Hierarchical Bayesian Modeling of Planet Populations

Angie Wolfgang
NSF Postdoctoral Fellow, Penn State
The Big Picture

We’ve found > 3000 planets, and counting.  
Earth’s place in the Universe . . .

We’re measuring their properties.  
(Mass, radius, atmospheres, orbits, host stars, …)

Habitable exoplanets?  
. . . so what’s the physics involved?  
(How do they form, interact with their surroundings, change via which physical processes?)
The Big Picture

\[ \text{Mass of Star [Solar Mass]} \]

\[ \begin{array}{cccc}
1 & 10 & 100 & 10^3 \\
10 & 10^3 & 10^4 & 10^5 \\
100 & 10^2 & 1 & 0.1 \\
1000 & 10 & 1 & 0.01 \\
\end{array} \]

\[ \text{Orbital Period [Days]} \]

\[ \text{Planet Mass [Jupiter Mass]} \]

The Observed Population

... so what’s the physics involved? (How do they form, interact with their surroundings, change via which physical processes?)
Deterministic planet formation model with physical parameters $\alpha$ (disk mass, viscosity...):

$$f(M,P,\ldots|\alpha)$$
The Big Picture

From the outset:
1) distributions of planet properties,
2) inference on \( \alpha \),
3) model comparison (\( f_1 \) vs \( f_2 \) vs \( f_3 \)).

Deterministic planet formation model with physical parameters \( \alpha \) (disk mass, viscosity...):
\[ f(M,P,...|\alpha) \]
Deterministic planet formation model with physical parameters \( \alpha \) (disk mass, viscosity…): 
\[ f(M, P, \ldots | \alpha) \]

From the outset:
1) distributions of planet properties,
2) inference on \( \alpha \),
3) model comparison (f₁ vs f₂ vs f₃).

The Big Picture
Large Measurement Uncertainty

From the outset:
1) distributions of planet properties,
2) inference on $\alpha$,
3) model comparison ($f_1$ vs $f_2$ vs $f_3$).

$M_{\text{true}}, P_{\text{true}}, \ldots$
Large Measurement Uncertainty

From the outset:
1) distributions of planet properties,
2) inference on $\alpha$,
3) model comparison ($f_1$ vs $f_2$ vs $f_3$).

Multiple levels $\rightarrow$ hierarchical modeling

$M_{\text{true}}, P_{\text{true}}, \ldots$
Large Measurement Uncertainty

From the outset:
1) distributions of planet properties,
2) inference on $\alpha$,
3) model comparison ($f_1$ vs $f_2$ vs $f_3$).

Multiple levels $\rightarrow$ hierarchical modeling

\[
f(M_{\text{true}}, P_{\text{true}}, \ldots | \alpha)\]
What is Hierarchical Bayesian Modeling (HBM)?

Arises naturally when want to make **scientific inferences** about a **population** based on many **individuals**.

“Regular” Bayes:

\[
p(\theta|x) \propto p(x|\theta) \ p(\theta)
\]

 posterior   likelihood   prior

“What is the probability that \( \theta \) has some value, given the data?”

Hierarchical Bayes:

\[
p(\alpha,\theta|x) \propto p(x|\theta,\alpha) \ p(\theta|\alpha) \ p(\alpha)
\]

 posterior   likelihood   prior

**Population Parameters**

**Individual Parameters**

\( M_{\text{true}}, P_{\text{true}}, \ldots \)

**Observables**

\( M_{\text{obs}}, P_{\text{obs}}, \ldots \)
HBM for Exoplanets

Hogg et al. 2010
(orbal eccenticities)

Morton & Winn, 2014
Campante et al. 2016
(angle betweeen stellar spin & planet’s orbit)

Foreman-Mackey et al, 2014
(Kepler occurrence rates)

Demory 2014
(geometric albedos)

Rogers 2015
(rocky-gaseous transition)

Wolfgang & Lopez, 2015
(super-Earth composition distribution)

Shabram et al. 2016
(short-period eccentricity distribution)

Wolfgang, Rogers, & Ford, 2016
Chen & Kipping, submitted
(mass-radius relationship)

2013 SAMSI workshop on analyzing Kepler data

All are some variation on this 3-level structure:
(from Wolfgang et al. 2016)

\[
\begin{align*}
\alpha & \quad C & \quad \gamma & \quad \sigma_M \\
R_t^{(i)} & \quad M_t^{(i)} & \quad R_{ob}^{(i)} & \quad M_{ob}^{(i)}
\end{align*}
\]
HBM for Exoplanets

Hogg et al. 2010
(orital eccentricities)
Morton & Winn, 2014
Campante et al. 2016
(angle between stellar spin & planet’s orbit)
Foreman-Mackey et al, 2014
(Kepler occurrence rates)
Demory 2014
(geometric albedos)
Rogers 2015
(rocky-gaseous transition)
Wolfgang & Lopez, 2015
(super-Earth composition distribution)
Shabram et al. 2016
(short-period eccentricity distribution)
Wolfgang, Rogers, & Ford, 2016
Chen & Kipping, submitted
(mass-radius relationship)

Already have posteriors for the observables?
Can use importance sampling in multi-level models (Hogg et al. 2010)

More recently: full HBM using JAGS or own hierarchical MCMC code;
many people moving to STAN
Example: Planet HBM Results

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016

Best–Fit M–R Relations

Weiss & Marcy, 2014

By Mass Measurement Method

By Radius Range of Input Data

Intrinsic Scatter $\sigma_M (M_{\text{Earth}})$

Power-law Index $\gamma$

Constant C ($M_{\text{Earth}}$)

Mass ($M_{\text{Earth}}$)

Radius ($R_{\text{Earth}}$)

< 1.6 $R_{\text{Earth}}$
< 4 $R_{\text{Earth}}$
< 8 $R_{\text{Earth}}$

Weiss & Marcy, 2014

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016

Best–Fit M–R Relations

Weiss & Marcy, 2014

By Mass Measurement Method

By Radius Range of Input Data

Intrinsic Scatter $\sigma_M (M_{\text{Earth}})$

Power-law Index $\gamma$

Constant C ($M_{\text{Earth}}$)

Mass ($M_{\text{Earth}}$)

Radius ($R_{\text{Earth}}$)

< 1.6 $R_{\text{Earth}}$
< 4 $R_{\text{Earth}}$
< 8 $R_{\text{Earth}}$

Weiss & Marcy, 2014

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016

Best–Fit M–R Relations

Weiss & Marcy, 2014

By Mass Measurement Method

By Radius Range of Input Data

Intrinsic Scatter $\sigma_M (M_{\text{Earth}})$

Power-law Index $\gamma$

Constant C ($M_{\text{Earth}}$)

Mass ($M_{\text{Earth}}$)

Radius ($R_{\text{Earth}}$)

< 1.6 $R_{\text{Earth}}$
< 4 $R_{\text{Earth}}$
< 8 $R_{\text{Earth}}$

Weiss & Marcy, 2014

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016

Best–Fit M–R Relations

Weiss & Marcy, 2014

By Mass Measurement Method

By Radius Range of Input Data

Intrinsic Scatter $\sigma_M (M_{\text{Earth}})$

Power-law Index $\gamma$

Constant C ($M_{\text{Earth}}$)

Mass ($M_{\text{Earth}}$)

Radius ($R_{\text{Earth}}$)

< 1.6 $R_{\text{Earth}}$
< 4 $R_{\text{Earth}}$
< 8 $R_{\text{Earth}}$

Weiss & Marcy, 2014

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016

Best–Fit M–R Relations

Weiss & Marcy, 2014

By Mass Measurement Method

By Radius Range of Input Data

Intrinsic Scatter $\sigma_M (M_{\text{Earth}})$

Power-law Index $\gamma$

Constant C ($M_{\text{Earth}}$)

Mass ($M_{\text{Earth}}$)

Radius ($R_{\text{Earth}}$)

< 1.6 $R_{\text{Earth}}$
< 4 $R_{\text{Earth}}$
< 8 $R_{\text{Earth}}$

Weiss & Marcy, 2014

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016

Best–Fit M–R Relations

Weiss & Marcy, 2014

By Mass Measurement Method

By Radius Range of Input Data

Intrinsic Scatter $\sigma_M (M_{\text{Earth}})$

Power-law Index $\gamma$

Constant C ($M_{\text{Earth}}$)

Mass ($M_{\text{Earth}}$)

Radius ($R_{\text{Earth}}$)

< 1.6 $R_{\text{Earth}}$
< 4 $R_{\text{Earth}}$
< 8 $R_{\text{Earth}}$
Example: Planet HBM Results

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016

So what’s next?
On the Theory Side

There are many competing theories; like to quantitatively compare which is a better fit to data.

Now: BIC, qualitative

Next: Nested Sampling? Need your help!

M_{true}, P_{true}, ...
More on the Theory Side

Planet formation is complicated, and $f(M_{\text{true}}, P_{\text{true}}, \ldots | \alpha)$ involves expensive computer simulations.

Now: parametric functions, little theory

Next: incorporate emulator functions
Emulators: An example

Sub-Neptune compositions: Wolfgang & Lopez, 2015

Wanted to understand BOTH:

- compositions of individual super-Earths (fraction of mass in a gaseous envelope: $f_{\text{env}}$)
- the distribution of this composition parameter over the Kepler population ($\mu$, $\sigma$).

**Now:** internal structure models described by power laws

**Next:** internal structure models described by nonparametric/marginally parametric distributions
On the Data Side

The Observed Population

$$f(M_{\text{true}}, P_{\text{true}}, \ldots | \alpha)$$
On the Data Side

Non-trivial detection functions are present in the observed population

Now: ignore \( p(\text{detect}) \) or cut stellar sample

Next: characterize and include \( p(\text{detect}) \)

Mass of Star [Solar Mass]

Orbital Period [Days]

Planet Mass [Jupiter Mass]

The Observed Population

\( f(M_{\text{true}}, P_{\text{true}}, \ldots | \alpha) \)
Example: Including $p(\text{det})$

Kepler occurrence rates: Foreman-Mackey et al. 2014

$p(\text{det})$ characterized by injecting synthetic transit signals in data and running detection algorithm on them (Petigura et al. 2014)

Grid of recovery fraction vs. radius and period

Incorporated with inferred occurrence rate (Poisson point process)
Example: Including $p(\text{det})$

Kepler occurrence rates: Foreman-Mackey et al. 2014

$p(\text{det})$ characterized by injecting synthetic transit signals in data and running detection algorithm on them (Petigura et al. 2014)

But $p(\text{det})$ not always known . . . or even characterizable!

Grid of recovery fraction vs. radius and period

Incorporated with inferred occurrence rate (Poisson point process)
More on the Data Side

Mass, radius, period is not what we actually observe, and current likelihoods $p(M_{\text{obs}}|M_{\text{true}})$ very simple.
More on the Data Side

Mass, radius, period is not what we actually observe, and current likelihoods $p(M_{\text{obs}}|M_{\text{true}})$ very simple

Next: inference on population directly from RVs vs. time, flux vs. time
Even more on the Data Side

But we don’t actually observe RVs vs. time or flux vs. time either . . . our real data is light on a detector.
Even more on the Data Side

But we don’t actually observe RVs vs. time or flux vs. time either . . . our real data is light on a detector

Next next: > 3 level HBMs, inference straight from the actual data
Next: Super-Earth Compositions

Understanding selection effects in mass-radius space: Wolfgang, Jontof-Hutter, & Ford, in prep.

But how exactly to evaluate the hierarchical model with unknown number of non-detections?
Next: Super-Earth Compositions

Characterizing Joint Mass-Radius-Flux Distribution:
Wolfgang, Jontof-Hutter, Rogers & Ford, in prep.
Next: Super-Earth Compositions

Characterizing Joint Mass-Radius-Flux Distribution:
Wolfgang, Jontof-Hutter, Rogers & Ford, in prep.

But many options for parameterizations …
hierarchical model comparisons?
Poisson point process?
Next: Super-Earth Compositions

Initial Sub-Neptune compositions: Wolfgang & Lopez, in prep.

Need emulator function more accurate than power laws but still computationally “easy”.

Population-wide Distributions

Internal Structure Models

Likelihood
Summary:

Where we are:

~ a dozen exoplanet astronomers working on very simple hierarchical models describing distributions of planet properties

Where we can go this year at SAMSI:

1) Incorporate survey detection efficiency
2) Develop emulator functions to include computationally expensive theoretical simulations directly into HBM
3) Compare different theoretical models via hierarchical model comparison
4) Implement more realistic likelihoods: inference from lower-level data