Bayesian Priors for Transits and RVs

David Kipping
Sagan Workshop 2016

but first, a brief advertisement

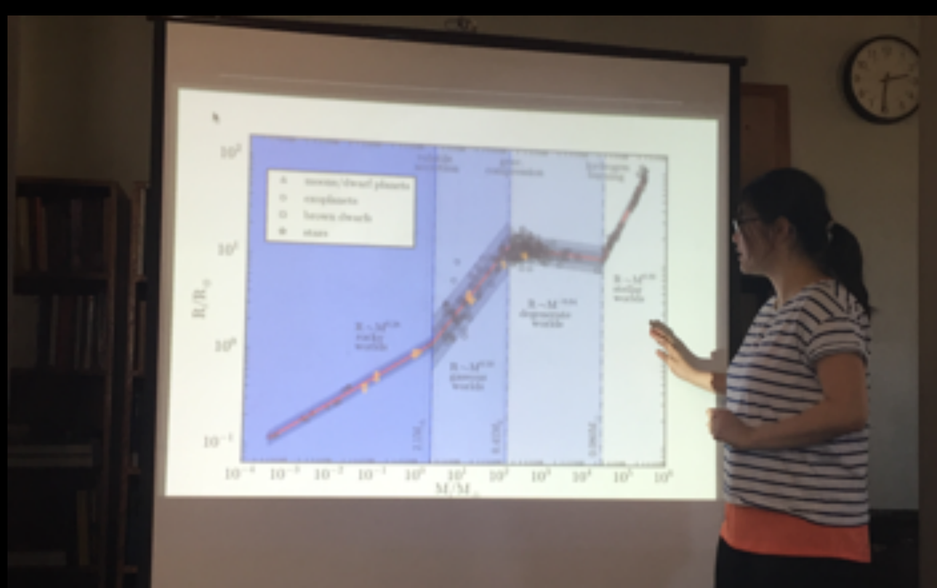
Cool Worlds Lab

Columbia University in the City of New York

http://coolworlds.astro.columbia.edu

Jingjing Chen - Wed morning POP

Probabilistic Forecasting of the Masses & Radii of Other Worlds



come talk to me to learn more about our group!

(some) things we are interested in...

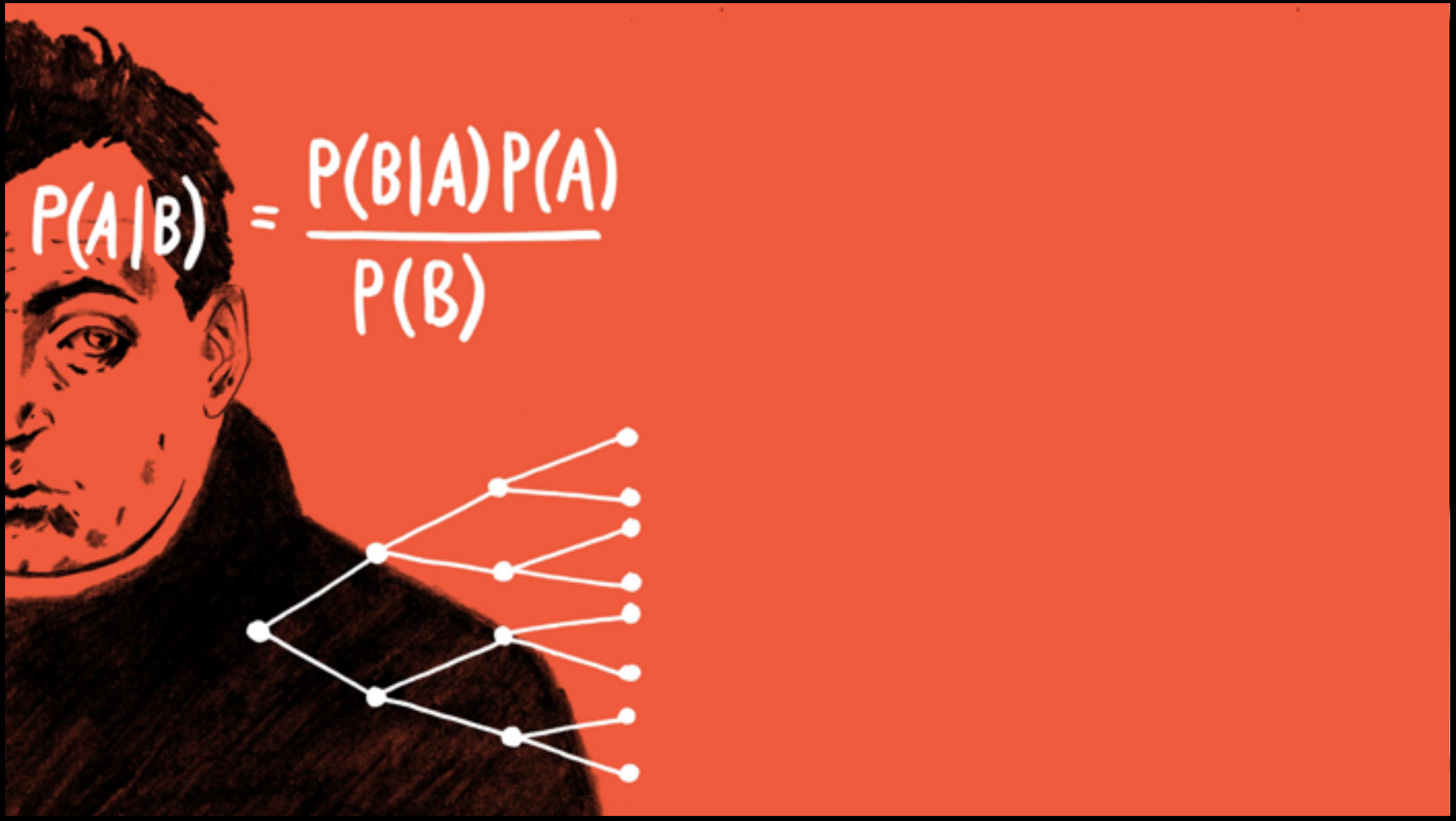
population modeling, neural networks, exomoons, exorings, long-period planets, single-transits, compact objects in photometry, SETI, rocky planet compositions, LSST, TESS, GAIA prior belief $\xrightarrow{\textit{data}}$ posterior belief

likelihood, $\mathscr{L}$

prior, π

posterior, $\mathscr{P}$

$$P(\boldsymbol{\Theta}|\mathcal{D},\mathscr{M}) \propto P(\mathcal{D}|\boldsymbol{\Theta},\mathscr{M})P(\boldsymbol{\Theta}|\mathscr{M})$$

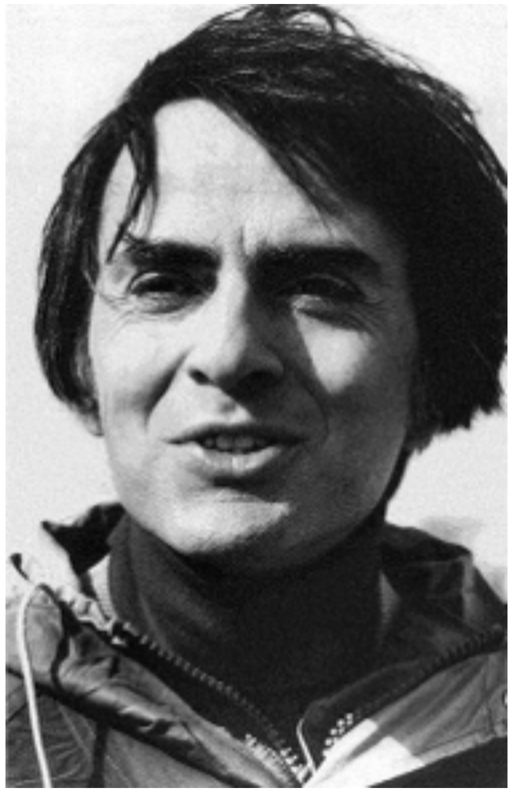


www.youtube.com/c/CoolWorldsLab

Think of some cases where...

a detection claim was made about something for which the world/scientific community has a very strong prior against

did it turn out right or wrong?

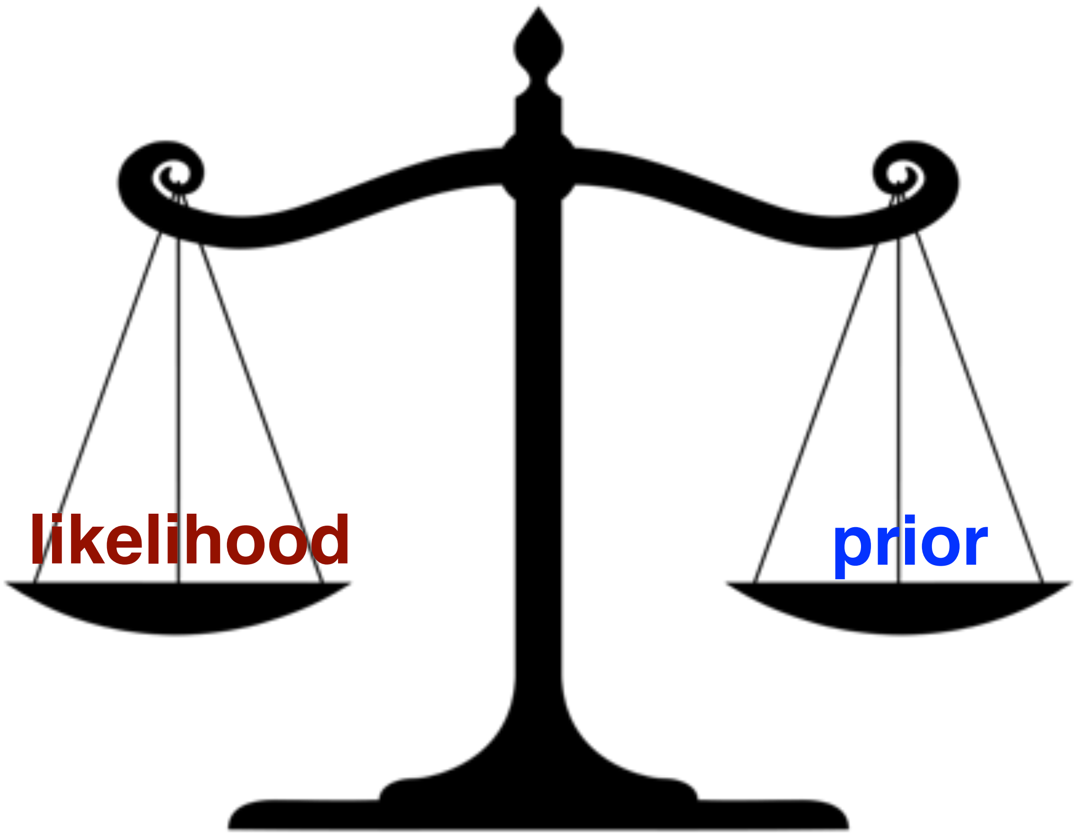


"extraordinary claims require extraordinary evidence"
(to overwhelm our prior belief)

Carl Sagan

$$P(\boldsymbol{\Theta}|\mathcal{D},\mathscr{M}) \propto P(\mathcal{D}|\boldsymbol{\Theta},\mathscr{M})P(\boldsymbol{\Theta}|\mathscr{M})$$

the end result, the posterior, is a balancing act
between the likelihood and the prior

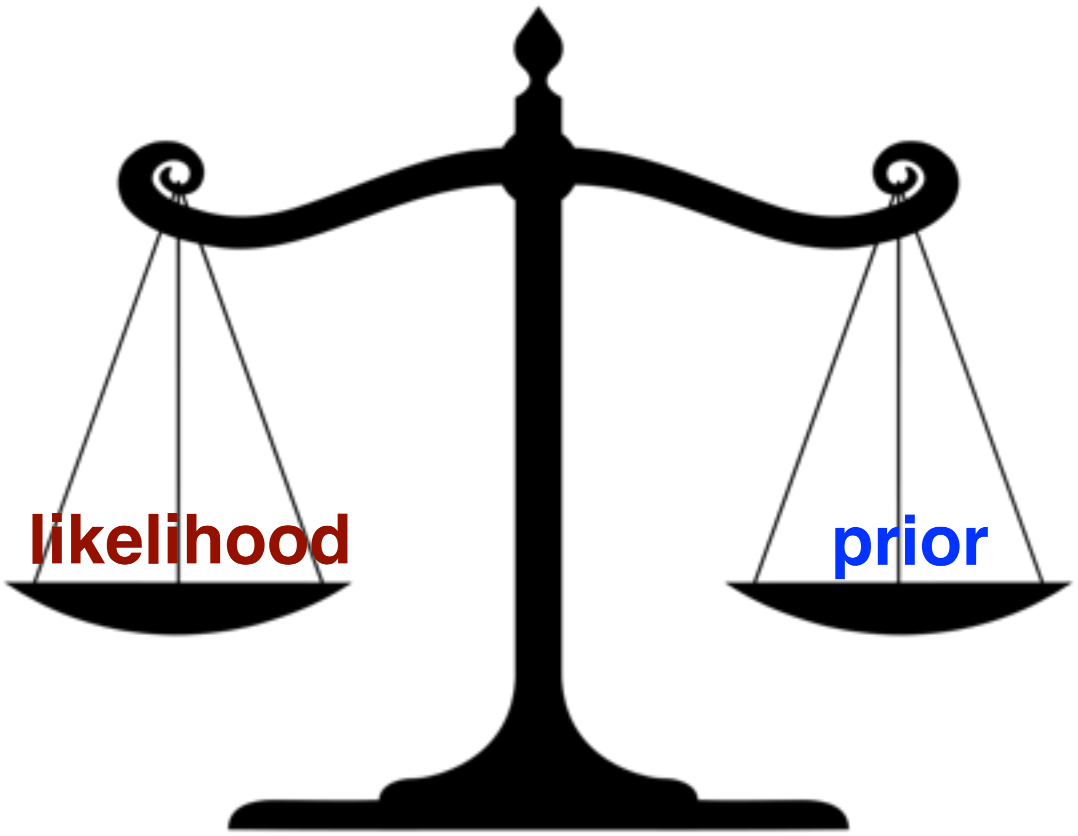


- Think about the outcome being affected by **both** the likelihood and the prior
- Posteriors from low signal-to-noise data (low likelihood) are strongly affected by the priors
- Posteriors from high signal-to-noise data (high likelihood) are weakly affected by the priors

some people despise Bayesian statistics because one needs to define a prior

this is not a weakness! it's a strength!

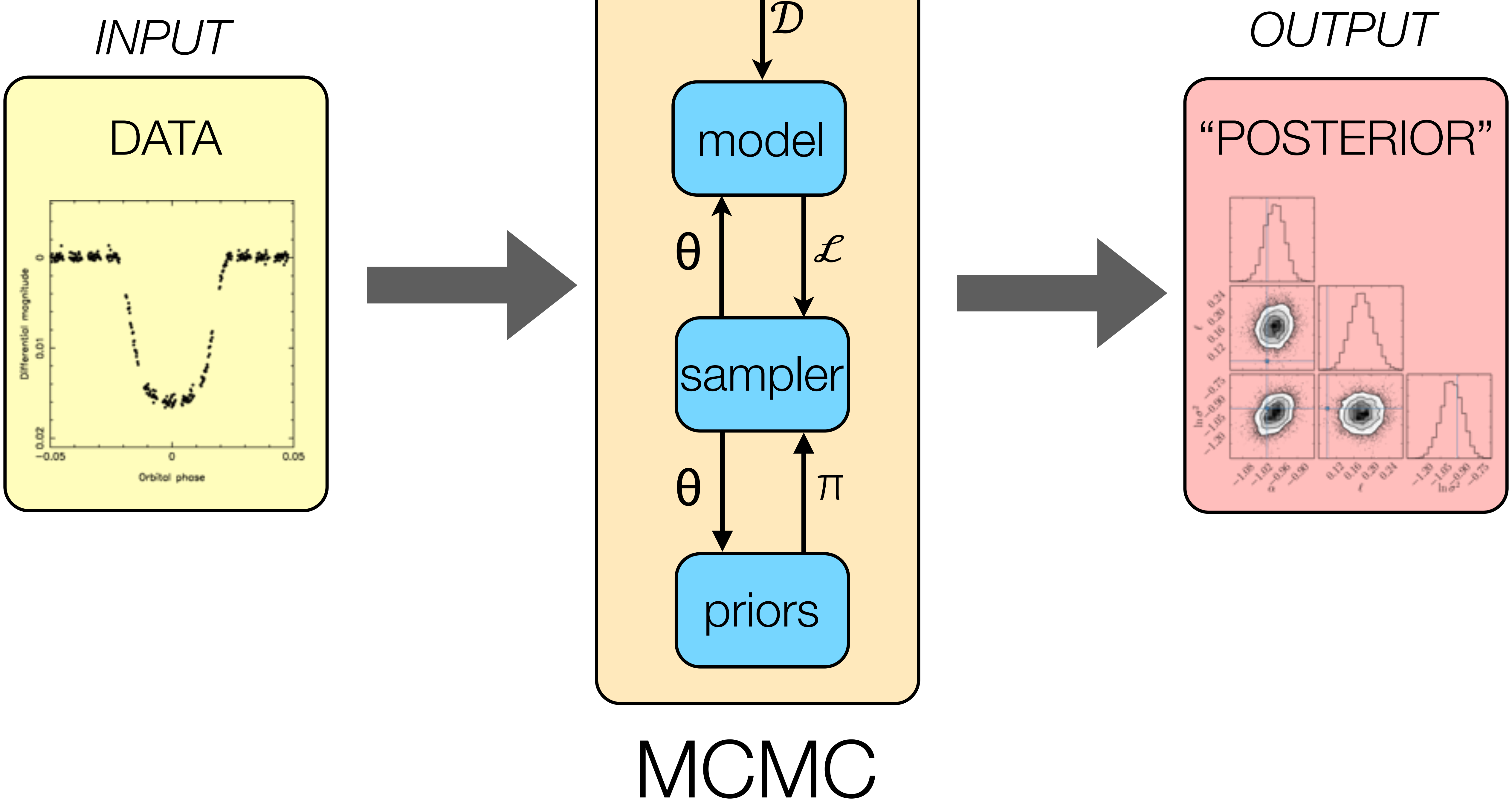
- If your result changes when you change between reasonable priors, then this is telling you that your data are crappy, which is useful information!!
- Your previous posterior can become the prior for the next experiment: "Bayesian learning"



the sampler "guesses" different θ vectors, calculates the posterior probability of that guess, and then makes small jumps

PROCESS

actually the point of the sampler is to make intelligent guesses with high posterior probabilities



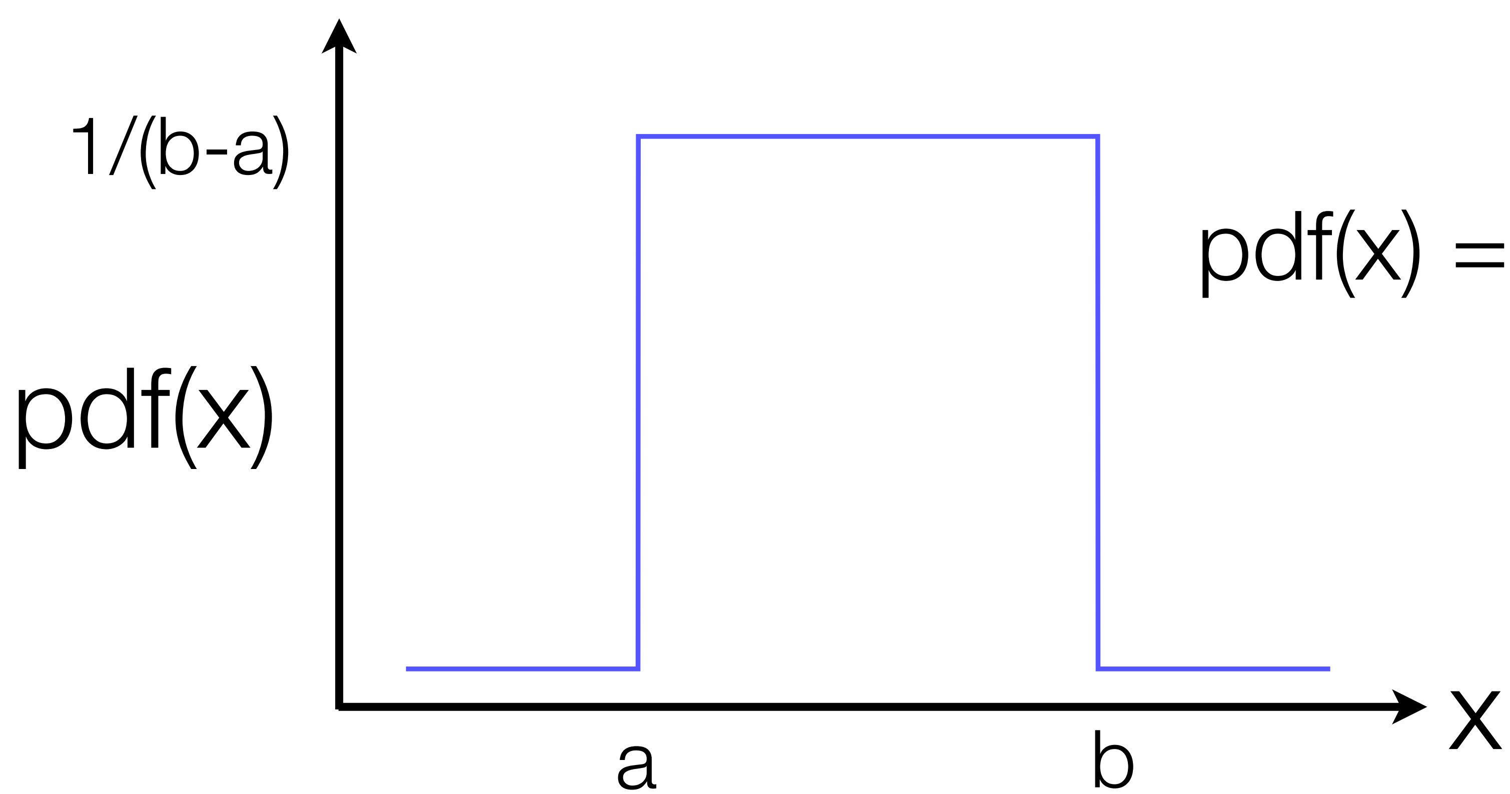
how to... choose a prior	how to... implement a prior
uninformative	log likelihood penalization
informative	
conjugate useful for analytic work, but not really used in practical exoplanet work	inverse transform sampling

how to... choose a prior	how to... implement a prior
uninformative	log likelihood penalization
informative	
conjugate useful for analytic work, but not really used in practical exoplanet work	inverse transform sampling

uninformative priors: *uniform*

bit of misnomer, really a prior which is not subjectively elicited

simplest rule is via the principle of indifference, which assigns equal probability to all possibilities = a uniform prior



$$\mathrm{pdf}(x) = \begin{cases} (b-a)^{-1} & \text{if } a<x<b \\ 0 & \text{otherwise} \end{cases}$$

uninformative priors: *uniform*

often exoplaneteers technically use an improper prior for this, since a and b are not formally defined in their paper or even code

here, the user is treating $a=-\infty$ and/or $b=+\infty$, but that leads to pdf(x) = 0 everywhere => you should be rejecting all MCMC trials!

in practice, this is generally OK though, since

π = a constant for a uniform prior

$\pi_{i+1} - \pi_i = 0$ for a uniform prior

thus the jump acceptance probability is insensitive to a or b, and thus a and b can be just very large numbers general case

METROPOLIS RULE

if $\mathscr{P}_{\text{trial}} > \mathscr{P}_i$,

accept the jump, so $\theta_{i+1} = \theta_{\text{trial}}$

if $\mathscr{P}_{\text{trial}} < \mathscr{P}_i$,

accept the jump with probability $\mathscr{P}_{\text{trial}}/\mathscr{P}_i$

someone ignoring priors

METROPOLIS RULE

if $\mathscr{L}_{\text{trial}} > \mathscr{L}_i$,

accept the jump, so $\theta_{i+1} = \theta_{\text{trial}}$

if $\mathscr{L}_{\text{trial}} < \mathscr{L}_i$,

accept the jump with probability $\mathscr{L}_{\text{trial}}/\mathscr{L}_i$

someone ignoring priors and assuming normal errors

METROPOLIS RULE

if $\chi^2_{\text{trial}} < \chi^2_i$,

accept the jump, so $\theta_{i+1} = \theta_{\text{trial}}$

if $\chi^2_{\text{trial}} > \chi^2_i$,

accept the jump with probability $\exp(-\Delta\chi^2/2)$

you are using unbounded uniform priors implicitly

uninformative priors: *uniform*

often, a or b or both can be set to some physical lower/upper bound

eccentricity, $e>0$ by definition and $e<1$ if the orbit is periodic

ratio-of-radii, $p>0$ by definition and $p<1$ if the object is smaller than the star

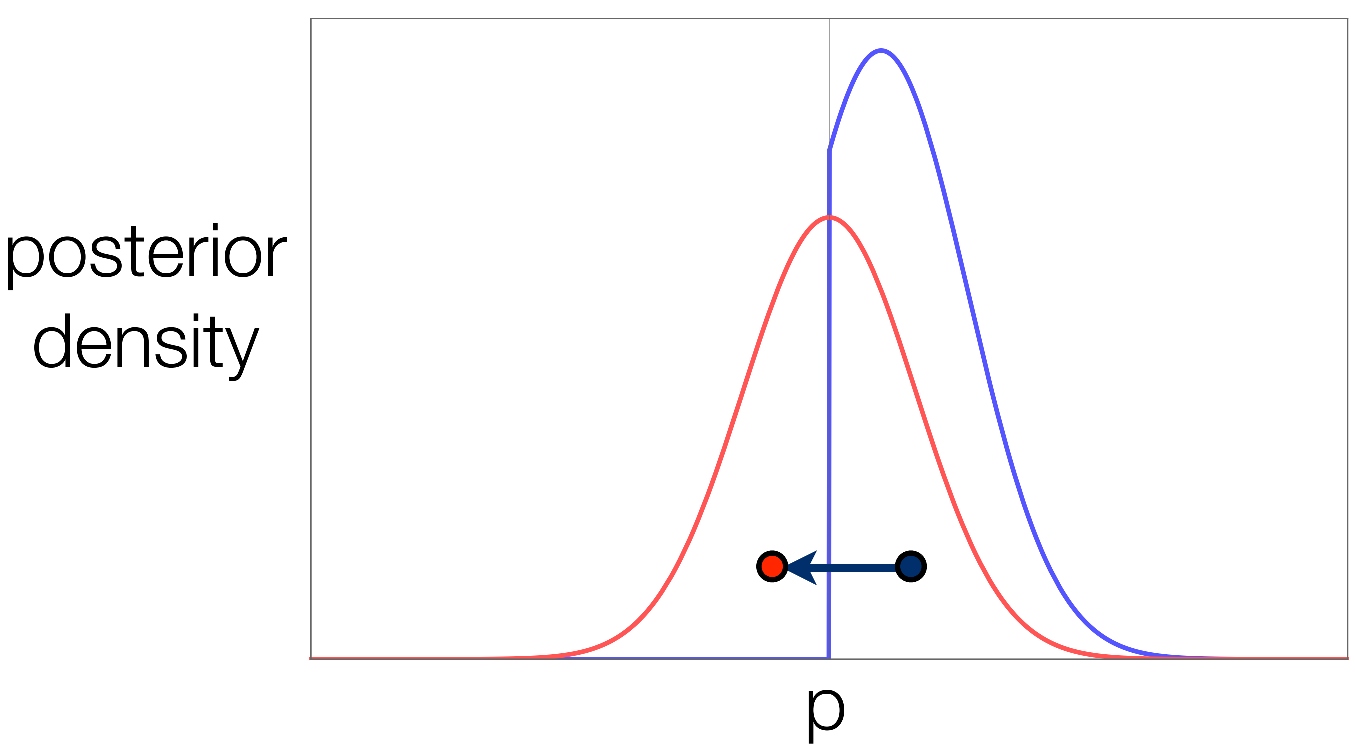
a/R^*, $a_R>(1+p)$ or else the planet is in contact with the star

uninformative priors: *uniform*

however, sometimes we deliberately explore unphysical solutions...

e.g. ratio-of-radii, -1<p<+1 and treat negative radii as being inverted transits

for *amplitude-like* parameters (e.g. p, e, K) near zero, helps avoid posterior bias due to boundary conditions...



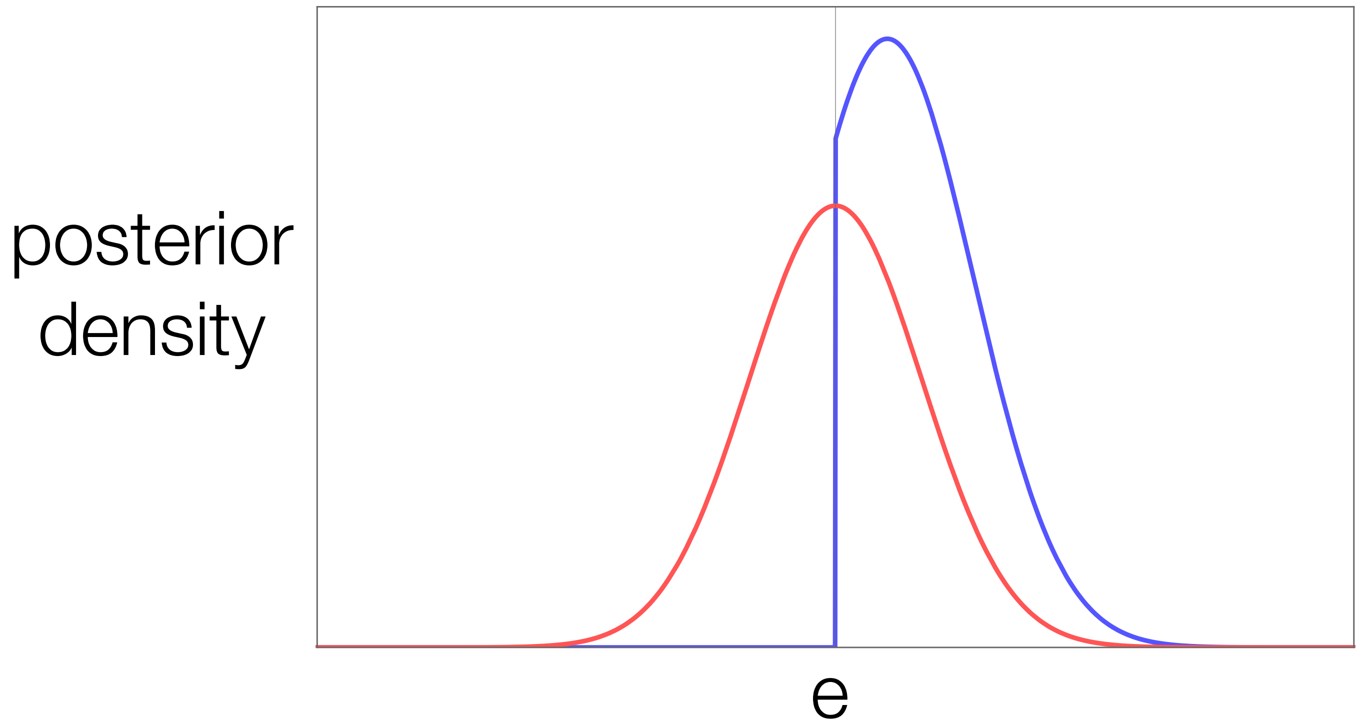
if we set a boundary condition at 0, MCMC posteriors get positively-skewed due to rejection bias of walkers

uninformative priors: *uniform*

this is particularly well-known for eccentricity, where even high SNR data with a truth of e=0 return posteriors positively biased if one fits for e directly

see Lucy & Sweeney (1971), Zakamska, Pan & Ford (2011), Lucy (2012)

for eccentricity, a good trick is to fit for $-1<e^{1/2}\sin\omega<+1$ and $-1<e^{1/2}\cos\omega<+1$

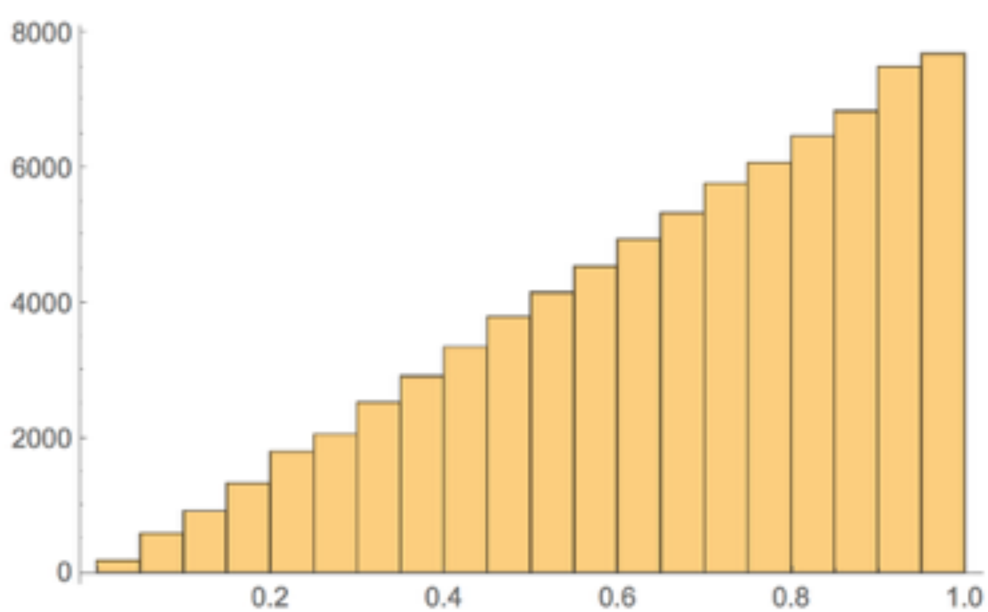


uninformative priors: *uniform*

if you have to calculate the parameter of interest from your fitted terms, check out what the prior on the parameters of interest is via Monte Carlo

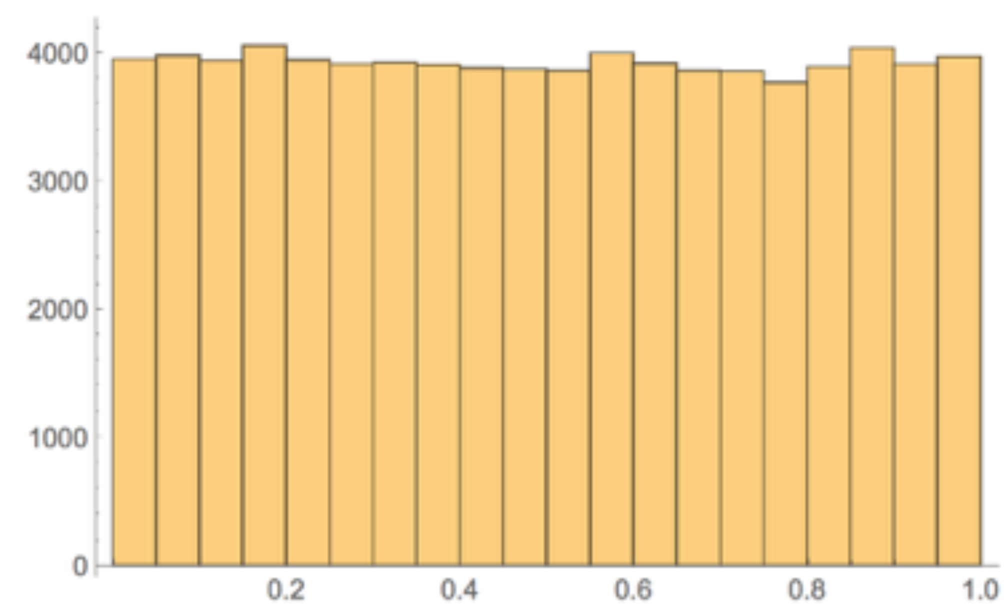
esinω & ecosω

```
n = 10^5;
h = RandomVariate[UniformDistribution[{-1, 1}], n];
k = RandomVariate[UniformDistribution[{-1, 1}], n];
e = Table[√(h[[i]]^2 + k[[i]]^2), {i, 1, n}];
Histogram[Select[e, # < 1 &]]
```



e$^{1/2}$sinω & e$^{1/2}$cosω

```
n = 10^5;
h = RandomVariate[UniformDistribution[{-1, 1}], n];
k = RandomVariate[UniformDistribution[{-1, 1}], n];
e = Table[h[[i]]^2 + k[[i]]^2, {i, 1, n}];
Histogram[Select[e, # < 1 &]]
```



on a related note, but beyond the scope of this priors lecture, the proposal function can be selected to minimize inter-parameter correlations. See Carter et al. (2008) for transits and Ford (2006) for RVs. Although, **emcee** would do this for free anyway

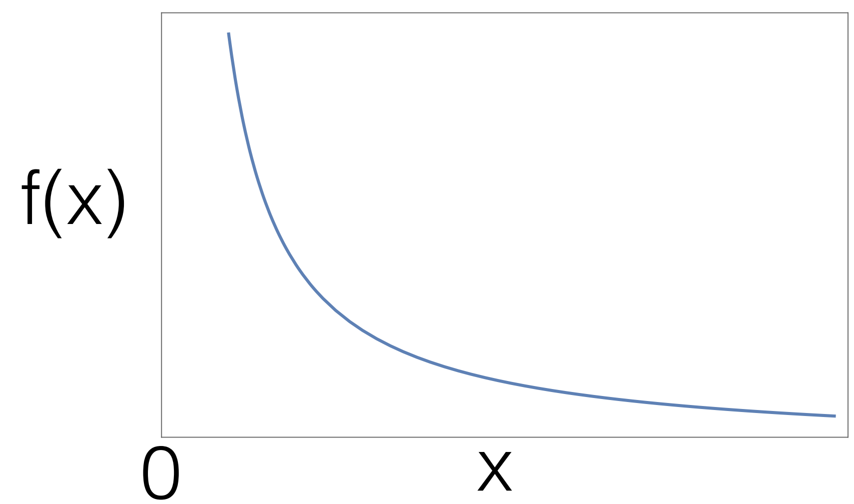
uninformative priors: *log-uniform*

for parameters which are scale-like and span orders-of-magnitude, a log-uniform distribution is usually considered "more uninformative"

e.g. K, P, a_R, ρ_*

not t_{mid} (can span a large range but is certainly a location-like parameter)

$$f(x) = \frac{1}{x} \frac{1}{\log(x_{max}/x_{min})}$$

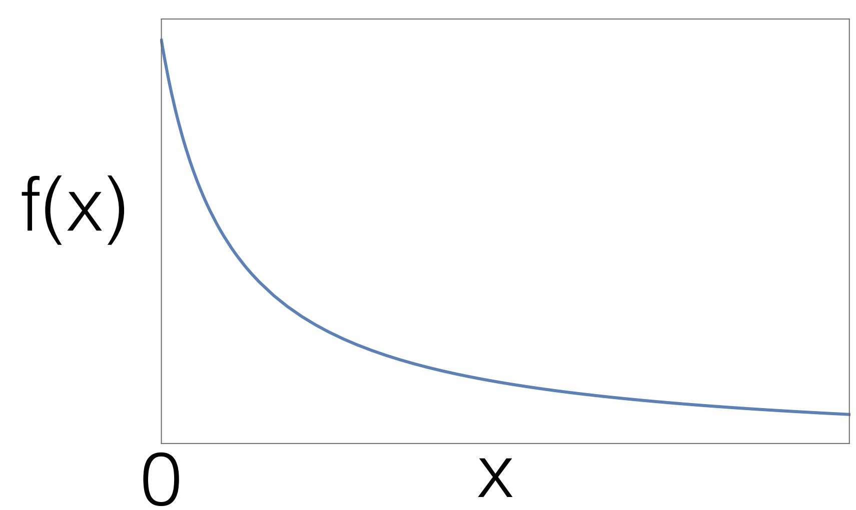


but be warned that $f(x) \rightarrow \infty$ as $x \rightarrow 0$

so not useful if you have a parameter which extend to 0

modified log-uniform can extend to 0, useful for K but not usually needed for P

$$f(x) = \frac{1}{x+x_0} \frac{1}{\log((x_0+x_{max})/x_0)}$$



see Ford & Gregory (2007)

how to...
choose a prior

uniform, think about boundary conditions

check priors on key parameters via Monte Carlo

uninformative

log-uniform for parameters scaling orders-of-magnitude

informative

conjugate

useful for analytic work, but not really used in practical exoplanet work

how to...
implement a prior

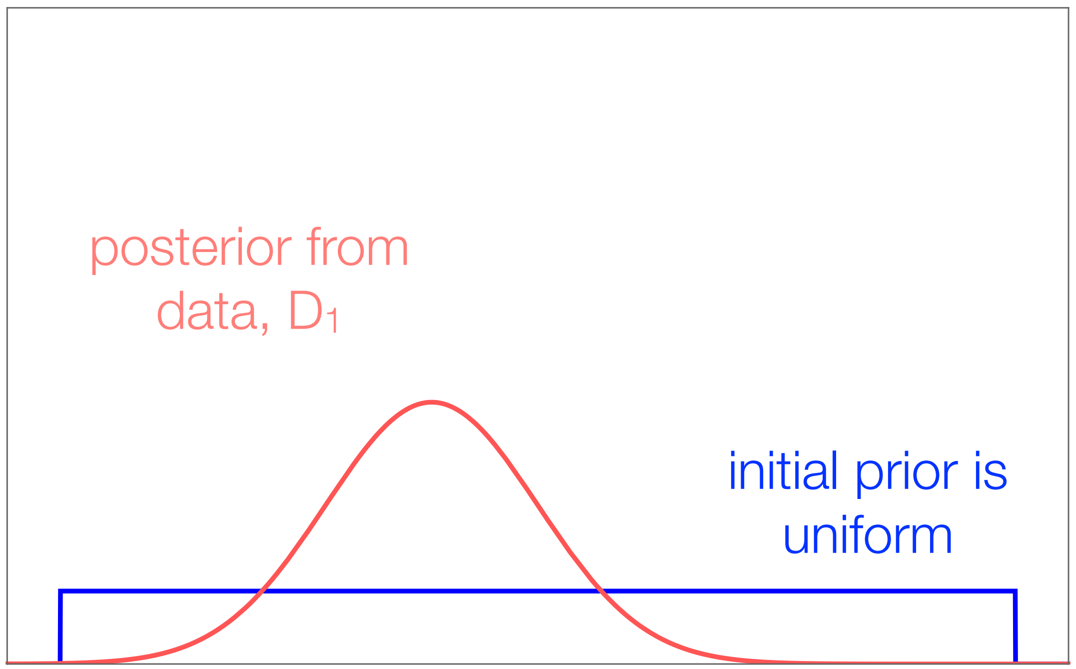
log likelihood penalization

inverse transform sampling

how to... choose a prior	how to... implement a prior
uniform, think about boundary conditions check priors on key parameters via Monte Carlo **uninformative** log-uniform for parameters scaling orders-of-magnitude	**log likelihood penalization**
informative	
conjugate useful for analytic work, but not really used in practical exoplanet work	**inverse transform sampling**

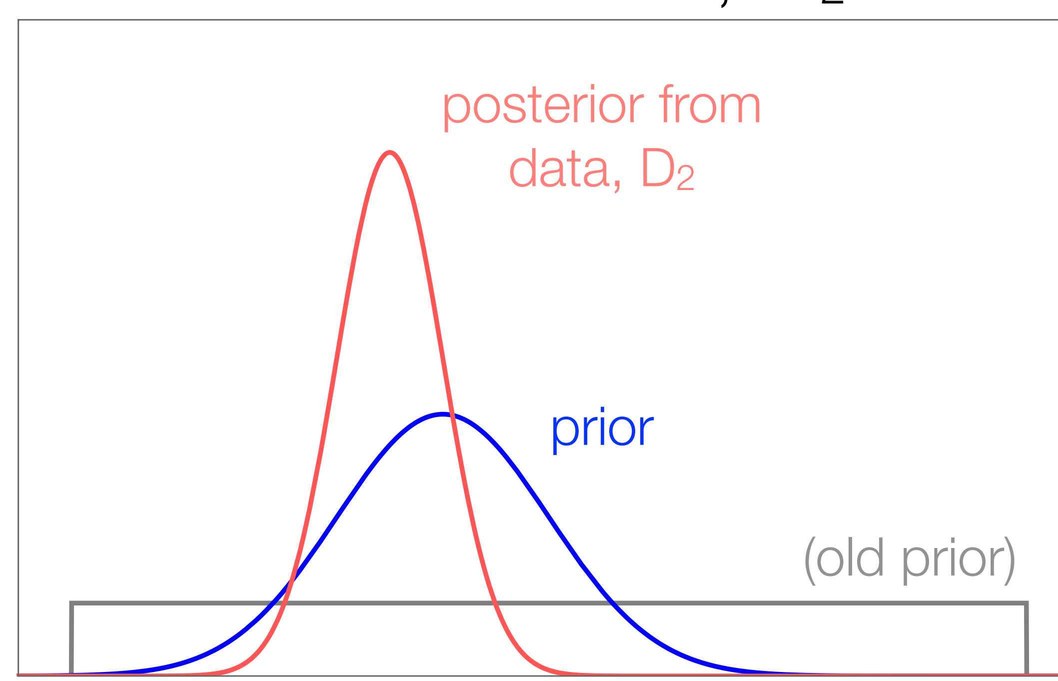
informative priors: *Bayesian learning*

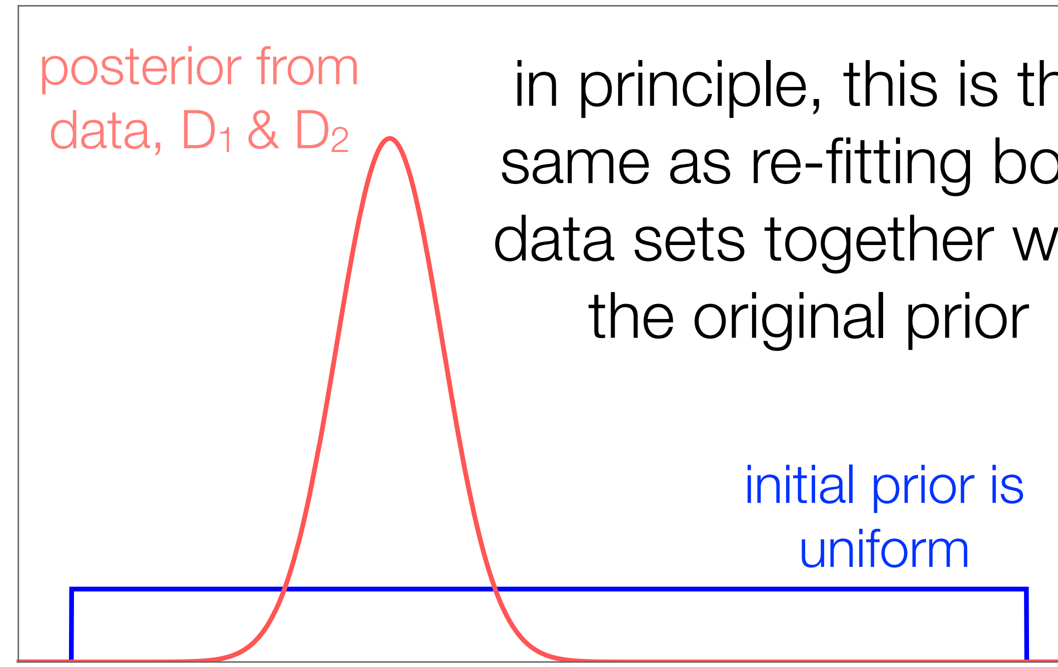
consider running MCMC on
some initial data, D_1



use posterior from here
as a prior here

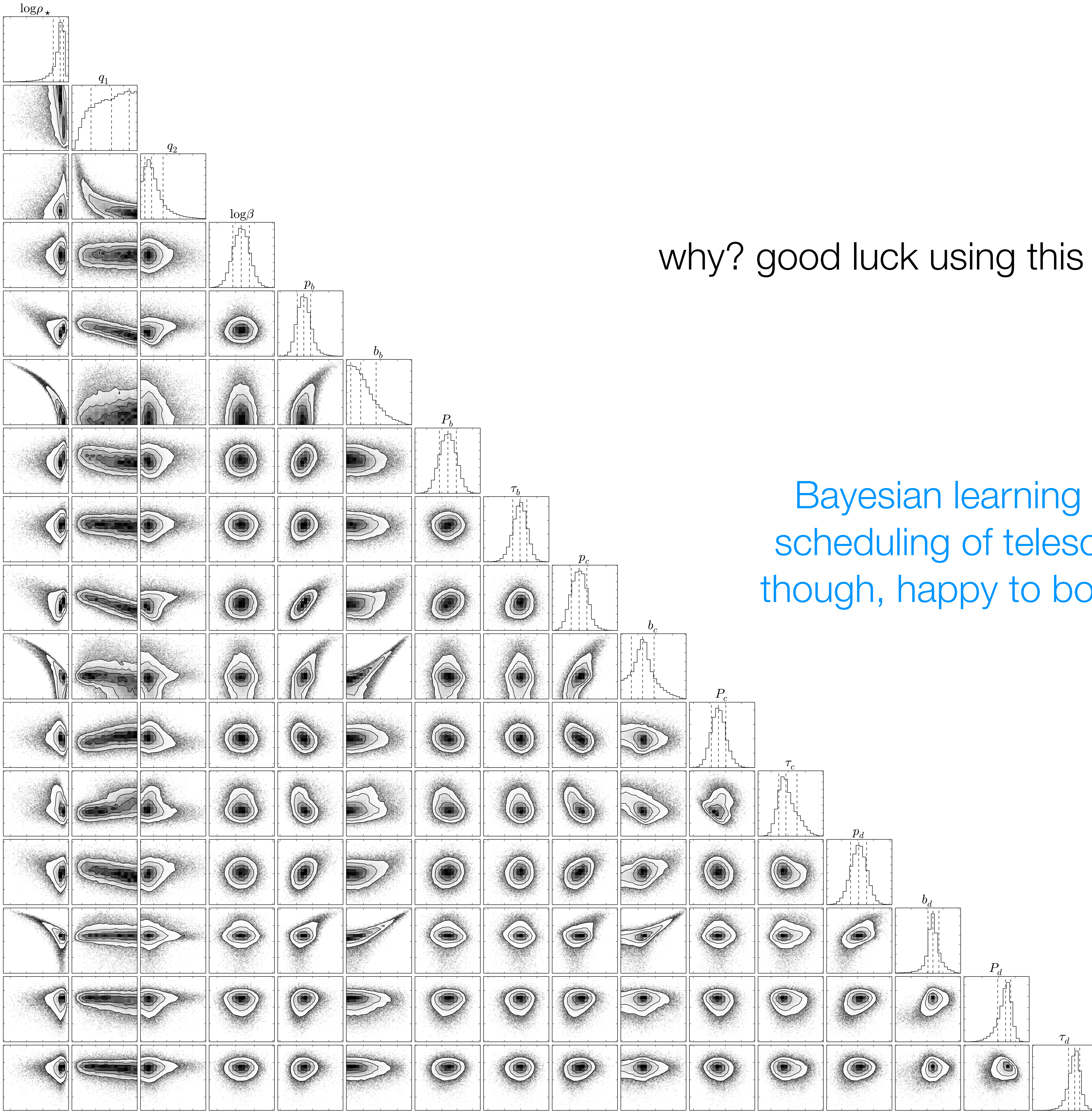
now run MCMC on
some new data, D_2





in principle, this is the
same as re-fitting both
data sets together with
the original prior

**exoplaneteers almost always
just re-fit everything rather than
trying to do Bayesian learning**



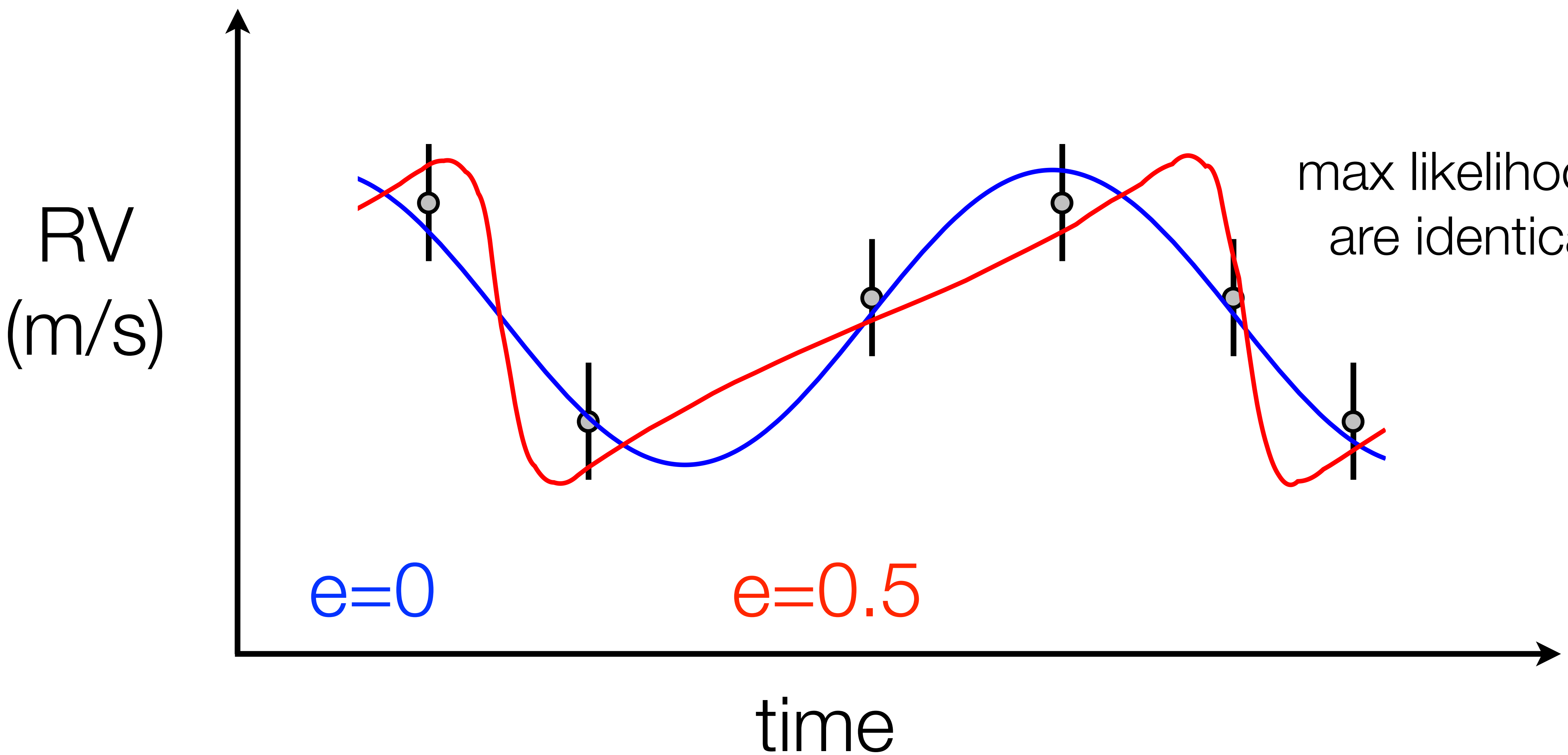
why? good luck using this as a prior...

Bayesian learning could be useful for scheduling of telescope time/resources though, happy to bounce ideas with you!

Kipping et al. (2016)

informative priors: *observed distributions*

someone shows you some RV data of a new planet...

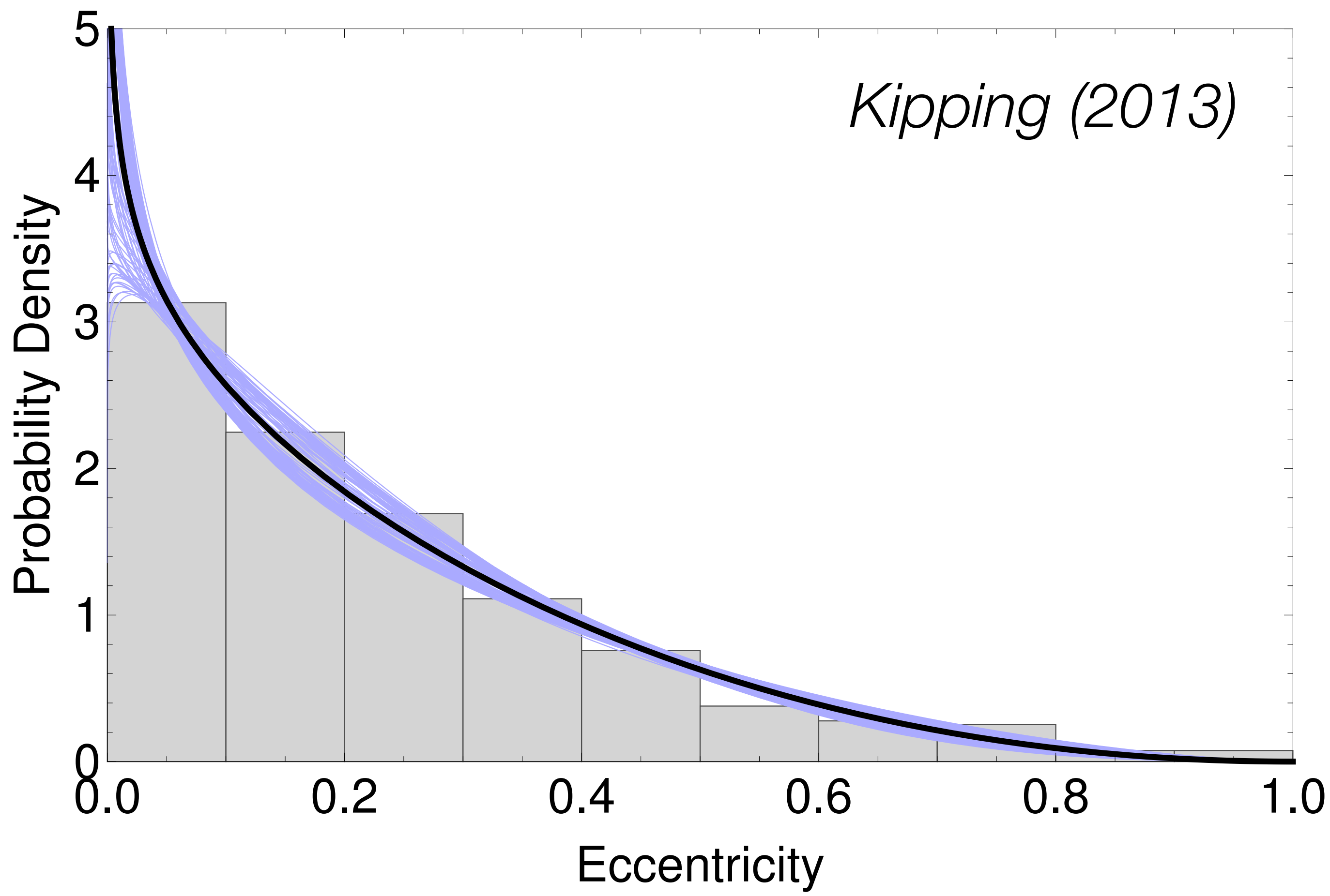


so which solution do you think is more likely to be the truth?

informative priors: *observed distributions*

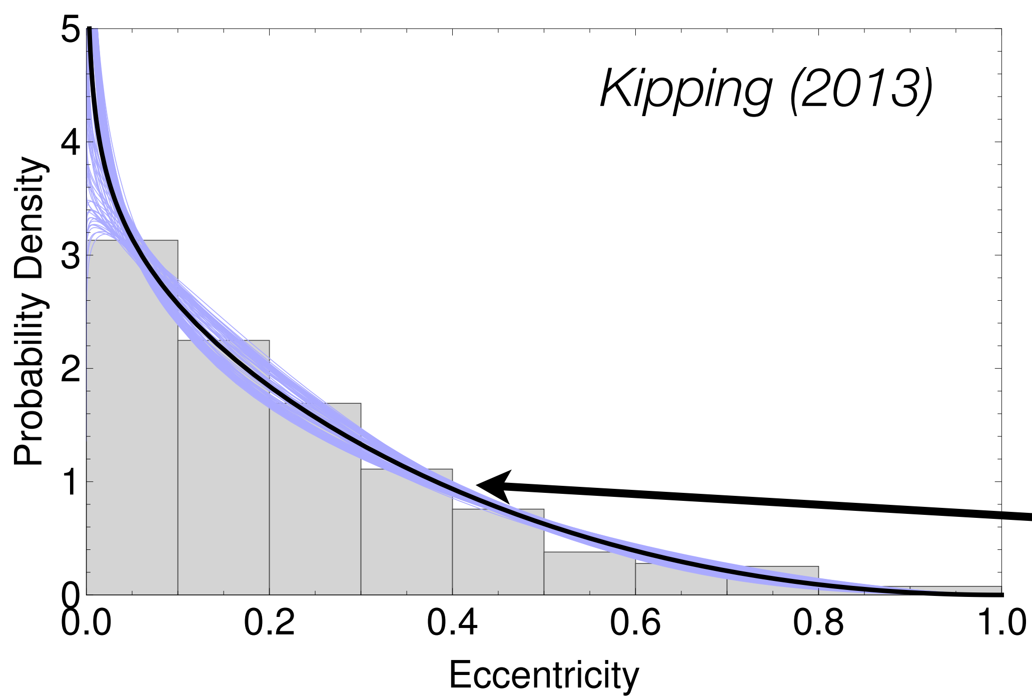
but you've seen hundreds of RV curves before, you know from experience that eccentric solutions are rarer than circular orbits

in fact the distribution of eccentricities of RV planets looks like this



we can encode the sage wisdom of the seasoned observer using an informative prior

informative priors: *observed distributions*



Beta distribution example

P(e) ~ Beta(0.867,3.03)

intrinsic distribution filtered though the detection biases of RVs

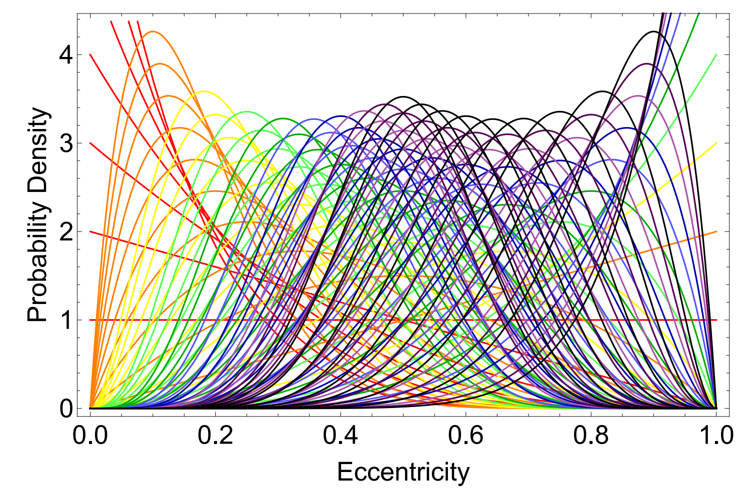
you can't just apply this to the sample of detected transiting planets though, transits have **different detection biases**

informative priors: *observational bias*

let's assume intrinsic is a Beta...

$$\mathrm{P}(e) = \frac{(1-e)^{\beta-1} e^{\alpha-1}}{\mathrm{B}[\alpha,\beta]} \qquad \mathrm{P}(\omega) = \frac{1}{2\pi}$$

$$\mathrm{P}(e,\omega) = \Big(\frac{1}{2\pi}\Big)\Big(\frac{(1-e)^{\beta-1} e^{\alpha-1}}{\mathrm{B}[\alpha,\beta]}\Big)$$

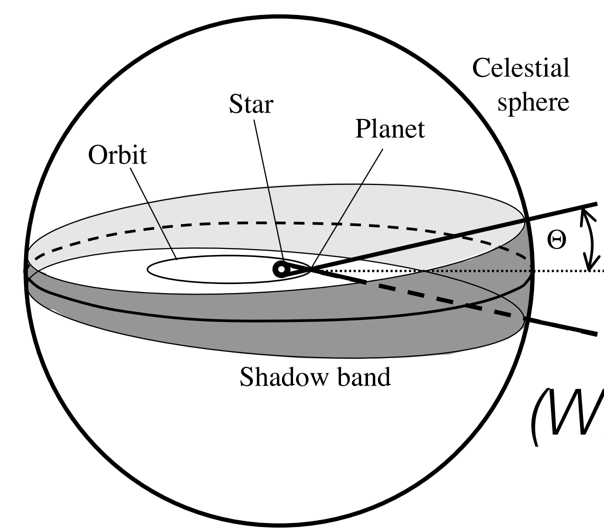


(Kipping 2013)

geometric transit probability is...

$$\mathrm{P}(\hat{b}|e,\omega) = \Big(\frac{1}{a_R}\Big)\Big(\frac{1+e\sin\omega}{1-e^2}\Big)$$

(Barnes 2007, Burke 2008)



(Winn 2011)

so eccentricity distribution conditioned on planet transiting is...

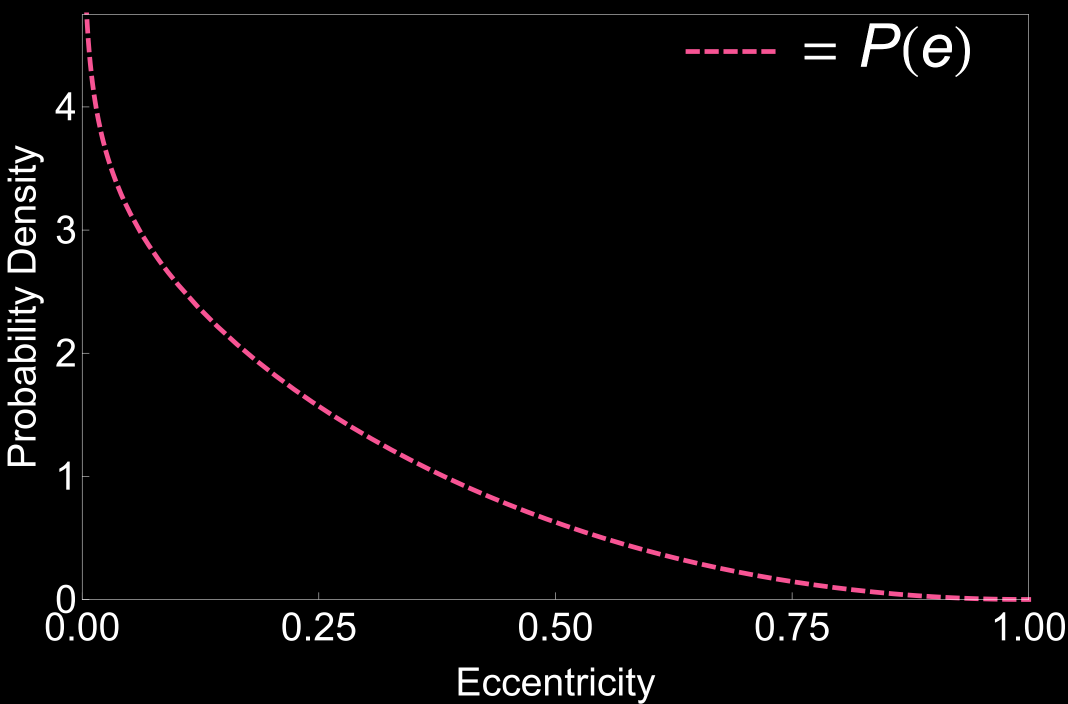
$$\mathrm{P}(e,\omega|\hat{b}) = \frac{\mathrm{P}(\hat{b}|e,\omega)\mathrm{P}(e,\omega)}{\int_{e=0}^{1}\int_{\omega=0}^{2\pi}\mathrm{P}(\hat{b}|e,\omega)\mathrm{P}(e,\omega)\ \mathrm{d}e\ \mathrm{d}\omega}$$

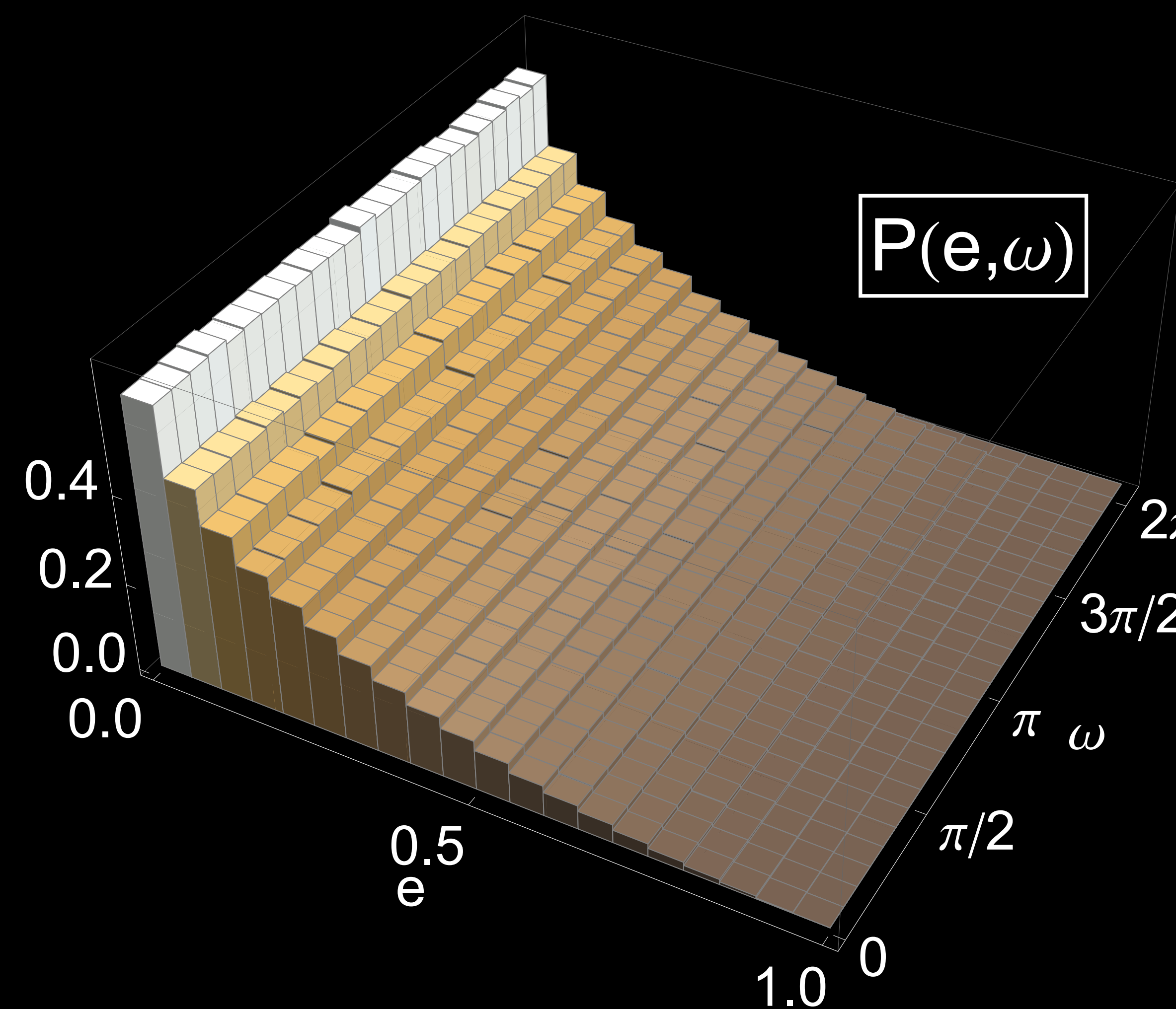
(Kipping 2014)

$$\mathrm{P}(e,\omega|\hat{b}) = \Big(\frac{\beta-1}{2\pi\tilde{\gamma}_1\Gamma[\alpha+\beta]}\Big)\Big(\frac{1+e\sin\omega}{1-e^2}\Big)\Big(\frac{(1-e)^{\beta-1} e^{\alpha-1}}{\mathrm{B}[\alpha,\beta]}\Big)$$

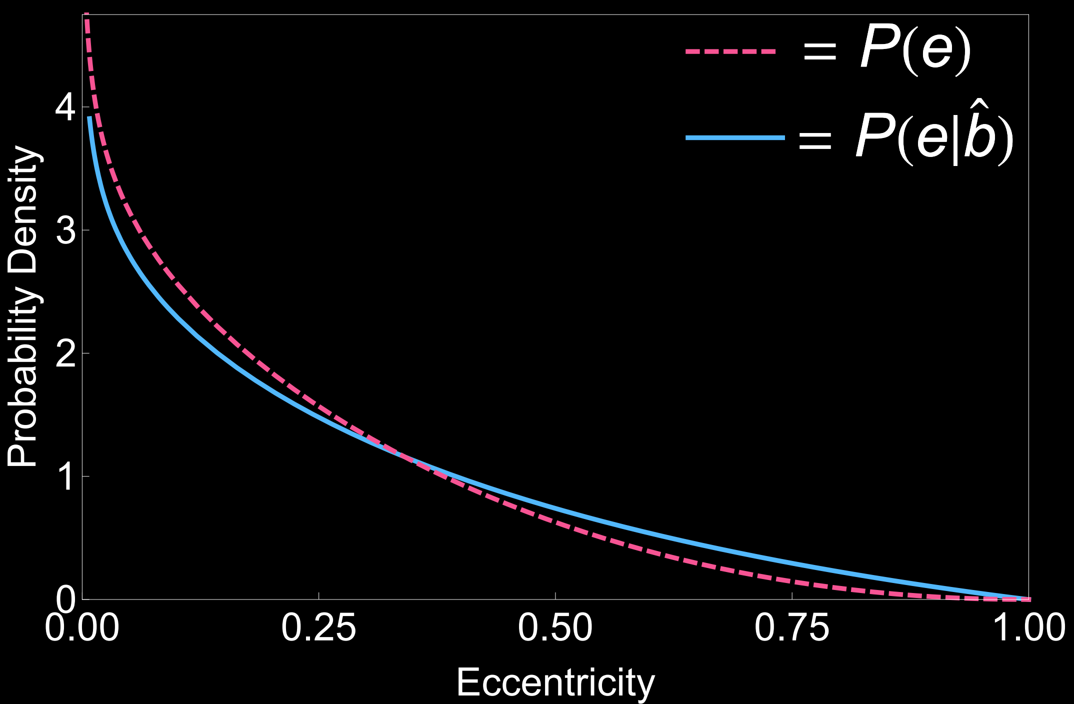
so... what did that do?

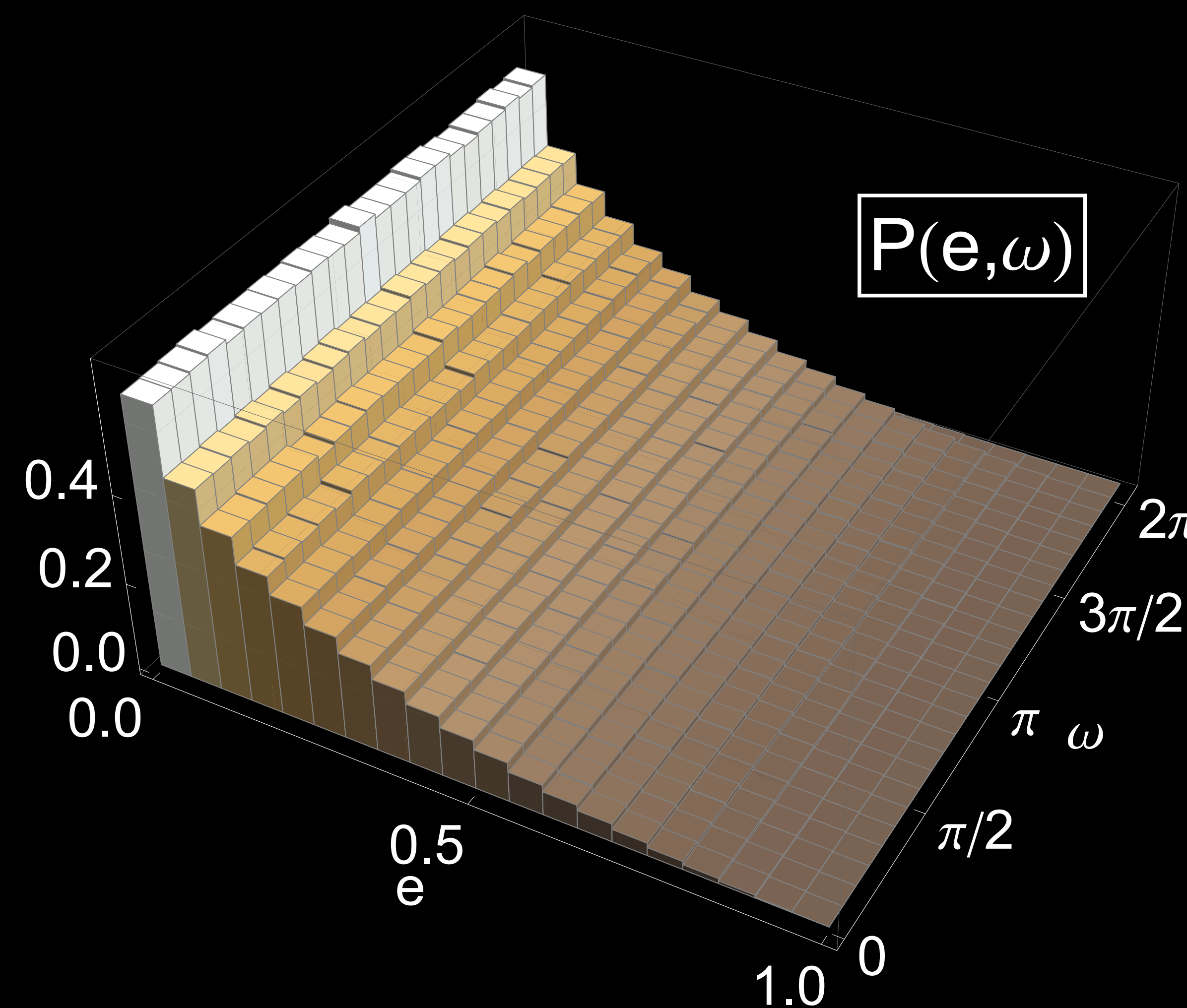
informative priors: *observational bias*



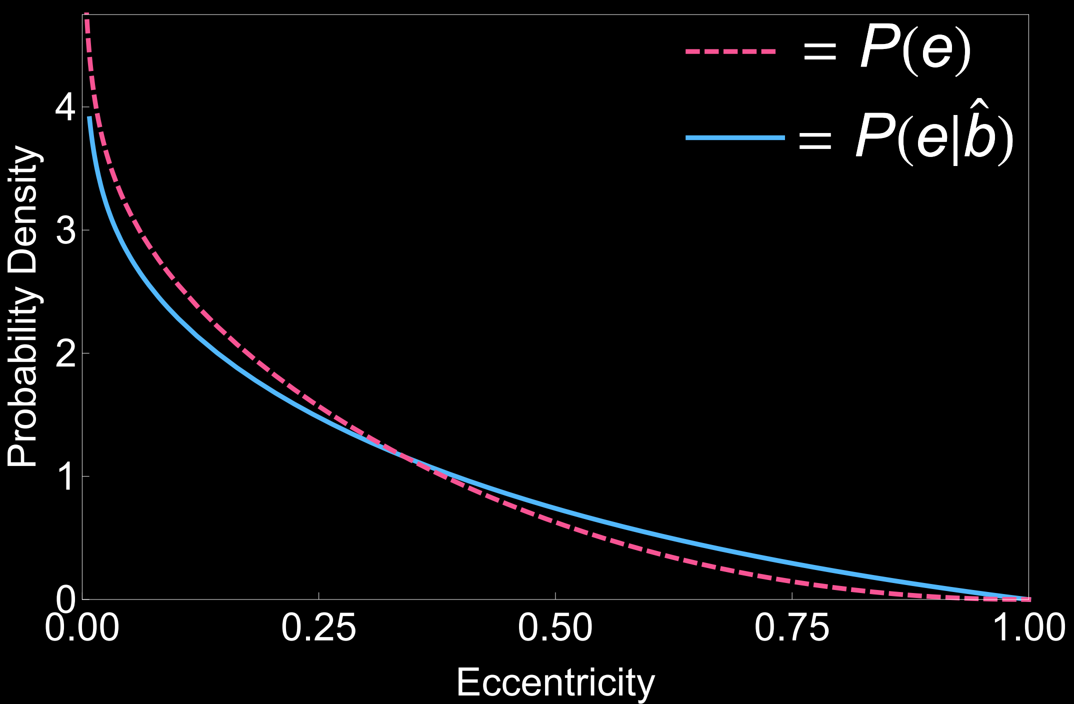


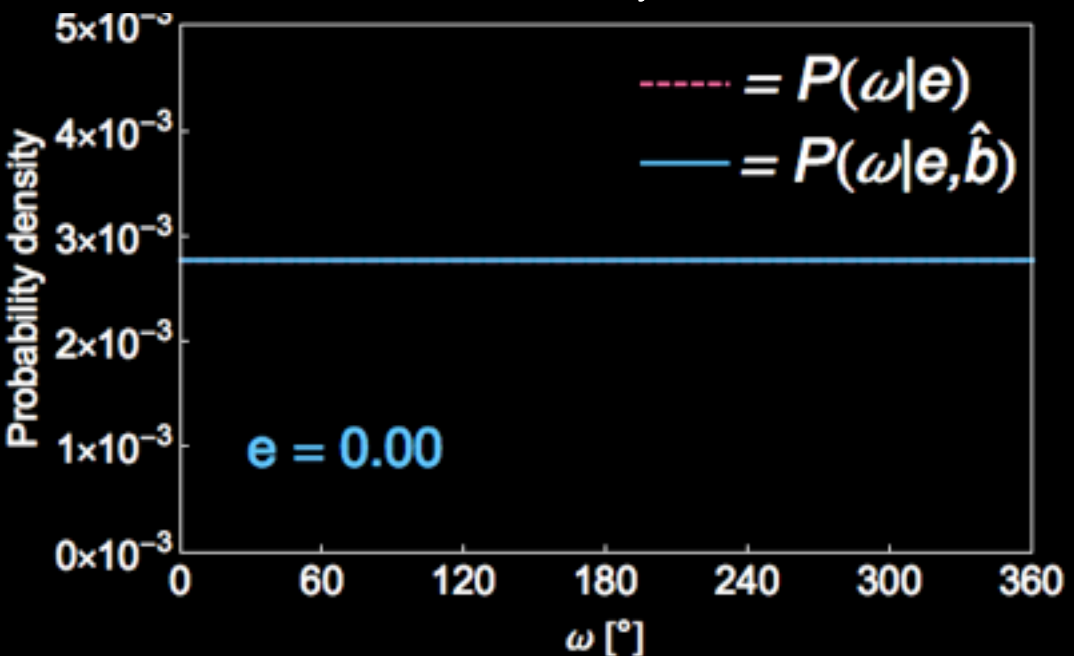
informative priors: *observational bias*

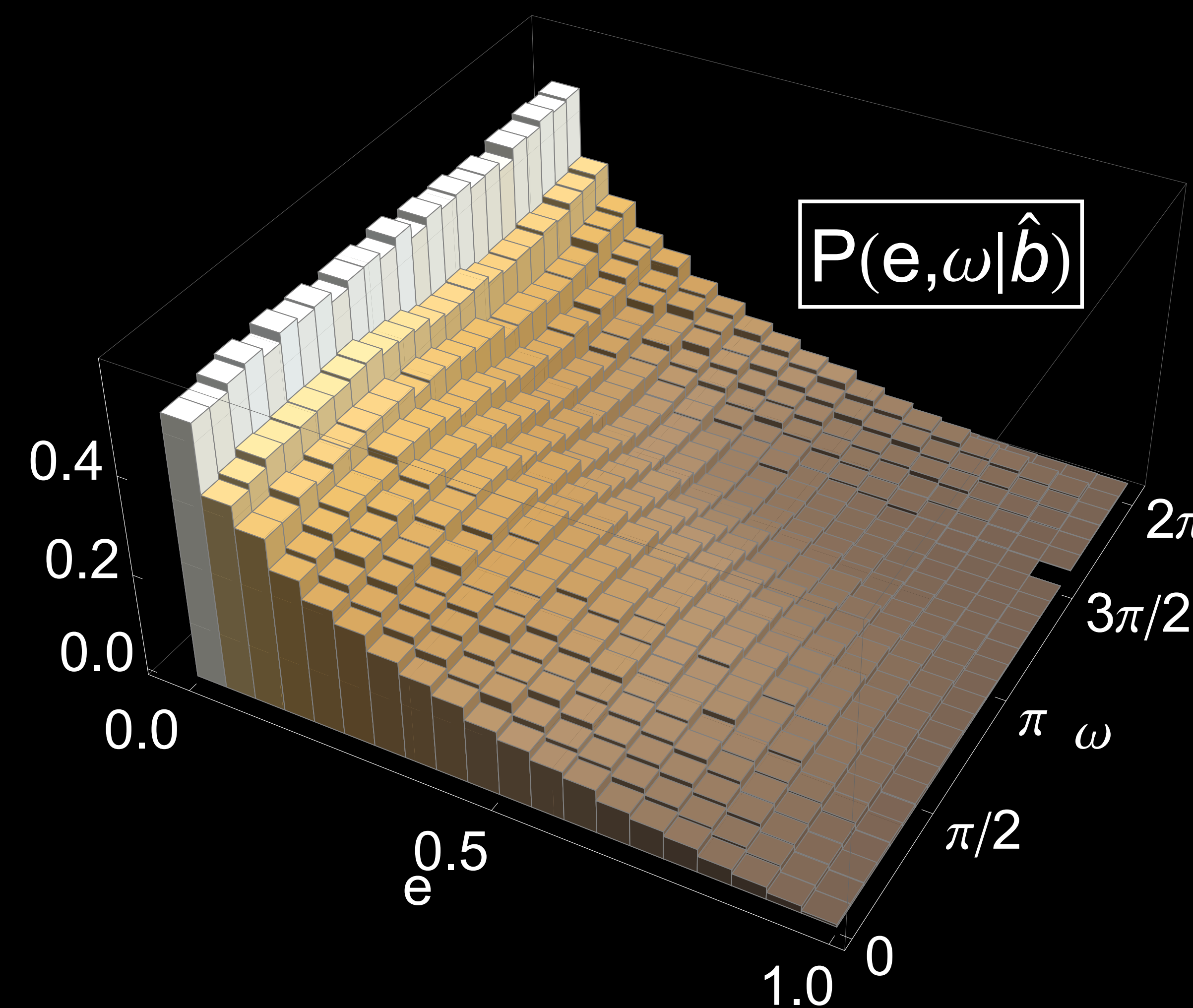




informative priors: *observational bias*







informative priors: *observational bias*

this gets even more tricky if we consider conditioning on both geometric bias and detection bias (e.g. apoapsis transits are longer => more detectable)

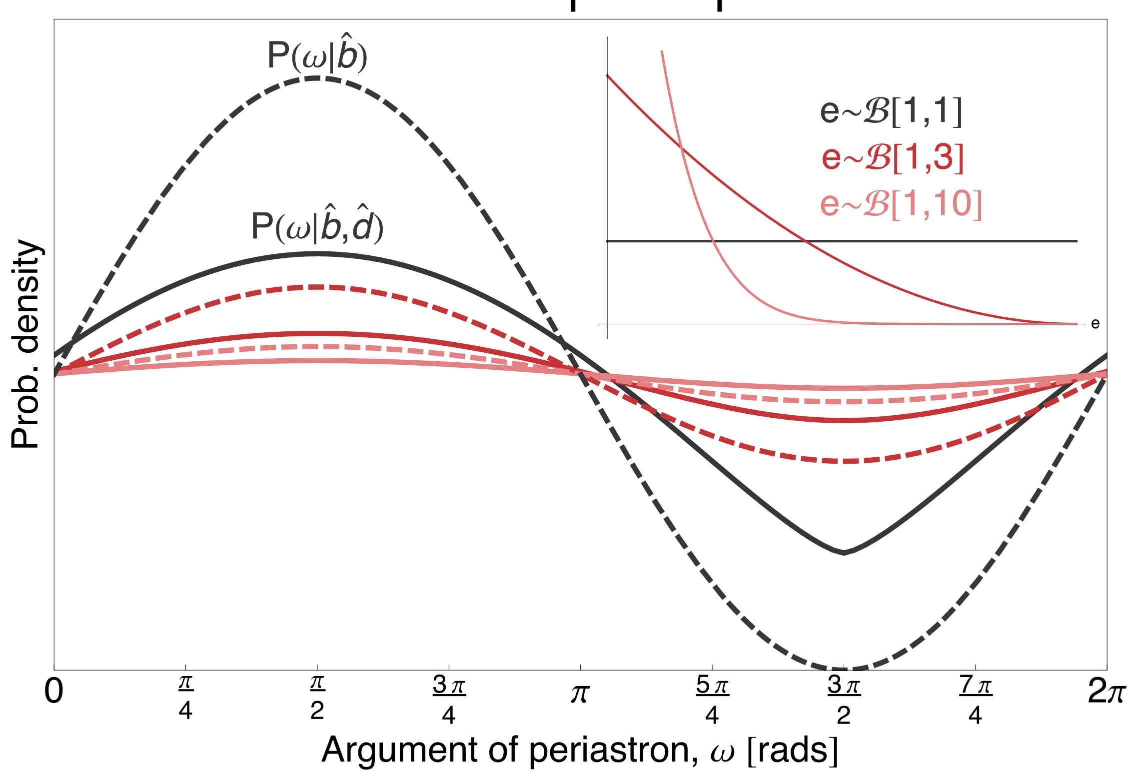
$$\Pr(e|\hat{d},\hat{b}) \propto \frac{\Pr(e)}{(1-e^2)^{3/4}} \left(\sqrt{1-e}E\left[\frac{2e}{e-1}\right] + \sqrt{1+e}E\left[\frac{2e}{e+1}\right] \right)$$

detection bias *geometric bias*

(Kipping & Sandford 2016)

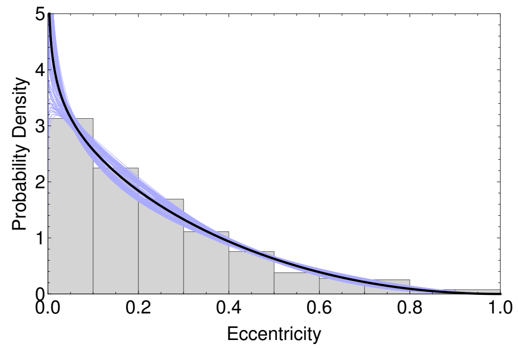
also: we don't know what observational bias of RVs are!!
(stay tuned via Chen & Kipping)

detection bias suppresses observational bias towards periapsis transits



informative priors: *observational bias*

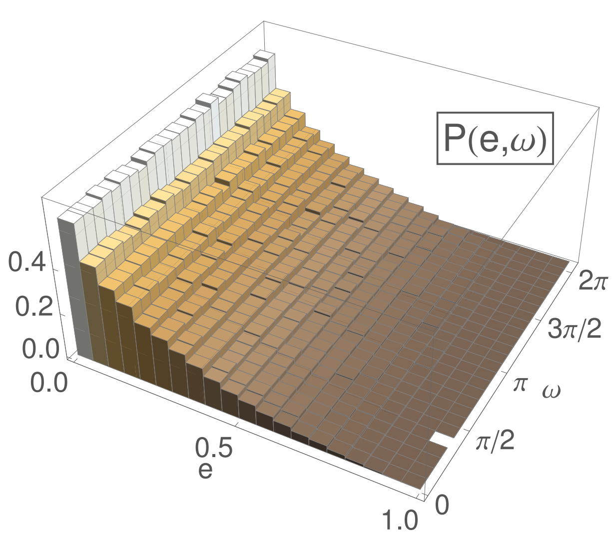
distribution of X from
detection method Y



this is ok

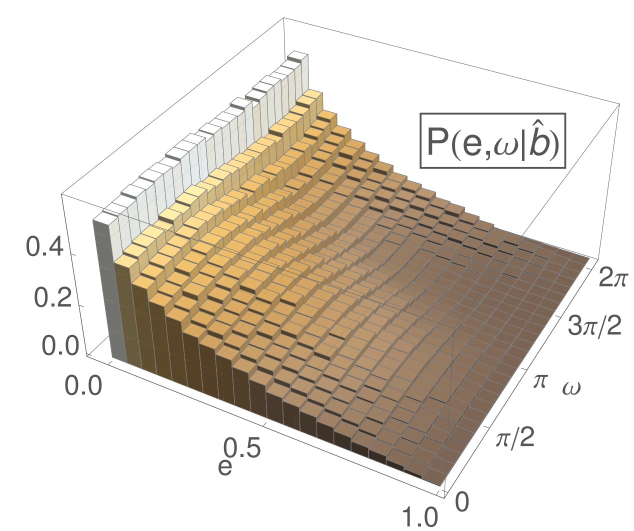
treat as a prior for P(X)
for analyzing data from
detection method Y

distribution of X from
detection method Y



don't do this
without thought

treat as a prior for P(X)
for analyzing data from
detection method Z



how to... choose a prior	how to... implement a prior
uniform, think about boundary conditions check priors on key parameters via Monte Carlo **uninformative** log-uniform for parameters scaling orders-of-magnitude	**log likelihood penalization**
Bayesian learning observational experience **informative** observational biases between detection techniques	
conjugate useful for analytic work, but not really used in practical exoplanet work	**inverse transform sampling**

how to... choose a prior	how to... implement a prior
uniform, think about boundary conditions check priors on key parameters via Monte Carlo **uninformative** log-uniform for parameters scaling orders-of-magnitude	log likelihood penalization
Bayesian learning observational experience **informative** observational biases between detection techniques	inverse transform sampling
conjugate useful for analytic work, but not really used in practical exoplanet work	

implementing priors: *log-like penalization*

$$P(\boldsymbol{\Theta}|\mathcal{D},\mathscr{M}) \propto P(\mathcal{D}|\boldsymbol{\Theta},\mathscr{M})P(\boldsymbol{\Theta}|\mathscr{M})$$

$$\mathscr{P} \propto \mathscr{L}\ \pi$$

$$\log\mathscr{P} \propto \log(\mathscr{L}\ \pi)$$

log probabilities more numerically stable

$$\log\mathscr{P} \propto \log\mathscr{L} + \log\pi$$

so just add on log(prior probability)

can think of as being loglike penalization

implementing priors: *log-like penalization*

example: a normal distribution prior, $N(\mu,\sigma)$

$$\pi(x) = \frac{\exp(-½(x-\mu)^2/\sigma^2)}{(2\pi)^{½}\sigma}$$

$$\log\pi = -½\log(2\pi) - ½\log(\sigma^2) - ½(x-\mu)^2/\sigma^2$$

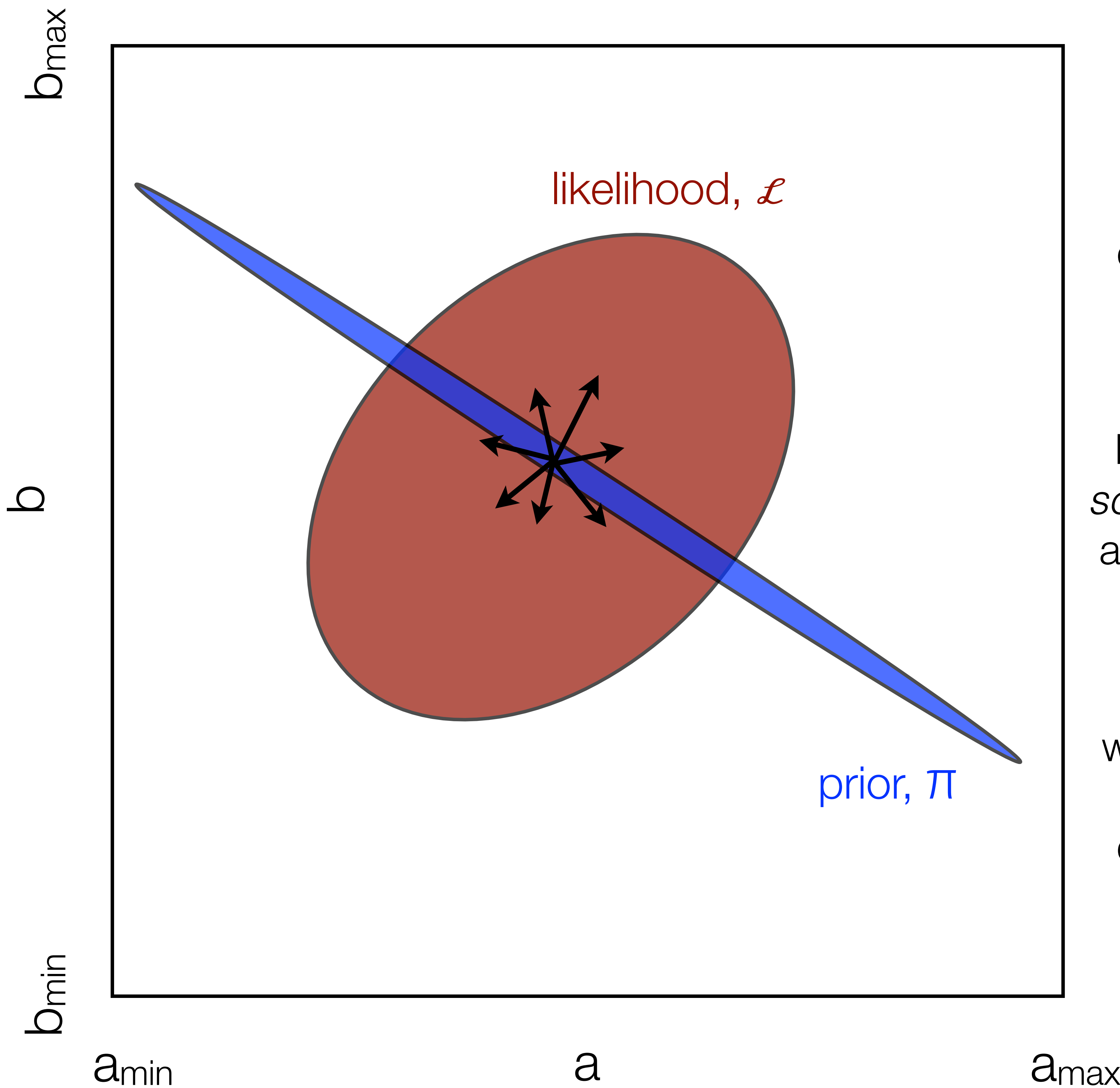
unless you want the evidence, can ignore constants

$$\log\pi = - ½(x-\mu)^2/\sigma^2$$

(highest prior probability occurs when $x=\mu$, as expected)

how to... choose a prior	how to... implement a prior
uniform, think about boundary conditions check priors on key parameters via Monte Carlo **uninformative** log-uniform for parameters scaling orders-of-magnitude Bayesian learning observational experience **informative** observational biases between detection techniques **conjugate** useful for analytic work, but not really used in practical exoplanet work	**log likelihood penalization** just need to know pdf of prior **inverse transform sampling**

implementing priors: *inverse sampling*



accepted jumps have to be a i) high $\mathcal{L}$ ii) high π

everytime we compute logπ and make essentially blind jumps

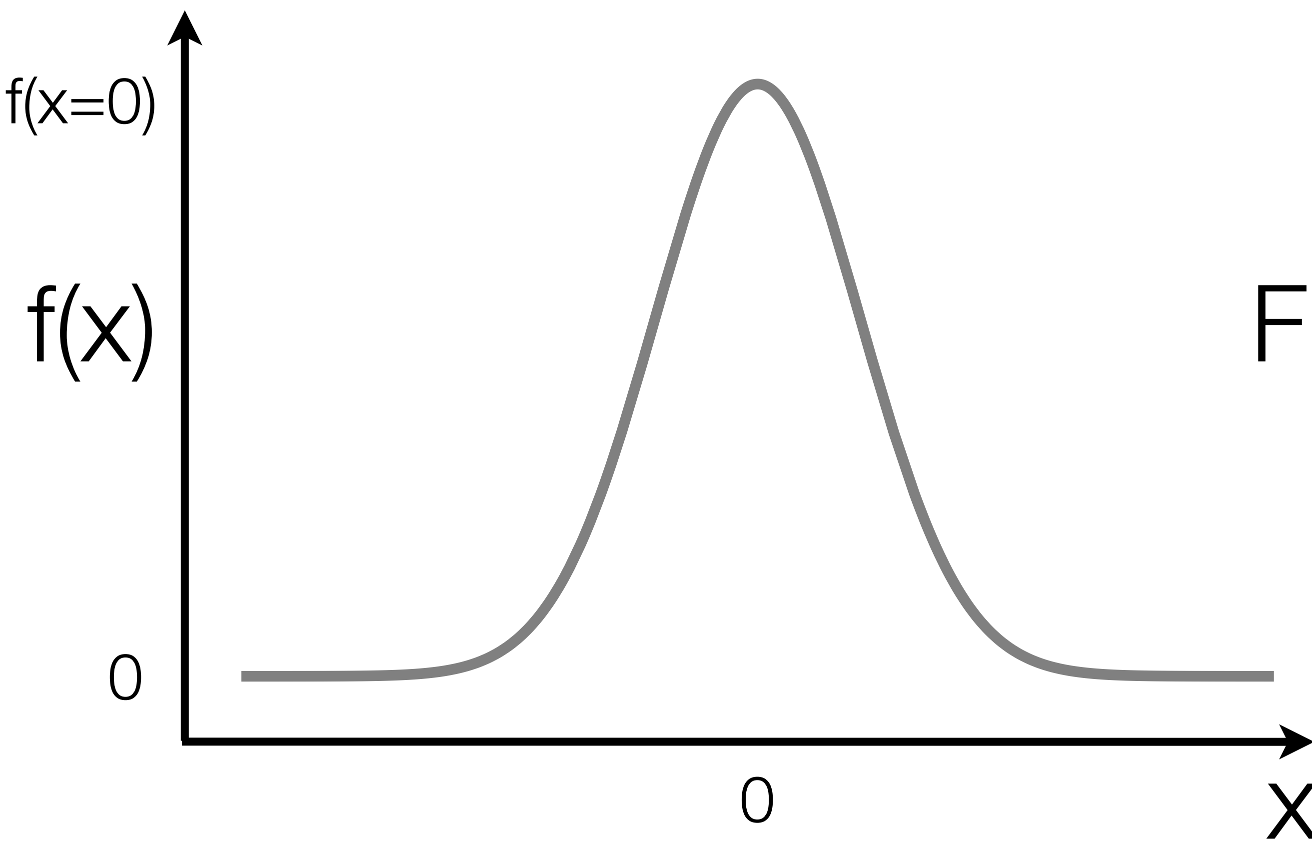
but a more elegant solution is *sometimes* possible, by drawing a sample directly from the prior

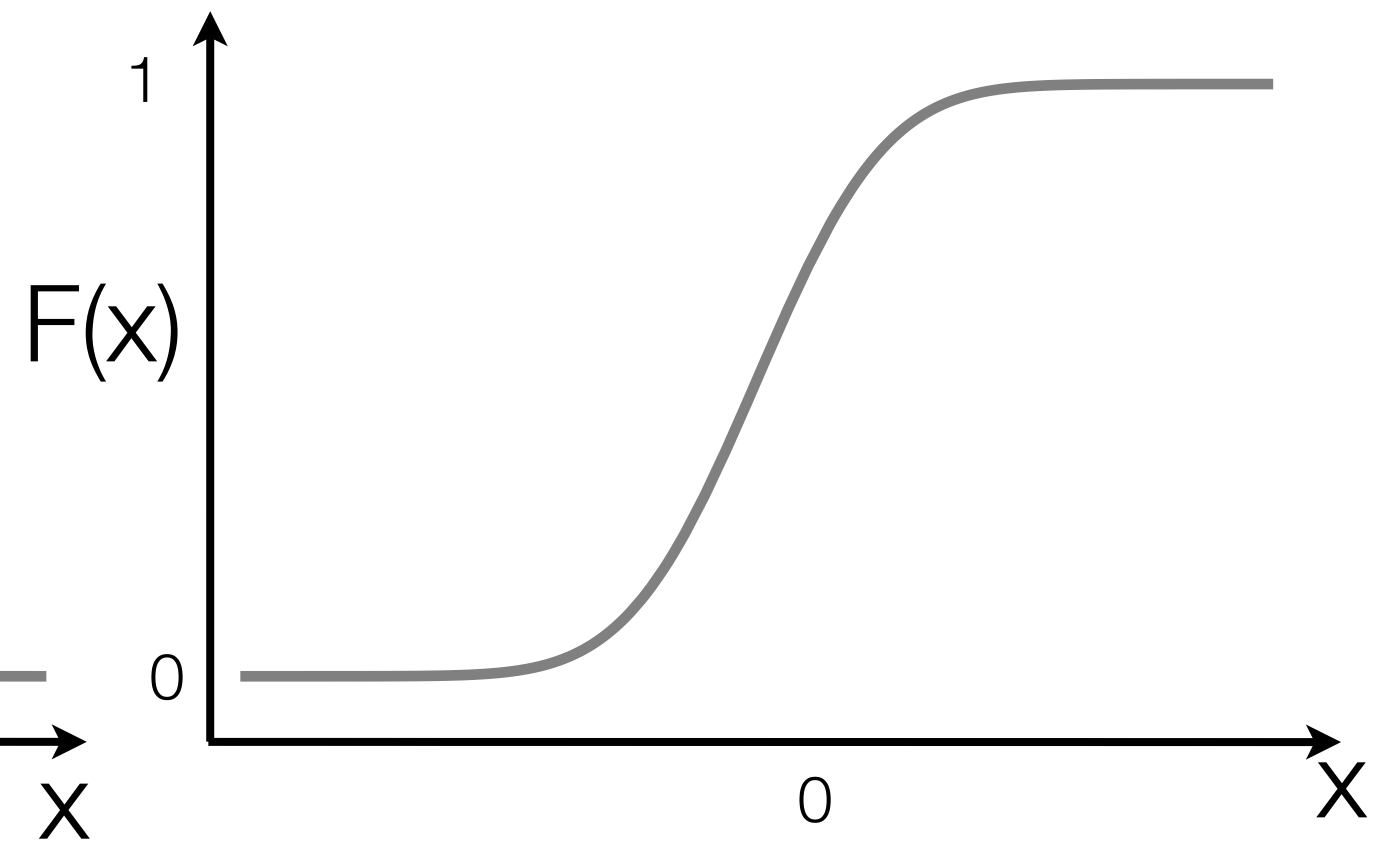
drawing a random.normal() won't work though, as we need to "walk" in the parameter in order to build a Markov chain

this is possible with *inverse transform sampling*

implementing priors: *inverse sampling*

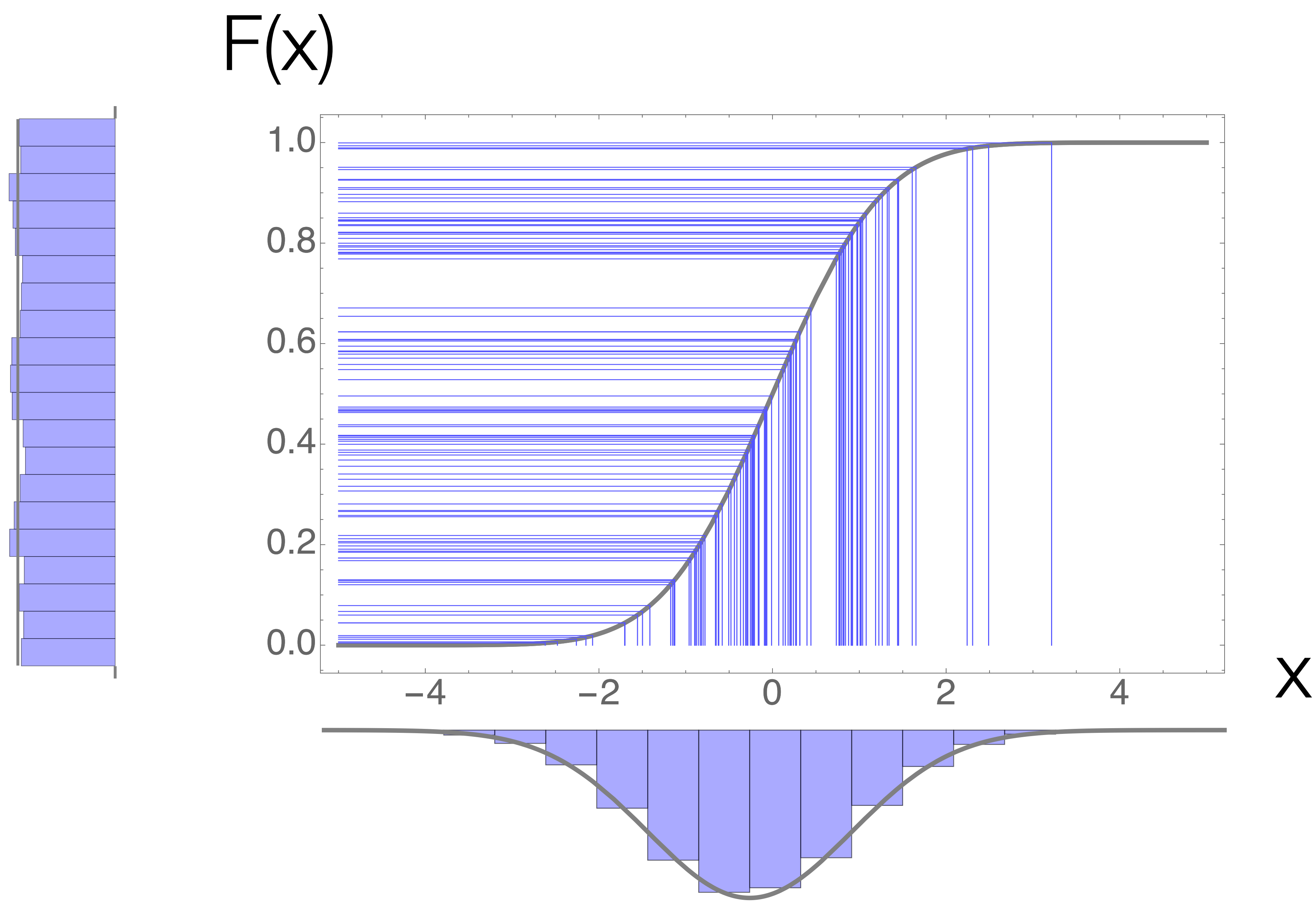
$$F(x) = \int_{-\infty}^{x} f(x')\, dx'$$





for a proper prior, 0<F(x)<1

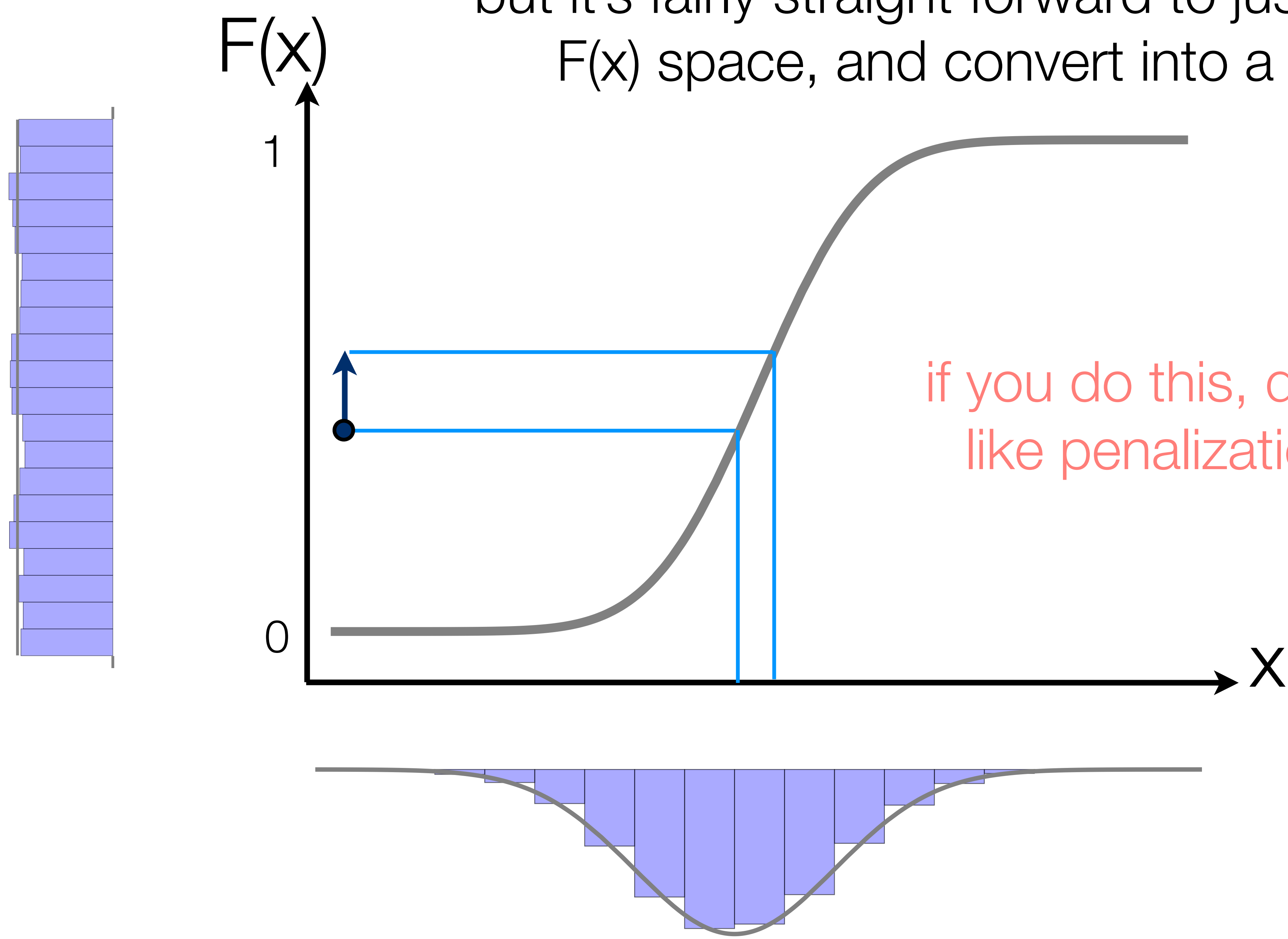
implementing priors: *inverse sampling*



but is just random samples, does not constitute walking

implementing priors: *inverse sampling*

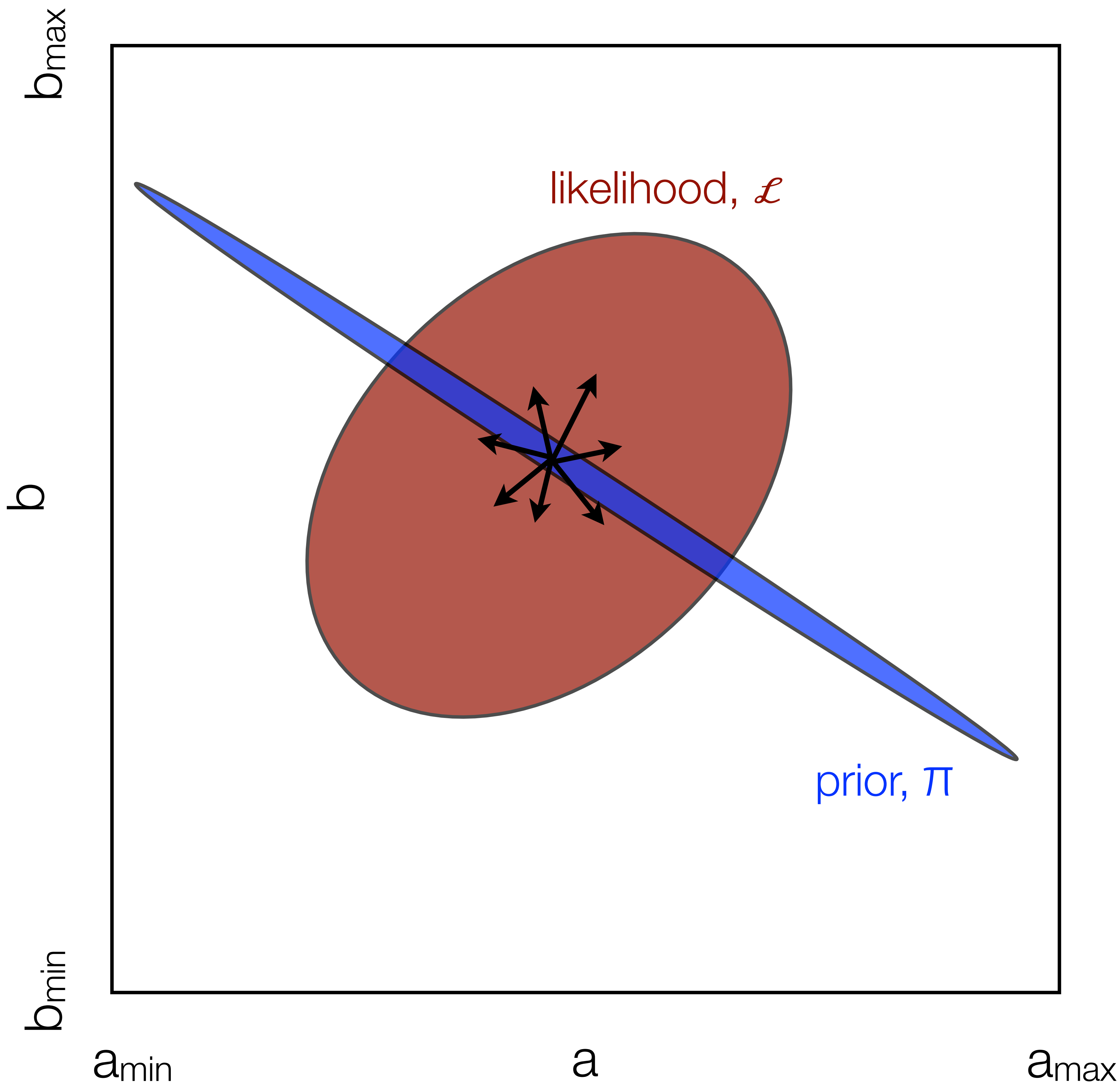
but it's fairly straight forward to just walk in the F(x) space, and convert into a x sample



if you do this, don't do log-like penalization as well!

in the limit of no likelihood, Markov chain will be a uniform chain in F(x) => normal dist in x, as required

implementing priors: *inverse sampling*



inverse transform sampling is optimally efficient for exploring the prior volume

easy to implement for standard 1D distributions

because of this, <u>some Bayesian inference packages sample from the priors exclusively in this way</u> e.g. MultiNest (Feroz 2008,2009)

but, 2D and non-standard distributions (that one might derive when doing observational priors) can be intractable

how to...
choose a prior

uniform, think about boundary conditions

check priors on key parameters via Monte Carlo

uninformative

log-uniform for parameters scaling orders-of-magnitude

Bayesian learning

observational experience

informative

observational biases between detection techniques

conjugate

useful for analytic work, but not really used in practical exoplanet work

how to...
implement a prior

log likelihood penalization

just need to know pdf of prior

optimally efficient for exploring prior volume, default for MultiNest

inverse transform sampling

non-standard and >1D distributions can be intractable

how to... choose a prior

uniform, think about boundary conditions

check priors on key parameters via Monte Carlo

uninformative

log-uniform for parameters scaling orders-of-magnitude

Bayesian learning

observational experience

informative

observational biases between detection techniques

conjugate

useful for analytic work, but not really used in practical exoplanet work

how to... implement a prior

log likelihood penalization

just need to know pdf of prior

optimally efficient for exploring prior volume, default for MultiNest

inverse transform sampling

non-standard and >1D distributions can be intractable

transits

(Foreman-Mackey+ 2013)

P ~ log-uniform

t_{mid} ~ uniform

$p=R_P/R_*$ ~ uniform or log-uniform

b or $\cos(i)$ ~ uniform

a/R_* or ρ_* ~ log-uniform

(Kipping 2014; Kipping & Sandford 2016)

e ~ Beta corrected & w ~ uniform

(Ford 2006)

or $e^{1/2}\sin w$ & $e^{1/2}\cos w$ ~ uniform

(Kipping 2013b/2016)

q_1/q_2 or $\alpha_h/\alpha_r/\alpha_\theta$ ~ uniform

RVs

(Ford & Gregory 2007, Balan & Lahav 2008)

K ~ modified log-uniform

if following up a known transiter

or K~ uniform to -ve

(Ford & Gregory 2007, Balan & Lahav 2008)

s ~ modified log-uniform

be warned that $t_{conj} \neq t_{mid}$ for $e>0$

t_{conj}~ uniform

not the "right" answer, just my personal recommendations, although I would always think about the specifics of my problem!

extra slides on limb darkening

$$I(\mu) = 1 - u_1(1-\mu) - u_2(1-\mu)^2$$

1] everywhere positive: $I(\mu)>0$

2] monotonically decreasing form surface to limb: $dI/d\mu>0$

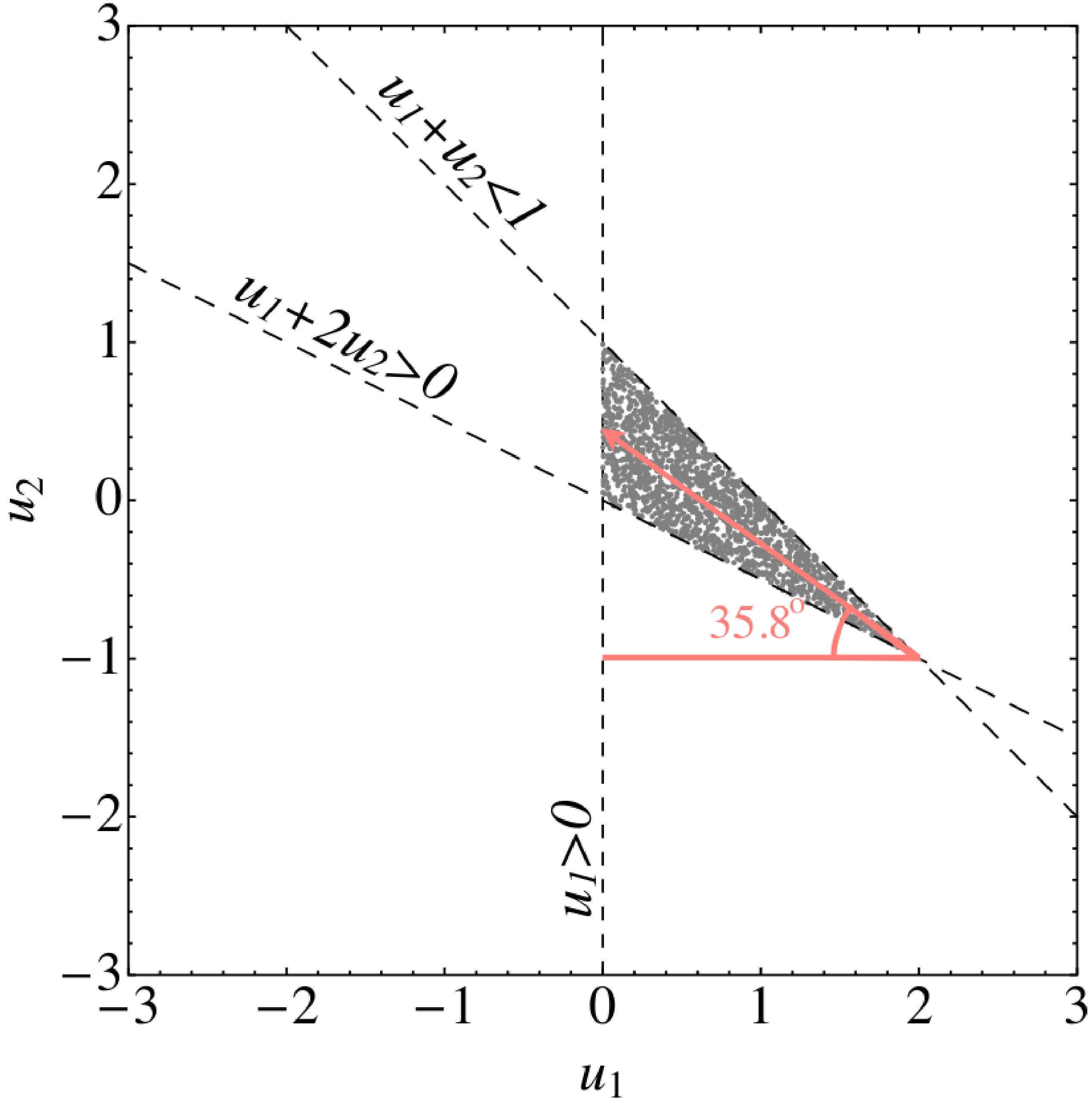
$$\boxed{\begin{aligned} q_1 &= (u_1+u_2)^2 \\ q_2 &= \tfrac{1}{2}u_1(u_1+u_2)^{-1} \end{aligned}}$$

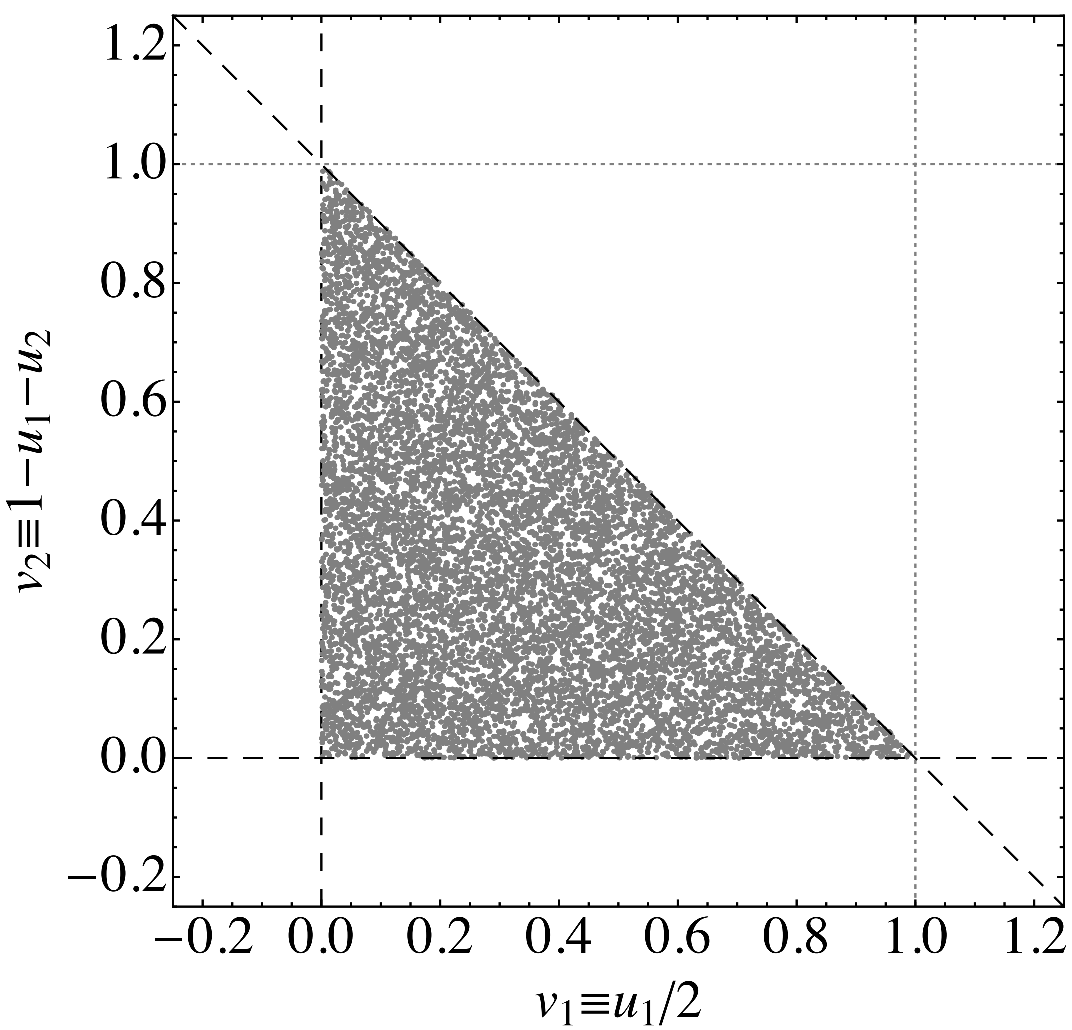
Kipping (2013b)

$u_1+u_2 < 1$

$u_1 > 0$

$u_1+2u_2 > 0$





$v_1 = q_1^{1/2}\, q_2$

$v_2 = 1-q_1^{1/2}$

thanks computer games!

$$I(\mu) = 1 - u_1(1-\mu) - u_2(1-\mu)^2$$

quadratic law

$$I(\mu) = 1 - \boxed{c_1(1-\mu^{1/2})} - c_2(1-\mu) - c_1(1-\mu^{3/2}) - c_2(1-\mu^2)$$

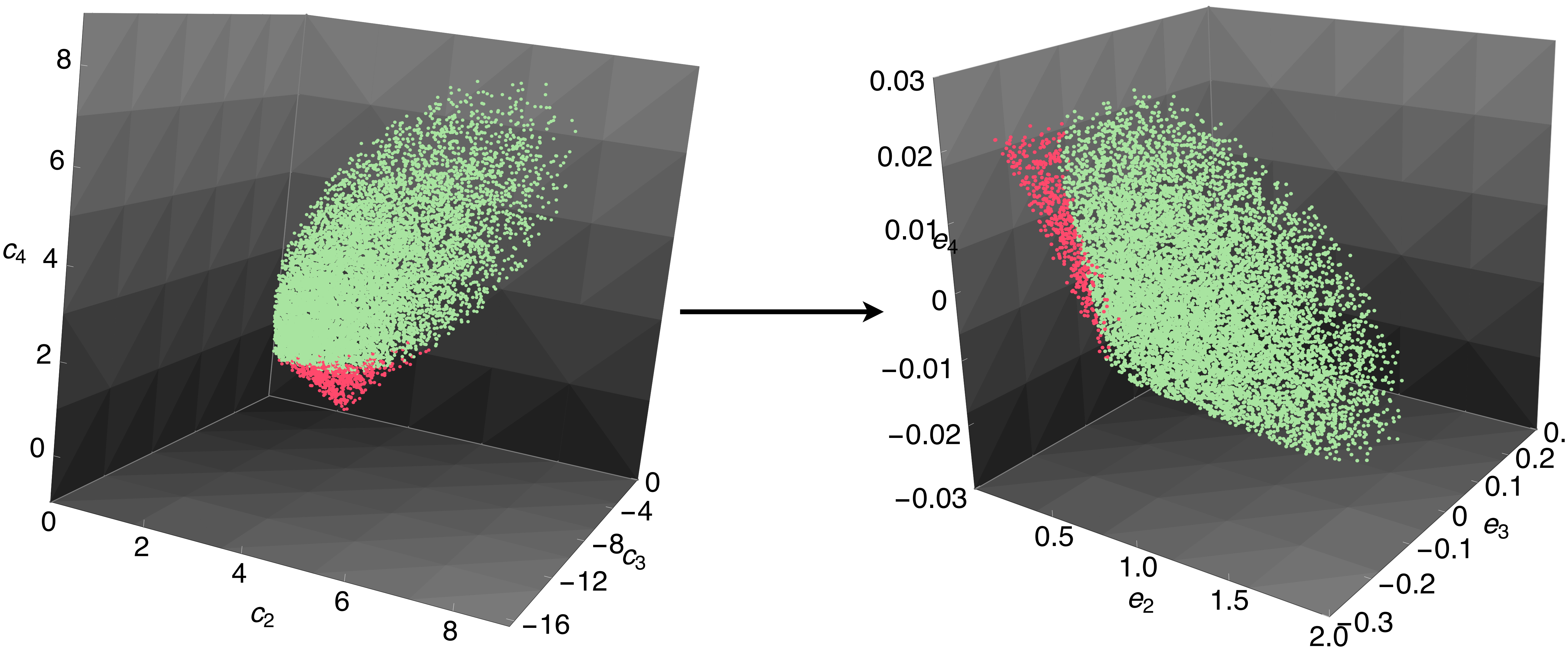
non-linear law

Sing (2010) argue that dropping the c_1 term is motivated by Solar data (Neckel & Labs 1994) and 3D stellar models (Bigot et al. 2006), which show that $I(\mu)$ varies smoothly at small μ, meaning that a $\mu^{1/2}$ term is superfluous

$$I(\mu) = 1 - c_2(1-\mu) - c_1(1-\mu^{3/2}) - c_2(1-\mu^2)$$

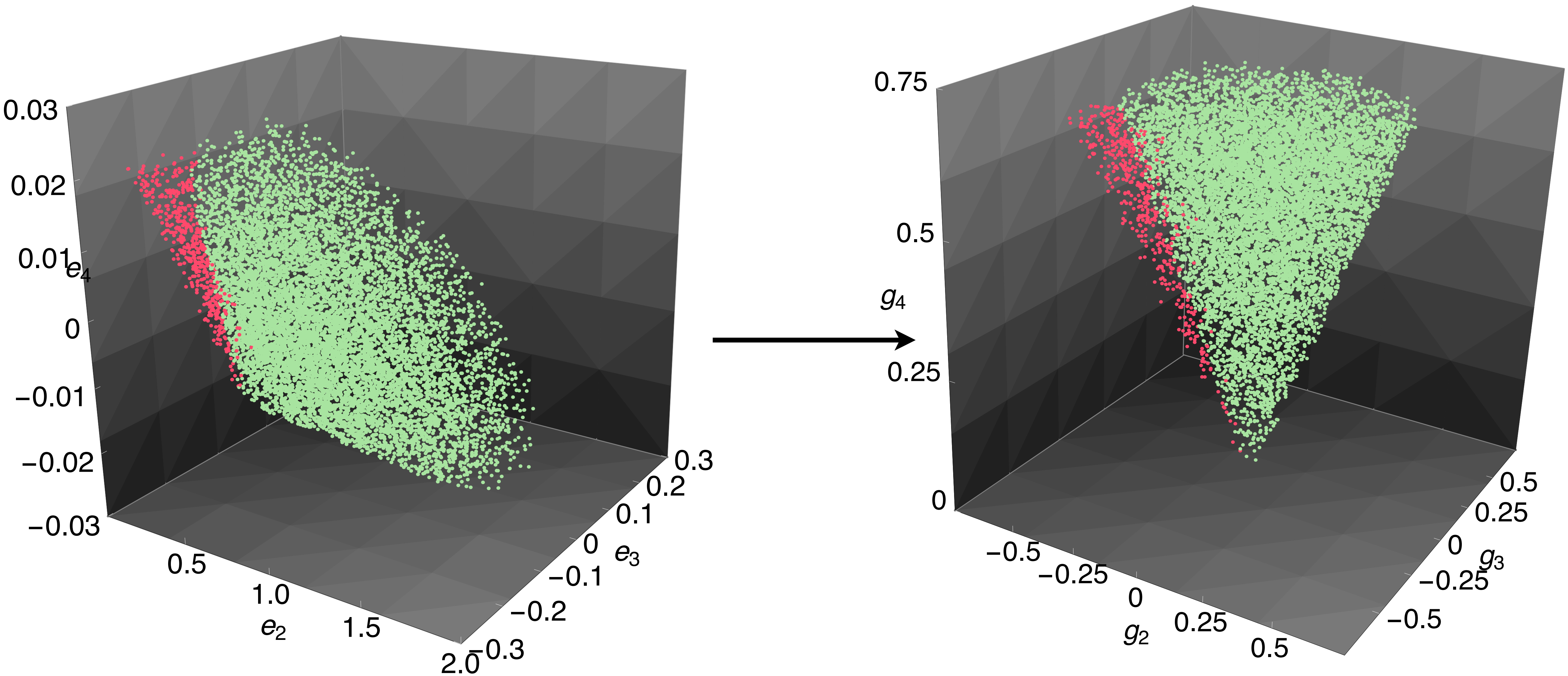
3-parameter law

I can't imagine 4-dimensions, so no.
I can imagine 3 though!



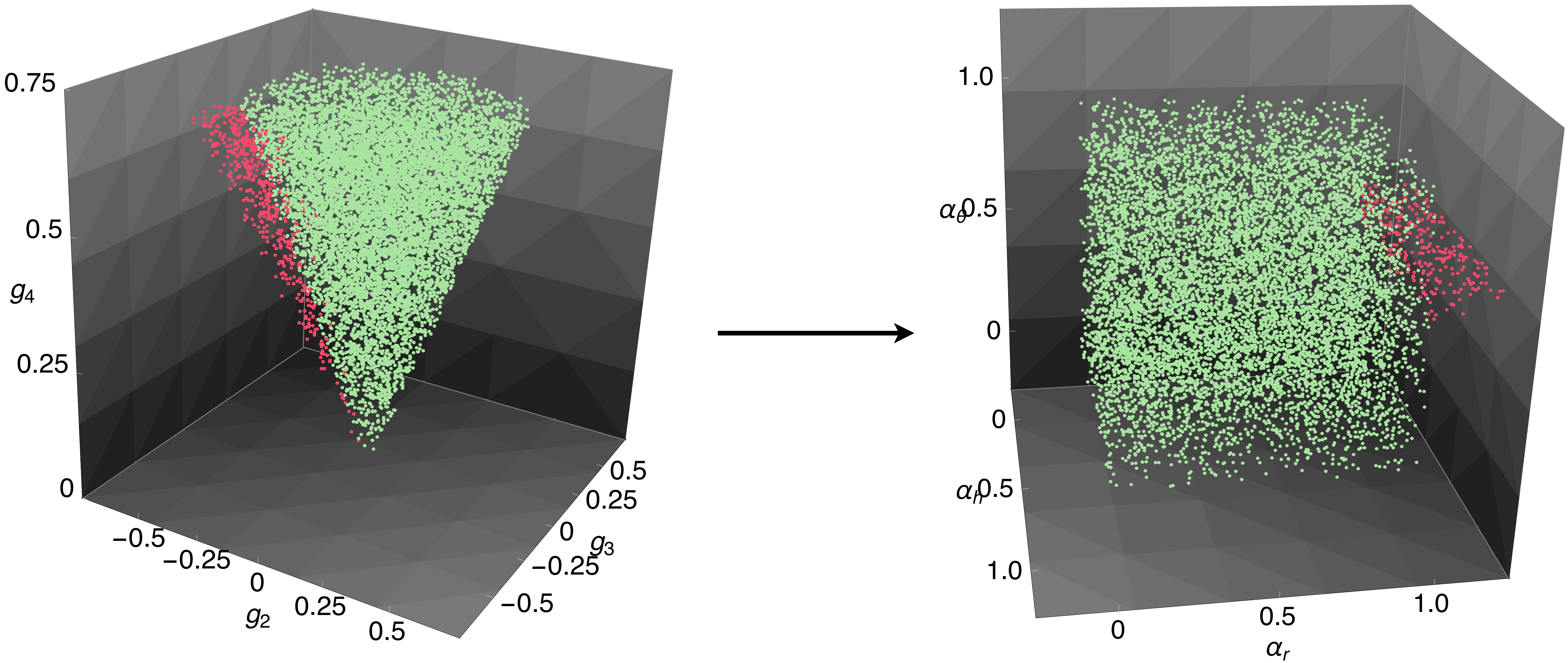
re-scale axes to push loci inside the unitary cube

then rotate the loci round aligning envelope with the

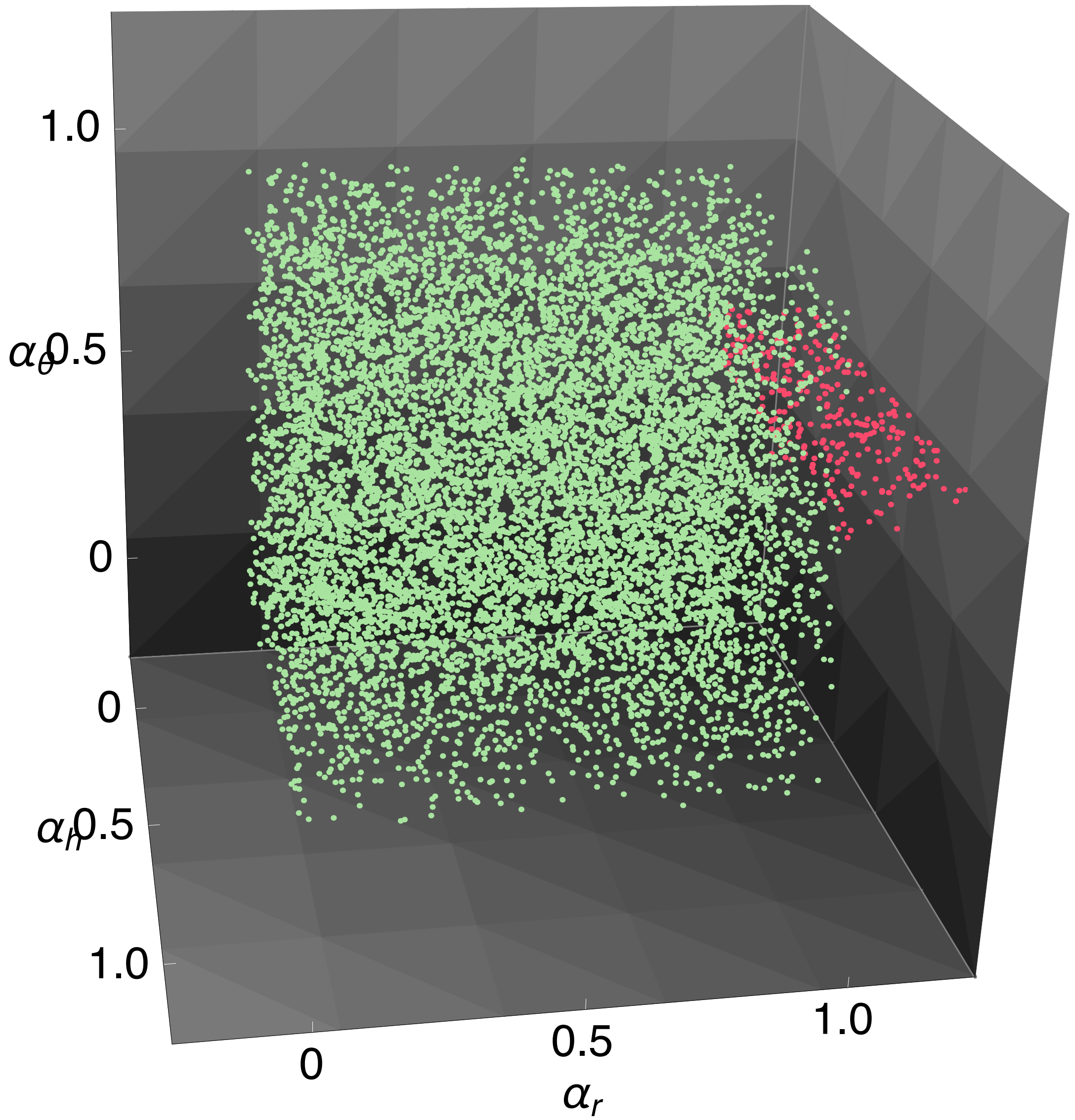


next again re-scale to inside the unitary cube

then re-align cone's apex to y-axis



finally use standard method for sampling from a cone to re-parameterize into alpha



$$c_2 = \frac{\alpha_h^{1/3}}{12}\Bigg(28(9-5\sqrt{2}) + 3\alpha_r^{1/2}\Big(-6\cos(2\pi\alpha_\theta) + (3+10\sqrt{2}\sin(2\pi\alpha_\theta)\Big)\Bigg),$$

$$c_3 = \frac{\alpha_h^{1/3}}{9}\Bigg(-632+396\sqrt{2} + 3\alpha_r^{1/2}(4-21\sqrt{2})\sin(2\pi\alpha_\theta)\Bigg),$$

$$c_4 = \frac{\alpha_h^{1/3}}{12}\Bigg(28(9-5\sqrt{2}) + 3\alpha_r^{1/2}\Big(6\cos(2\pi\alpha_\theta) + (3+10\sqrt{2}\sin(2\pi\alpha_\theta)\Big)\Bigg).$$

https://github.com/davidkipping/LDC3

94.4% completeness

using the alpha parameterization, we draw the green samples, which encompass 94.4% of the total allowed region

97.3% validity

due to slight modification of assuming a perfect cone, 97.3% of the samples drawn using the alpha-parameterization satisfy the initial conditions

ensuring 100% validity

the remaining 2.7% unphysical samples can be easily removed with a rejection algorithm check afterwards (this check is fully analytic!)