_p')} = \frac{f_{c,d1}(R_p')}{f_{c,s}(R_p')} \cdot \frac{f_{c,s}(R_p')}{f_c(R_p')} \lesssim 1 \cdot (1+\langle\gamma_R\rangle)^3, \tag{80}$$

where the first part of the inequality is obvious, but the second is by assuming it's $\lesssim f_{s,c}/f_{d,c}$, which might be wrong because of the radius dependence.

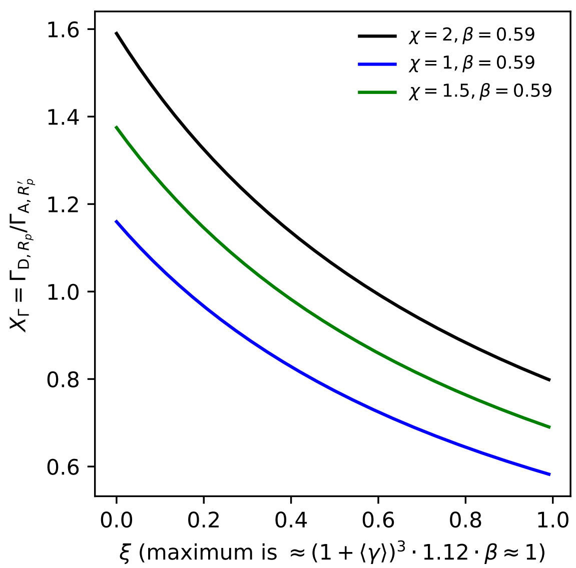

FIG. 5.— *Y axis*: correction factor for the astronomer who ignores binarity but counts planetary radii that are "close enough" (A') and the astronomer who knows everything (D). *X axis*: χ as defined in Eq. 79.

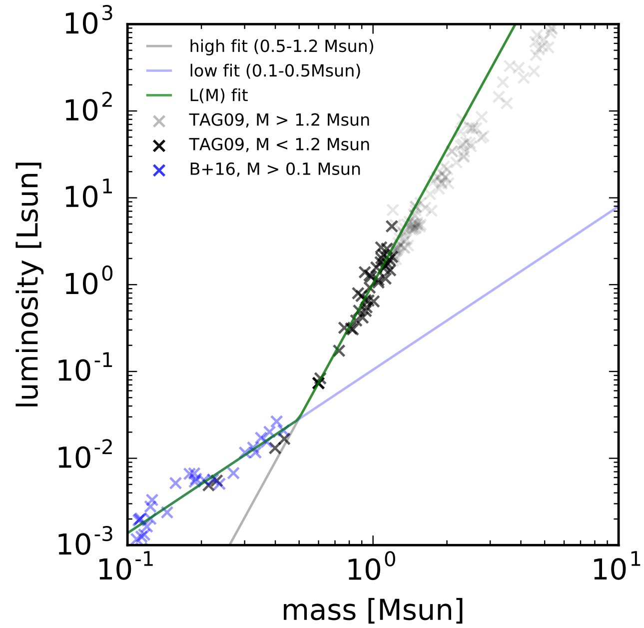

FIG. 6.— Empirical fit to main sequence dwarf mass luminosity data compiled from Torres et al. [2009] and Benedict et al. [2016]. The "low fit" is a least squares fit to data from $0.1 - 0.5 M_\odot$, and the "high fit" is to data above that, and below the Kraft break ($1.2 M_\odot$). The $L(M)$ relation taken for numerics is the maximum of the two fits. For analytics, we assume $L \propto M^{3.5}$.

A 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. 5, which shows that the correction factor varies from 0.6 to 1.6, depending on what fraction of secondaries you assume host planets.

I could run numerics for this to make the result more interesting.

4. MODEL # 3: A SYNTHETIC *KEPLER* ANALOG

If we assume every KIC star to be single, what order of magnitude of error do we make in occurrence rates estimated for planets of different sizes (and periods)?

We can attempt an answer with Monte Carlo simulations of something like the *Kepler* field. I say "something like" because the binarity properties of the *Kepler* field (or more specifically the stars selected in the KIC) are not known.

but it might be better to do something directly on the KIC's selected target stars, like what Ciardi did! The following is a good example, but at most it can "suggest" that a similar level of error has been made from the *Kepler* survey.
But it has MAJOR benefits compared to the GIC method. It eliminates the need to simulate the biases of the KIC (since they're already baked in). The process of "adding" binaries to the KIC is the closest thing to REALITY, since in reality the KIC is missing binary companion information. It BAKES IN all the real biases that come with mis-estimating stellar radii/masses when they are actually binaries (because it's already been done). It also has literature precedent in Ciardi's work, so you don't need to cook up a crazy new procedure – can just take his with the flaws baked in.
So this method is best if we care about what errors were made in Kepler/the KIC. It's not as nice if we want something "general".

The plan is as follows:

- Verify Galaxia can produce something resembling the KIC.
- Introduce binaries (Galaxia does not, by default, have any).
- Use Galaxia [Sharma et al 2010] to construct an analog of the KIC.
- Introduce a planet population.
- Numerically assess what occurrence rates would be derived, and compare to previous sections.

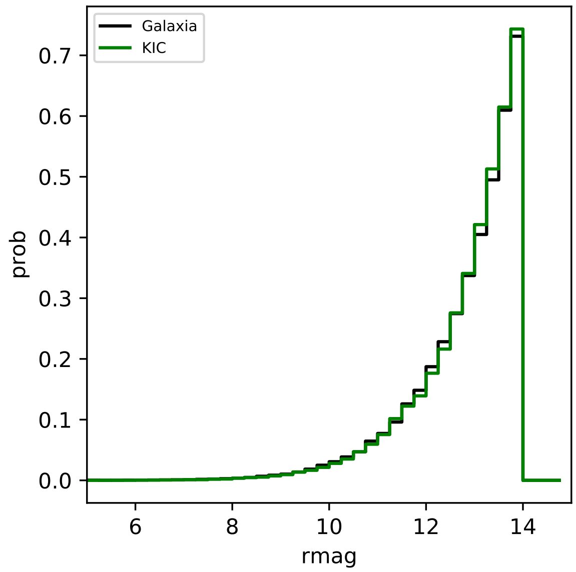

FIG. 7.— Probably density distributions of r magnitude for $r <$ 14 stars within 7.5° of the center of the *Kepler* field, for Galaxia and the KIC.

4.1. *Verifying Galaxia can produce a KIC analog*

Galaxia [Sharma et al 2010] is a stellar population synthesis code for creating surveys of the Milky Way. It samples an assumed analytic distribution function. In other words, it assumes a number density of stars as a function of position, velocity, age, metallicty, and mass, from which it then samples to construct a "realistic" stellar population. The assumed model includes:

- A metallicity distribution (log-normal),
- A velocity distribution (triaxial Gaussian),
- The stellar density and galactic components specified by Robin et al [2003] (the Besançon model) – see Sharma et al [2010] Table 1. This includes the IMF for a young and old thin disk, a thick disk, a spheroidal component, a bulge component, as well as an ISM and dark halo. It also includes the star-formation rate.

While this model of course lacks some detail, Sharma et al [2010] show that agreement with the volume-limited Hipparcos sample is probably good enough for our purpose of estimating the number of "solar-like" (by definition hereafter, $0.7M_\odot < M_\star < 1.3M_\odot$) stars in something like the *Kepler* field.

We then run the Galaxia in its "circular survey" mode, towards the direction of the *Kepler* field[6]. We apply the Schlegel, Finkbeiner et al [1999] extinction map when converting absolute to apparent magnitudes.

To verify the results are in reasonable agreement with the stars that actually exist in that direction, we compare the output with the KIC[7] [Brown et al 2011]. Following Sharma et al [2016, ApJ 822:15], for each catalog

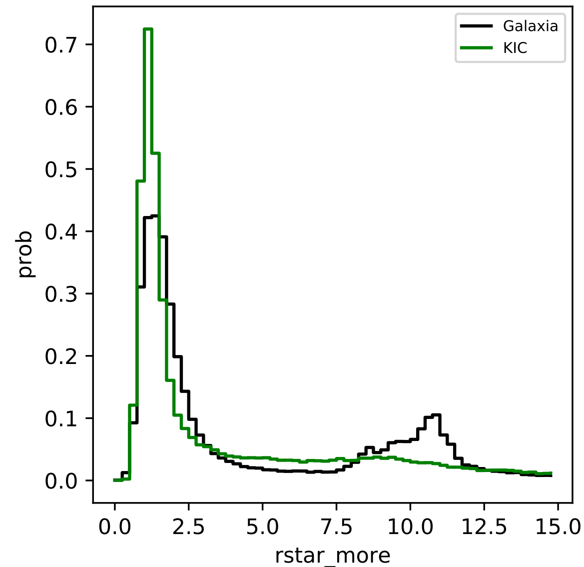

FIG. 8.— Same as Fig. 7, for $R_\star$.

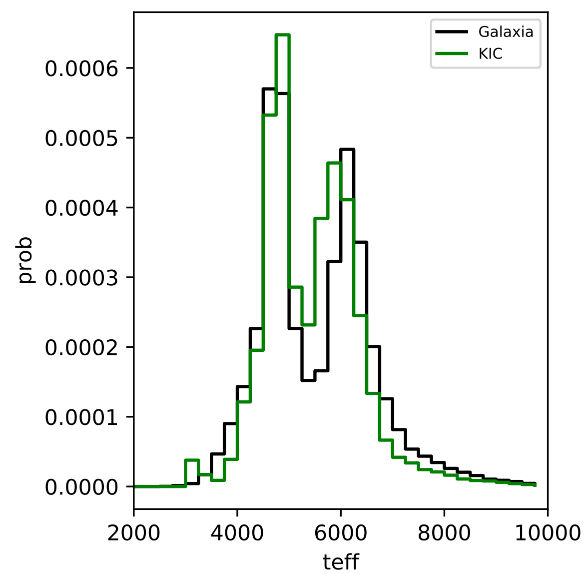

FIG. 9.— Same as Fig. 7, for $T_{\rm eff}$.

we selected all stars within 7.5° of the center of the *Kepler* field, and with Sloan $r < 14$. The KIC is complete to magnitudes fainter than $r = 14$, so the comparison should not be affected by its completeness.

A few distributions are shown in Figs. 7- 9. From Fig. 7, we can see that the stellar counts as a function of Sloan r magnitude are in close agreement with the KIC. From Fig. 8, the KIC reports relatively more dwarf stars, and fewer giants than Galaxia does. This is likely related to KIC's inability to distinguish between dwarfs and giants with only limited photometry – although Brown et al [2011]'s idea of using a specific "D51" photometric band as a $\log g$ diagnostic was reported to have helped with this problem. In Fig. 9, we see something resembling the K dwarf desert again (5000K $\approx$ K3V dwarf).

[6] galactic longitude 76.532562, galactic latitude 13.289502.
[7] From http://archive.stsci.edu/kepler/kic.html, we downloaded the 13.1 million row "—"-delimited gzipped ASCII file containing the complete Kepler Input Catalog (version 10).

Although there are a few slightly worrying discrepancies (notably in Fig. 8), we get the idea that Galaxia at least roughly resembles "reality" (inasmuch as the reported KIC parameters represent reality!).

4.2. *Introduce binaries*

By default, Galaxia does not include binaries. However, from Fig. 7, we've seen that its number counts as a function of apparent magnitude are good. So we assign binarity to the Galaxia stars as follows.

- Assume a binary fraction BF = 0.45 [Raghavan et al 2010].
- Ignore higher order multiples.
- If a star is drawn to be a binary:

  Its reported magnitude becomes a system magnitude.

  Draw the binary's mass ratio q from the probability distribution function for a magnitude limited survey, i.e. from Fig. 2. This is sensible so long as the pdf is independent of the primary's mass. We assume this to be the case over $0.7 - 1.3M_{\odot}$[8].

  For $q \geq 5/7$, it turns out the above is sufficient information to analytically specify the individual stellar masses, and thus luminosities with the relation shown in Fig. 6.

  For $q < 5/7$, the individual stellar masses are not uniquely specified[9]. Thus for this latter fraction of the population, we draw the primary mass from the pdf of single star primary masses.

  With the individual masses known, use Eq. 24 to assign a radius to each star in the binary system.

Benefits of the above procedure include that it does not change Galaxia's star counts as a function of magnitude. It produces the correct distribution of the mass ratio for the binaries in a magnitude limited sample, and thus a reasonable light ratio distribution.

The drawbacks are that the individual component masses this procedure produces are not drawn from the same distribution throughout. This might bias any subsequent synthetic transit survey. From Fig. 12, it's likely not excessively important (it's more important to get the mass ratio and light ratio distributions correct).

In addition, the procedure produces a different mass-radius relation for stars in binary systems, vs. stars in single star systems. This might also bias any subsequent synthetic transit survey. The mass-radius relation for single stars comes from the Padova isochrones [Marigo et al 2008, Marigo & Girardi 2007, Girardi et al. 2000, Bertelli et al 1994] used in Galaxia. Although Galaxia does not directly report radii, it reports luminosities and effective temperatures, from which we compute the radii. The alternative choice would be to use the reported Galaxia masses with our Eq. 24 to produce radius estimates. This would produce consistency in the mass-radius relation, but would break $L = 4\pi R^2 \sigma T_{\rm eff}$ for single star systems. The different mass-radius relations are shown in Fig. 14. Generally, it seems that the at a given mass, the stars in binary systems are biased to have slightly greater radii. (Ignoring the evolved stars). However this is a small effect – perhaps of order 10% in the radius. This will affect the transit depths in a synthetic transit survey, making it more difficult to detect planets around binaries.

[8] It was not immediately self-evident to me that this distribution was the correct choice, since for fixed mass ratio as the primary mass changes so does the detectable volume. However, as long as prob(q) is independent of the primary's mass, the average distribution over all the possible detectable volumes will remain the same.

[9] See the derivation of 2017/08/25. This is because of the mass luminosity relation we specified, and the fact that we have q, and L_d, but want individual masses, and individual luminosities.

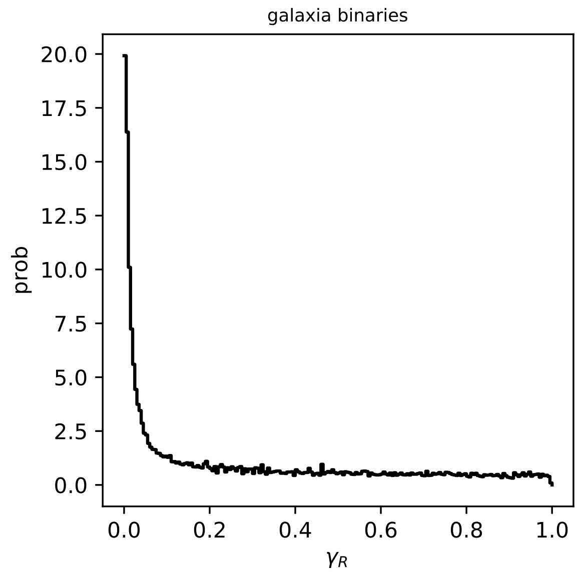

FIG. 10.— Light ratio distribution for the $r < 17$ Galaxia sample discussed in Sec. 4.2.

4.3. *Using Galaxia to construct a KIC analog*

With glowing confidence, we then proceed to construct a KIC analog as follows. We take all Galaxia stars within 7.5 degrees of the center of the *Kepler* field, and take stars with $r < 17$.

a different magnitude cut, e.g. $r < 16$, might matter?

This is now beyond where the KIC is complete, so direct comparisons of distributions with the KIC will be confused. We then apply an analog of the prioritization scheme used by the *Kepler*mission to select target stars, described by Batalha et al [2010]'s Table 1.

For each star, we compute the minimum detectable planet radius at three semimajor axes: (1) the inner radius of the so-called "habitable" zone,

$$a_{\rm HZ} = 0.95\,{\rm AU}\left(\frac{L}{L_{\odot}}\right)^{1/2}, \quad (81)$$

(2) $0.5a_{\rm HZ}$, and (3) $5R_{\star}$. We make the same assumptions as Batalha et al [2010] did about the transit duration. We assume a noise model of pure shot noise, rather than using any CDPP estimates. We then apply the Batalha et

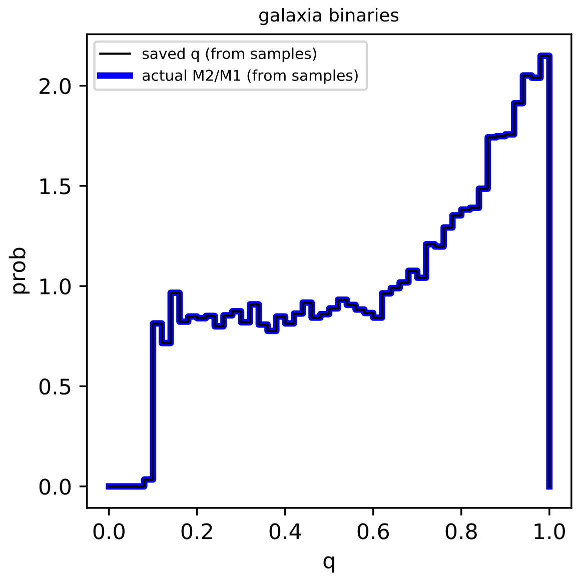

FIG. 11.— Mass ratio distribution for the $r < 17$ Galaxia sample discussed in Sec. 4.2.

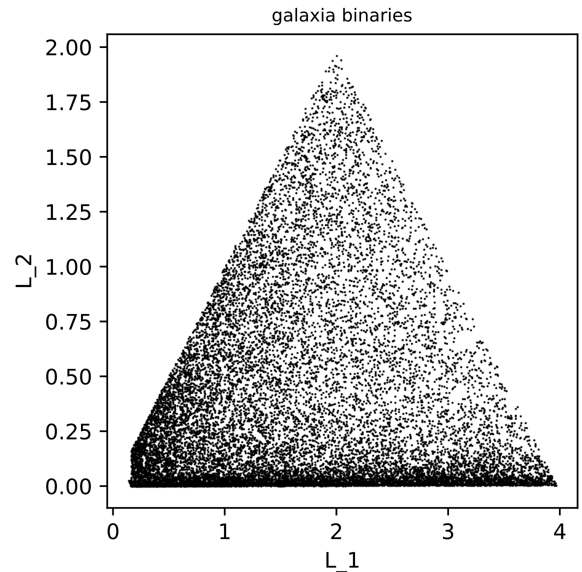

FIG. 13.— Scatter plot of primary and secondary luminosity (solar units).

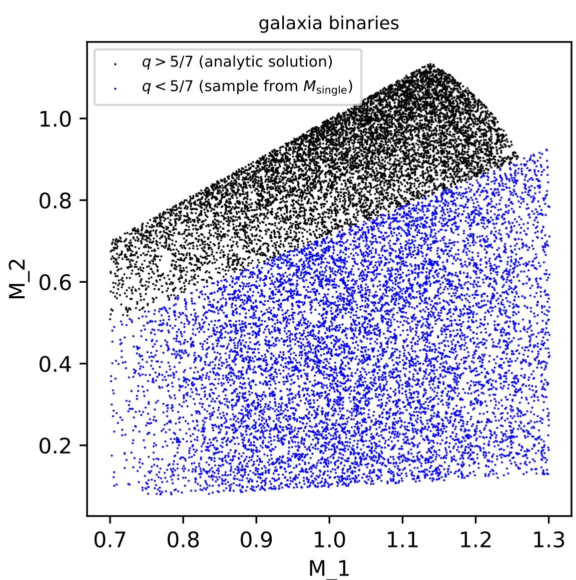

FIG. 12.— Scatter plot of primary and secondary mass (solar units).

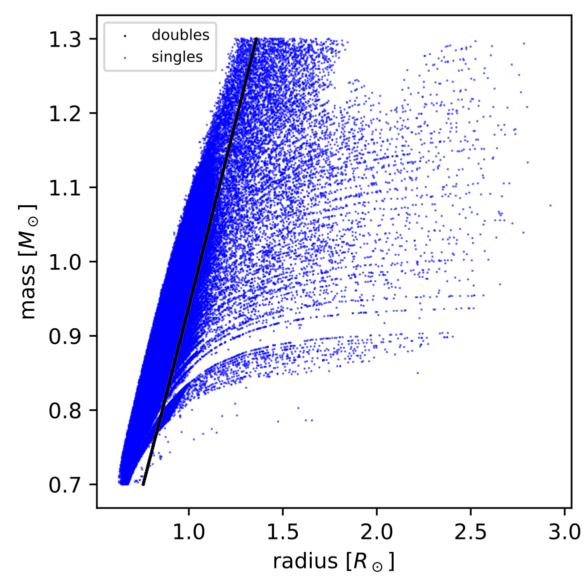

FIG. 14.— Scatter plot of mass vs radius for singles and binaries. For binaries, different points are shown for the primary and secondary of a given system. Only points with $0.7 < M_\star/M_\odot < 1.3$ are selected.

al [2010] Table 1 prioritization, except that we use SDSS r magnitudes instead of Kepler Kp magnitudes. The rankings are shown in Fig. 15. We get similar (within a factor of two) counts as they did towards the end of the table, though at the beginning we seem to have fewer stars[10].

Note that, since we assigned binarity to Galaxia stars, applying the Batalha et al [2010] prioritization procedure requires estimating *incorrect* single star parameters for what are really binary star systems! We do this by keeping the total system luminosity, inverting Fig. 6 to find an incorrect mass, and then applying Eq. 24 to get an incorrect radius. We can then calculate the minimum detectable planet radius as

$$R_{p,\min} = \left(\frac{7.1\sigma_{\rm tot}}{r}\right)^{1/2} R_\star, \tag{82}$$

for $r = 1$ the dilution that we ignore, and

$$\sigma_{\rm tot} \equiv {\rm N} = (F_\gamma^{\rm N} A N_{\rm tra} T_{\rm dur})^{-1/2} \tag{83}$$

In applying the prioritization, we omit the latter two

[10] I suspect this is because of the slightly different noise calculation – and perhaps Batalha using $N_{\rm sample}$ rather than just counting photons matters

Criteria
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 1R_e$; $a = $ HZ; $Kp < 13$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 2R_e$; $a = $ HZ; $Kp < 13$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 1R_e$; $a = $ HZ; $Kp < 14$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 2R_e$; $a = $ HZ; $Kp < 14$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 1R_e$; $a = $ HZ; $Kp < 15$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 1R_e$; $a = $ HZ; $Kp < 16$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 1R_e$; $a = \frac{1}{2}$HZ; $Kp < 14$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 2R_e$; $a = \frac{1}{2}$HZ; $Kp < 14$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 2R_e$; $a = 5R_*$; $Kp < 14$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 2R_e$; $a = $ HZ; $Kp < 15$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 2R_e$; $a = $ HZ; $Kp < 16$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 2R_e$; $a = 5R_*$; $Kp < 15$
$N_{\rm tr} \geqslant 3$; $R_{p,\rm min} \leqslant 2R_e$; $a = 5R_*$; $Kp < 16$

FIG. 15.— Prioritization applied by Batalha et al [2010], from whom I copy-pasted this Table. I did the same, but using r magnitudes instead of Kp.

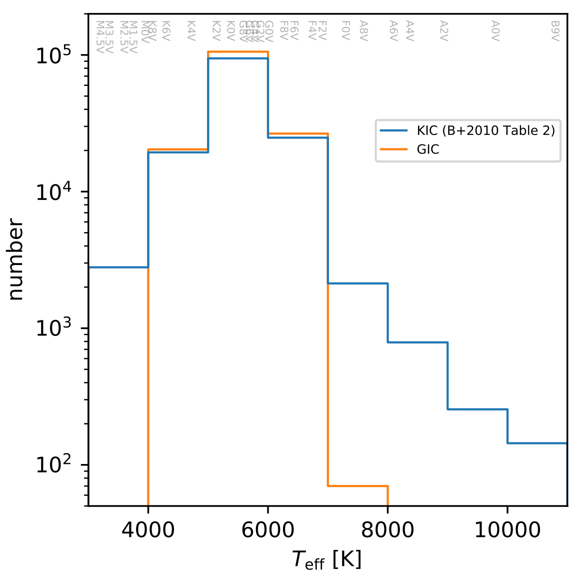

FIG. 16.— Effective temperature distribution of the stars in our synthetic Kepler input catalog analog – the Galaxia input catalog. Note that the effective temperatures of binary star systems in the GIC are intentionally wrongly estimated, as if they were single star systems. The listed $T_{\rm eff}$ to spectral type correspondence is from Mamajek's tables.

priority classes of Fig. 15, to obtain a catalog of 152682 stellar systems, which we hereafter refer to as the *Galaxia Input Catalog (GIC)*. We compare the effective temperature distribution of the GIC with that of the KIC, reported by Batalha et al [2010] Table 2, in Fig. 16. Evidently, the GIC is dominated by sun-like stars, in a very similar manner to the exoplanet target stars of the KIC.

4.4. *Introduce a planet population*

As in Sec. 3, we allow for three different occurrence rates: $\Gamma_{t,s}$, the fraction of stars in single systems with a planet of radius R_p and orbital period P, and also $\Gamma_{t,d1}$ and $\Gamma_{t,d2}$ for the fraction per sun-like primary and secondary (respectively) of double star systems with a planet of (R_p, P).

We then randomly select which stars get a planet, and using the true stellar masses and radii compute the corresponding impact parameters for the planets as

$$b = \frac{a}{R_\star}\cos i, \quad \cos i \sim \mathcal{U}(0,1). \tag{84}$$

Any planet with $|b| < 1$ is transiting, and has its transit duration evaluated as

$$T_{\rm dur} = 13\,{\rm hr}\left(\frac{P}{\rm yr}\right)^{1/3}\left(\frac{\rho_\star}{\rho_\odot}\right)^{-1/3}\sqrt{1-b^2}, \tag{85}$$

which assumes circular orbits.

4.5. *Run the survey*

We run the following survey: $A = 0.708\,{\rm m}^2$ (*Kepler* effective area), $T_{\rm obs} = 20\,$years (to detect more Earth-like planets than *Kepler* did), and $x_{\rm min} = 7.1$, the same signal to noise threshold as *Kepler*.

For simplicity, we assign a r band zero point of $10^6\,{\rm ph/s/cm}^2$ to correspond to an $r = 0$. We use this to compute the observed photon number fluxes for every system, regardless of whether it has a planet. We then compute the signal to noise distribution of transit events, and the number of detections, according to Eq. **??**.

4.6. *Numerically assess what occurrence rates would be derived, and compare to previous models*

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

This is similar to what we did in Sec. 3.10.2, with the exception that we need to specify the calculation more precisely.

Specifically, our "Astronomer A" in this case will estimate an occurrence rate for planets of radius R_p and orbital period P as

$$\Gamma_{{\rm A},R_p} = \frac{N_{\rm det,s}}{Z}, \tag{86}$$

for

$$Z \approx (N_s + N_d)\frac{1}{J}\sum_{j=1}^{J} Q_j, \tag{87}$$

for Q_j the total detection efficiency, indexed over single and double systems, given as the product of the geometric transit probability and the system-specific completeness:

$$Q_j \equiv f_g^{(j)} f_c^{(j)}. \tag{88}$$

Keep in mind that for Astronomer A, we need to compute the *incorrect* transit probability for double star systems (since Astronomer A is assuming that they are single stars).

To evaluate the system-level completeness, we assume that Astronomer A performs something equivalent to perfect injection-recovery. For a stellar system with known properties (stellar and planetary), and a perfectly known noise model, the probability of detection is a binary function: either $S/N > (S/N)_{\rm min}$, or it is not[11]. So

[11] There are probabilities involved if we allow for any uncertainty in the parameters, but this is too complicated for our current goal.

we compute $f_c^{(j)}$ as Astronomer A would: an array (over star systems) of zeros wherever the (R_p, P) planet could not be detected, and ones where it could. We then evaluate the total detection efficiency Q_j for each system, and compute the occurrence rate as in Eq. 86. The resulting occurrence rates are only slightly smaller than those in Sec. 4.6.2 discussed below and shown in Fig. 17.

4.6.2. *What if we ignore binarity, but count planets whose derived radii are "close enough" to R_p?*

As in Sec. 3, we now count planets with observed radii from $(1-x)R_\oplus < R_p^{\rm obs} < R_\oplus$. We'll use $x = 0.1$, but it can be whatever we want. The occurrence rate in this case will be

$$\Gamma_{A,R_p'} = \frac{N_{\rm det,s} + N_{\rm det,d}(R_p')}{Z'}, \tag{89}$$

for

$$Z' \approx (N_s + N_d)\frac{1}{J}\sum_{j=1}^{J} Q_j', \tag{90}$$

where the geometric transit probability is the same as before, but the system-specific completeness is now a function of the desired radius interval:

$$Q_j' = f_g^{(j)} f_c^{(j)}(R_p'). \tag{91}$$

There are three possible cases for the completeness, specified by the minimum detectable planet radius $R_{p,\rm min}$ (computed according to Eq. 82):

$$f_c^{(j)}(R_p') = \begin{cases} 1 & \text{if } R_{p,\rm min} > R_p, \\ 0, & \text{if } R_{p,\rm min} < (1-x)R_p, \\ \frac{R_p - R_{p,\rm min}}{R_p - (1-x)R_p}, & \text{otherwise.} \end{cases} \tag{92}$$

The latter term above is the case in which the minimum detectable planet radius happens to be between $(1-x)R_p$ and R_p. In that case the completeness becomes the fraction of planets in that interval that would be detected. Assuming a uniform prior for the planet radius distribution $R_p \sim \mathcal{U}((1-x)R_p, R_p)$, this becomes the expression given in Eq. 92.

These occurrence rates can then be compared to the "true occurrence rate" derived by the all-knowing Astronomer D:

$$\Gamma_{{\rm D},R_p} = \frac{\Gamma_{t,s}N_s + (\Gamma_{t,d1} + \Gamma_{t,d2})N_d}{N_s + 2N_d}. \tag{93}$$

A note in passing: the more information-rich parameters for Astronomer D to report would be $(\Gamma_{t,s}, \Gamma_{t,d1}, \Gamma_{t,d2})$, rather than $\Gamma_{{\rm D},R_p}$. But I digress.

Similar to Fig. 5, Fig. 17 shows that the magnitude of the correction needed by the $\Gamma_{{\rm A},R_p'}$ estimate depends strongly on the fraction of secondaries that are assumed to host the (R_p, P) planets. The simplest assumption is that the occurrence rate is the same for all solar type stars: any "Sun-like" $(0.7 < M_\star/M_\odot < 1.3)$ star hosts planets at the "true rate" Γ_t, and any other mass of star does not. Given how we have defined the GIC, this means that all single stars and primaries of double star systems are sun-like. The fraction of Sun-like secondaries in this case is 37.4% of all secondaries. Reading the appropriate value off Fig. 17, this corresponds to $\Gamma_{\rm D} = 1.39\Gamma_{\rm A'}$. This gives a relative percentage error of $\delta\Gamma_{{\rm D},R_p} \equiv (\Gamma_{\rm A,R_p'} - \Gamma_{\rm D,R_p})/\Gamma_{\rm D,R_p} = -28\%$.

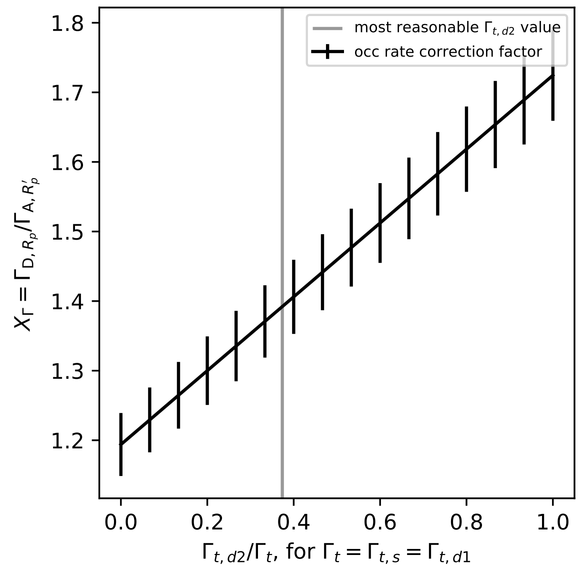

FIG. 17.— The "most reasonable" $\Gamma_{t,d2}$ value is obtained by assuming that the occurrence rate of (R_p, P) planets is the same for all solar type stars, and zero otherwise. This applies most immediately to the question of Earth-like planets around Sun-like stars.