

基于Kepler Q1–Q17数据对类太阳型恒星的行星生成率估算*

张青欣^{1,2,3} 暴春晖^{1,2,3} 季江徽^{1,2,3†}

(1 中国科学院紫金山天文台 南京 210023)

(2 中国科学技术大学天文与空间科学学院 合肥 230026)

(3 中国科学院行星科学重点实验室 南京 210023)

摘要 近30年来的系外行星探测揭示了行星在宇宙中普遍存在的事实. 为了深入研究适宜生命居住行星的普遍性, 一方面需要了解宜居行星的特性; 另一方面可以通过分析已发现系外行星的分布特征, 推算该类行星在恒星周围的存在几率. 在目前已发现的系外行星中, 凌星法发现的占据了绝大多数, 如Kepler空间望远镜所观测的系外行星共有2344颗. 2018年Kepler正式退役, 其科学团队发布了最终版的Kepler Data Release (DR25), 包含观测季度Q1–Q17的恒星共198709颗. 通过对Kepler数据的分析, 使用逆检测效率法和最大似然分析法两种不同的方法对系外行星半径周期参数空间内的行星生成率进行了估算, 同时将计算样本根据恒星的光谱类型进行分类, 分别估算得到了F、G、K型的Kepler恒星周围的行星生成率及其整体的生成率. 对于半径范围 $1\text{--}20 R_{\oplus}$ (R_{\oplus} 为一个地球半径), 轨道周期范围0.4–400 d的Kepler凌星系外行星, 宿主恒星为F型时逆检测效率法和最大似然法估算得到的行星生成率分别为 0.36 ± 0.02 和 0.47 ± 0.02 , 宿主恒星为G型时的行星生成率分别为 1.62 ± 0.05 和 1.23 ± 0.04 , 宿主恒星为K型恒星时的行星生成率分别为 2.61 ± 0.12 和 2.73 ± 0.13 . 而逆检测效率法和最大似然法估算F、G、K型恒星周围存在这类行星的整体生成率分别为 1.16 ± 0.03 和 0.90 ± 0.02 . 最后, 对比分析了不同方法以及不同光谱型恒星间的行星生成率差异, 并结合前人研究讨论了该估算结果的可靠性.

关键词 行星与卫星: 普通, 凌星系外行星, 行星生成率, 宜居行星, 开普勒空间望远镜

中图分类号: P145; 文献标识码: A

1 引言

截至2022年5月8日, 约有5000颗已被证实的系外行星被人类发现, 包括810个多行星系统, 其中由Kepler空间望远镜所发现的系外行星共有2344颗. 这些发现告诉我们系外行星在宇宙中普遍存在的

事实. 根据太阳系行星的质量大小与形态, 通常可将行星分为岩石行星(类地行星)、气态巨行星(类木行星)和冰巨星(类海王星)等类型; 在系外行星中发现了很多轨道特殊的族群—热木星、超级地球、温暖海王星、超短周期行星等. 由Kepler发现的不

2022-05-10收到原稿, 2022-05-23收到修改稿

*国家自然科学基金项目(12033010、11773081), 中国载人航天工程巡天空间望远镜专项(CMS-CSST-2021-B12), 中科院空间科学先导专项(XDA15020800)资助

†jjjh@pmo.ac.cn

同类型宿主恒星的凌星系外行星的周期分布如图1所示, 半径分布如图2所示. 首先, 多样的系外行星族群的不断发现为了解宜居行星是否存在提供了可能, 探测和发现其他恒星周围的宜居行星是系外行星研究重要的科学问题, 因此我们需要回答不同行星族群在宇宙中的分布与存在概率; 其次, 宜居行星的深入研究也将回答是否存在“另外一个地球”或者“人类是否孤独”这样的哲学命题. 研究不同行星族群的在其他恒星周围分布规律和存在几率有助于深入了解一般行星系统中行星的形成与演化, 最终回答“太阳系是否特殊”这样的科学问题. 大量的系外行星的观测样本使从统计上了解行星存在概率成为可能, 即行星生成率的估计, 且由于不同的探测方法与探测任务对不同类型的系外行星的探测灵敏度不同, 如凌星法与视向速度法都更容易发现质量较大、周期较短的行星, 因此独立分析不同轨道周期半径行星的生成率显得十分必要.

早期对巨行星生成率的研究往往是利用视向速度探测项目^[1-2]及Kepler空间望远镜^[3]发现的巨行星样本开展, 研究认为类太阳恒星周围轨道周期短于数年的巨行星的行星生成率约为0.10, Cumming等^[1]对于围绕类太阳恒星运行的最小质量在 $0.3-10 M_{\text{Jup}}$ (M_{Jup} 为一个木星质量)及轨道周期范围在2-2000 d的巨行星的生成率估计为0.105. 还有研究认为巨行星的行星生成率与轨道周期成幂律关系, 但这种单一的幂律周期分布通常会过度预测距离较远(≥ 10 AU)的巨行星数量. 为了更好地匹配观测到的分布, Fernandes等^[4]用分段幂律代替单一幂律且在轨道距离 $\sim 2-3$ AU时发现了巨行星的行星生成率存在一个潜在的峰值^[5], 分段幂律的实现将分布函数拓展到100 AU, 对于质量范围 $0.1-20 M_{\text{Jup}}$ 和 $1-20 M_{\text{Jup}}$ 的系外行星生成率分别为 $0.27^{+0.08}_{-0.05}$ 和 $0.062^{+0.015}_{-0.012}$.

而对于较小的行星, 大约一半的类太阳恒星拥有至少一个轨道周期小于100 d且大小介于地球与海王星之间的行星. 行星形成理论通常没有预言如

此丰富的近距离轨道行星, 甚至一些理论详细地预测了近距离轨道的超级地球或亚海王星是特别罕见的^[6]. 小质量行星生成率高得惊人的观测结果推动了新的行星形成理论的出现, 即小质量行星可以在短周期轨道上原位形成, 而不是先形成于距离恒星较远的地方而后向内轨道迁移^[7-8]. 对于轨道周期小于50 d、最小质量 $3-30 M_{\oplus}$ 的行星(M_{\oplus} 为一个地球质量), 两个独立的视向速度观测项目得出的行星生成率分别为 0.15 ± 0.05 ^[9]和 0.27 ± 0.05 ^[2]. 然而, 利用Kepler数据, 在相同的轨道周期范围内, 半径 $2-4 R_{\oplus}$ 的行星生成率略小, 为 0.13 ± 0.008 ^[10].

观测结果表明在各种类型恒星周围都有系外行星, 而不仅仅局限于类太阳型恒星, 还包括质量非常小的恒星^[11]、低金属丰度恒星^[12]、巨星^[13]及其他演化阶段的恒星(白矮星、脉冲星等)^[14]. 因此, 恒星属性对于行星生成率存在一定影响, 例如伴星、金属丰度、质量等.

对于Kepler观测的候选行星, 约有10% (30%)的主恒星被观测到角距 $1''$ ($4''$)内存在伴星, 其中大部分相对于目标恒星是相当暗淡的^[15]. 其对于凌星信号的稀释效应总体上平均只影响行星半径几个百分点^[15], 在半径推导的不确定性范围内, 因此不会对一般的行星统计产生重大影响. 但是行星半径 $R_p \leq 2R_{\oplus}$ 的系外行星的凌星信号更容易受到稀释, 相关统计数据可能受到更显著的影响^[15-16]. 除了凌星信号稀释效应, 伴星也可以通过许多动力学过程影响行星的存在. 基于视向速度和高分辨率成像观测的研究表明, 近距离恒星质量伴星的存在通常与环绕主恒星行星的生成率较低有关^[17-18].

通常主恒星的整体金属丰度, 与可供行星形成组分的总质量有关, 因此认为行星的生成率和性质可能与主恒星的金属丰度有关是合理的. 对于通过视向速度观测发现的巨行星, 其生成率与恒星金属丰度存在密切联系^[19-20]. 较小的行星, 尤其是那些半径 $R_p \leq 4R_{\oplus}$ 的系外行星, 其行星生成率与主恒星金属丰度无明显相关性^[21-22].

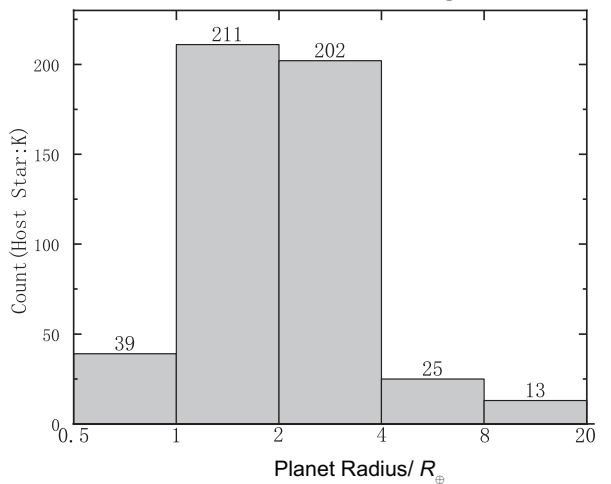
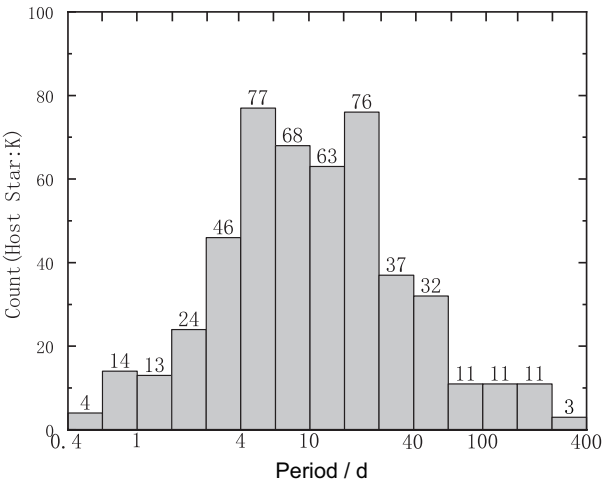
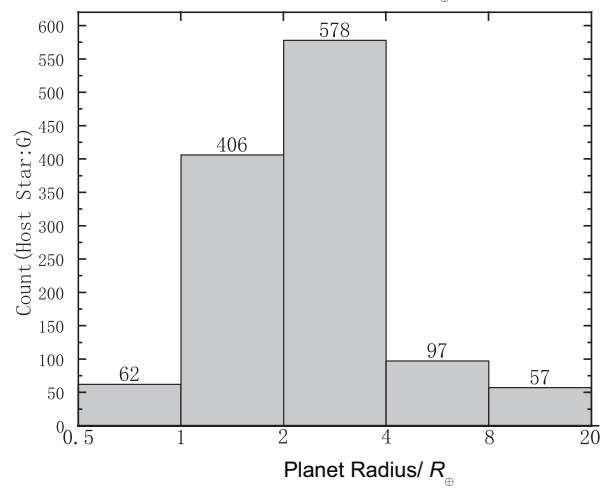
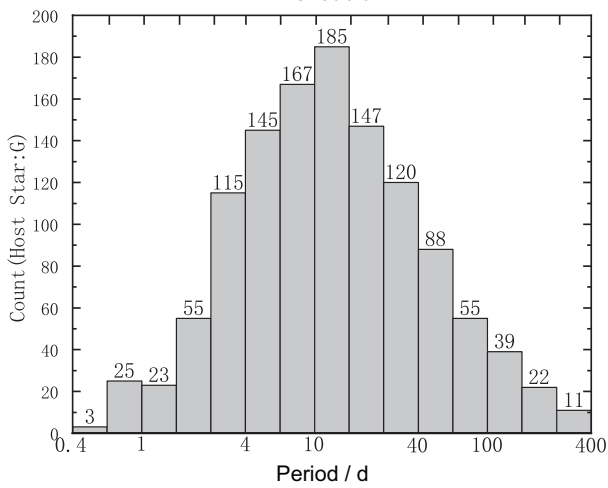
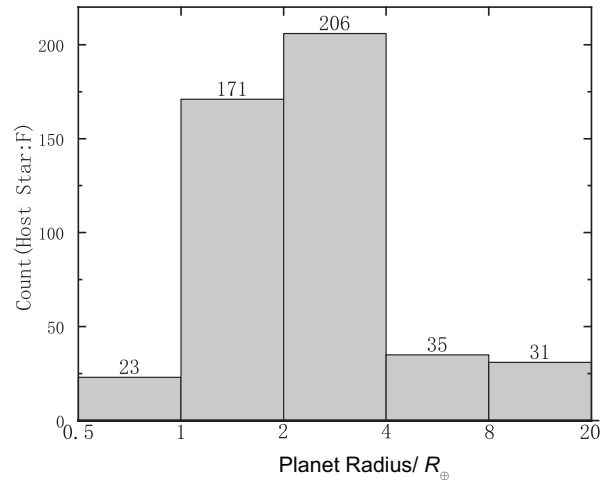
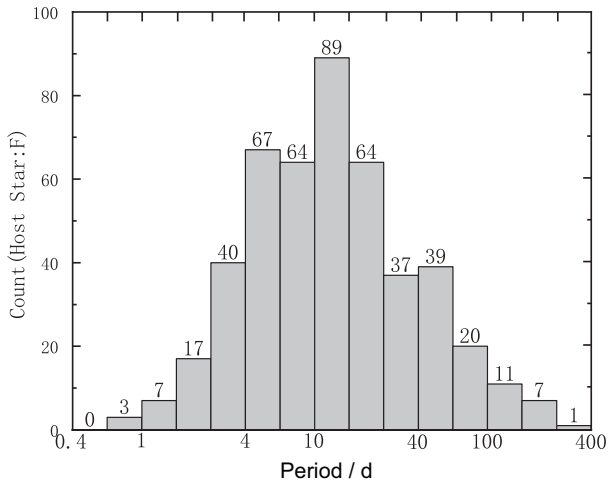


图 1 宿主恒星为F、G、K型的Kepler系外行星周期分布

图 2 宿主恒星为F、G、K型的Kepler系外行星半径分布

Fig. 1 Period distribution of Kepler exoplanets whose host stars are F, G, K spectral types

Fig. 2 Radius distribution of Kepler exoplanets whose host stars are F, G, K spectral types

行星生成率与主恒星质量也可能存在相关性. 质量较低的恒星周围不容易形成行星, 特别是质量较大的行星. 一种理论上的可能性是恒星的质量与原行星盘的总质量相关, 从而与可供形成行星的固体物质的质量相关, 这与在(亚)毫米波长直接对原行星盘的观测一致^[23-24]. 巨行星对恒星质量的依赖性已经用不同的探测方法进行了许多研究^[10, 25]. 对长达3 yr的视向速度观测的1266颗巨行星样本进行分析, 发现在其生成率与恒星质量之间存在线性关系^[26]. 对于较小的行星, Kepler为研究它们与主恒星质量之间的依赖关系提供了最好的样本. 研究表明, 开普勒参数空间中的行星生成率与恒星质量是反相关的^[10, 27].

本文通过分析Kepler的全部的Q1-Q17的最大恒星样本与Kepler已发现确认的系外行星, 使用了逆检测效率法(Inverse Detection Efficiency Method, IDEM)和纳入贝叶斯(Bayesian)框架的最大似然法(Maximum Likelihood Analysis, ML)两种方法, 对Kepler样本的行星生成率进行了估算, 分别讨论了F、G、K型恒星周围的行星生成率, 比较了两种估算方法在F、G、K型恒星周围的估算结果的差异. 同时, 在使用纳入贝叶斯框架的最大似然法时, 我们对净探测效率的估算采用了由Kepler团队开发的估算工具Kepler Planet Occurrence Rate Tools (KeplerPORTs), 提高了对于Kepler空间望远镜进行凌星观测时各种选择效应的估算精度. 在本文第2节将介绍估算行星生成率的两种模型方法; 在第3节介绍具体的数据来源与分析方法, 并给出估算结果; 第4节中我们讨论了两种方法的估算结果, 比较了F、G、K型恒星的行星生成率, 并做了对比分析.

2 估算行星生成率的模型方法

多数与行星生成率相关的研究中, 都认为由于宇宙中行星数量与恒星数量的比值随着观测样本的扩大会趋于一个定值而更有研究意义, 因此本文研究的行星生成率也即每颗恒星拥有行星的平均数量, 称为行星的固有生成率 n_p , 我们令 N_p 为观测样本的行星数量, N_* 为观测样本的恒星数量, n_p 表示为:

$$n_p = \frac{N_p}{N_*}. \quad (1)$$

由于目前观测技术的限制, 对于不同类型行星的探测能力具有差异, 例如Kepler更容易发现周期在1-50 d的行星, 在实际对行星生成率进行估算时, 我们可以将行星限制在被预先定义参数空间内, 通常为周期-半径平面(对于凌星法观测到的系外行星). 通过将参数空间人为划分为多个网格区域^[10], 可以假设单个区域内部行星生成率是一致的, 且不同区域的行星生成率互相独立. 在此基础上, 对每一个网格区域内部的行星生成率 n_{cell} 进行估算是更为合理的, 可以最大限度降低观测灵敏度对行星样本带来的影响. 类似地, 本文主要研究F、G、K型的主序星, 因此恒星样本也被预先限制为同一类型的主序星.

下面介绍本文使用的两种推算行星生成率的理论模型.

2.1 逆检测效率法(IDEM)

IDEM是一项计算便捷且有效的快速估算行星生成率的方法, 在很多行星生成率的相关研究中被采用^[10]. IDEM通过对观测到的行星出现概率的计算来修正尚未被Kepler发现的可能存在行星的数量. 网格内的行星生成率 n_{cell} 作为给定区域内行星半径和轨道周期的函数, 实际与分布在该区域的行星样本数相关. 在每个区域中, 统计Kepler探测到的行星的数量, 同时筛选出有足够观测精度的恒星子集, 在给定行星半径与周期区域内, n_{cell} 与对应的误差 $\sigma_{n_{\text{cell}}}$ 分别为^[10]:

$$n_{\text{cell}} = \sum_{j=1}^{N_{p,\text{cell}}} \frac{1/p_j}{N_{*,j}}, \quad (2)$$

$$\sigma_{n_{\text{cell}}} = \frac{n_{\text{cell}}}{\sqrt{N_{p,\text{cell}}}}, \quad (3)$$

其中, $N_{p,\text{cell}}$ 为给定区域内的行星样本数, 分子上的 p_j 为行星 j 发生凌星现象的概率. 通过对每颗行星凌星概率的计算修正来近似获得区域内真实的行星数量, 在实际计算时, 使用其特定值 $1/p_j = a_j/R_*$ (a_j 为行星半长轴, R_* 为主恒星半径), 其值直接由开普勒测光法测量, 而非使用比值计算, 因

为单独测量恒星半径和半长轴精度较低. 在分母上, $N_{*,j}$ 为若存在行星 j 时满足凌星信噪比大于阈值的恒星数. 公式中的所有参数均可在DR25中获得.

上文提到IDEM的优点在于计算便捷, 可以快速得到估算结果, 但该方法并没有利用统计学理论作为支撑. 另一个缺点在于忽略了凌星观测参数中的不确定性, 由于测量误差会导致估计的凌星特性相关参数(例如, 凌星深度、持续时间等参数)偏离真实的凌星参数, 特别是对于在凌星探测信噪比阈值附近的样本, 真实的探测概率与估计之间存在较大的差异.

2.2 基于简化贝叶斯模型的最大似然法

鉴于IDEM忽略了探测中不确定性的缺点, ML法被提出, 该方法可以综合考虑观测凌星系外行星时的各种探测效率, 得到行星生成率的ML估计, 记为 n_{ML} . 对于一个给定区域, 假设存在 N_p 颗系外行星与 N_* 颗恒星, n_{ML} 为^[28]:

$$n_{\text{ML}} = \frac{N_p}{\sum_{j=1}^{N_*} p_j} = \frac{N_p}{N_* \langle p \rangle} = \frac{N_p}{N_{*,\text{eff}}}, \quad (4)$$

这里 $N_{*,\text{eff}} \equiv N_* \langle p \rangle$ 即恒星有效样本数为恒星样本与净探测效率 $\langle p \rangle$ 的乘积.

Kepler望远镜观测时的净探测效率 $\langle p \rangle$, 即探测事件发现准确行星的比例. 对于凌星观测, 影响探测效率的选择效应主要有3种, 分别是在随机轨道的行星发生凌星事件的随机概率 $p_{\text{tr}} \leq 1$ 、观测时该行星正处于凌星时刻的概率 $p_{\text{disc}} \leq 1$ 以及Kepler观测假阳性事件的比例. 不同的不相关选择效应相乘后便可得出净探测效率 $\langle p \rangle$.

为了对每颗恒星进行详细的探测效率计算, 在本文工作中考虑使用简化贝叶斯模型(Simplified Bayesian Model, SBM), 该方法在近似贝叶斯模型的基础上进行简化, 有效降低原本所需的大量计算资源. 在贝叶斯模型中, 将宇宙中每一颗行星的形成视为泊松过程, 在预先划分的行星周期半径参数空间的网格区域内其泊松速率参数恒定, 则网格区域内的 n_{cell} 分布也为泊松随机变量.

当我们观测到 N_* 个恒星且探测到这些恒星周围所有的行星时, 单一网格区域内所有行星数量

($N_{p,\text{cell,all}}$)是具有速率参数($n_{\text{cell}}N_*$)的泊松随机变量(N_* 此时已确定, 速率参数简化为 n_{cell}). 如果我们让 n_{cell} 的先验分布为具有形状参数 α_0 和逆尺度参数 β_0 的伽马分布, 则 n_{cell} 的后验分布为形状参数 $\alpha_0 + N_{p,\text{cell,all}}$ 和逆尺度参数 $\beta_0 + N_*$ 的伽马分布^[29]. Hsu等^[30]研究表明, 对于大多数的周期半径范围, 先验分布的选择不会对后验分布的生成率结果产生明显影响, 基本被限制在几个百分点的误差范围内. 只有当网格内的约束数据非常少时, 如M型恒星周围的行星样本或Kepler样本中周期较长半径较小的区域, 先验分布的选择才会显著影响后验分布的估算结果.

实际上由于凌星几何概率和探测效率, 现在并不能探测到所有的行星. 相反, 只对目标恒星的一个子集周围的行星敏感, 这主要取决于行星轨道平面的方向和每颗恒星的光度测量精度. 假设在目标恒星上探测到其行星的综合概率 $\langle p \rangle$ 在区域内是几乎一致的, 则有效恒星数 $N_{*,\text{eff}}$ 是净探测概率 $\langle p \rangle$ 的伯努利随机变量 N_* 的总和. 在这个模型中 n_{cell} 的后验分布表示为形状参数为 $\alpha_0 + N_{p,\text{cell}}$ 和逆尺度参数为 $\beta_0 + N_{*,\text{eff}}$ 的伽马分布^[29]:

$$P(n_{\text{cell}} | N_{p,\text{cell}}, N_{*,\text{eff}}) \sim \Gamma(\alpha_0 + N_{p,\text{cell}}, \beta_0 + N_{*,\text{eff}}). \quad (5)$$

通过用期望值 $E[N_{*,\text{eff}}]$ 取代随机变量 $N_{*,\text{eff}}$ 来近似 n_{cell} 的后验分布. 利用蒙特卡罗方法, 对探测的有效恒星数的期望值进行了估计. 对于每颗恒星和每个行星半径周期范围, 绘制 N_{samp} 颗行星(均匀分布在区间), 这里令 $N_{\text{samp}} = 100$, 每个目标恒星的 p_j 的值就等于 N_{samp} 颗模拟行星计算出探测概率 $p_{i,j}$ 的平均值, 即 $p_j \simeq (1/N_{\text{samp}}) \sum_{i=1}^{N_{\text{samp}}} p_{i,j}$. 有效恒星数的期望 $E[N_{*,\text{eff}}]$ 是所有目标恒星概率总和, 即 $E[N_{*,\text{eff}}] = \sum_{j=1}^{N_*} p_j$. 在此基础上, 对各行星大小和轨道周期范围内的行星固有发生率进行了后验近似^[29]:

$$P(n_{\text{cell}} | N_{p,\text{cell}}, E[N_{*,\text{eff}}]) \sim \Gamma(\alpha_0 + N_{p,\text{cell}}, \beta_0 + E[N_{*,\text{eff}}]). \quad (6)$$

该后验分布的均值和误差分别为^[29]:

$$\mu_{n_{\text{cell}}} = \frac{\alpha_0 + N_{\text{p,cell}}}{\beta_0 + E[N_{*,\text{eff}}]}, \quad (7)$$

$$\sigma_{n_{\text{cell}}} = \frac{\mu_{n_{\text{cell}}}}{\sqrt{\beta_0 + E[N_{*,\text{eff}}]}}. \quad (8)$$

在这里可假设 α_0 和 β_0 , 当两者都远小于 $N_{\text{p,cell}}$ 和 $E[N_{*,\text{eff}}]$ 时, 该后验均值本质上是探测到的行星数量与有效样本大小的比值, 即(4)式所示.

3 Kepler数据来源与处理

3.1 数据来源

本文所使用的Kepler恒星观测数据集来自2016年发布的Kepler DR25, 也是通过Kepler数据

处理管道在完整的17个季度中搜索凌星信号的任务数据集的最终版本. 该任务共包含了198709个恒星目标, 其中112046个目标在全部17个季度都有观测, 另86663个目标的被观测季度少于17个, 这些恒星样本的有效温度与质量分布如图3所示. DR25可在米库尔斯基太空望远镜档案库¹ (Mikulski Archive for Space Telescopes, MAST)或NASA (National Aeronautics and Space Administration)系外行星档案库²下载(米库尔斯基太空望远镜档案库是一个天文数据档案, 聚焦于光学、紫外线和近红外, 拥有包括哈勃、开普勒、凌星系外行星巡天卫星(Transiting Exoplanet Survey Satellite, TESS)和詹姆斯韦布空间望远镜(James Webb Space Telescope, JWST)等十多个任务的数据.)

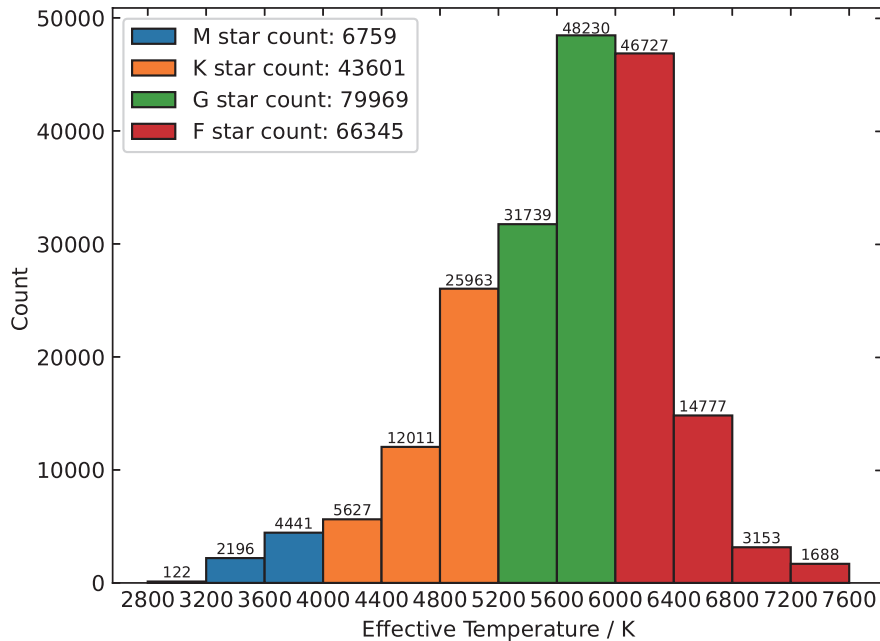


图 3 Kepler DR25恒星样本不同类型恒星的分布

Fig. 3 Distribution of stars with different spectral types in Kepler DR25

在Kepler DR25所释放的目标恒星数据列表中, 第1列即为目标恒星的Keplerid, 可在simbad数据库³中查询对应ID得到恒星更详细的观测数据

(如Gaia数据等), 其中有确切光谱型的恒星总数为31612, 其中F型共10756颗, 占比34.03%; G型11843颗, 占比37.46%; K型共6097颗, 占比19.29%. 另外

¹<https://archive.stsci.edu/>.

²<https://exoplanetarchive.ipac.caltech.edu/>.

³<http://simbad.u-strasbg.fr/simbad/>.

根据DR25中目标恒星有效温度(Teff)的数据, 不同类型的恒星分布如图3所示, 考虑K型恒星有效温度范围^[31]为3900–5200 K, 共有43601颗; G型恒星有效温度范围^[31]为5200–6000 K, 共有79969颗; F型恒星有效温度范围^[31]为6000–7600 K, 共有66345颗. 最后为了确保恒星样本凌星时的观测精度足够高, 在以上基础上, 我们将筛选条件定为开普勒光度Kmag < 15且表面重力加速度的对数值lg($g/(cm \cdot g \cdot s^{-2})$)范围在4.0–4.9. 最后得到恒星样本总数为77678, 其中F型恒星有38384颗, G型恒星32433颗, K型恒星6861颗. 本文所采用的部分恒星参数如表1所示.

本文使用的Kepler发现的已确认行星和候选行星相关数据, 是通过MSAT下载, 获得原始数据之后, 根据行星的宿主恒星光谱型对行星进行分类. 最终获得围绕F型恒星的已发现系外行星样本个数为472颗; 围绕G型恒星的已发现系外行星样本个数为1212颗; 围绕K型恒星的已发现系外行星样本个数为496颗.

3.2 KeplerPORTs的应用

行星生成率的计算必须知道Kepler数据处理管道完整性的精确模型^[32], 开普勒团队于2017年发布了基于Burke等^[33]的检测轮廓计算方法, 且采用了使用了更大数据库进行凌星注入与回收测试以提高准确性的开普勒管道最终版本, 并给出了计算工具KeplerPORTs. 该程序可在Github上下载⁴. 在KeplerPORTs中输入预设的行星半径与周期参数, 给定恒星的窗函数(Window Function)以及凌星深度函数(One-sigma Depth Function), 可输出若给定恒星周围存在预设行星时, Kepler空间望远镜的检测效率.

3.3 不同光谱型恒星的行星生成率估算

通过上面的统计样本, 我们分别采用IDEM和ML方法来推导出行星的生成率. 首先将行星半径周期平面上的参数空间分成等距的对数单元区域, 由于行星周期和半径的精度远小于单元区域的尺寸, 我们忽略了行星参数的不确定性. 在每个区域中, 独立计算出探测到的行星数量 N_p .

表1 Kepler DR25恒星样本中部分恒星参数
Table 1 Part of the stellar parameters in the Kepler DR25

Type	Kepid	Teff /K	lg g	Mass / M_{\odot}	Star radius/ R_{\odot}	Kmag
K	1435229	3927	4.774	0.467	0.465	14.586
	5123323	4191	4.654	0.593	0.6	14.833
	11342883	4298	4.805	0.464	0.447	13.343
	9787380	4437	4.681	0.551	0.562	13.397
	5287944	4493	4.689	0.58	0.57	12.344
	5440787	4530	4.689	0.548	0.554	13.24
	11028278	4729	4.769	0.566	0.514	13.735
	8776258	4755	4.718	0.541	0.533	13.45
	5687809	4807	4.719	0.614	0.568	14.908
	8880808	4994	4.617	0.81	0.724	14.527
G	5270719	5074	4.474	0.809	0.863	14.836
	1161345	5296	4.576	0.911	0.815	11.631
	3096946	5344	4.417	0.876	0.959	14.106
	5563980	5421	4.525	0.736	0.776	13.438
	10515282	5633	4.312	0.905	1.1	13.191
	1161137	5782	4.076	1.078	1.574	12.479
	10006971	5823	4.042	1.011	1.586	13.22
	8733356	5853	4.529	0.99	0.895	14.835
	11091454	5912	4.51	0.965	0.904	14.474
	9290949	5971	4.333	0.973	1.113	14.463
F	3216708	6029	4.32	0.903	1.089	14.841
	7207570	6263	4.46	1.068	1.009	14.552
	4365363	6331	4.328	1.025	1.152	13.778
	6949314	6453	4.443	1.076	1.033	14.419
	9757312	6545	4.194	1.246	1.476	11.702
	4568094	6816	4.338	1.277	1.262	14.225
	5369419	6971	4.206	1.097	1.368	12.973
	8093177	7108	4.11	1.218	1.61	12.083
	7007103	7171	4.121	1.646	1.848	11.237
	8316195	7230	4.216	1.432	1.543	12.766
12166194	7334	4.01	1.605	2.076	10.766	

⁴<https://github.com/nasa/KeplerPORTs>.

对于IDEM法, 我们首先计算每颗候选行星的凌星几何概率 $p_{g,i} = R_*/a$. 接着计算每颗候选行星 i 经过目标恒星 j 周围被探测到的概率 $p_{d,i,j}$, 假设这颗行星有相同的轨道周期, 对于目标恒星 j , 若存在这样一颗行星, 观测到的凌星信号的信噪比为^[10]

$$\text{SNR} = \frac{\delta}{\sigma_{\text{CDPP}}} \sqrt{\frac{n_{\text{tr}} t_{\text{dur}}}{3h}}, \quad (9)$$

这里的 $\delta = R_p^2/R_*^2$ 为半径 R_p 的行星在半径 R_* 的恒星前发生凌星时, 中心凌星的光变深度. n_{tr} 为90 d, 即一个季度内, 发生凌星的次数, t_{dur} 为凌星持续时间. σ_{CDPP} 为Kepler观测凌星时量化的系统噪声.

我们设定信噪比的阈值为10, 只统计在给定行星的凌星深度和周期时, $\text{SNR} > 10$ 的恒星数量. 由(2)式可以估算得到每个单元区域内的生成率. 区域内生成率的误差可表示为 $\delta = n_{\text{cell}}/\sqrt{N_{r,p}}$, 式中 $N_{r,p}$ 为此半径周期区域内的行星数目. 图4到图7分

别为用IDEM方法估算得到的类太阳主序星的整体行星生成率的分布以及F、G、K型恒星各自的行星生成率分布图. 对于本文选择的行星半径范围 $0.5\text{--}20 R_{\oplus}$, 轨道周期 $0.4\text{--}400$ d的范围内, 每颗F、G、K型恒星周围存在这类行星的数量分别为 $n_{\text{IDEM.F}} = 0.40 \pm 0.02$ 、 $n_{\text{IDEM.G}} = 1.76 \pm 0.05$ 和 $n_{\text{IDEM.K}} = 3.03 \pm 0.14$. 同时F、G、K型恒星周围存在这类行星的整体生成率为 $n_{\text{IDEM}} = 1.41 \pm 0.03$.

对于ML法的行星生成率(4)式, 将该公式纳入SBM框架. 对于恒星样本 j , 随机生成100个行星粒子, 其半径、周期分别为 R_{pjk}, P_{jk} , k 取值范围为1–100的自然数, 则该恒星样本在区域内的平均检测效率为 $1/100 \sum_{k=1}^{100} S(P_{jk}, R_{pjk})$, 这里 $S(P_{jk}, R_{pjk})$ 是KeplerPORTs内置函数, 变量为 P_{jk} 和 R_{pjk} . 对于所有的恒星样本求和取平均便可得到区域内的平均行星探测效率.

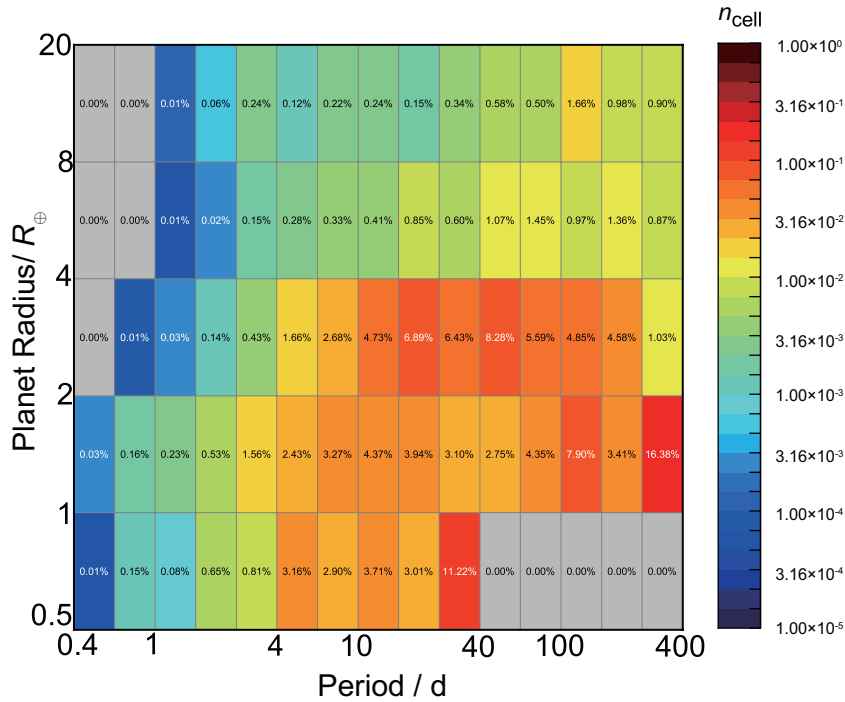


图 4 IDEM估算得到的F、G、K型恒星周围整体生成率的分布

Fig. 4 Estimated distribution of the total planetary occurrence rates around F, G, K-type stars by IDEM

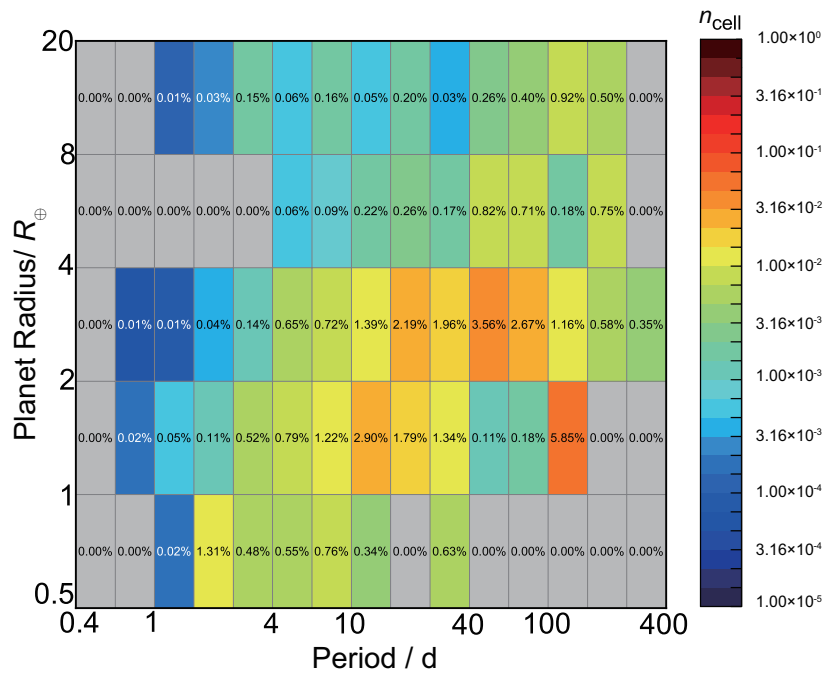


图 5 IDEM估算得到的F型恒星周围生成率的分布

Fig. 5 Estimated distribution of planetary occurrence rates around F-type stars by IDEM

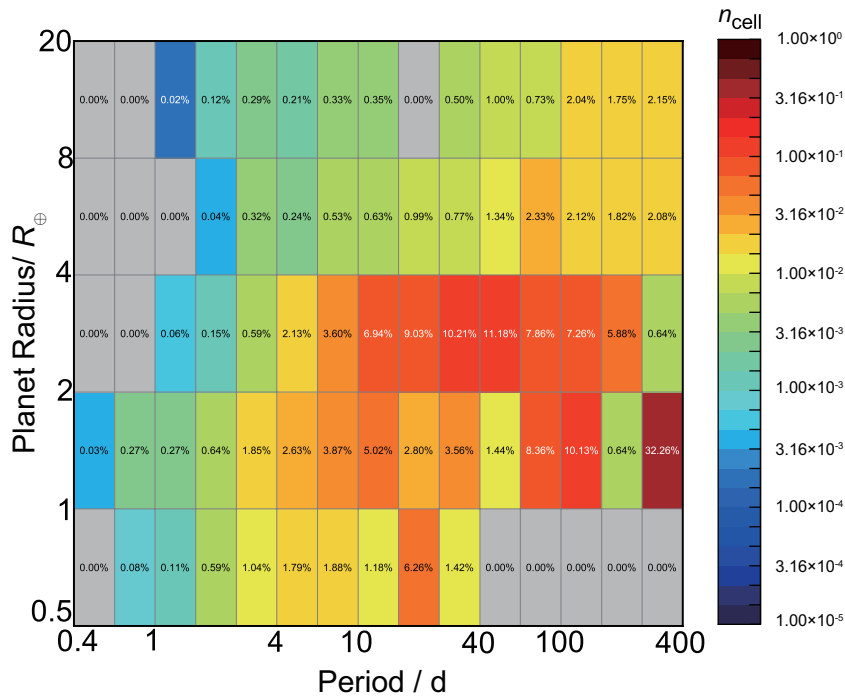


图 6 IDEM估算得到的G型恒星周围生成率的分布

Fig. 6 Estimated distribution of planetary occurrence rates around G-type stars by IDEM

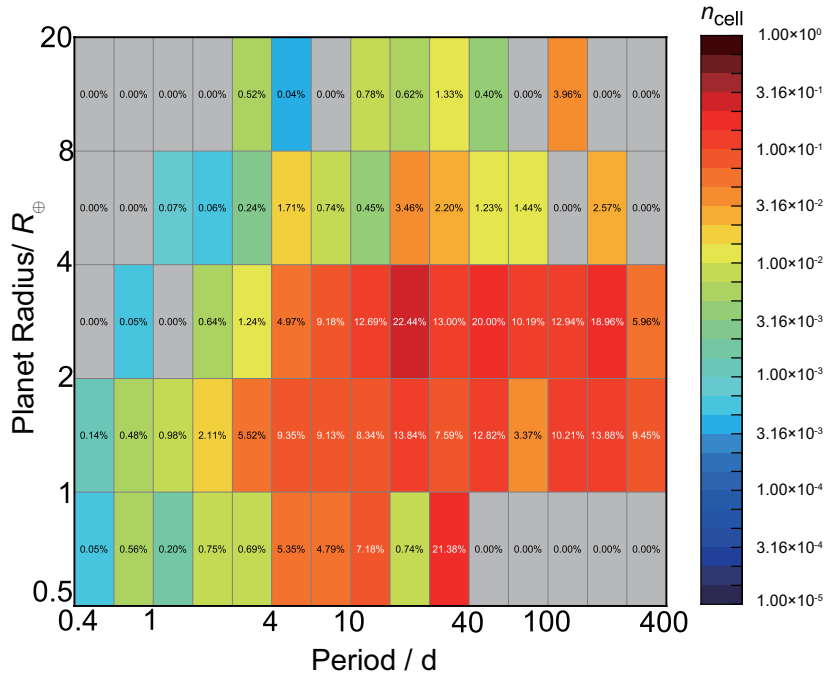


图 7 IDEM估算得到的K型恒星周围生成率的分布

Fig. 7 Estimated distribution of planetary occurrence rates around K-type stars by IDEM

考虑行星生成率为扁平先验, 其后验分布可以描述为 $\alpha_0 = 1$ 和 $\beta_0 = 0$ 的伽马分布. 每个区域内伽马分布的平均值和标准偏差由(7)式和(8)式给出, 分别作为估算值和相关的不确定度. 图8到图11分别为用ML方法估算得到的类太阳主序星的整体行星生成率的分布以及F、G、K型恒星各自的行星生成率分布图. 由于仅当 $\alpha_0 \ll N_p$ 时, 泊松分布的均值才可近似为我们所估算的生成率. 在行星半径小于 $1R_{\oplus}$ 、周期大于40 d的范围内, 行星生成率

的估算不确定性很大. 我们考虑估算半径范围 $1-20 R_{\oplus}$, 轨道周期小于400 d的系外行星的生成率. 在F型恒星周围存在这类行星的生成率 $n_{ML,F} = 0.47 \pm 0.02$, 对应的 $n_{IDEM,F} = 0.36 \pm 0.02$; G型恒星周围 $n_{ML,G} = 1.23 \pm 0.04$, 对应的 $n_{IDEM,G} = 1.62 \pm 0.05$; K型恒星周围 $n_{ML,K} = 2.73 \pm 0.13$, 对应的 $n_{IDEM,K} = 2.61 \pm 0.12$. 同时F、G、K型恒星周围存在这类行星的整体生成率为 $n_{ML} = 0.90 \pm 0.06$, 对应的 $n_{IDEM} = 1.16 \pm 0.03$.

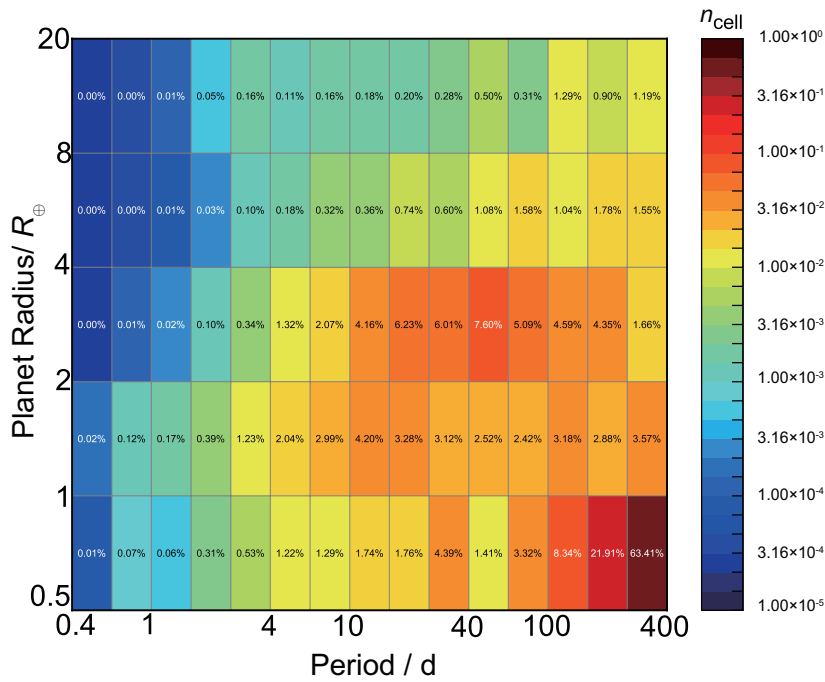


图 8 ML估算得到的F、G、K型恒星周围整体的生成率分布

Fig. 8 Estimated distribution of the total planetary occurrence rates around F, G, K-type stars by ML

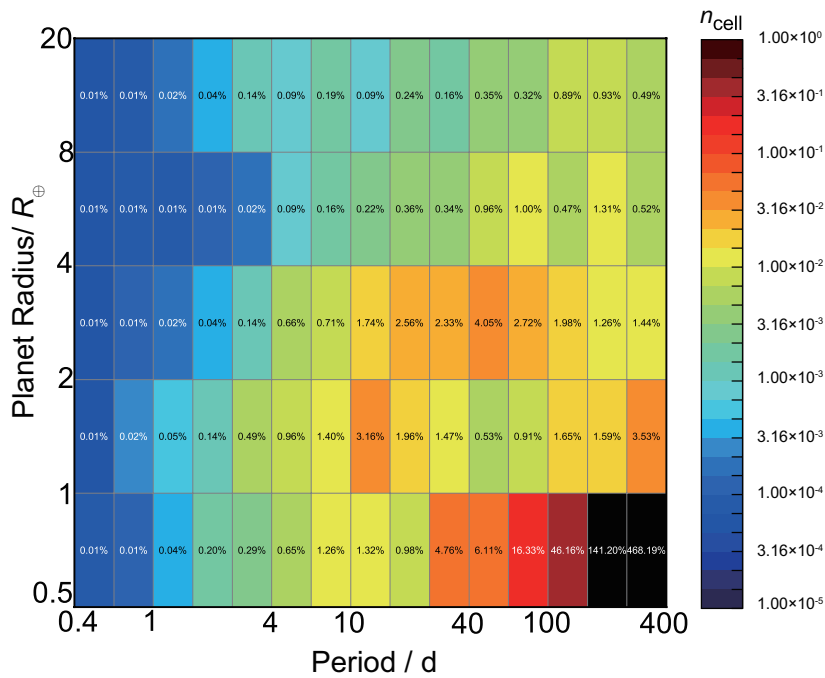


图 9 ML估算得到的F型恒星周围生成率的分布

Fig. 9 Estimated distribution of planetary occurrence rates around F-type stars by ML

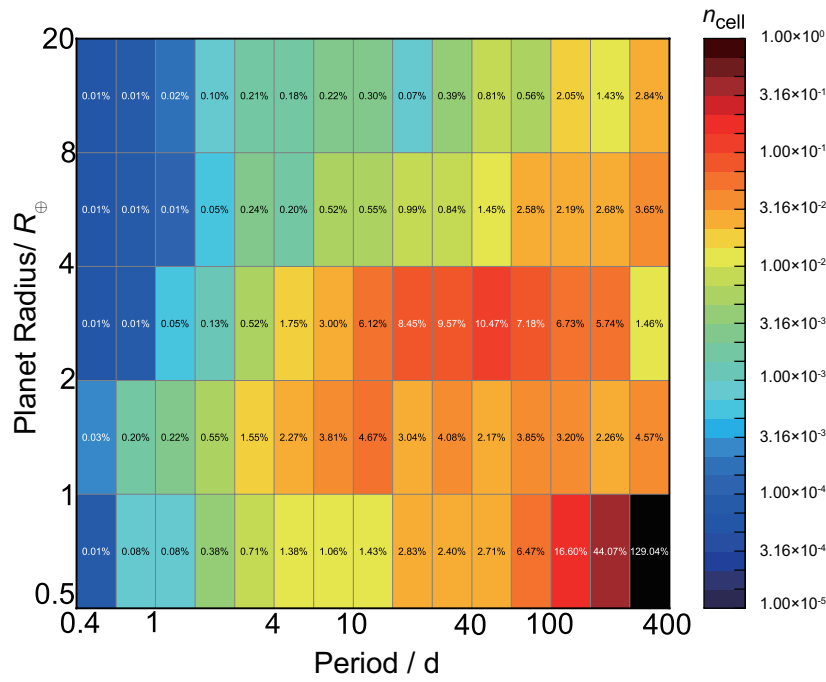


图 10 ML估算得到的G型恒星周围生成率的分布

Fig. 10 Estimated distribution of planetary occurrence rates around G-type stars by ML

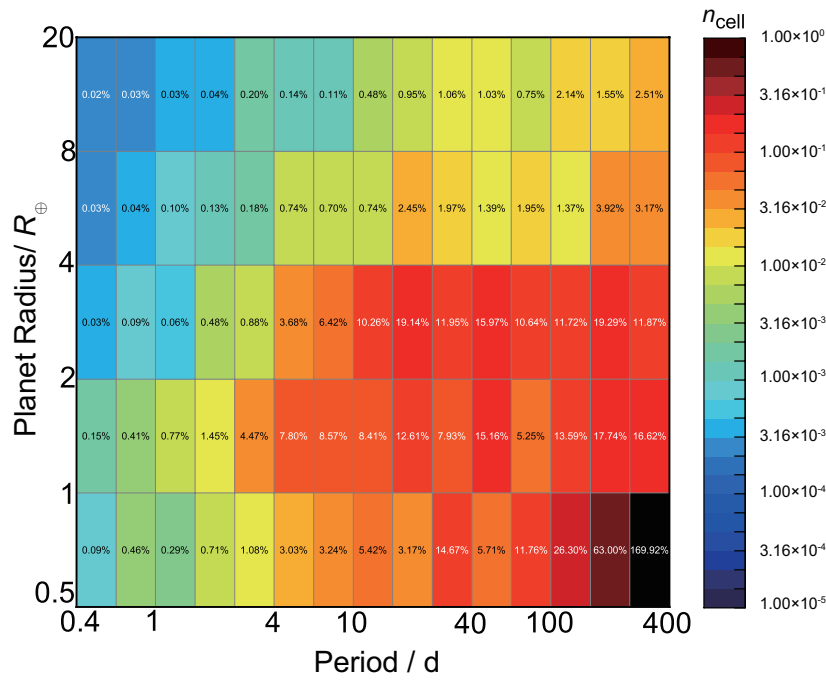


图 11 ML估算得到的K型恒星周围生成率的分布

Fig. 11 Estimated distribution of planetary occurrence rates around K-type stars by ML

4 讨论

本文通过对Kepler观测数据中已确认行星与DR25中的目标恒星样本进行筛选分析, 按照恒星的光谱型对样本进行分类. 同时使用两种不同的方法, 分别估算了F型、G型、K型恒星周围的系外行星生成率, 两种方法的估算结果在一定程度上相互匹配, 且都可看出随着恒星有效温度的升高, 其周围的行星生成率呈下降趋势. 对于半径范围 $1-20 R_{\oplus}$, 轨道周期小于400 d的系外行星, K型恒星周围的生成率远高于F型恒星周围.

由于目前系外行星观测技术的局限性, Kepler使用的凌星观测技术更偏向探测到半径更大、周期更短的系外行星, 即对于不同周期半径的系外行星, Kepler的探测效率是不同的. 通过将行星的周期半径参数空间划分为多个网格空间(本文为 5×15), 并假设位于同一网格区域内的行星其探测效率是相同的, 在这一基础上, 我们通过先对单一区域内求得相对可靠的区域生成率 n_{cell} , 再求和获得参数空间整体的生成率^[10].

对于IDEM, 通过假设已知的行星存在于样本恒星周围时, 计算凌星信号的信噪比来判断, 是否能够发现该行星. 当信噪比超过设定的阈值时(本文为10)即计入有效恒星样本. 网格区域内的行星样本数和有效恒星样本数的比值就是该片区域的行星生成率^[10]. 我们估算出的对于 $1-20 R_{\oplus}$ 、周期小于400 d的整体行星生成率 $n_{\text{IDEM}} = 1.16 \pm 0.03$, 这一结果也符合之前的一些研究结论^[28].

另一方面, 在使用纳入贝叶斯框架的最大似然法时, 生成率的整体计算公式依旧是行星样本数除以有效恒星样本数. 此时对于有效恒星样本数并不依赖于所选取的行星样本, 而是通过模拟注入100次凌星信号^[29], 经由KeplerPORTs程序求得对应恒星平均探测效率^[33]. ML相比IDEM具有统计基础, 该模拟过程将行星的形成视为泊松过程, 且不同区域相互独立. 对于有效恒星样本的取值并不再依赖行星样本的相关参数, 进一步降低样本选择对估算的影响. 本文估算出的对于 $1-20 R_{\oplus}$ 、周期小于400 d的整体行星生成率 $n_{\text{ML}} = 0.90 \pm 0.02$.

Zhu等^[28]给出的同样周期半径范围的行星生成率 $n_p = 1.23 \pm 0.06$, 本文估算结果相对偏小的原因在于对DR 25的恒星样本进行了更精细的筛选. 在考虑了恒星样本的Kepler光度以及表面重力加速度 $\lg g$ 的基础上, 有效温度选择范围扩大为3900-7600 K, 使得样本中F型恒星占比更多. 因此F型恒星周围较低的行星生成率拉低了整体生成率的估算值, 同时K型恒星由于数量占比仅约9%, 对整体生成率的贡献不大.

在系外行星研究中, 适宜生命生存的宜居行星是最关键的科学问题之一. 目前唯一已知的宜居行星只有地球, 因此我们更关注位于类太阳恒星宜居带内的类地行星的生成率. 近邻宜居行星巡天计划(Closeby Habitable Exoplanet Survey, CHES)所关注的便是这些靠近太阳系的(距离太阳系约10 pc)类太阳型恒星的宜居带内的类地行星, 其主要的科学目标是通过天体测量法, 在100颗近邻的F、G、K恒星周围寻找地球2.0或宜居带内的岩石行星^[34]. 基于上述估算, 得到对于轨道周期160-400 d、半径范围 $1-2 R_{\oplus}$ 的Kepler凌星系外行星, 在F、G、K型恒星周围的行星生成率分别为 0.051 ± 0.008 、 0.068 ± 0.009 和 0.344 ± 0.085 . 在有效温度3900-7600 K的F、G、K型恒星周围的宜居带类地行星的整体生成率为 0.064 ± 0.007 . Bryson等^[35]得出Kepler样本下G、K型周围保守宜居带内的 $0.5-1.5 R_{\oplus}$ 的行星的生成率在 $0.37_{-0.21}^{+0.48}$ 到 $0.60_{-0.36}^{+0.90}$ 之间, 本文估算的K型恒星周围的宜居带行星生成率为0.344, 在其误差范围内.

行星生成率的估算结果与所选取的行星和恒星样本关系甚大, 本文仅选用了Kepler空间望远镜观测到的凌星系外行星作为分析样本. 一方面, TESS作为Kepler任务的后继者, 自2018年发射升空以来已经观测到5000多颗凌星系外行星候选目标, 同时计划于2026年发射的PLATO (PLANetary Transits and Oscillations)也将搜索更多恒星的凌星现象, 并重点关注类太阳恒星周围存在的岩石行星. 凌星系外行星和恒星样本的扩展会得到更准确的凌星行星种群的分布规律.

另一方面, 凌星法以及当前同样发现较多行星

的视向速度法更容易探测到短周期大质量的系外行星,而对于参数空间外围的系外行星,如轨道周期大于1 yr的系外行星,Kepler空间望远镜对其并不敏感。尽管可以通过统计方法进行估算,外围区域估算的行星生成率误差范围也是相当之大。未来随着凌星法和视向速度法灵敏度的提高,可以更容易探测到半径地球大小甚至更小的系外行星,同时在足够长的观测时间覆盖下,可拓展长周期行星的样本范围。多种系外行星探测方法的结合(例如可探测地球大小系外行星的微引力透镜法)也将可能更好地改善对行星生成率的估算结果。

参 考 文 献

- [1] Cumming A, Butler R P, Marcy G W, et al. *PASP*, 2008, 120: 531
- [2] Mayor M, Marmier M, Lovis C, et al. *arXiv:1109.2497*
- [3] Santerne A, Moutou C, Tsantaki M, et al. *A&A*, 2016, 587: A64
- [4] Fernandes R B, Mulders G D, Pascucci I, et al. *ApJ*, 2019, 874: 81
- [5] Bryan M L, Knutson H A, Howard A W, et al. *ApJ*, 2016, 821: 89
- [6] Ida S, Lin D N C. *ApJ*, 2008, 685: 584
- [7] Hansen B M S, Murray N. *ApJ*, 2012, 751: 158
- [8] Chiang E, Laughlin G. *MNRAS*, 2013, 431: 3444
- [9] Howard A W, Marcy G W, Johnson J A, et al. *Science*, 2010, 330: 653
- [10] Howard A W, Marcy G W, Bryson S T, et al. *ApJS*, 2012, 201: 15
- [11] Hirano T, Dai F, Gandolfi D, et al. *AJ*, 2018, 155: 127
- [12] Santos N C, Israelian G, Mayor M, et al. *A&A*, 2003, 398: 363
- [13] Ran I W, Lee B C, Kim K M, et al. *Journal of The Korean Astronomical Society*, 2008, 41: 59
- [14] Wolszczan A, Frail D A. *Nature*, 1992, 355: 145
- [15] Furlan E, Ciardi D R, Everett M E, et al. *AJ*, 2017, 153: 71
- [16] Bouma L G, Masuda K, Winn J N. *AJ*, 2018, 155: 244
- [17] Kraus A L, Ireland M J, Huber D, et al. *AJ*, 2016, 152: 8
- [18] Moe M, Kratter K M, Badenes C. *ApJ*, 2019, 875: 61
- [19] Santos N C, Israelian G, Mayor M. *A&A*, 2001, 373: 1019
- [20] Fischer D A, Valenti J. *ApJ*, 2005, 622: 1102
- [21] Sousa S G, Santos N C, Mayor M, et al. *A&A*, 2008, 487: 373
- [22] Buchhave L A, Latham D W, Johansen A, et al. *Nature*, 2012, 486: 375
- [23] Andrews S M, Rosenfeld K A, Kraus A L, et al. *ApJ*, 2013, 771: 129
- [24] Ansdell M, Williams J P, van der Marel N, et al. *ApJ*, 2016, 828: 46
- [25] Nielsen E L, De Rosa R J, Macintosh B, et al. *AJ*, 2019, 158: 13
- [26] Johnson J A, Aller K M, Howard A W, et al. *PASP*, 2010, 122: 905
- [27] Mulders G D, Pascucci I, Apai D. *ApJ*, 2015, 798: 112
- [28] Zhu W, Dong S B. *ARA&A*, 2021, 59: 291
- [29] Hsu D C, Ford E B, Ragozzine D, et al. *AJ*, 2018, 155: 205
- [30] Hsu D C, Ford E B, Ragozzine D, et al. *AJ*, 2019, 158: 109
- [31] Habelts G M H J, Heintze J R W. *A&AS*, 1981, 46: 193
- [32] Burke C J, Christiansen J L, Mullally F, et al. *ApJ*, 2015, 809: 8
- [33] Burke C J, Catanzarite J. *Planet Detection Metrics: Per-Target Detection Contours for Data Release 25: KSCI-19111-002. Moffett Field: NASA Ames Research Center*, 2017: 19
- [34] Ji J H, Li H T, Zhang J B, et al. *RAA*, 2022, 22: 072003
- [35] Bryson S, Kunitomo M, Kopparapu R K, et al. *AJ*, 2021, 161: 36

Estimations of Planetary Occurrence Rates for Solar-type Stars with Data Release 25 from Kepler Q1–Q17 Observations

ZHANG Qing-xin^{1,2,3} BAO Chun-hui^{1,2,3} JI Jiang-hui^{1,2,3}

(1 Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023)

(2 School of Astronomy and Space Sciences, University of Science and Technology of China, Hefei 230026)

(3 CAS Key Laboratory of Planetary Sciences, Chinese Academy of Sciences, Nanjing 210023)

ABSTRACT The detection of exoplanets in the past three decades has revealed the fact that planets are ubiquitous in the universe. In order to deeply study the ubiquity of habitable planets, on one hand, we need to understand the characteristics of habitable planets; on the other hand, we can analyze the the distribution characteristics of exoplanets have been found, and the probability of occurrence of such planets around stars is calculated. Among the exoplanets that have been found so far, most of them are discovered by the transit method. For example, the number of the planets detected by the Kepler space telescope is 2344. Kepler telescope officially retired in 2018, and the Kepler team released the final version of Kepler Data Release (DR25), including a total of 198709 stars observed quarterly Q1–Q17. Here we analyze the Kepler data by using two different methods, Inverse Detection Efficiency Method (IDEM) and Maximum Likelihood Analysis (ML), to estimate planet occurrence rates in the space of the parameters of radius and orbital period. At the same time, the samples were classified according to the spectral types of stars, and the planet occurrence rates around F, G and K Kepler stars as well as its overall formation rate were estimated respectively. We estimate the planetary occurrence rates for planets among radius range of 1–20 R_{\oplus} and orbital period range of 0.4–400 days by IDEM and ML, for which around F stars are 0.36 ± 0.02 and 0.47 ± 0.02 . The rates around G stars by IDEM and ML are 1.62 ± 0.05 and 1.23 ± 0.04 . The rates around K stars by IDEM and ML are 2.61 ± 0.12 and 2.73 ± 0.13 . And the overall occurrence rate of such planets around F, G, K stars by IDEM and ML are 1.16 ± 0.03 and 0.90 ± 0.02 . According to our estimation, we further show the results for the planet occurrence rates around stars with different spectral types by different methods, and discuss the reliability of the results in comparison with the previous studies.

Key words planets and satellites: general, transiting exoplanet, planetary occurrence rate, habitable planet, Kepler space telescope