Estimating Exoplanet Occurrence Rates Using Approximate Bayesian Computation

Danley Hsu

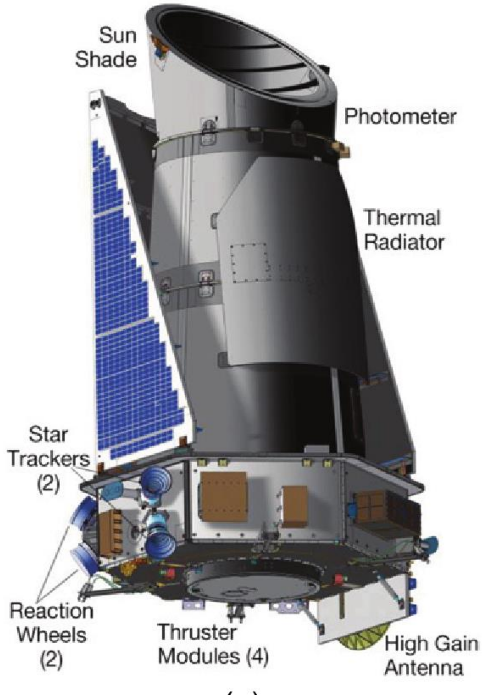
Collaborators: Eric Ford, Darin Ragozzine, Keir Ashby, Robert Morehead

Likelihood-Free Inference Workshop – March 19, 2019

PennState

Exoplanets are a hot topic in astronomy right now

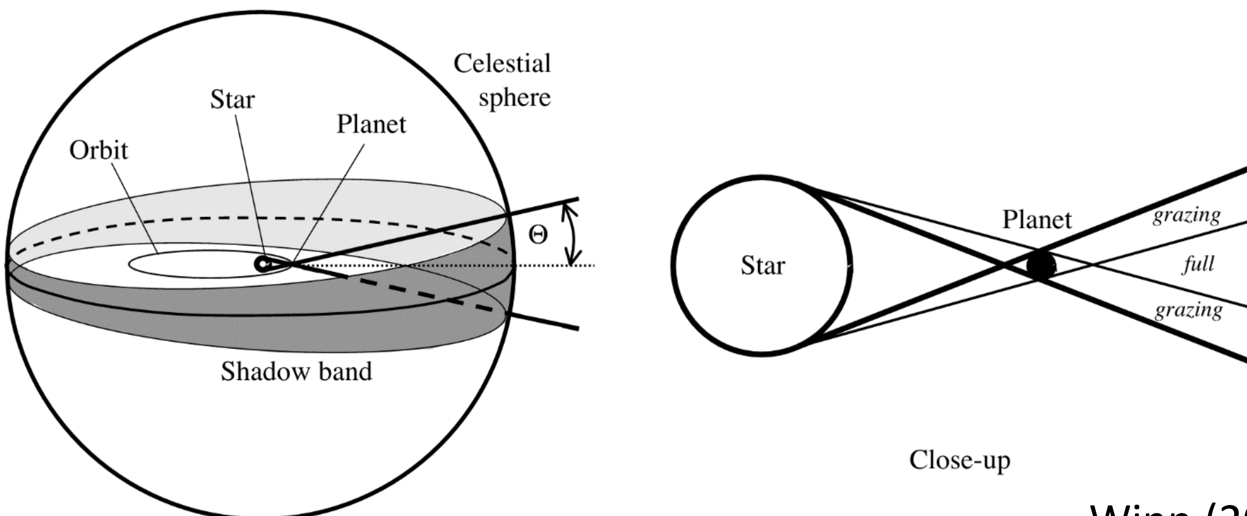
- Helps us get closer to answering whether there exists other life in the universe
- Occurrence rate studies help inform design studies for future survey missions
- Robust, uniformly observed sample of exoplanets: Kepler spacecraft



Borucki (2016)

Geometric transit probability truncates observed planet population

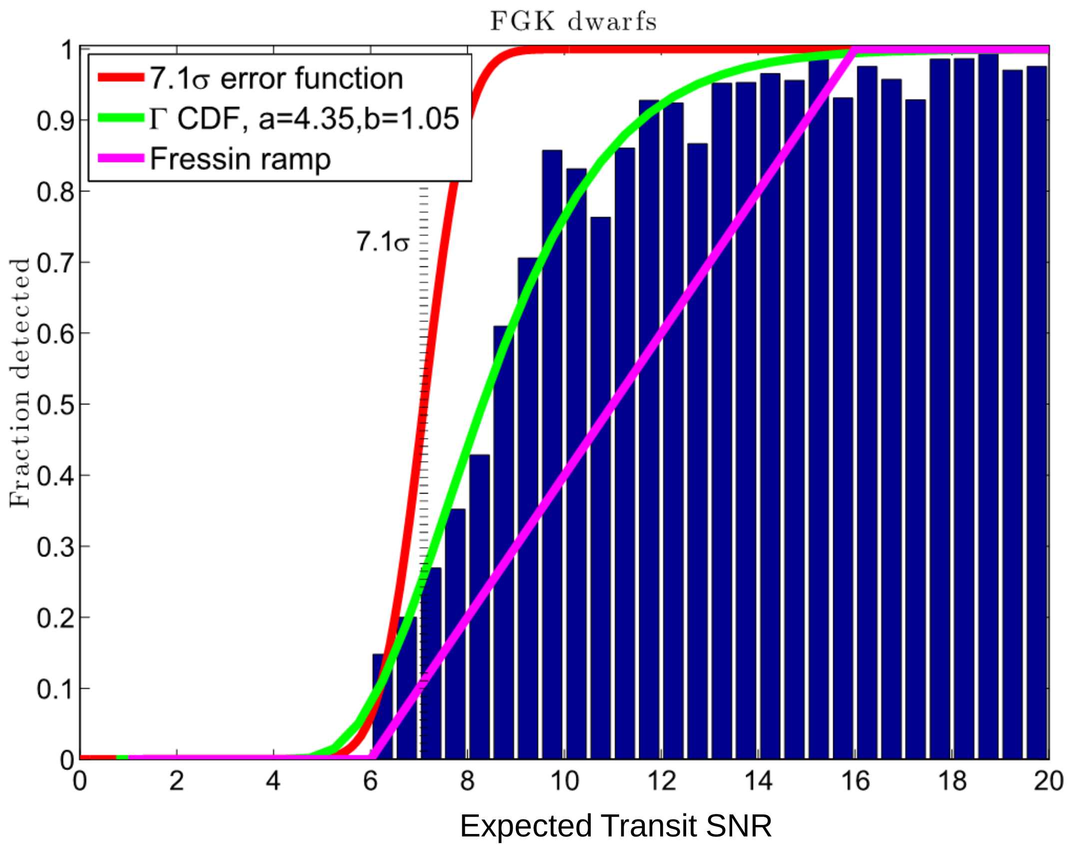
- Single planet, circular orbit: $P \sim \sin^{-1}(R_*/a)$
- Multiple planets, eccentric orbits: complicated



Winn (2010)

Observed planet population truncated due to transit SNR

- More sensitive to shorter orbital periods
- SNR depends on planet & stellar radius, intrinsic photometric variability of host star



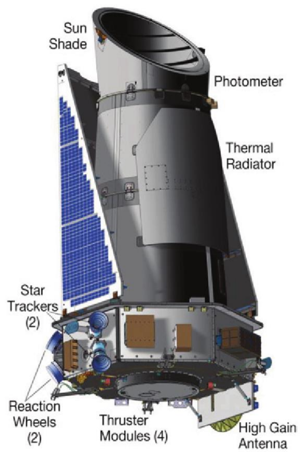
Christiansen et al. (2015)

Various models for occurrence rate estimates

- Parametric:
  - Power law
  - Broken power law
  - Exponential Cut-Off

- Non-parametric:
  - Grid of bins (2-D orbital period vs. planet radius)
  - Gaussian Process

Approximate Bayesian Computation (ABC) can be applied to Kepler data



Borucki (2016)

$$\frac{N_{r,p}}{N_{\rm targ}}$$

$$\cancel{\left| \left(\frac{n_{r,p}}{N_\star}\right)_i - \frac{N_{r,p}}{N_{\rm targ}} \right| > \epsilon_{\rm goal}}$$

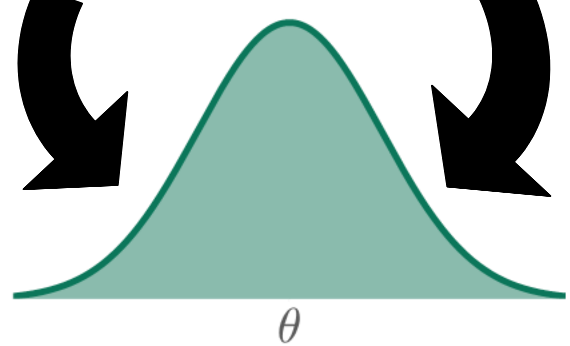
$$\left| \left(\frac{n_{r,p}}{N_\star}\right)_i - \frac{N_{r,p}}{N_{\rm targ}} \right| \leq \epsilon_{\rm goal}$$

$$\theta \in [0, 1)$$

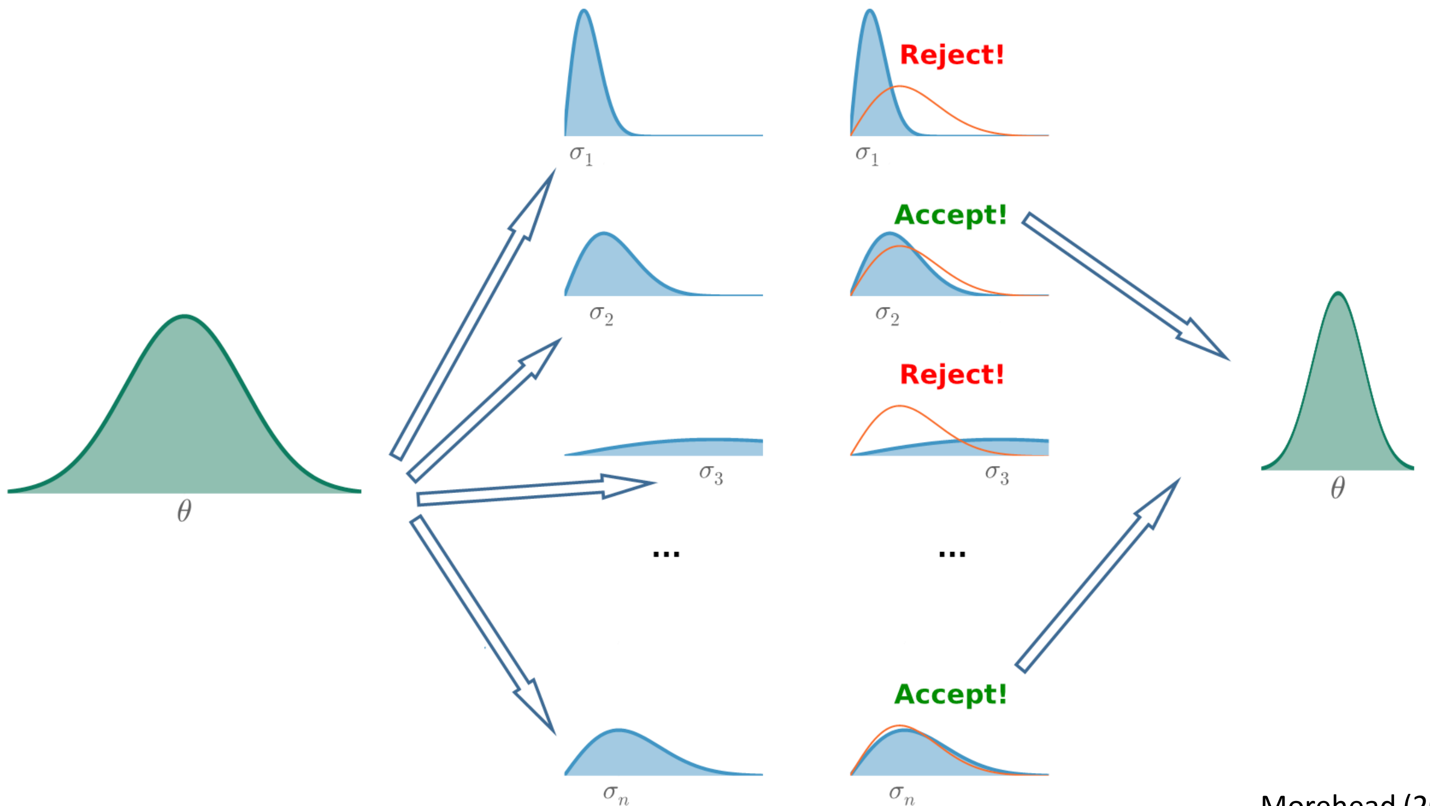
$$\left(\frac{n_{r,p}}{N_\star}\right)_1$$

$$\cdots$$

$$\left(\frac{n_{r,p}}{N_\star}\right)_{N_{\rm part}}$$



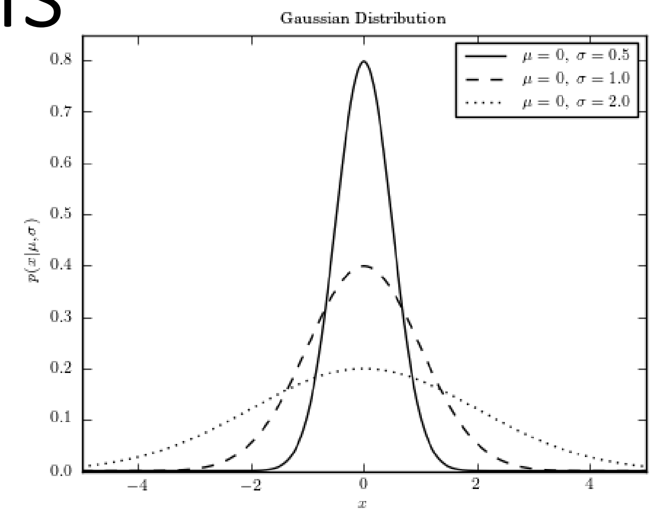
Population Monte Carlo – Importance sampling to improve efficiency



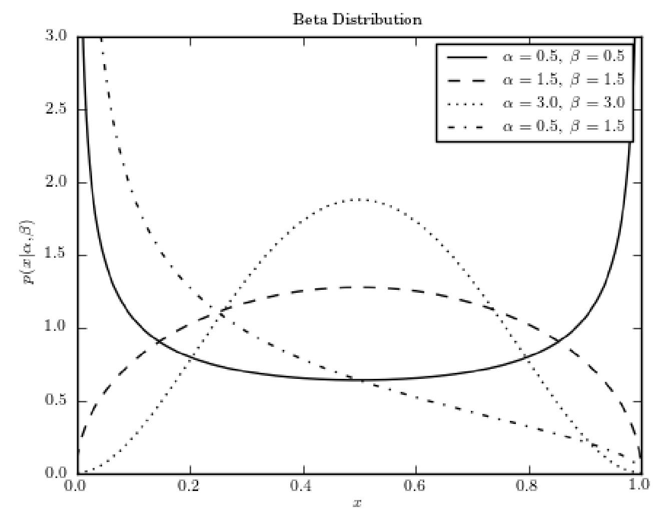
Morehead (2016)

Improving proposal efficiency with Beta distributions

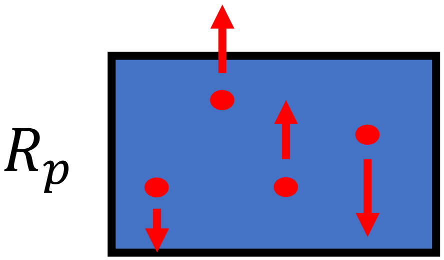
- Gaussian:
  - Negative (unphysical) rates proposed
  - Lots of rejections for initial ABC generations
- Beta:
  - Bound between [0, 1]
  - Transformed Beta allows adjustment of upper limit



astroML

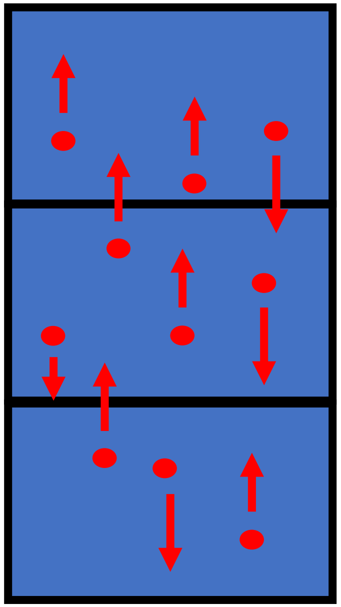


Stellar uncertainties and long-term stability motivate multi-bin fit

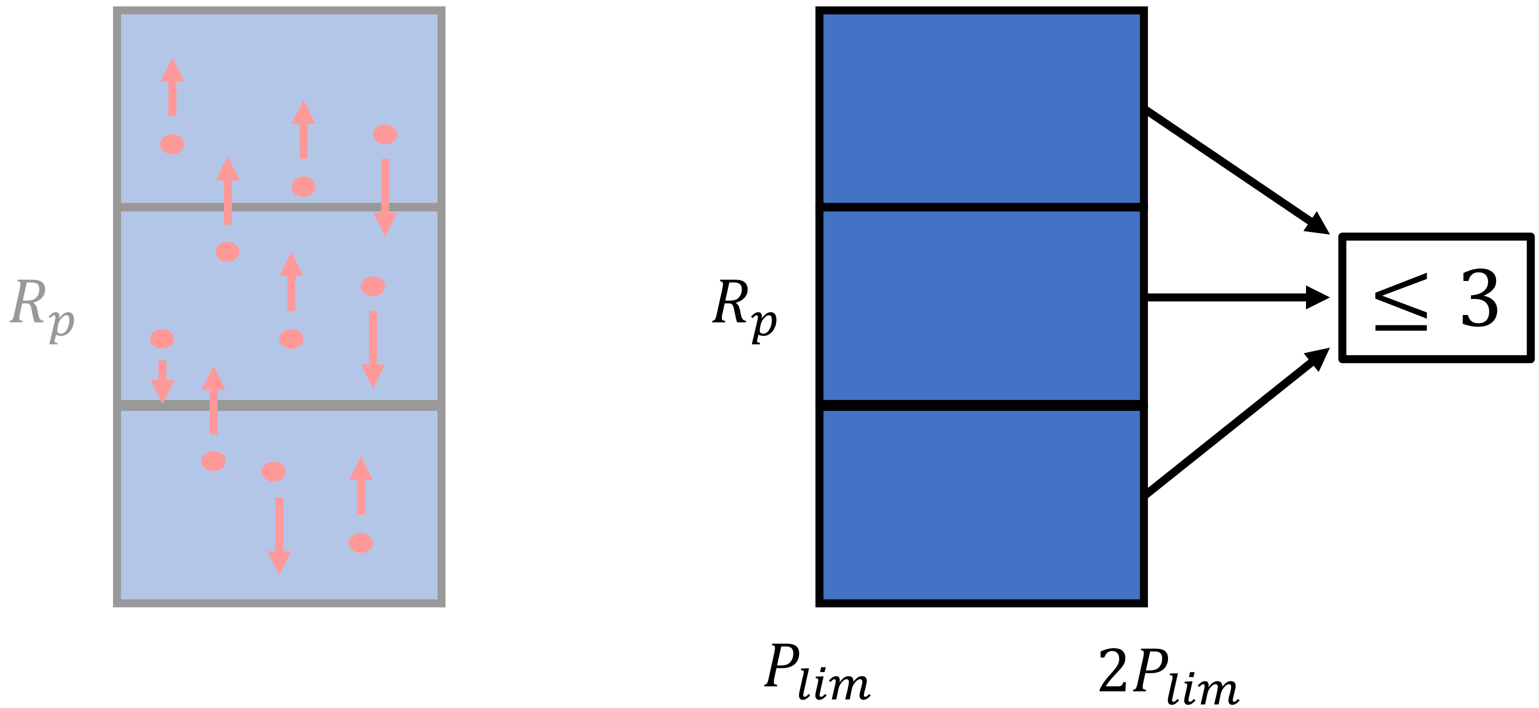


Stellar uncertainties and long-term stability motivate multi-bin fit

R_p



Stellar uncertainties and long-term stability motivate multi-bin fit



Multi-bin parameterization informs distance function

- Customized distance:

$$\rho = \sum_{n=1}^{N_r} \frac{\left|x_{n,obs} - x_{n,sim}\right|}{\sqrt{\left|x_{n,obs}\right| + \left|x_{n,sim}\right|}}$$

- Different bins have different amounts of data
  - Using L1 or L2 norm distances result in highly disparate contributions to the total distance between different bins
  - Custom distance scales contribution to total distance by magnitude of data in each bin

Multi-bin parameterization informs prior choice

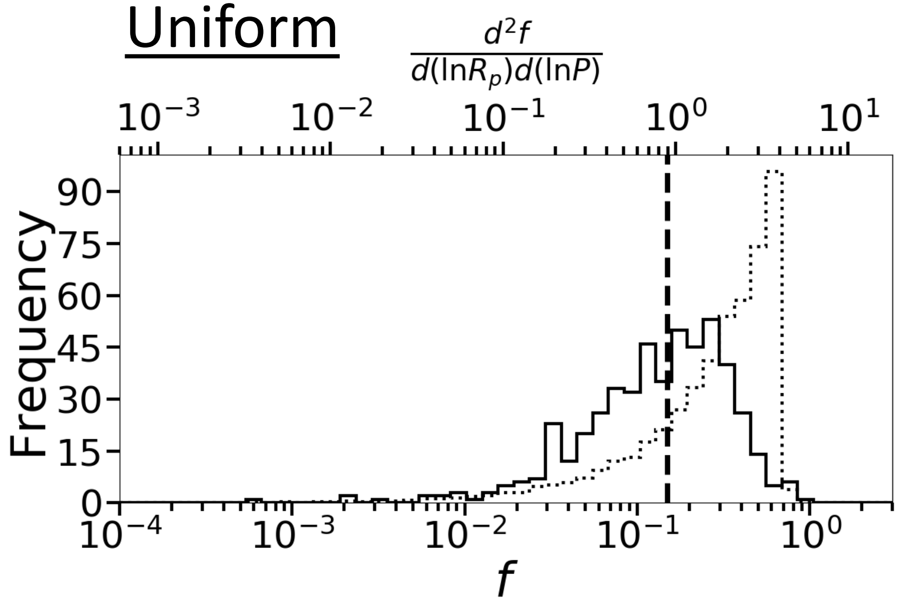
- Uniform – Full grid
  - Each bin = [0, <upper limit>]
  - Upper limit scaled by bin width (log-space) and constant factor

- Dirichlet – Small radius/long period bins (with little data where prior informs upper limits)
  - Total rate per set of bins = [0, 3]
  - Fraction of total rate per bin = [0, 1] with requirement that they sum to 1

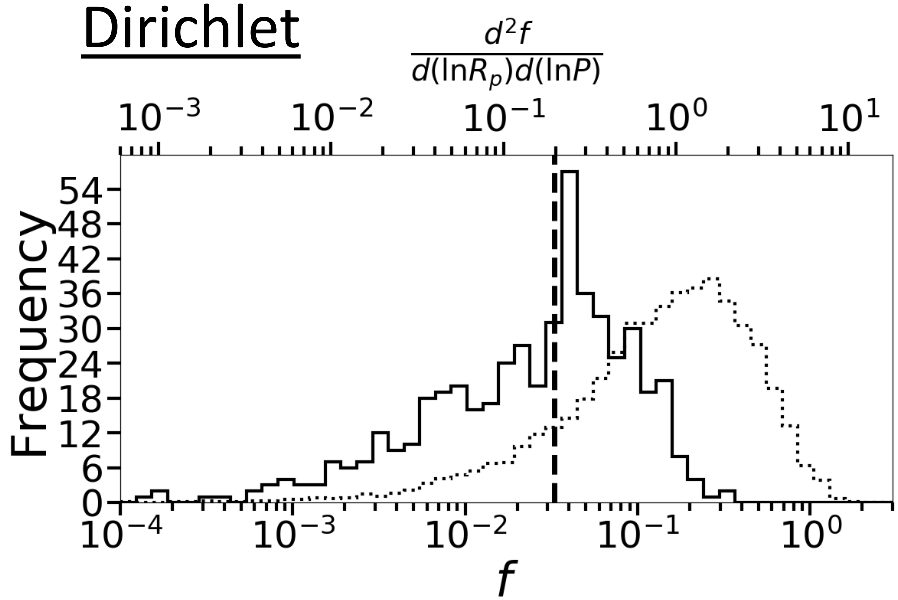
Multi-bin parameterization informs prior choice

- Two prior choices:
  - Uniform – Full grid
  - Dirichlet – Small radius/long period bins (with little data where prior informs upper limits)

Uniform



Dirichlet

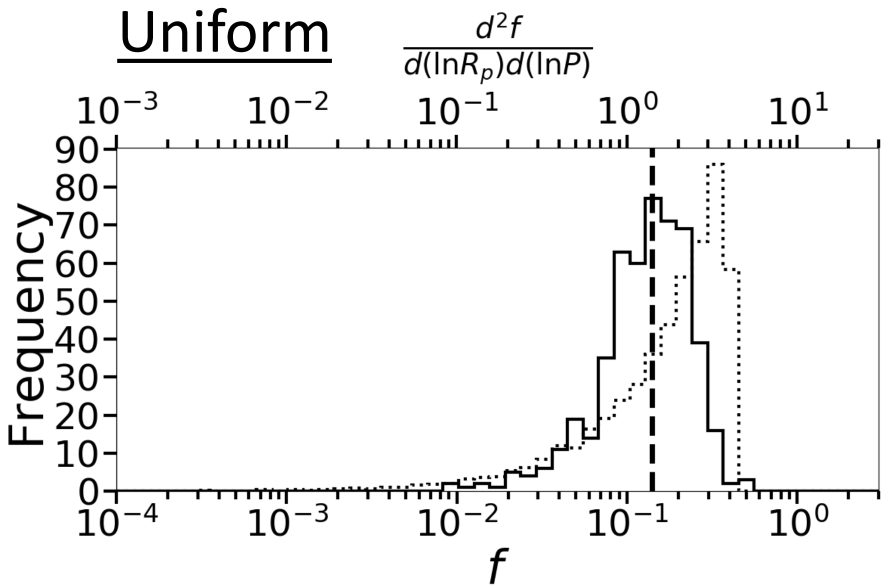


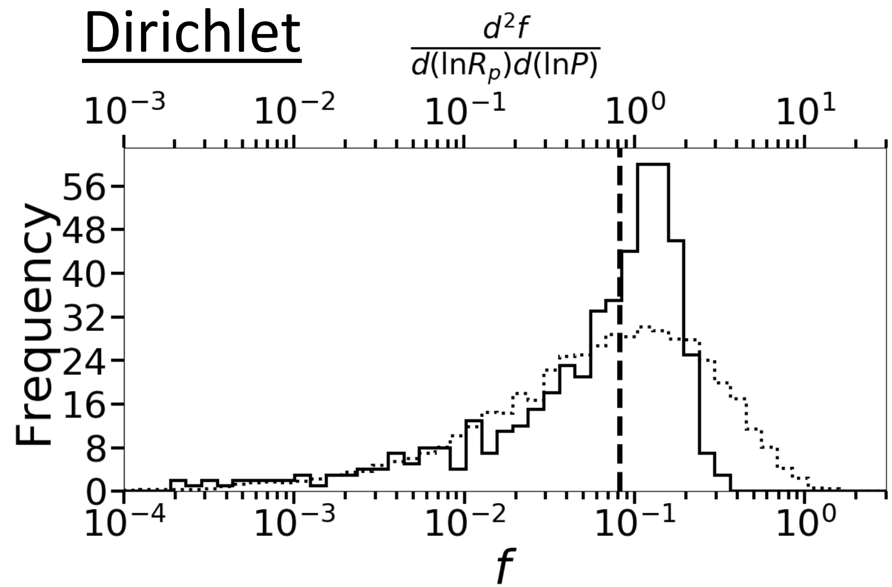
PennState

$P = 237 - 500$ d, $R_p = 1.25 - 1.5\ R_\oplus$

Multi-bin parameterization informs prior choice

- Two prior choices:
  - Uniform – Full grid
  - Dirichlet – Small radius/long period bins (with little data where prior informs upper limits)

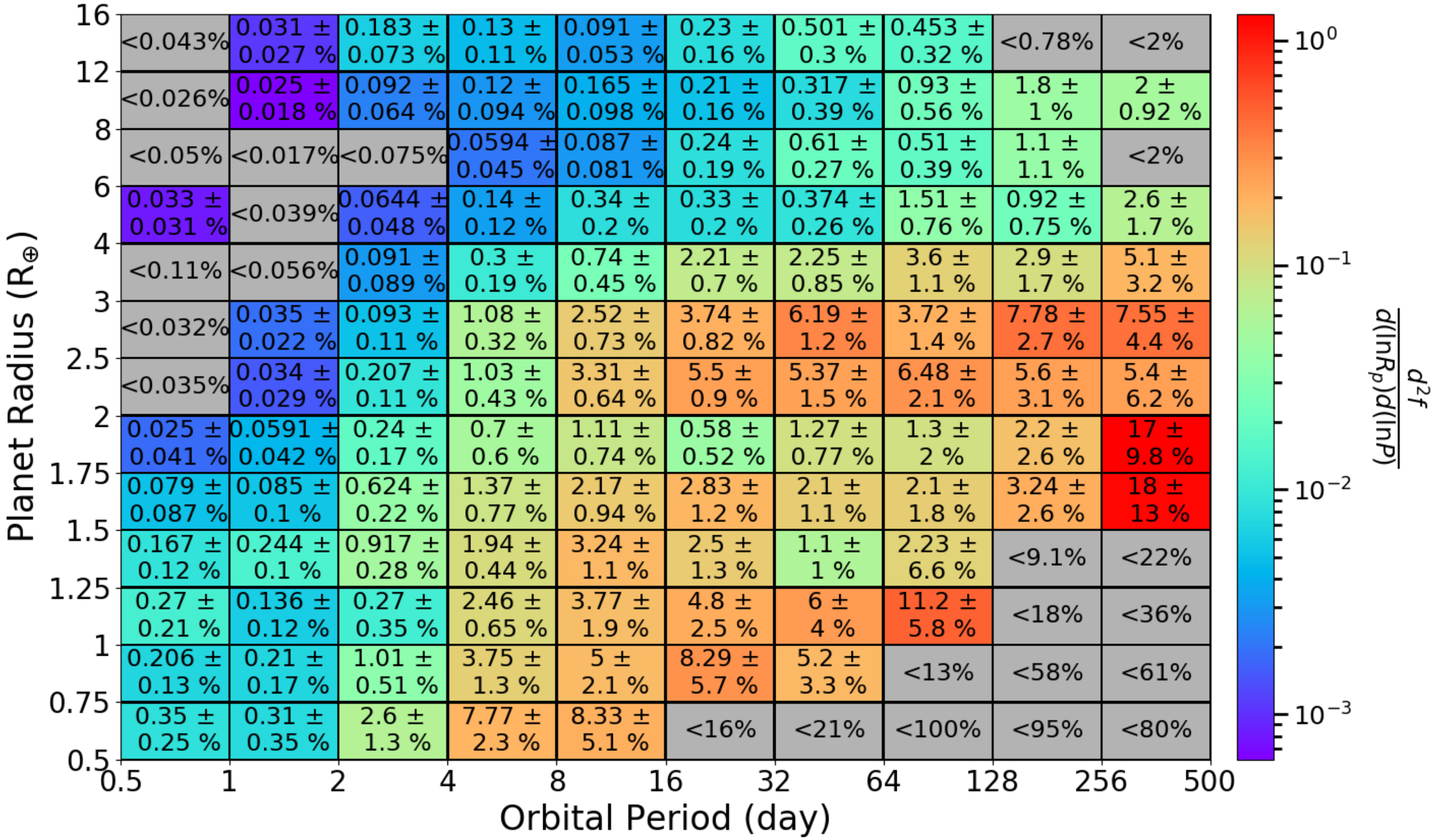




$$P = 237 - 500 \text{ d}, R_p = 1.75 - 2\ R_\oplus$$

PennState

Estimated Occurrence Rates for final Kepler catalog (DR25)



PennState

Lots of Further Work using ABC in Exoplanet Studies

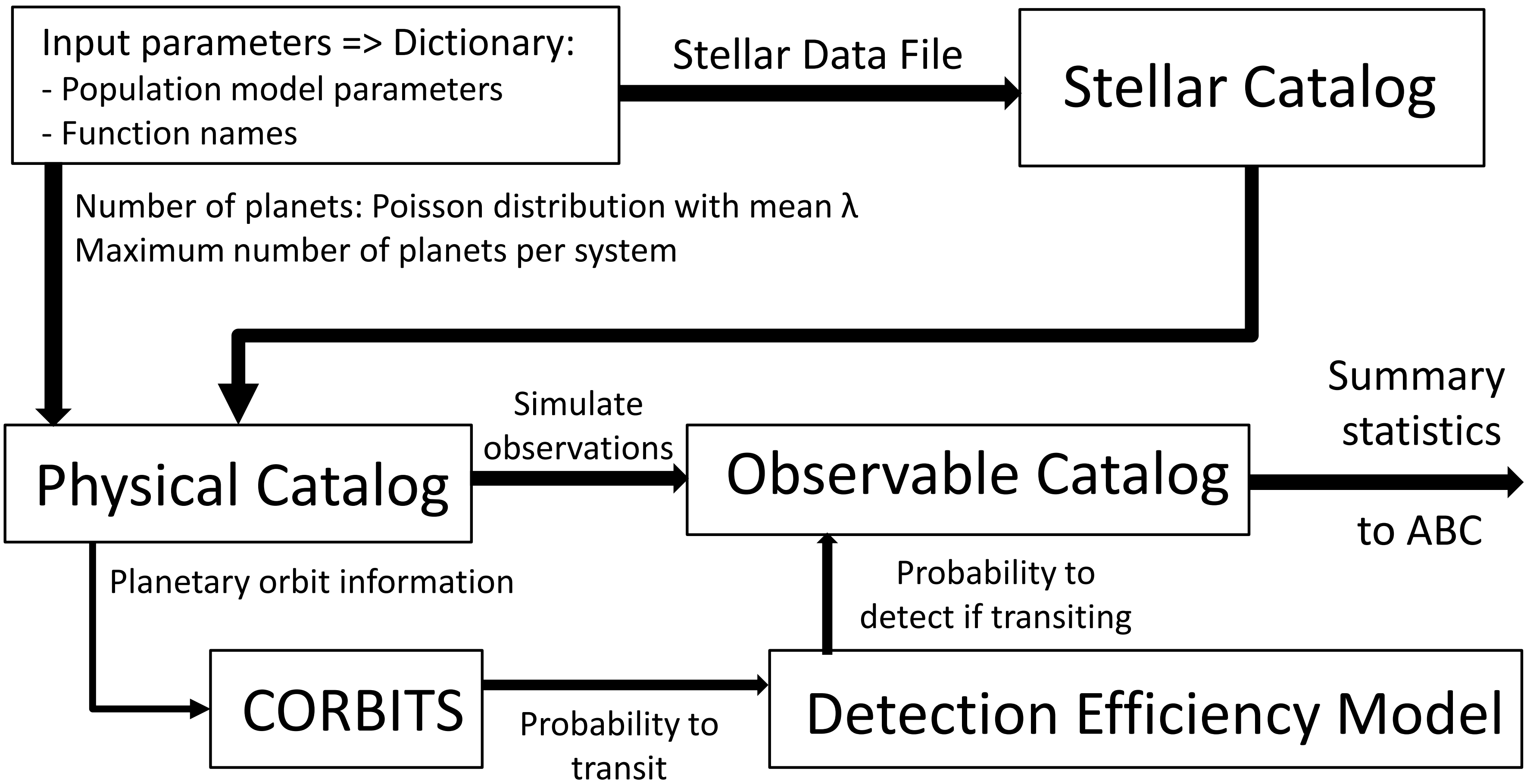
- Model improvements:
  - Reliability of planet candidate identification
  - Stellar multiplicity
- Applications:
  - Fraction of stars with at least one planet (dependent on size/period)
  - Characterize planetary system architecture

Summary

- Observed exoplanet populations truncated by geometric & detection probabilities, which cannot be adequately described by a simple likelihood function
- ABC provides a statistically rigorous method for occurrence rate estimates
- ABC model informed by parameterization and prior knowledge

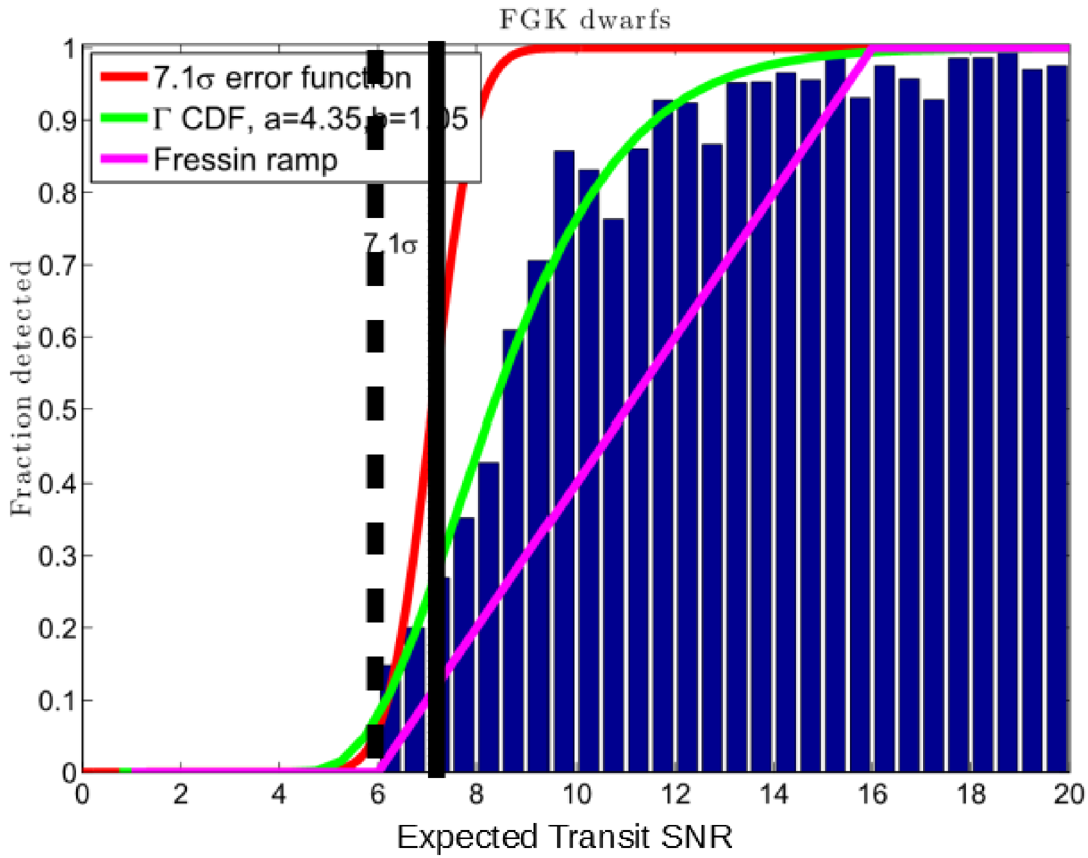
- ABC: https://github.com/eford/ABC.jl (Julia v0.6) / https://github.com/eford/ApproximateBayesianComputing.jl (Julia v1.0)
- SysSim: https://github.com/eford/ExoplanetsSysSim.jl

SysSim – Catalog Generation



PennState

Previously used calculation method (IDEM) didn't account for uncertainties



Christiansen et al. (2015)

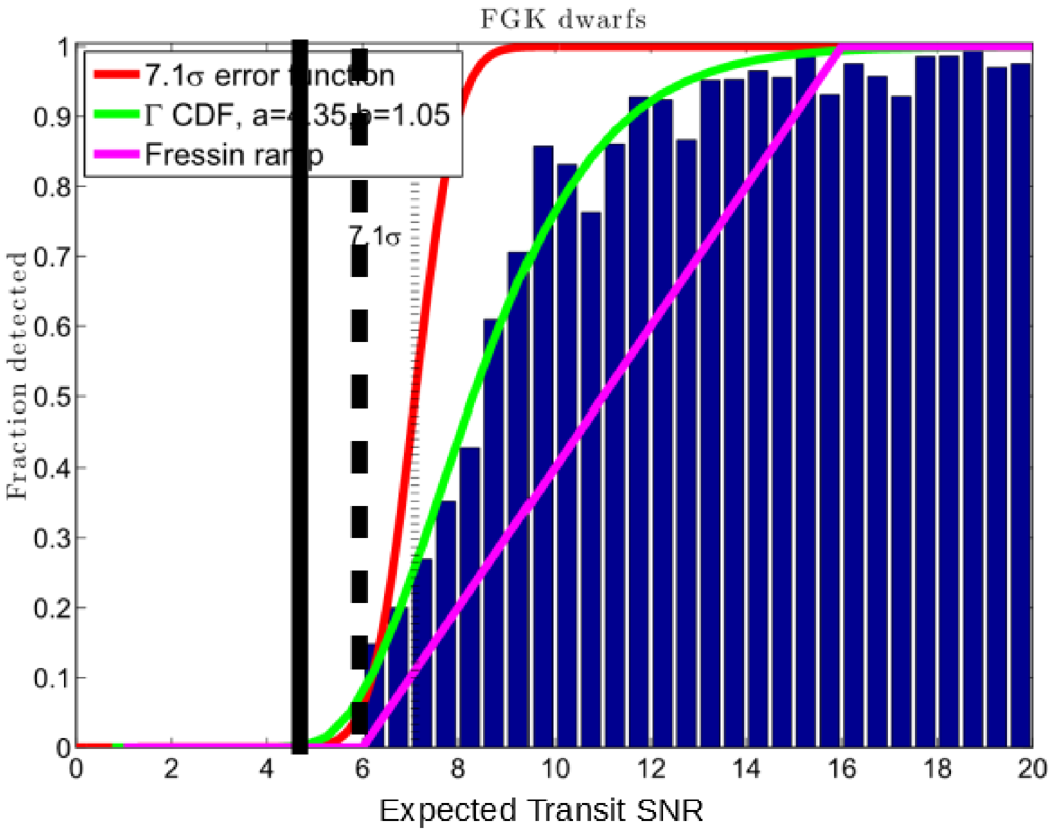
$p_{\mathrm{det},i}$

C_i

$f_{r,p}$

PennState

Previously used calculation method (IDEM) didn't account for uncertainties



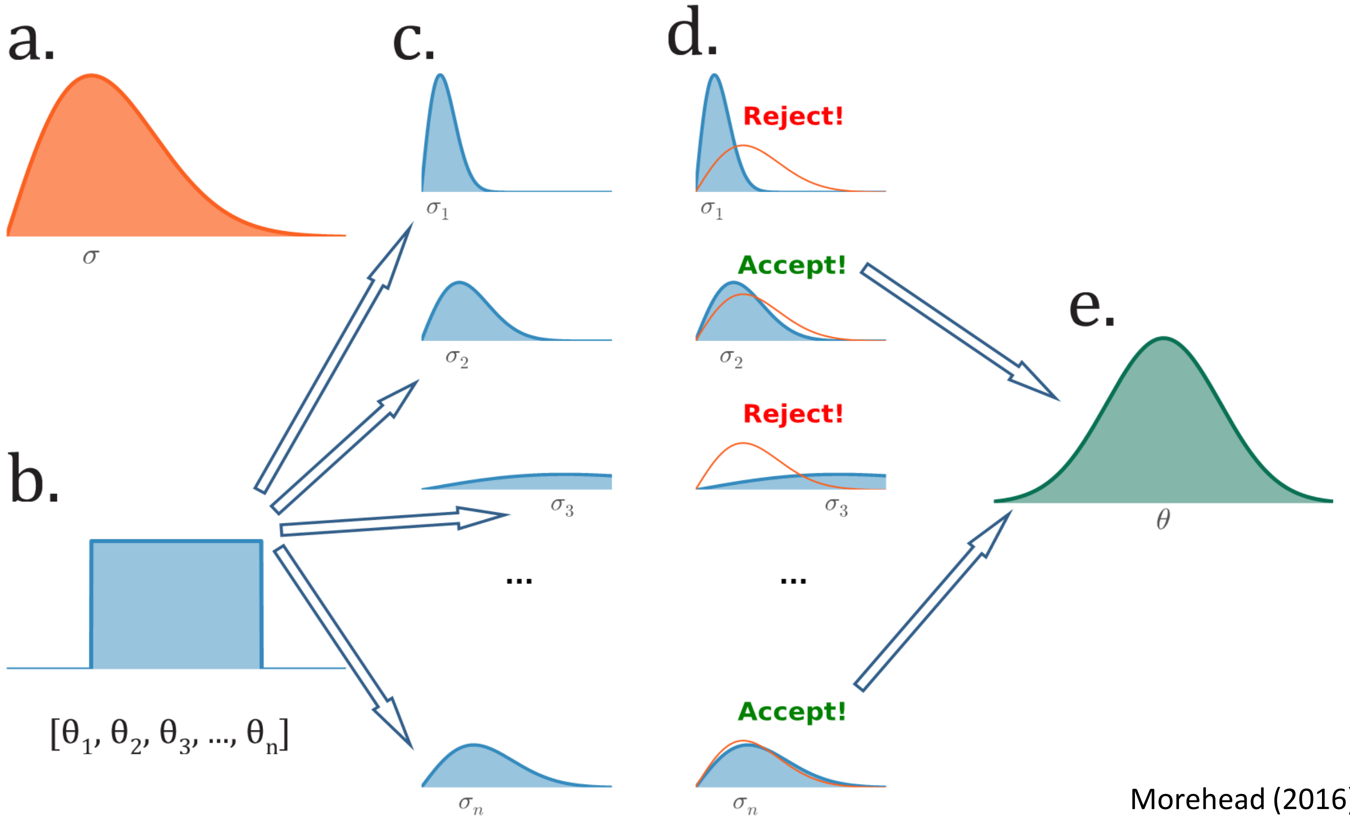
Christiansen et al. (2015)

$$p_{\mathrm{det},i} \sim 0$$

No Detection

$$f_{r,p}$$

Approximate Bayesian Computation (ABC) enables rigorous statistical modeling



Morehead (2016)