Hierarchical Bayesian Modeling of Planet Populations

You've found planets in your data (or not) . . .

. . . now what?!

Angie Wolfgang
NSF Postdoctoral Fellow, Penn State

Why Astrostats?

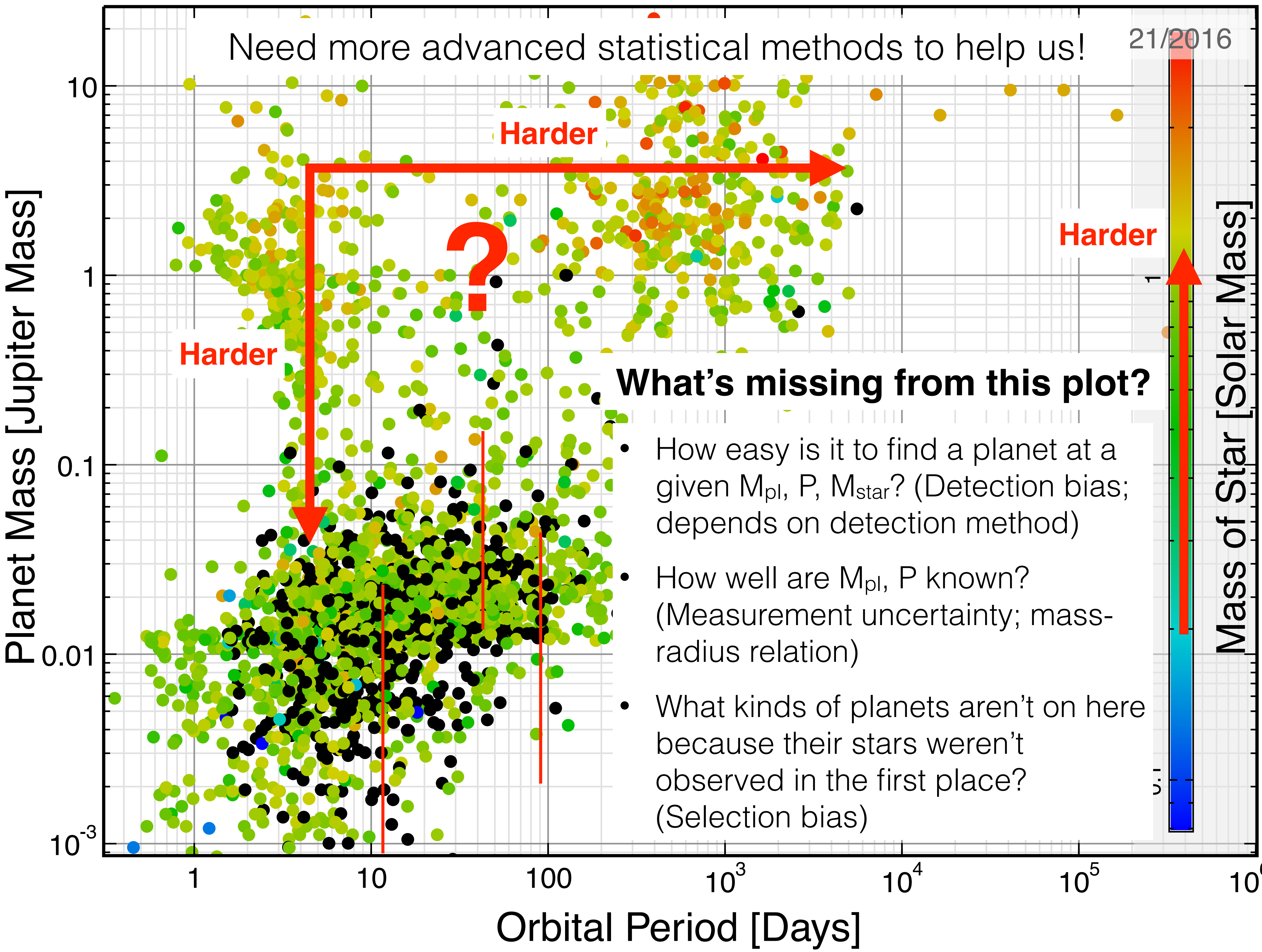
- This week: a sample of the richness of our research problems and the wide range of statistical tools to be considered. Stats + exoplanets: a burgeoning field! Follow your confusion - you will be among the first to apply new tools to the BIG data ahead!

- Stats is not something you do after the physics, or to supplement the physics . . . it is what ENABLES the physics!! Stats gives us efficient tools to explore the data and tells us how to deal with uncertainty so we can make accurate inferences and conclusions. (Cosmologists understand this, but many galactic, stellar, exoplanet astronomers still don't.)

- We come from a culture of spherical cows: many astronomers choose to ignore (or don't fully understand) the assumptions that underly the (few) statistical tools they use . . . and that can get us into trouble (retracted detections, biased results). Join me in showing our field how we can do better science!

Observed Exoplanet Population

Lots of (hetereogenous) planet detections ≠ an unbiased sample ready to compare to theory



What's missing from this plot?

- How easy is it to find a planet at a given M_{pl}, P, M_{star}? (Detection bias; depends on detection method)
- How well are M_{pl}, P known? (Measurement uncertainty; mass-radius relation)
- What kinds of planets aren't on here because their stars weren't observed in the first place? (Selection bias)

Hierarchical Modeling

is a statistically rigorous way

to make **scientific inferences**

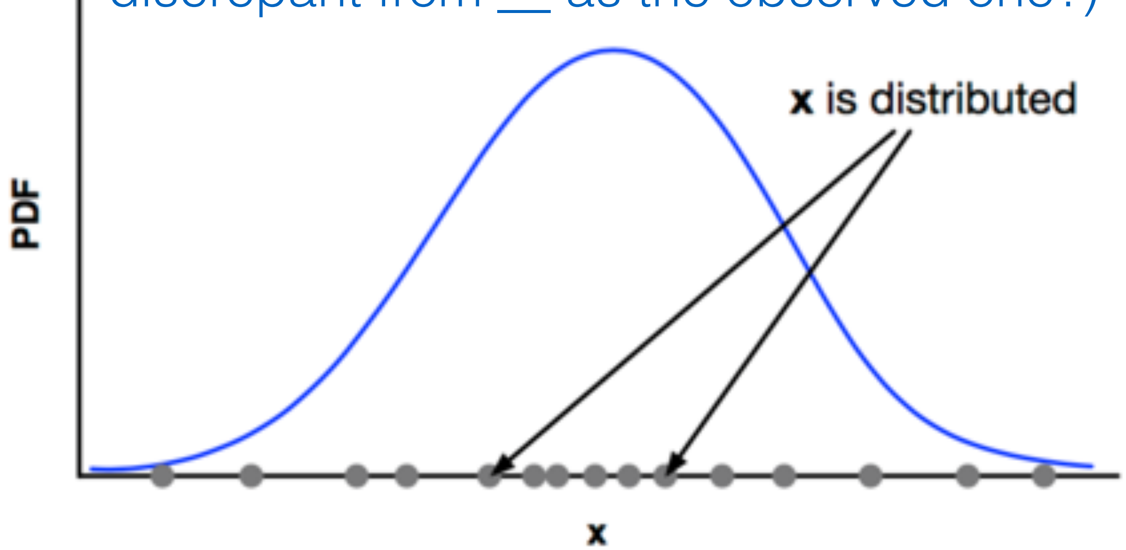
about a **population** (or specific object)

based on many **individuals** (or observations).

Frequentist multi-level modeling techniques exist,
but we will discuss the Bayesian approach today.

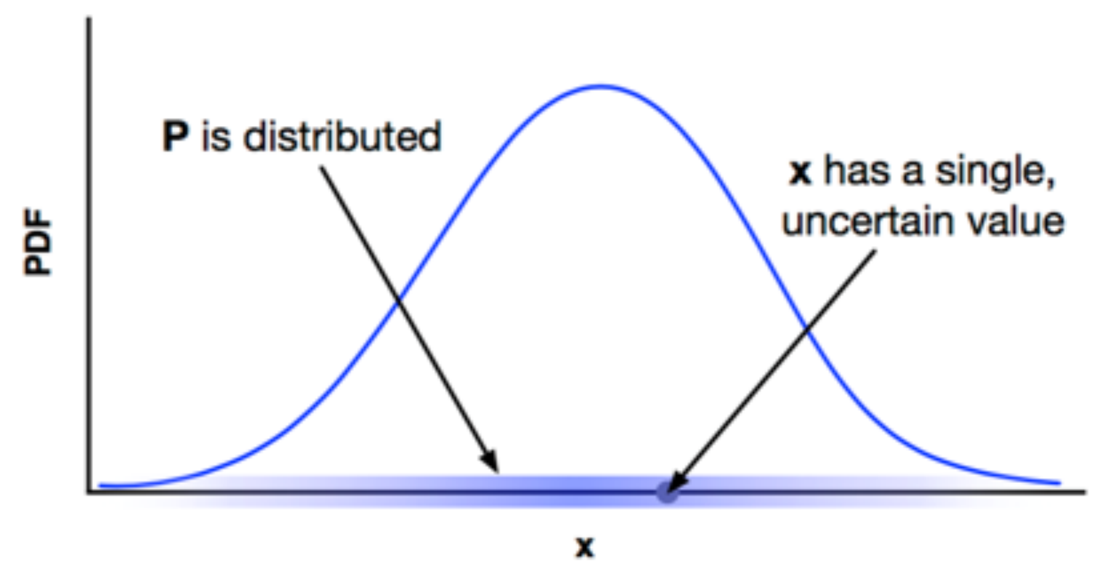
Frequentist: variability of sample

(If __ is the true value, what fraction of many hypothetical datasets would be as or more discrepant from __ as the observed one?)



Bayesian: uncertainty of inference

(What's the probability that __ is the true value given the current data?)



Understanding Bayes

Bayes' Theorem

(straight out of conditional probability)

(On Tuesday we learned how to evaluate **p(θ|x)** numerically with MCMC to infer **θ** from **x** ...)

$$p(\theta|x) \propto p(x|\theta)\, p(\theta)$$

posterior likelihood prior

(But let's get a better intuition for the statistical model itself.)

x = data
θ = the parameters of a model that can produce the data
p() = probability density distribution of
| = "conditional on", or "given"

p(θ) = prior probability
(How probable are the possible values of θ in nature?)

p(x|θ) = likelihood, or sampling distribution
(Ties your model to the data probabilistically: how likely is the data you observed given specific θ values?)

p(θ|x) = posterior probability
(A "new prior" distribution, updated with information contained in the data: what is the probability of different θ values given the data and your model?)

Applying Bayes

p(θ|x) ∝ p(x|θ) p(θ)

posterior likelihood prior

Example (1-D): Getting orbital parameters from RV data

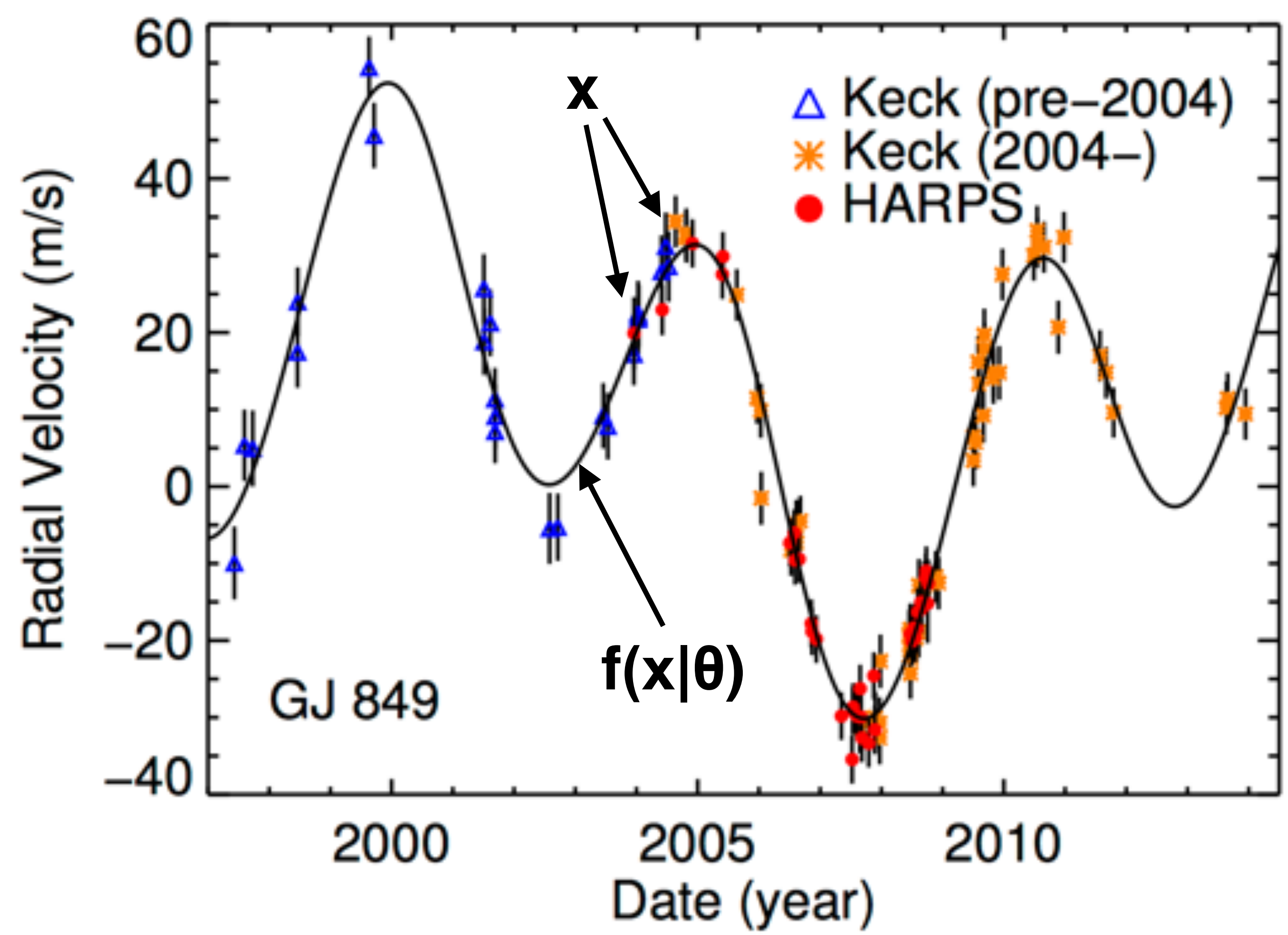
x = 117 RV measurements

Model: 2-body Keplerian orbit

θ = orbital period (P),
time of pericenter (T_p),
eccentricity (e),
argument of periastron (w),
RV semiamplitude (K),
stellar jitter

Model can be summarized as **f(x|θ)**:
Maps **θ** → **x.**

But this is NOT **p(x|θ)** because
f(x|θ) is not a probability distribution!!



Feng et al. 2015, using RUN DMC
(Differential Evolution MCMC)
by Nelson et al. 2014a

Applying Bayes

p(θ|x) ∝ p(x|θ) p(θ)

posterior likelihood prior

Example (1-D): Getting orbital parameters from RV data

x = 117 RV measurements

Model: 2-body Keplerian orbit

θ = orbital period (P),
time of pericenter (T_p),
eccentricity (e),
argument of periastron (w),
RV semiamplitude (K),
stellar jitter

Model can be summarized as **f(x|θ)**:
Maps **θ** → **x.**

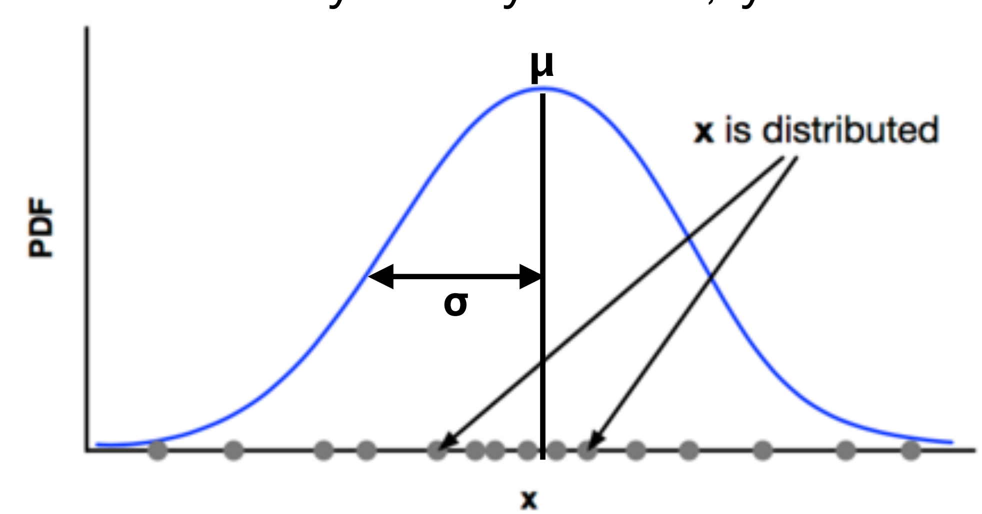
But this is NOT **p(x|θ)** because **f(x|θ)** is not a probability distribution!!

If use X^2 for fitting, then you are implicitly assuming that:

$$p(x_i|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where **μ = f(x_i|θ)**
and **σ** = "statistical measurement error"

i.e. you are assuming "Gaussian noise"
(if you could redo a specific x_i the same way many times, you'd find:)



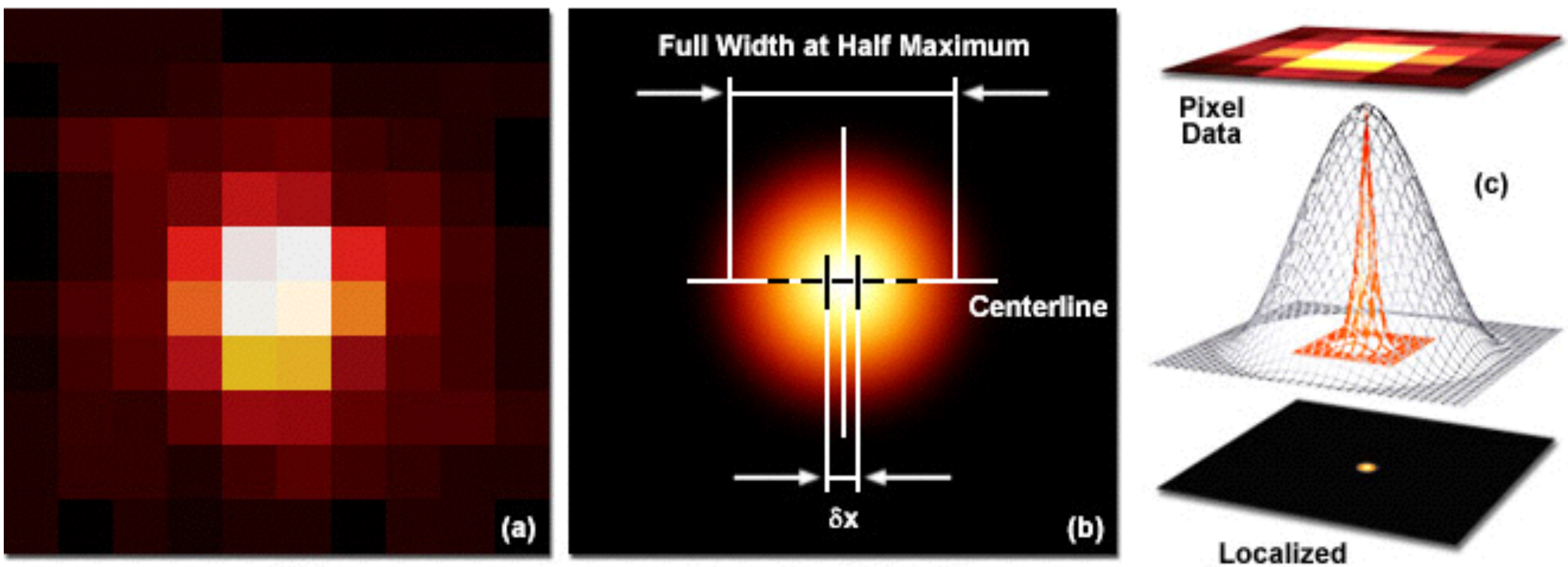
Applying Bayes

p(θ|x) ∝ p(x|θ) p(θ)

posterior likelihood prior

Example (2-D): Fitting a PSF to an image

Both likelihood **and** model are Gaussian!!



Raw Data Gaussian Fit Function

x = matrix of pixel brightnesses

θ = **μ, σ** of Gaussian (location, FWHM of PSF)

f(x|θ) = 2-D Gaussian

p(x|θ) $= \frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$

where **μ = f(x|θ)** and **Σ** = noise (possibly spatially correlated)

Applying Bayes

$$\mathbf{p(\theta|x) \propto p(x|\theta)\ p(\theta)}$$

posterior likelihood prior

Example (1-D): Getting orbital parameters from RV data

x = 117 RV measurements

Model: 2-body Keplerian orbit

θ = orbital period (P),
time of pericenter (T_p),
eccentricity (e),
argument of periastron (w),
RV semiamplitude (K),
stellar jitter

Model can be summarized as **f(x|θ)**:
Maps **θ** → **x.**

But this is NOT **p(x|θ)** because **f(x|θ)** is not a probability distribution!!

Ok, now we know of one way to write **p(x|θ)**.

What about **p(θ)?**

1) If we have a previous measurement/inference of that object's metallicity, redshift, etc., use it ***with its error bars*** as **p(θ).** (Usually "measured" via X^2, so **p(θ)** is Gaussian with μ = measurement and σ = error. BUT full posteriors from previous analysis is better.)

2) Choose wide, uninformative distributions for all the parameters we don't know well.

3) If analysis is analytical (vs numerical), use conjugate prior.

4) Use distributions in nature from previous observations of similar objects.

Going Hierarchical

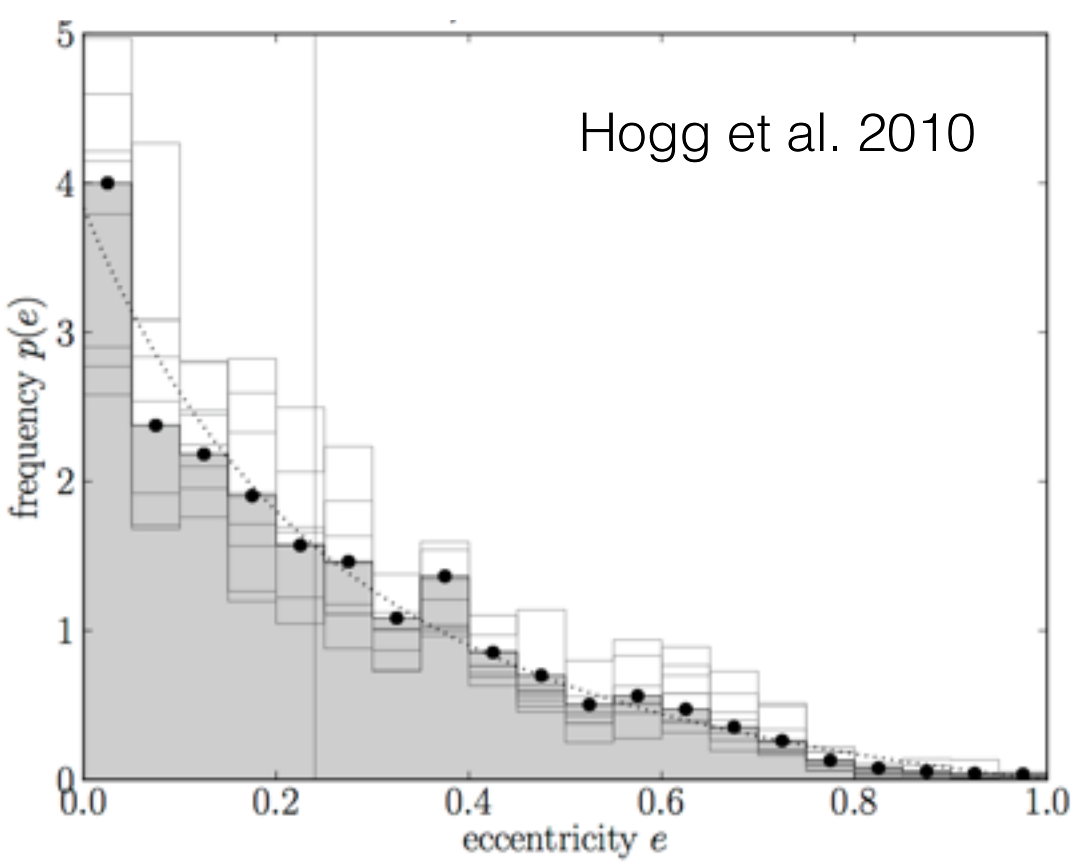
Option #3 for **p(θ):**

Use distributions in nature from previous observations of similar objects.

Histograms of population properties, when normalized, can be interpreted as probability distributions for individual parameters:

$$p(\theta) = n(\theta|\alpha) / \int n(\theta|\alpha)d\theta = p(\theta|\alpha)$$

where **n(θ|α)** is the function with parameters **α** that was fit to the histogram (or even the histogram itself, if you want to deal with a piecewise function!)



For example, **e** was part of the **θ** for RV fitting.

One could use the dashed line (parametric form below) as

$$p(e) = p(e|\alpha) = n(e|\alpha) / \int n(e|\alpha)de$$

with **n(e|α)** $= \left[\frac{1}{[1+e]^b} - \frac{e}{a^b}\right]$ and **α** = **{a=2,b=4}.**

But BE CAREFUL of detection bias, selection effects, upper limits, etc.!!!!!!

Going Hierarchical

Option #3 for **p(θ):**

Use distributions in nature from previous observations of similar objects.

Histograms of population properties, when normalized, can be interpreted as probability distributions for individual parameters:

$$\mathbf{p(\theta) = n(\theta|\alpha) / \int n(\theta|\alpha)d\theta = p(\theta|\alpha)}$$

where **n(θ|α)** is the function with parameters **α** that was fit to the histogram (or even the histogram itself, if you want to deal with a piecewise function!)

Abstracting again ….

$$\mathbf{p(\theta|x) \propto p(x|\theta)\ p(\theta)}$$

posterior likelihood prior

$$\mathbf{p(\theta|x) \propto p(x|\theta)\ p(\theta|\alpha)}$$

(Almost there!!)

posterior likelihood prior

Population helps make inference on individual …

Going Hierarchical

. . . but what if we want to use the individuals to infer things (the α's) about the population?

i.e., **$p(\theta|\alpha)$** contains some interesting physics and getting values for **α** given the data can help us understand it.

$$p(\theta|x) \propto p(x|\theta)\, p(\theta|\alpha)$$

posterior likelihood prior

$$p(\alpha,\theta|x) \propto p(x|\theta,\alpha)\, p(\theta|\alpha)\, p(\alpha)$$

posterior likelihood prior

If you truly don't care about the parameters for the individual objects, then you can marginalize over them:

$$p(\alpha|x) \propto \left[\int p(x|\theta,\alpha)\, p(\theta|\alpha)\, d\theta\right] p(\alpha)$$

posterior likelihood prior

Graphically:

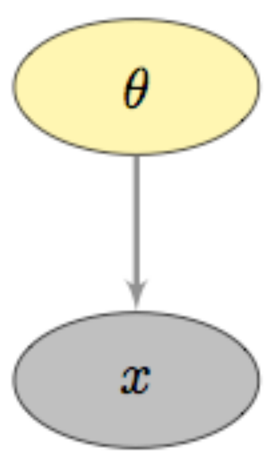
"Regular" Bayes:

$$p(\theta|x) \propto p(x|\theta)\, p(\theta)$$

posterior likelihood prior

Parameters

Observables



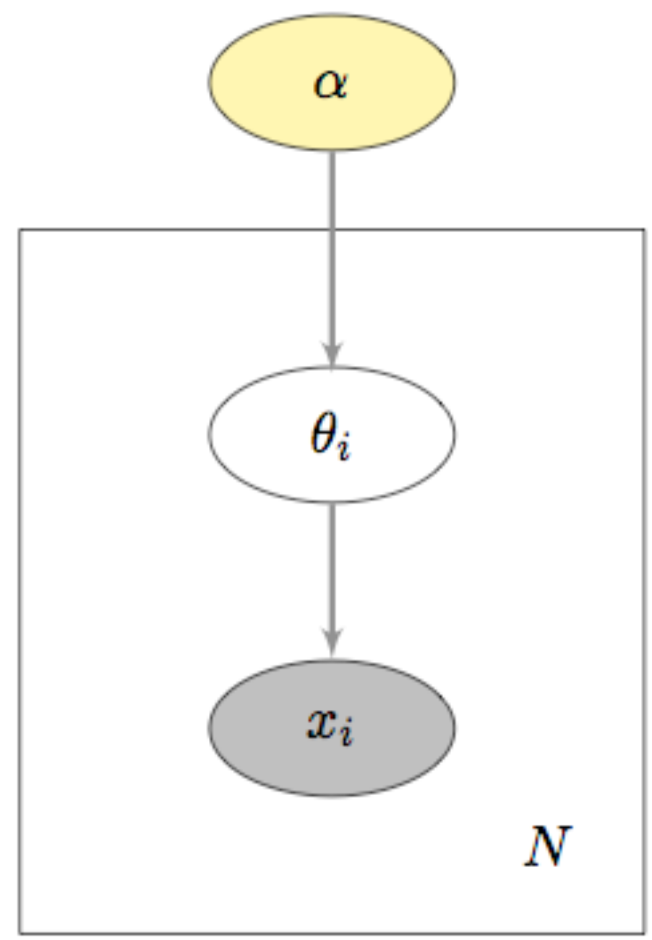
Tilde notation:

$x \sim p(x|\theta)$
$\theta \sim p(\theta)$

Hierarchical Bayes:

$$p(\alpha,\theta|x) \propto p(x|\theta,\alpha)\, p(\theta|\alpha)\, p(\alpha)$$

posterior likelihood prior



Population Parameters

Individual Parameters

Observables

$x \sim p(x|\theta,\alpha)$
$\theta \sim p(\theta|\alpha)$
$\alpha \sim p(\alpha)$

Graphically:

"Regular" Bayes:

$$p(\theta|x) \propto p(x|\theta)\, p(\theta)$$

posterior likelihood prior

Hierarchical Bayes:

$$p(\alpha,\theta|x) \propto p(x|\theta,\alpha)\, p(\theta|\alpha)\, p(\alpha)$$

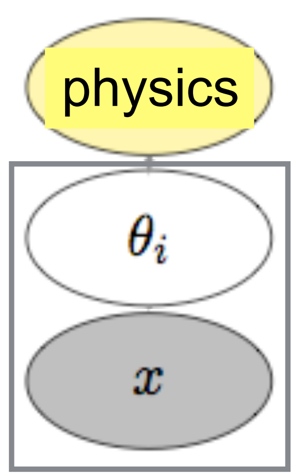
posterior likelihood prior

Even for an individual object, connection between parameters and observables can involve several layers. (Example: measuring mass of a planet)

Parameters

Latent Variables

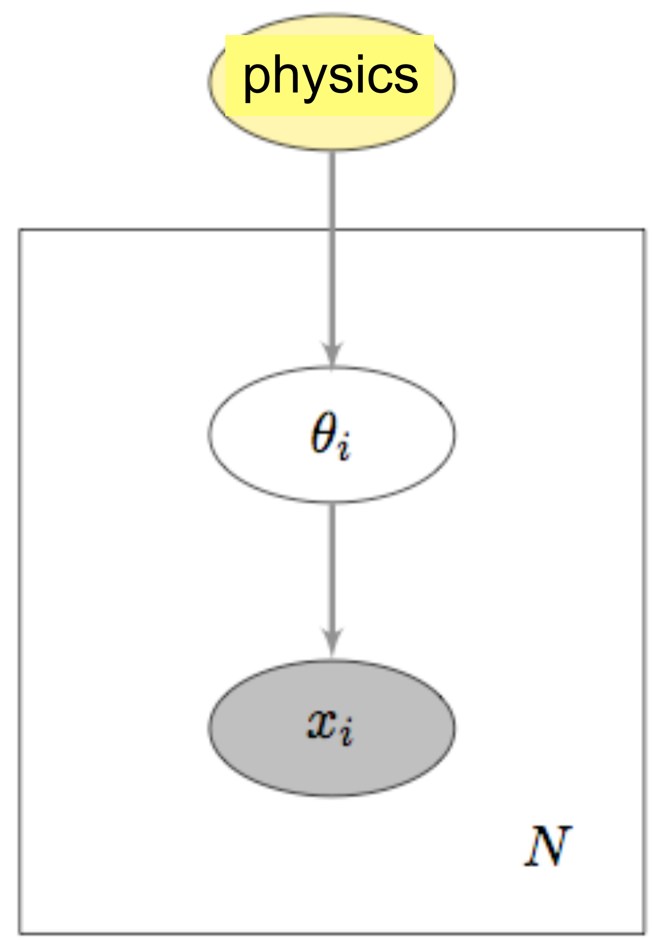
Observables



M_{pl}

RVs

spectra

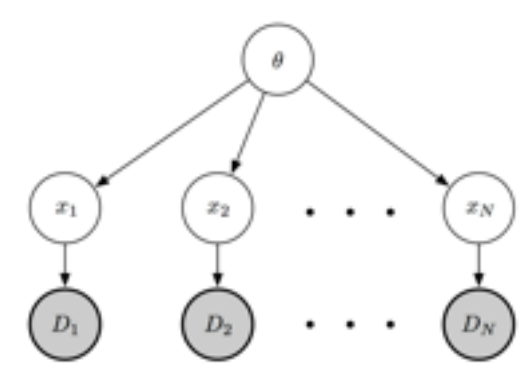


Population Parameters

Individual Parameters

Observables

Conditional independence between individuals:



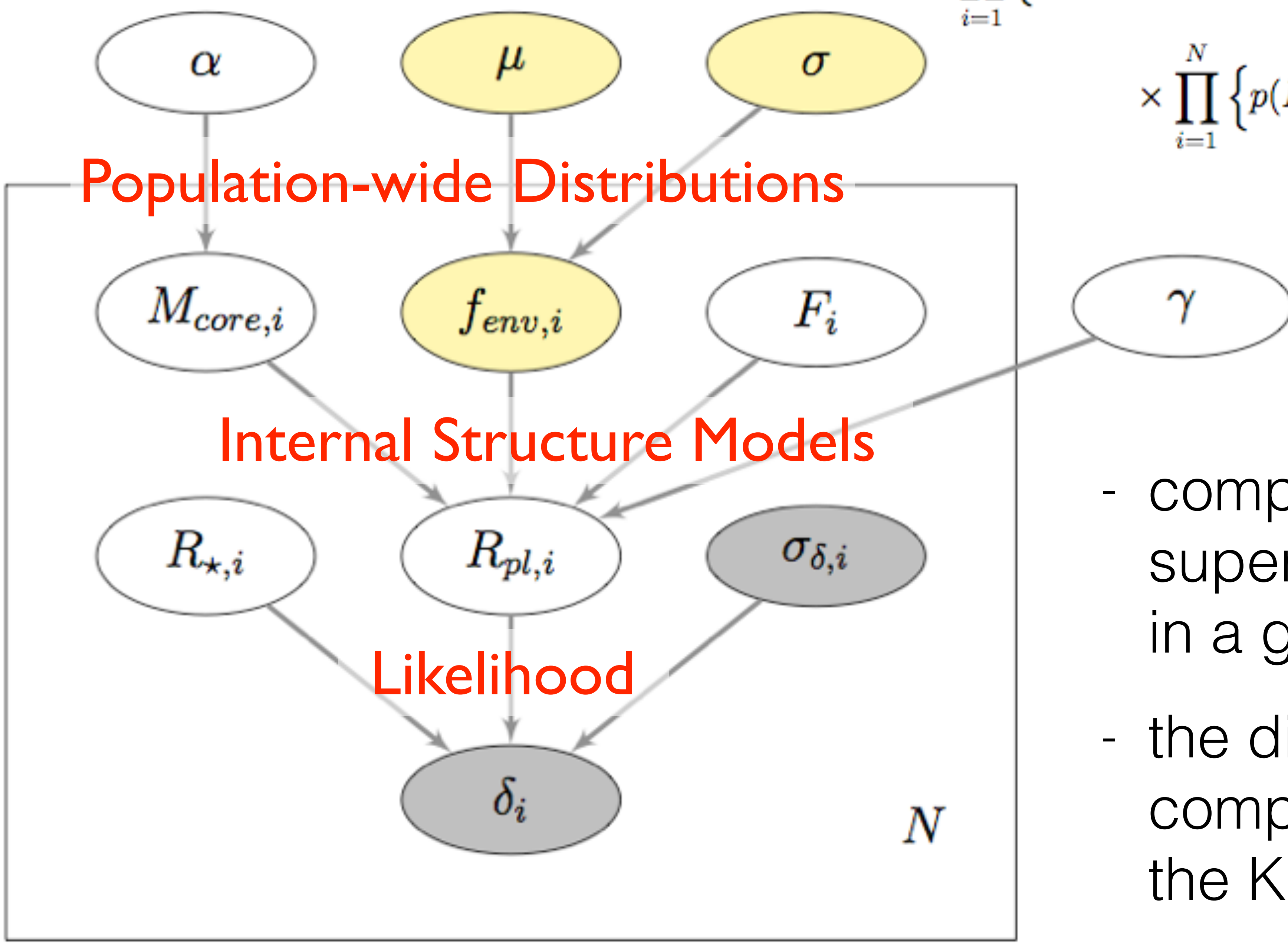
HBM in Action: Model

Exoplanet compositions: Wolfgang & Lopez, 2015

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{X}) = p(\{R_{pl,i}, M_{core,i}, f_{env,i}\}, \{\alpha, \mu, \sigma, \gamma\} | \delta_i, \sigma_{\delta,i}, F_i) \propto$$

$$\prod_{i=1}^{N} \left\{ p(\delta_i | \sigma_{\delta,i}, R_{pl,i}, R_{\star,i}, M_{core,i}, f_{env,i}, F_i, \alpha, \mu, \sigma, \gamma) \right\}$$

$$\times \prod_{i=1}^{N} \left\{ p(R_{\star,i}) p(M_{pl,i} | \alpha) p(f_{env,i} | \mu, \sigma) \right\} p(\alpha) p(\mu) p(\sigma) p(\gamma)$$



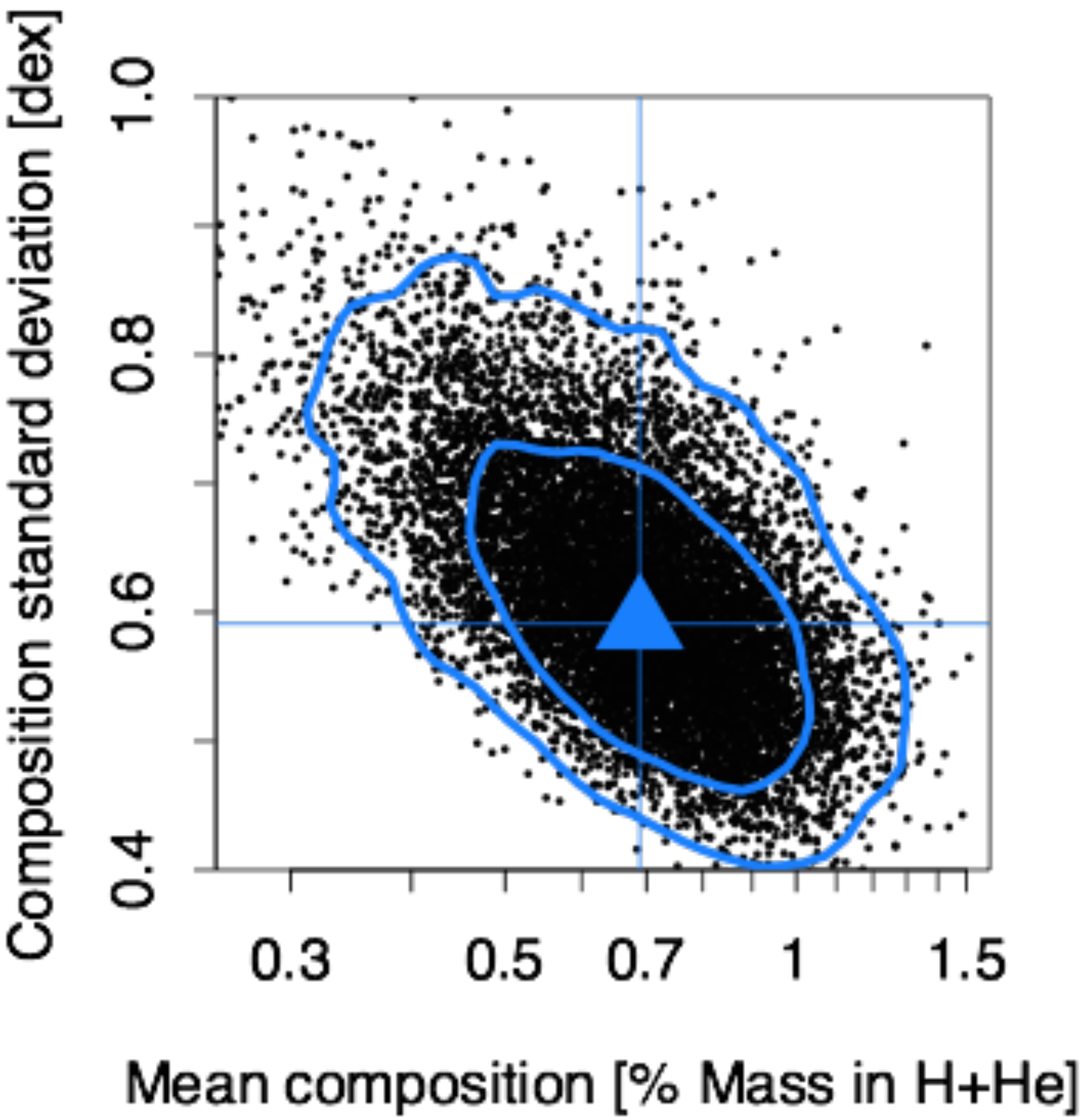
Wanted to understand BOTH:

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

HBM in Action: Results

Exoplanet compositions: Wolfgang & Lopez, 2015

Posterior on population parameters:

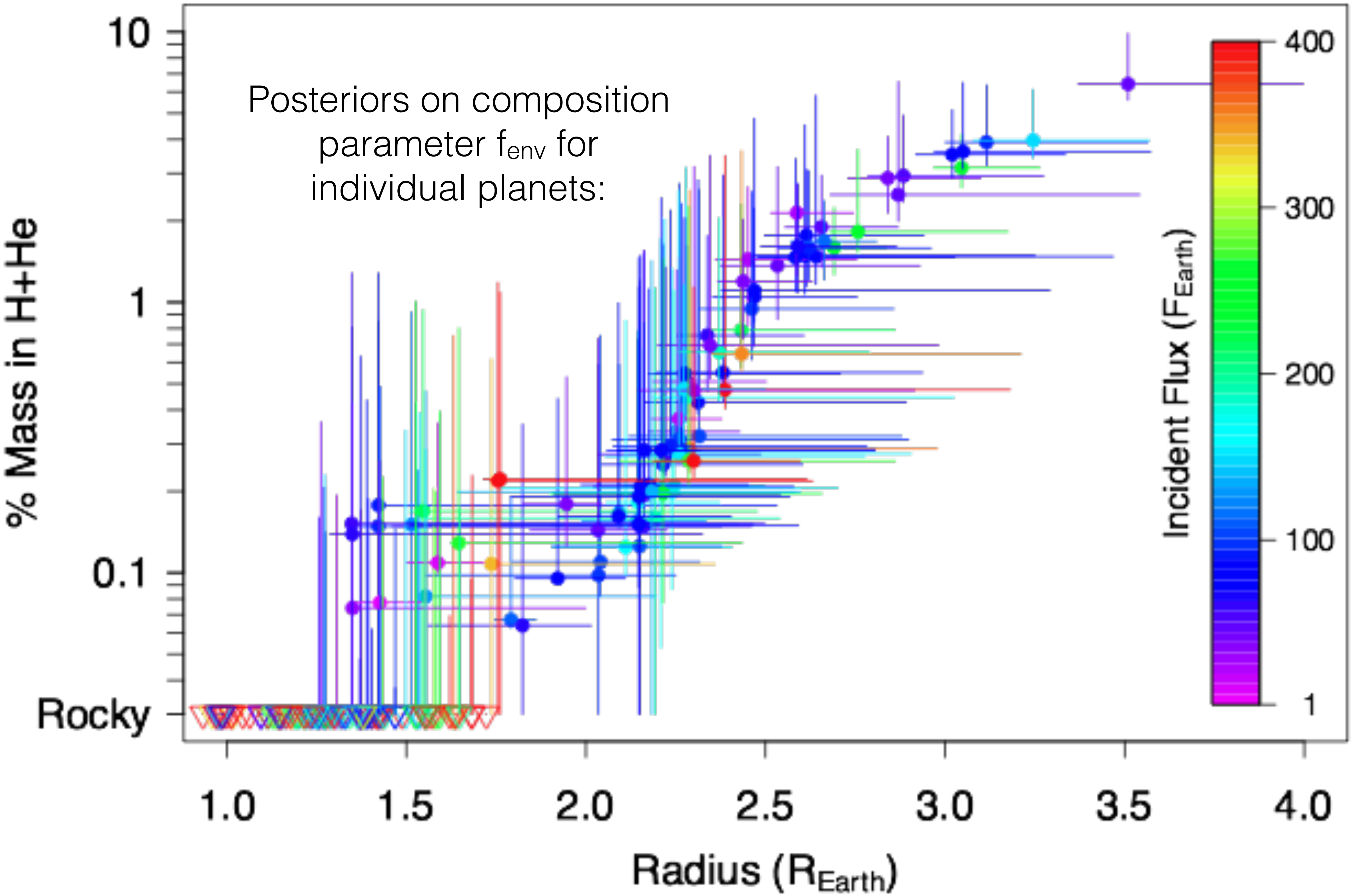


Marginal composition distribution:

Width of distribution had not been previously characterized.

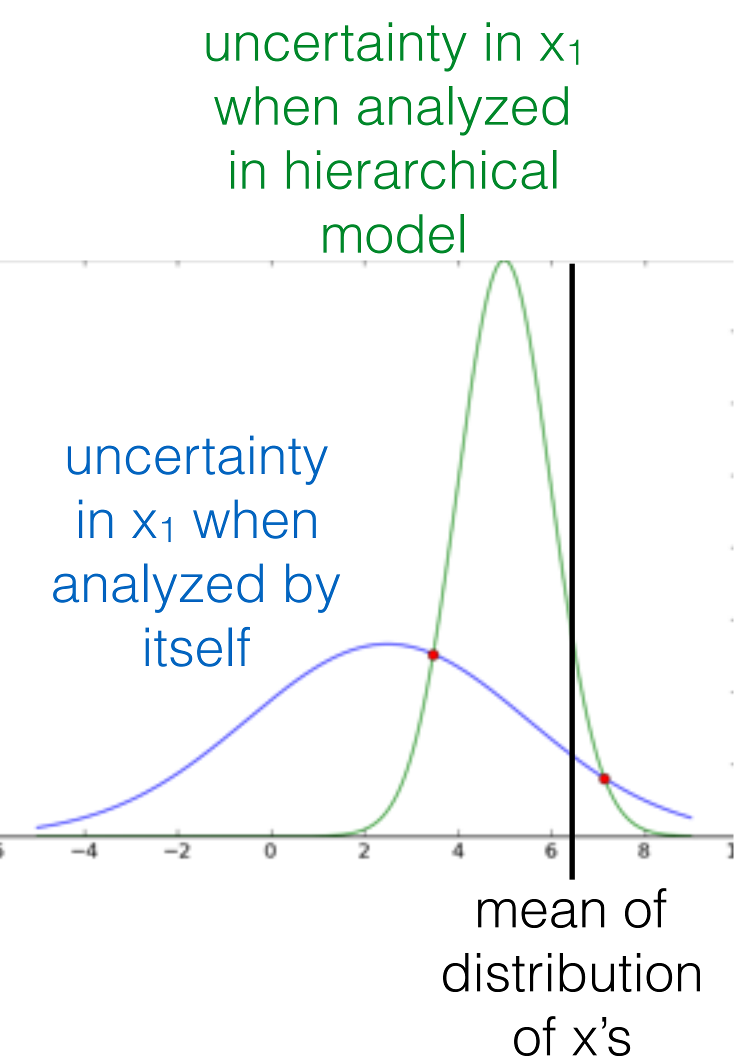
HBM in Action: Results

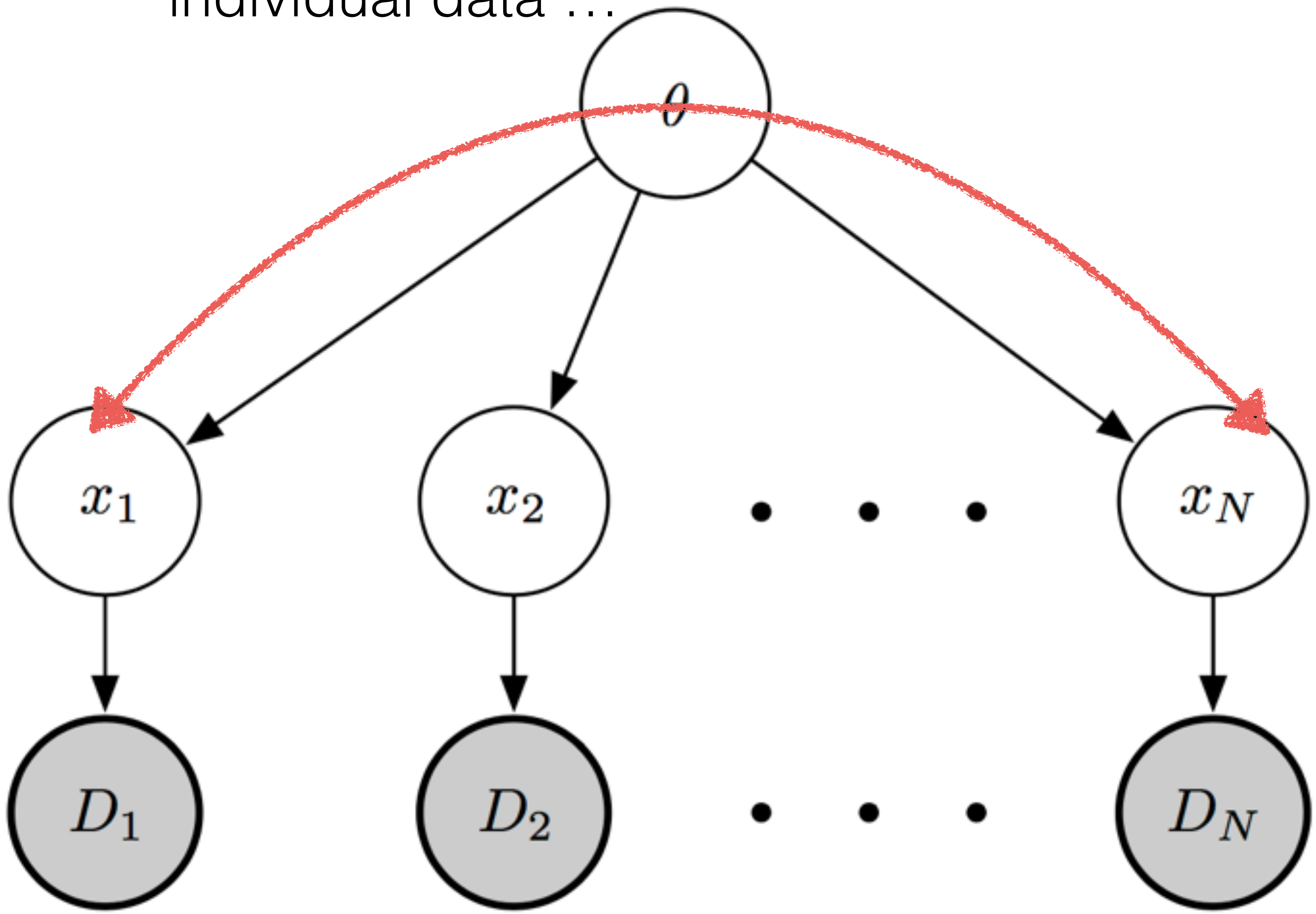
Exoplanet compositions: Wolfgang & Lopez, 2015



A Note About Shrinkage

Hierarchical models “pool” and share the information in the individual data …

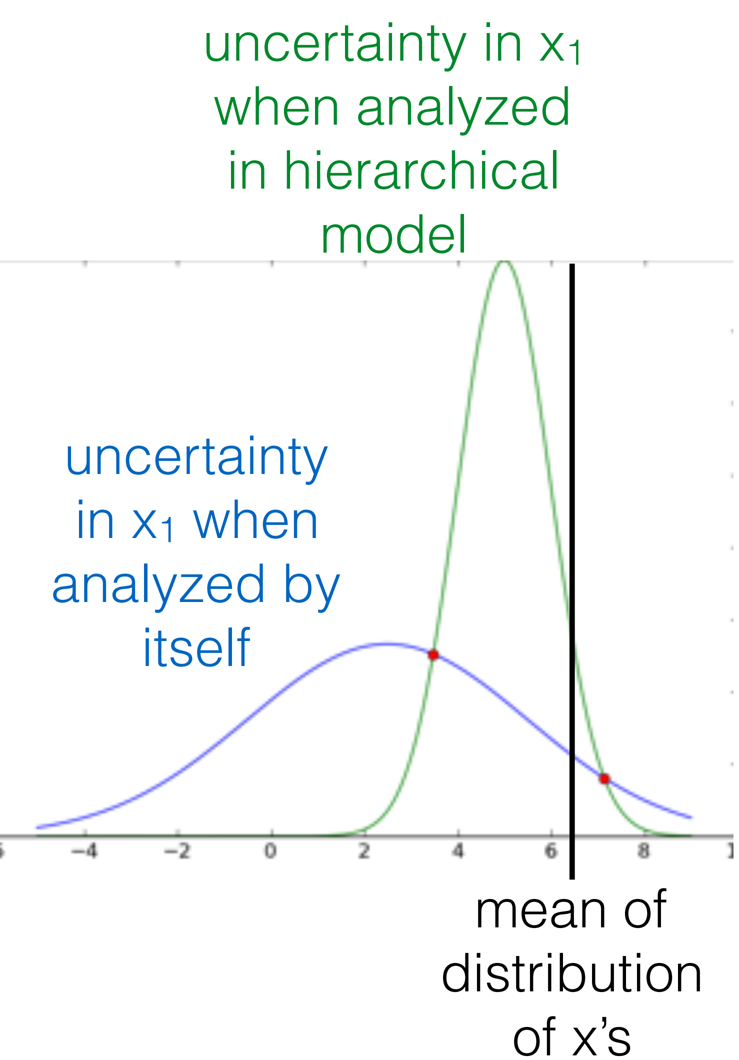


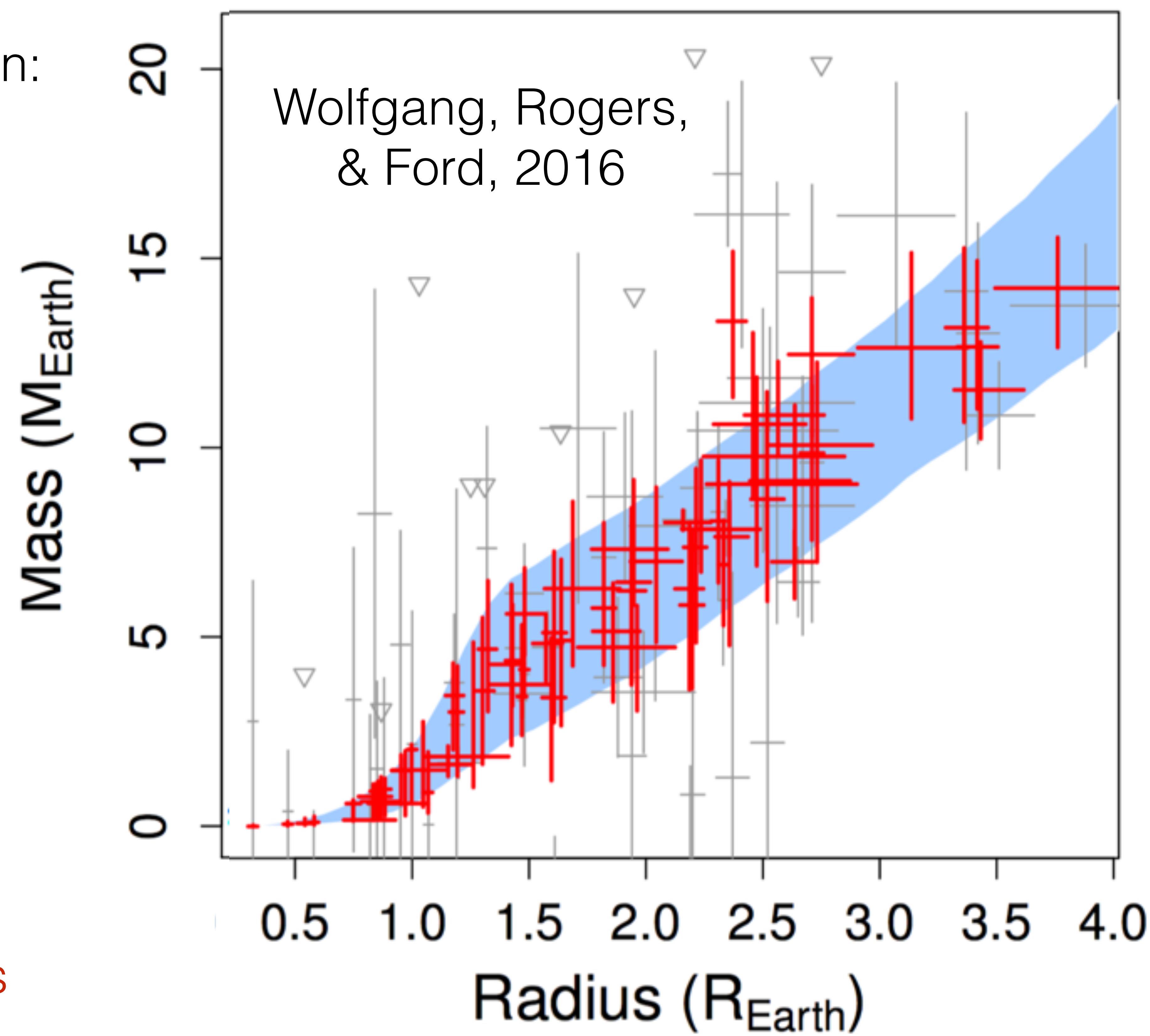


… which shrinks individual estimates together and lowers overall RMS error. (A key feature of any multi-level modeling!)

A Note About Shrinkage

Shrinkage in action:





Gray = data

Red = posteriors

Practical Considerations

1) Pay attention to the structure of your model!!
    - Did you capture the important dependencies and correlations?
    - Did you balance realism with a small number of population-level parameters?

2) Evaluating your model with the data (performing hierarchical MCMC):
    - JAGS (http://mcmc-jags.sourceforge.net; can use stand-alone binary or interface with R)
    - STAN (http://mc-stan.org/documentation/; interfaces with R, Python, Julia, MATLAB)
    - Or write your own hierarchical MCMC code

3) Spend some time testing the robustness of your model: if you generate hypothetical datasets using your HBM and then run the MCMC on those datasets, how close do the inferences lie to the "truth"?

4) Developing and validating these models takes time and is a significant result in and of itself: if you use or build on one in the literature, reference it!

In Sum, Why HBM?

After all, people have corrected for survey biases before without all of this work . . .

- Readily quantify uncertainty in model parameters, and in the same way we tend to ask our science questions.
- Obtain simultaneous posteriors on individual and population parameters: self-consistent constraints on the physics
- Naturally deals with large measurement uncertainties and upper limits (censoring)
- Similarly, can account for selection effects *within* the model, simultaneously with the inference
- Enables direct, probabilistic relationships between theory and observations
- Framework for model comparison

Further Reading

Introductory/General:

DeGroot & Schervish, *Probability and Statistics*
(Solid fundamentals)

Gelman, Carlin, Stern, & Rubin, *Bayesian Data Analysis*
(In-depth; advanced topics)

Loredo 2013; arXiv:1208.3036
(Few-page intro/overview of multi-level modeling in astronomy)

B.C. Kelly 2007
(HBM for linear regression, also applied to quasars)

Some exoplanet applications:

Hogg et al. 2010
(HBM with importance sampling for exoplanet eccentricities)

Morton & Winn, 2014
(HBM for stellar spin-planet orbit obliquity)

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

Rogers 2015
(HBM for rocky-gaseous transition)

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

Wolfgang, Rogers, & Ford 2016
(full HBM for mass-radius relationship)

Want to learn more?

2016-2017: SAMSI Program on Statistical, Mathematical and Computational Methods for Astronomy

Working Group IV: Population Modeling & Signal Separation for Exoplanets & Gravitational Waves

Purpose: To bring astronomers and statisticians together to work on mutually interesting problems.

August 22-26, 2016: Opening Workshop; weekly telecons start

October 17-28, 2016: Workshop on Hierarchical Bayesian Modeling of Exoplanet Populations

Spring 2017: Extended work on hierarchical modeling

May 8-10, 2017: Transition/closing Workshop

Come talk to me if you'd like to participate!