CONSTRUCTION OF COUPLED PERIOD–MASS FUNCTIONS IN EXTRASOLAR PLANETS THROUGH A NONPARAMETRIC APPROACH

ING-GUEY JIANG1, LI-CHIN YEH2, YEN-CHANG CHANG3, AND WEN-LIANG HUNG3

1 Department of Physics and Institute of Astronomy, National Tsing Hua University, Hsin-Chu, Taiwan
2 Department of Applied Mathematics, National Hsinchu University of Education, Hsin-Chu, Taiwan
3 Graduate Institute of Computer Science, National Hsinchu University of Education, Hsin-Chu, Taiwan

Received 2008 May 8; accepted 2008 October 20; published 2008 December 15

ABSTRACT

Using the period and mass data of 279 extrasolar planets, we have constructed a coupled period–mass function through a nonparametric approach. This analytic expression of the coupled period–mass function has been obtained for the first time in this field. Moreover, due to a moderate period–mass correlation, the shapes of mass/period functions vary as a function of period/mass. These results of mass and period functions give way to two important implications: (1) the deficit of massive close-in planets is confirmed, and (2) the more massive planets have larger ranges of possible semimajor axes. These interesting statistical results will provide important clues to the theories of planetary formation.

Key words: methods: data analysis – methods: numerical – methods: statistical – planetary systems

1. INTRODUCTION AND MOTIVATION

After the first detection of an extrasolar planet (exoplanet) around a millisecond pulsar in 1992 (Wolszczan & Frail 1992), it was soon reported that another exoplanet, the first one around a Sun-like star, i.e., 51 Pegasi b, was found (Mayor & Queloz 1995). Since then, there has been a continuous flood of discoveries of extrasolar planets. As of 2008 February, more than 200 planets have been detected around solar-type stars. These discoveries have led to a new era in the study of planetary systems. For example, the traditional theory for the formation of the solar system does not likely explain certain structures of extrasolar planetary systems. This is due to the properties discovered in extrasolar planetary systems being quite unlike our own. Many detailed simulations and mechanisms have been proposed to explore these important issues (Jiang & Ip 2001; Kinoshita & Nakai 2001; Armitage et al. 2002; Ji et al. 2003; Jiang & Yeh 2004a; Jiang & Yeh 2004b; Boss 2005; Jiang & Yeh 2007; Rice et al. 2008).

As the number of detected exoplanets keeps increasing, the statistical properties of exoplanets have become more meaningful. For example, assuming that the mass and period distributions are two independent power-law functions, Tabachnik & Tremaine (2002) used the maximum likelihood method for determining the best power index. However, the possibility of a mass–period correlation is not addressed in their work. Zucker & Mazeh (2002) determined the correlation coefficient between mass and period in logarithmic space and concluded that the mass–period correlation is significant.

On the other hand, a clustering analysis of the data we have on exoplanets also gives some interesting results. Jiang et al. (2006) took the first step into clustering analysis and found that the mass distribution is continuous, and the orbital population could be classified into three clusters that correspond to the exoplanets in the regimes of tidal, ongoing tidal, and disk interaction. Marchi (2007) also worked on clustering through different methods.

To take things a step further from the mass–period distribution function of Tabachnik & Tremaine (2002) and the mass–period correlation of Zucker & Mazeh (2002), Jiang et al. (2007; hereafter JYCH07) employed an algorithm to numerically construct a coupled mass–period function. They were able to include the possible correlation between mass and period into the distribution function for the first time in this field and obtained a distribution function that found the correlation to be consistent. In fact, the mass–period distribution obtained by JYCH07 should be called the mass–period probability density function (pdf) in statistics. The integral of pdf is then called the cumulative distribution function (cdf).

Although JYCH07 successfully numerically constructed the coupled mass–period pdf, due to constraints in the algorithm they employed, they were forced to use the parametric approach of beta distribution on the pdf fitting. The pdf is a basic characteristic describing the behavior of random variables, i.e., mass and period, and is so important that one has to carefully choose the underlying functional form. One possibility to address this problem is to use a nonparametric approach. This is because a nonparametric approach is a distribution-free inference, i.e., an inference that is made without any assumptions regarding the functional form of the underlying distribution. In addition, the most valuable indication of a nonparametric approach is to let the data speak for itself. We therefore see no other reasonable course of action than to use a nonparametric approach in this paper.

Moreover, we still consider the period–mass coupling even while the pdf and cdf are being constructed. In order to make it possible to proceed, we will employ a method called “Copula Modeling” for obtaining the coupled pdf and cdf on the period and mass of exoplanets. This method is more general than the one used in JYCH07 so that a nonparametric approach can be used to obtain the coupled pdf. Copula modeling has a long history of development and was too complicated to be used with real data, in practical terms, until Trivedi & Zimmer (2005) clearly demonstrated a standard modeling procedure.

In Section 2, we briefly describe the data and in Section 3, an estimation of the nonparametric approach is presented. In Section 4, we introduce the method of copula modeling and demonstrate its credibility. The copula modeling is then directly applied to the data of exoplanets. The results are described and discussed in Section 5. Our main conclusions are found in Section 6.
Figure 1. Cumulative distribution function (cdf) and probability density function (pdf) of planetary period and mass. (a) The period cdf. (b) The mass cdf of the minimum-mass model (solid curve) and the guess-mass model (dotted curve). (c) The histogram of planets in p space and also the period pdf $\hat{f}_p(p)$ (solid curve). (d) The histogram of planets in m space of the minimum-mass model (solid line) and the guess-mass model (dotted line), and also the mass pdf of the minimum-mass model $\hat{f}_m(m)$ (solid curve) and the guess-mass model $\hat{f}_g(m)$ (dotted curve).

2. THE DATA

We took samples of exoplanets from The Extrasolar Encyclopedia (http://exoplanet.eu/catalog-all.php) on 2008 April 10. Our samples do not include OGLE235-MOA53b, 2M1207b, GQ Lupb, AB Pic b, SCR 1845b, UScoCTIO108b, or SWEEPS-04 because either their mass or their period data was not listed. The outlier, PSR B1620-26b, with a huge period (100 years), is also excluded.

The data of orbital periods is taken directly from the table in The Extrasolar Encyclopedia. As a result, only the values of projected mass ($m \sin i$) are listed and only a small fraction of exoplanets’ inclination angles $i$ are known, so we decided to provide two models of planetary mass in this paper. For the “minimum-mass model,” we simply set $\sin i = 1$ for all planetary systems in the data. For the “guess-mass model,” an inclination angle $i$ within the observational constraint is assigned to a planetary system through a random process and the mass is then determined accordingly. In this case, if the inclination angle $i$ is given in The Extrasolar Encyclopaedia for a particular planet, we simply use its value. If there is no mention of observational constraints, the angle $i$ will be randomly chosen between 0° and 90°. Please note that the unit of period is days, and the unit of mass is Jupiter Mass ($M_J$).

3. THE NONPARAMETRIC APPROACH

Considering $n$ data points of extrasolar planets in period and mass spaces, i.e., $(p_1, m_1), (p_2, m_2), \ldots, (p_n, m_n)$, $F_P(p)$ and $F_M(m)$, are the cdfs of period and mass and $f_P(p)$ and $f_M(m)$ are the pdfs of period and mass, respectively. An estimate of the cdf, $F_P(p)$, at the point $p$ is the proportion of samples that are less than or equal to $p$

$$\hat{F}_P(p) = \frac{1}{n+1} \sum_{j=1}^{n} I(p_i \leq p),$$

(1)

where $I(\cdot)$ is the indicator function defined by

$$I(p_i \leq p) = \begin{cases} 1, & \text{if } p_i \leq p, \\ 0, & \text{if } p_i > p. \end{cases}$$

Similarly, the nonparametric estimate of the cdf, $F_M(m)$, at the point $m$ is

$$\hat{F}_M(m) = \frac{1}{n+1} \sum_{j=1}^{n} I(m_i \leq m).$$

(2)

The solid curves in Figures 1(a) and 1(b) are $\hat{F}_P(p)$ and the minimum-mass model’s $\hat{F}_M(m)$, respectively. The dotted curve in Figure 1(b) is the guess-mass model’s $\hat{F}_M(m)$.

To obtain the analytic expressions for the pdfs $f_P(p)$ and $f_M(m)$, we first plot the histograms in $p$ and $m$ spaces, as shown in Figures 1(c) and 1(d). In these two histograms, we choose the bandwidths $h_P$ and $h_M$ of $p$ and $m$ as follows (Silverman 1986, page 47):

$$h_P = 0.9 A_P n^{-1/5}, \quad h_M = 0.9 A_M n^{-1/5},$$
where

\[ A_P = \min \left\{ S_P, \frac{IQR_P}{1.34} \right\}, \quad A_M = \min \left\{ S_M, \frac{IQR_M}{1.34} \right\}, \]

\( S_P \) and \( IQR_P \) \((IQR_M)\) are the standard deviation
and interquartile range of \( p_1, \ldots, p_n \) \((m_1, \ldots, m_n)\), respectively. Here the interquartile range is the difference
between the first and third quartiles (also see this definition in Section 5.1). In
our data, \( S_P = 896.464, \ IQR_P = 846.360, \) and \( S_M = 3.499 \) \((S_M = 5.235)\), \( IQR_M = 2.520 \) \((IQR_M = 3.694)\) for the
minimum-mass model (guess-mass model).

We then use the adaptive kernel method (Silverman 1986, page 101) to estimate
the pdfs \( f_P(p) \) and \( f_M(m) \) as follows.

1. Finding the pilot estimates,

\[
\tilde{f}_P(p) = \frac{1}{nh_P} \sum_{j=1}^{n} K\left( \frac{p - p_j}{h_P} \right),
\]

\[
\tilde{f}_M(m) = \frac{1}{nh_M} \sum_{j=1}^{n} K\left( \frac{m - m_j}{h_M} \right),
\]

where \( K(\cdot) \) is the Gaussian kernel, i.e.,

\[ K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \]

2. Defining local bandwidth factors \( \lambda^P_i \) and \( \lambda^M_i \) by

\[
\lambda^P_i = \left( \frac{\tilde{f}_P(p_i)}{g_P} \right)^{-1/2}, \quad \lambda^M_i = \left( \frac{\tilde{f}_M(m_i)}{g_M} \right)^{-1/2},
\]

\( i = 1, \ldots, n, \)

where \( g_P (g_M) \) is the geometric mean of \( \tilde{f}_P(p_i) (\tilde{f}_M(m_i)) \),

\[
\ln g_P = \frac{1}{n} \sum_{i=1}^{n} \ln \tilde{f}_P(p_i), \quad \ln g_M = \frac{1}{n} \sum_{i=1}^{n} \ln \tilde{f}_M(m_i).
\]

3. Obtaining the adaptive kernel estimate \( \hat{f}_P(p) (\hat{f}_M(m)) \) of the
pdf \( f_P(p) (f_M(m)) \) by

\[
\hat{f}_P(p) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\lambda^P_j h_P} K\left( \frac{p - p_j}{\lambda^P_j h_P} \right),
\]

\[
\hat{f}_M(m) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\lambda^M_j h_M} K\left( \frac{m - m_j}{\lambda^M_j h_M} \right).
\]

Thus, the analytic expressions are obtained. In order to
compare these with the histograms, we define \( \tilde{f}^h_P(p) \equiv \text{Area}_P \times \tilde{f}_P(p) \),
where \( \text{Area}_P = 51.24068 \) is the area under the
histogram of the period in Figure 1(c). Similarly, we set
\( \tilde{f}^h_M(m) \equiv \text{Area}_M \times \tilde{f}_M(m) \), the minimum-mass model,
where \( \text{Area}_M = 153.066 \), and set \( \tilde{f}^g_P(p) \equiv \text{Area}_P \times \tilde{f}_P(m) \),
the guess-mass model, where \( \text{Area}_M = 217.20 \). \( \tilde{f}^h_P(p) \) is plotted
as the solid curve in Figure 1(c), \( \tilde{f}^h_M(m) \) is the solid curve in
Figure 1(d), and \( \tilde{f}^g_P(p) \) is the dotted curve in Figure 1(d).

4. THE COPULA MODELING METHOD

In this section, we will describe the procedure to construct
a new period–mass pdf, in which the possible period and mass

\[ C(u_1, u_2; \theta) = \frac{-1}{\theta} \ln \left[ 1 + \left( \frac{e^{-\theta u_1} - 1}{e^{-\theta} - 1} \right)^{-1} \right]. \]

where \( u_1, u_2 \) \((0 \leq u_1, u_2 \leq 1)\) are two marginal distribution
functions and \( \theta (-\infty < \theta < \infty) \) is the dependence parameter.
Positive, zero, and negative values of \( \theta \) correspond to the positive
dependence, independence, and negative dependence between
two marginal variables, respectively.

For our work here, \( u_1 \) is the cdf of period, \( F_P(p) \), and \( u_2 \) is the
cdf of mass, \( F_M(m) \). In copula modeling, the pdf of the coupled
period–mass distribution is

\[
\frac{1}{\partial F_P \partial F_M} \left( \frac{\partial^2 C(F_P(p), F_M(m); \theta)}{\partial F_P \partial F_M} \right) = \frac{1}{e^{-\theta} - 1 + (e^{-\theta} - 1)(e^{-\theta} - 1)} f_P(p) f_M(m). \]

We now have an analytic form of the coupled period–mass
pdf where the parameter \( \theta \) is to be determined through the maximum
likelihood method.

The log–likelihood function of \( \theta \) for the samples \( (p_i, m_i), i = 1, \ldots, n \) can be written as

\[ \ell(\theta) = \ell_1 + \ell_2(\theta), \]

where

\[
\ell_1 = \sum_{i=1}^{n} [\ln f_P(p_i) + \ln f_M(m_i)], \quad \ell_2(\theta) = n \ln[-(e^{-\theta} - 1)] - \sum_{i=1}^{n} \left\{ \theta [f_P(p_i) + F_M(m_i)] \right. \]

\[ + 2 \ln(e^{-\theta} - 1 + (e^{-\theta} - 1)(e^{-\theta} F_M(m_i) - 1)) \right\}. \]
Differentiating $\ell(\theta)$ with respect to $\theta$, we obtain
\[
\frac{\partial \ell(\theta)}{\partial \theta} = \frac{\partial \ell_2(\theta)}{\partial \theta} = \sum_{i=1}^{n} \left\{ e^{-\theta} - 1 - \theta e^{-\theta} - [F_M(m_i) + F_P(p_i)] \ight\}
\]
\[
+ 2 \cdot (e^{-\theta} + F_M(m_i)e^{-\theta}F_u(m_i)(e^{-\theta}F_P(p_i) - 1)) / (e^{-\theta} - 1)
\]
\[
+ (e^{-\theta}F_u(m_i) - 1)(e^{-\theta}F_P(p_i) - 1)\right\}.
\]

After the estimates of cdfs $\hat{F}_P(p)$ and $\hat{F}_M(m)$ have been substituted into $\partial \ell(\theta)/\partial \theta$, the estimate of $\theta$ is obtained by solving
\[
\frac{\partial \ell(\theta)}{\partial \theta} = 0.
\]

Moreover, according to Genets (1987), the parameter $\theta$ in copula modeling is related to the Spearman rank–order correlation coefficient ($\rho_G$) through the following formula:
\[
\rho_S \approx \rho_G \equiv (1 - \theta e^{-\theta/2} - e^{-\theta})(e^{\theta/2} - 1)^{-2}.
\]  

We call this $\rho_G$ the Genets correlation coefficient in this paper.

4.2. The Credibility Test

Since this is the first time that copula modeling has been introduced and employed in astronomy, we shall demonstrate its credibility. We will generate four sets of 279 artificial data points of uniform random variables $x$ and $y$, with different strength of $x$–$y$ correlations as presented in Figures 2(a)–2(d). The Spearman correlation coefficients $\rho_S$ (see Section 5.2 for the definition) between $x$ and $y$ are in Table 1. We apply the copula modeling on these four sets of experimental data, where the nonparametric approach is used to obtain the cdfs of $x$ and $y$. Finally, the coupled $x$–$y$ pdf, the coupling parameter $\theta$, and $\rho_G$ are obtained. We also calculate the bootstrap confidence interval (C.I.) for $\theta$ and $\rho_G$ with the number of bootstrap replications $B = 2000$ (JYCH07). These results are all listed in Table 1.

Because the values of Spearman correlation coefficient $\rho_S$ are close to $\rho_G$ and within $\rho_G$’s 95% confidence intervals, we confirm that copula modeling gives the correct coupling parameter $\theta$ and the Genets correlation coefficients $\rho_G$. Thus, the coupling between $x$ and $y$ can be correctly included when the pdf is constructed for any given strength of correlation.

5. RESULTS

In this section, the results of the coupled period–mass distribution and the correlation coefficients will be presented.

5.1. The Coupled Period–Mass Distribution

Using the copula modeling, the estimate of $\theta$ is $\hat{\theta} = 2.3826$ for the minimum-mass model. Through the bootstrap algorithm as described in JYCH07 with the number of bootstrap replications $B = 2000$, the standard error of $\hat{\theta}$ is 0.3669. In order
Figure 3. Three-dimensional view of the coupled period–mass pdf, $f_{P,M}(p, m \mid \theta)$, of the guess-mass model.

to properly understand the dependence parameter $\theta$, we also obtain the 95% bootstrap C.I. for $\theta$, which is $(1.6514, 3.1190)$.

For the guess-mass model, the estimate of $\theta$ is $\hat{\theta} = 2.4565$ and its 95% bootstrap C.I. is $(1.7282, 3.1633)$. 

Figure 4. Color contour of the coupled period–mass pdf, $f_{P,M}(p, m \mid \theta)$, of the guess-mass model.
Furthermore, in order to check the stability of the guess-mass model, we repeat the random process to generate 100 guess-mass models and apply copula modeling on them. The average value of $\hat{\theta}$ is 2.9249 with the standard deviation 0.3349. We then employ the interquartile range (Turky 1977) to check for any outliers of $\hat{\theta}$ from these 100 guess-mass models. The interquartile range is the difference between the first quartile $Q_1$ and the third quartile $Q_3$, i.e., $IQR = Q_3 - Q_1$. Inner fences
are left and right from the median at a distance of 1.5 times the IQR. Outer fences are at a distance of three times the IQR. The values lying between the inner and outer fences are called suspected outliers and those lying beyond the outer fences are called outliers (Hogg & Tanis 2006).

The smallest, first quartile, median, third quartile, and largest of these 100 $\hat{\theta}$ values, denoted by Min, $Q_1$, Me, $Q_3$, Max, respectively, are

$$
\text{Min} = 2.3730, \quad Q_1 = 2.6297, \quad \text{Me} = 2.8833, \quad Q_3 = 3.1968, \quad \text{Max} = 3.5776.
$$

Therefore, IQR = 0.5671 and cutoffs for outliers are

$$
Q_3 + 1.5\text{IQR} = 4.0475, \quad Q_3 + 3\text{IQR} = 4.8981, \quad Q_1 - 1.5\text{IQR} = 1.7791, \quad Q_1 - 3\text{IQR} = 0.9284.
$$

Furthermore, we find that

$$
Q_1 - 1.5\text{IQR}(= 1.7791) < \text{Min}(= 2.3730) < \text{Max}(= 3.5776) < Q_3 + 1.5\text{IQR}(= 4.0475).
$$

Thus, all 100 $\hat{\theta}$ values of the guess-mass model lie within the inner fences. It means that no outliers exist in these 100 values and so the stability of the guess-mass model is confirmed.

Figure 3 shows the three-dimensional view of the coupled period–mass pdf, $f_p(m,p,|\theta|)$, of the guess-mass model. The contour of Figure 3 is presented in Figure 4. The plots of the minimum-mass model’s $f_{p,M}(p,m|\theta)$ are very similar to the above, so we have not shown them.

We know that when the period and mass are completely independent, $f_{p,M}(p,m|\theta) = f_p(p)f_M(m)$. Thus the term

$$
-\theta(e^{-\theta} - 1)e^{-\theta}f_p(p)e^{-\theta}f_M(m)
$$

in Equation (4) is the one to take the period–mass coupling into account. We will hereafter call it the coupling factor. To make it clear how the coupling factor behaves, its value as a function of $p$ and $m$ of the guess-mass model is plotted in Figure 5. Figure 6 is the color contour plot. It clearly shows that the coupling factor becomes larger than one when both period and mass are very small or when both of them are large (area of red). It also shows that the coupling factor is less than one in the blue area.

5.2. The Correlation Coefficients

JYCH07 calculated the linear correlation coefficients (also called Pearson’s correlation coefficients) in both $m − p$ and $\ln(m) − \ln(p)$ spaces and found a weak correlation in $m−p$ and a moderate correlation in $\ln(m)−\ln(p)$ space. In order to maintain a consistent determination on the correlation coefficients, we
now calculate the Spearman rank-order correlation coefficients (Press et al. 1992), which are invariant under strictly increasing nonlinear transformations (Schweizer and Sklar 2005).

For pairs of quantities \((x_i, y_i), i = 1, \ldots, n\), the linear correlation coefficient is given by

\[
r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}},
\]

where \(\bar{x} = \sum_{i=1}^{n} x_i / n\) and \(\bar{y} = \sum_{i=1}^{n} y_i / n\). The Spearman rank-order correlation coefficient is calculated by the above formula with \(x_i\) and \(y_i\) replaced by their ranks. Given \(R(x_i)\) the rank of \(x_i\) and \(R(y_i)\) the rank of \(y_i\), the Spearman rank-order correlation coefficient can be written as

\[
\rho_S = 1 - \frac{6}{n^3 - n} \sum_{i=1}^{n} [R(x_i) - R(y_i)]^2. \tag{10}
\]

We note that \(|\rho_S| = 1\) indicates a perfect dependence and \(\rho_S = 0\) means no dependence. When \(\rho_S = 1\) there is a direct perfect dependence and when \(\rho_S = -1\) there is an inverse perfect dependence. Furthermore, according to Cohen (1988), \(0.1 < |\rho_S| < 0.3\) means the correlation is weak, \(0.3 < |\rho_S| < 0.5\) indicates a moderate correlation, and \(0.5 < |\rho_S| < 1.0\) is indicative of strong correlation.

For the minimum-mass model, the Spearman rank-order correlation coefficient is obtained as \(\rho_S = 0.3769\). Through copula modeling, we also find the estimate of \(\rho_G\), which is \(\hat{\rho}_G = 0.3792\). It is obvious that the Spearman rank-order correlation coefficient \(\rho_S = 0.3769\) is very close to \(\hat{\rho}_G\). Moreover, the 95% bootstrap C.I. with the number of bootstrap replications \(B = 2000\) for \(\rho_G\) is \((0.2691, 0.4811)\). For the guess-mass model, we have \(\hat{\rho}_G = 0.3899\) with a 95% bootstrap C.I. \((0.2811, 0.4869)\). These results are all consistent and confirm that there is a positive period–mass correlation for exoplanets.

6. CONCLUSIONS

Using the data of exoplanets, for the first time in this field we have constructed an analytic coupled period–mass function through a nonparametric approach. Moreover, we calculate the Spearman rank-order correlation coefficient, which gives the same results for linear and logarithmic spaces, and the results in the previous section show that there is a moderate positive period–mass correlation.

In order to comprehend the implication of our results, in Figures 7(a) and 7(b), we plot \(f_{p,M}(p, m|\theta)\) with \(m = 1, 5, 10, \) and \(15 M_J\) (i.e., the period functions given different masses), and also \(f_{p,M}(p, m|\theta)\) with \(p = 1, 50, 100,\) and 150 days (i.e., the mass functions given different periods) in logarithmic spaces. For purposes of comparing, \(f_P(p) \times f_M(m)\) with \(m = 1, 5, 10, \) and \(15 M_J\) (the independent period functions) and \(f_P(p) \times f_M(m)\) with \(p = 1, 50, 100,\) and 150 days (the independent mass functions) are also plotted in Figures 7(c) and 7(d). Of course, the shapes of independent period functions with \(m = 1, 5, 10, \) and \(15 M_J\) are all the same, and the shapes of independent mass functions given different periods are all exactly as well.

We find that the period function of \(m = 1 M_J\) is very similar with the independent period functions. However, the period functions of \(m = 5, 10,\) and \(15 M_J\) are different from the independent ones, in a way that the functions are lower at the smaller \(p\) end and slightly higher at the larger \(p\) end. Thus, the overall period functions of massive planets (say \(m = 5, 10,\) and \(15 M_J\)) at large \(p\) and small \(p\) ends are closer than that of lighter planets (say \(m = 1 M_J\)). Therefore, the fractions of larger and smaller \(p\) (or semimajor axis) planets are closer for those planets with mass \(m = 5, 10,\) and \(15 M_J\).

This implies that the more massive planets have larger ranges of possible semimajor axes. This interesting statistical result will provide important clues to the theories of planetary formation.

On the other hand, the mass functions of \(p = 50, 100,\) and 150 days are all very similar with the independent mass functions. However, the mass function of \(p = 1\) day is different from the independent one in a way that the function is higher at the smaller \(m\) end and lower at the larger \(m\) end. Thus, the mass function of short period planets (say \(p = 1\) day) is steeper than that of long period planets (say \(p = 50, 100,\) and 150 days). This implies that the percentage of massive planets are relatively small for the short-period planets. This result reconfirms the deficit of massive close-in planets due to tidal interaction as studied in Jiang et al. (2003).

We are grateful to the referee’s suggestions. This work is supported in part by the National Science Council, Taiwan.

REFERENCES

Armitage, P. J., Livio, M., Lubow, S. H., & Pringle, J. E. 2002, MNRAS, 334, 248
Boss, A. P. 2005, ApJ, 629, 535
Cohen, J. 1988, Statistical Power Analysis for the Behavioral Sciences (Philadelphia, PA: Lawrence Erlbaum Associates)
Frank, M. J. 1979, Aequationes Math., 19, 194
Frees, E. W., & Valdez, E. A. 1998, J. North Am. Actuarial, 2, 1
Genest, C., & MacKay, J. 1986, The American Statistician 4, 280
Genets, C. 1987, Biométrieka, 74, 549
Hogg, R. V., & Tanis, E. A. 2006, Probability and Statistical Inference (Upper Saddle River, NJ: Pearson Prentice Hall)
Ji, J., Kinoshita, H., Liu, L., & Li, G. 2003, ApJ, 585, L139
Jiang, I.-G., & Yeh, L.-C. 2003, ApJ, 582, 449
Jiang, I.-G., & Yeh, L.-C. 2004a, MNRAS, 355, L29
Jiang, I.-G., & Yeh, L.-C. 2004b, Int. J. Bifurcation Chaos, 14, 3153
Jiang, I.-G., & Yeh, L.-C. 2007, ApJ, 656, 534
Jiang, I.-G., & Yeh, L.-C., Chang, Y.-C., & Hung, W.-L. 2007, AJ, 136(1JYCH07)
Jiang, I.-G., & Yeh, L.-C., Hung, W.-L., & Yang, M.-S. 2006, MNRAS, 370, 1379
Kinoshita, H., & Nakai, H. 2001, PASI, 53, L25
Klugman, S. A., & Pusa, P. 1999, Insurance: Mathematics and Economics, 24, 139
Marchi, S. 2007, ApJ, 666, 475
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
Press, W. H., et al. 1992, Numerical Recipes in Fortran (Cambridge: Cambridge Univ. Press)
Rice, W. K. M., Armitage, P. J., & Hogg, D. F. 2008, MNRAS, 384, 1242
Schweizer, B. W. 1986, Density Estimation for Statistics and Data Analysis (New York: Chapman and Hall)
Tabachnik, S., & Tremaine, S. 2002, MNRAS, 335, 151
Trivedi, P. K., & Zimmer, D. M. 2005, Foundations & Trends in Econometrics, 1, 1
Tukey, W. J. 1977, Exploratory Data Analysis (Reading, MA: Addison-Wesley)
Venter, G., Barnett, Kreps, R., & Major, J. 2007, Variance, 1, 103
Wolszczan, A., & Frail, D. A. 1992, Nature, 355, 145
Zucker, S., & Mazeh, T. 2002, ApJ, 568, L113