Bayesian Model Selection

Globular cluster mass profiles

— Dec.2020 Zhiya Lou

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **462,** 696–714 (2016) doi:10.1093/mnras/stw1104
Advance Access publication 2016 May 11

Testing lowered isothermal models with direct *N*-body simulations of globular clusters

Alice Zocchi,[1]★ Mark Gieles,[1] Vincent Hénault-Brunet[1] and Anna Lisa Varri[2]
[1]*Department of Physics, University of Surrey, Guildford GU2 7XH, UK*
[2]*School of Mathematics and Maxwell Institute of Mathematical Sciences, University of Edinburgh, Kings Buildings, Edinburgh EH9 3JZ, UK*

LIMPEY

Data:
- 3D Position (x,y,z -> r)
- 3D Velocity (xy, xz, yz -> v)

Model parameters:
- m (mass)
- r (half light radius)

scale

- g (truncation energy)
- phi (potential)

shape gravitational potential

- r_a (anisotropy radius can -> infinity, become isotropic)

Assume we have the <u>complete</u> dataset of the cluster's dynamic information (position and velocity)

Goal: Get the estimated of true parameters, posterior distribution
And, Which model we should choose?

use limpey generate dataset

Model Selection

1. fit isotropic data with the correct isotropic model
2. fit isotropic data with a more expansive anisotropic model

3. fit anisotropic data with the incorrect isotropic model
4. fit anisotropic data with the correct anisotropic model (the anisotropy does not go to 0)

Data X: true parameter {m5r3g1.5phi3.0}

Steps:

1. Assume the models

   Model1: isotropic (expected to prefer)

   Model2: anisotropic

2. Set up Bayes Method (Why not p-value?)

$$\text{Posterior Prob} = P(\Theta|X) = \frac{P(\Theta)P(X|\Theta)}{P(X)} = \frac{\text{Prior} \cdot \text{Likelihood}}{\text{Evidence}}$$

$$P(X|\Theta) = \prod_{i}^{N_*} \Lambda_i(X_i|\Theta)$$

$$\Lambda_i = \frac{f(X_i|\Theta)}{\text{Normalizing term}}$$

N #of stars

f distribution function in LIMPEY

$$f(E, J^2) = A \exp\left(-\frac{J^2}{2r_a^2 s^2}\right) E_\gamma\left(g, \frac{\phi(r_t) - E}{s^2}\right)$$

For Model Comparison, We evaluate K =

$$\frac{P(M1|X)}{P(M2|X)} = \boxed{\frac{P(X|M1)}{P(X|M2)}} \frac{P(M1)}{P(M2)}$$

A value of logK > 1 means that M1 is more strongly supported by the data than M2

3.Calculate Likelihood/ Choice of Sampling Methods

$$\text{Marginal likelihood} = P(X|M) = \int P(X|\theta, M)P(\theta|M)d\theta = E_{prior}[P(X|M)]$$

Naive Monte Carlo Estimator

$$\hat{p}(X|\theta) = \frac{1}{N}\sum_{i=1}^{N} p(X|\hat{\theta}_i), \text{ where } \hat{\theta}_i \sim p(\theta) \text{ prior distribution}$$

- Importance Sampling

$$\int p(X|\theta)p(\theta)d\theta = \int p(X|\theta)p(\theta)\frac{g(\theta)}{g(\theta)}d\theta = E_g(\frac{p(X|\theta)p(\theta)}{g(\theta)})$$

$$\hat{p} = \frac{1}{N}\sum_i \frac{p(X|\hat{\theta})p(\hat{\theta})}{g(\hat{\theta})}$$

drawing sample from proposal g instead of prior

importance weight = posterior/ proposal

Application: Estimation of Euler's constant

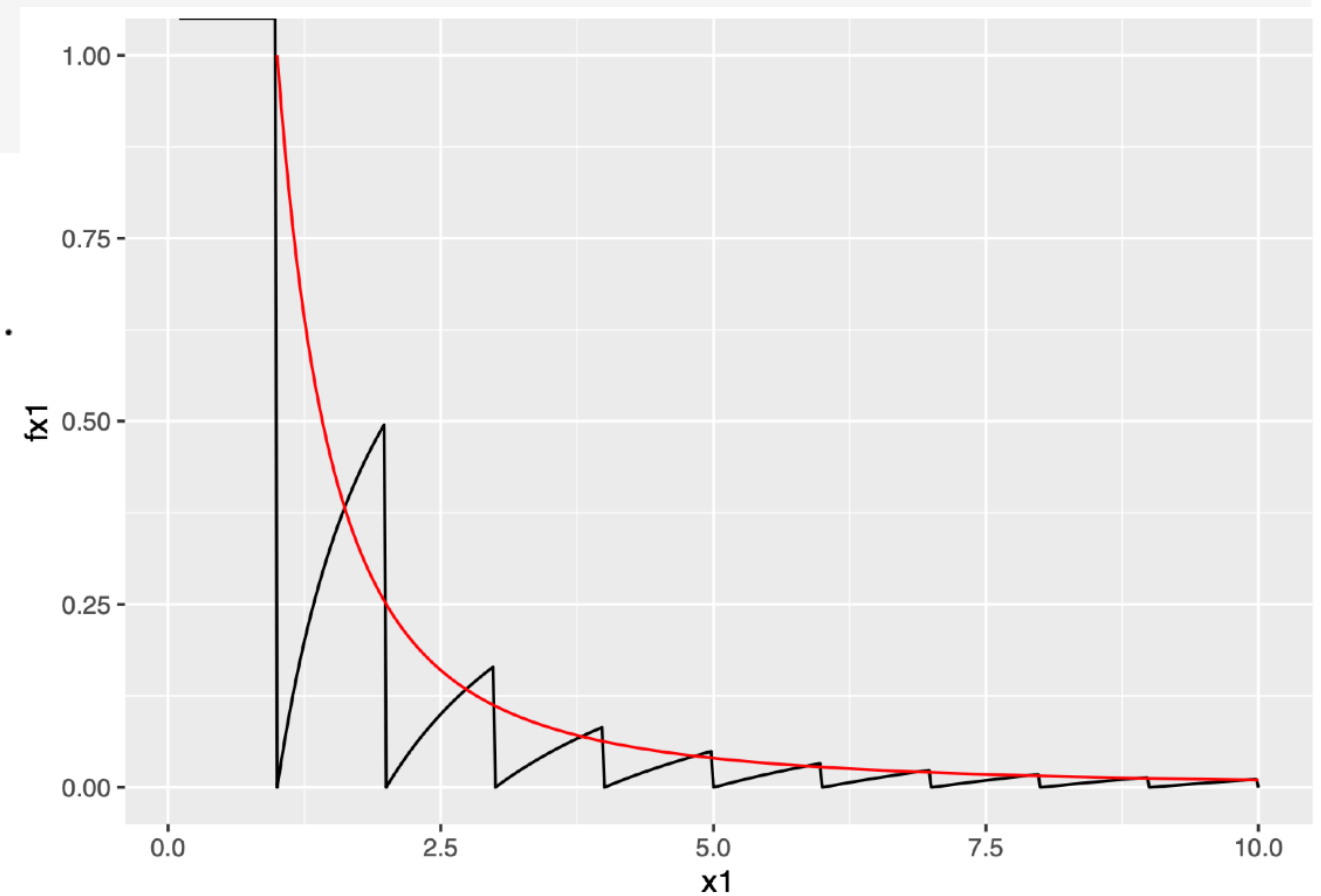
Let importance density $g(x) = x^{-2}$

$$\begin{aligned}\gamma &= \int_1^\infty \left(\frac{1}{\lfloor x \rfloor} - \frac{1}{x}\right)dx \\ &= \int_1^\infty x^2\left(\frac{1}{\lfloor x \rfloor} - \frac{1}{x}\right)x^{-2}dx \\ &= \int_1^\infty x^2\left(\frac{1}{\lfloor x \rfloor} - \frac{1}{x}\right)g(x)dx\end{aligned}$$

```
N=10000000
gammalist = c(rep(0,N))
for (i in 1:N){
  unif = runif(1)
  x <- 1/(1-unif)
  gammalist[i] = (x^2)*(1/floor(x)-1/x)
}
mean(gammalist)
```

```
## [1] 0.5769798
```

The true value of Euler constant is 0.57721566...

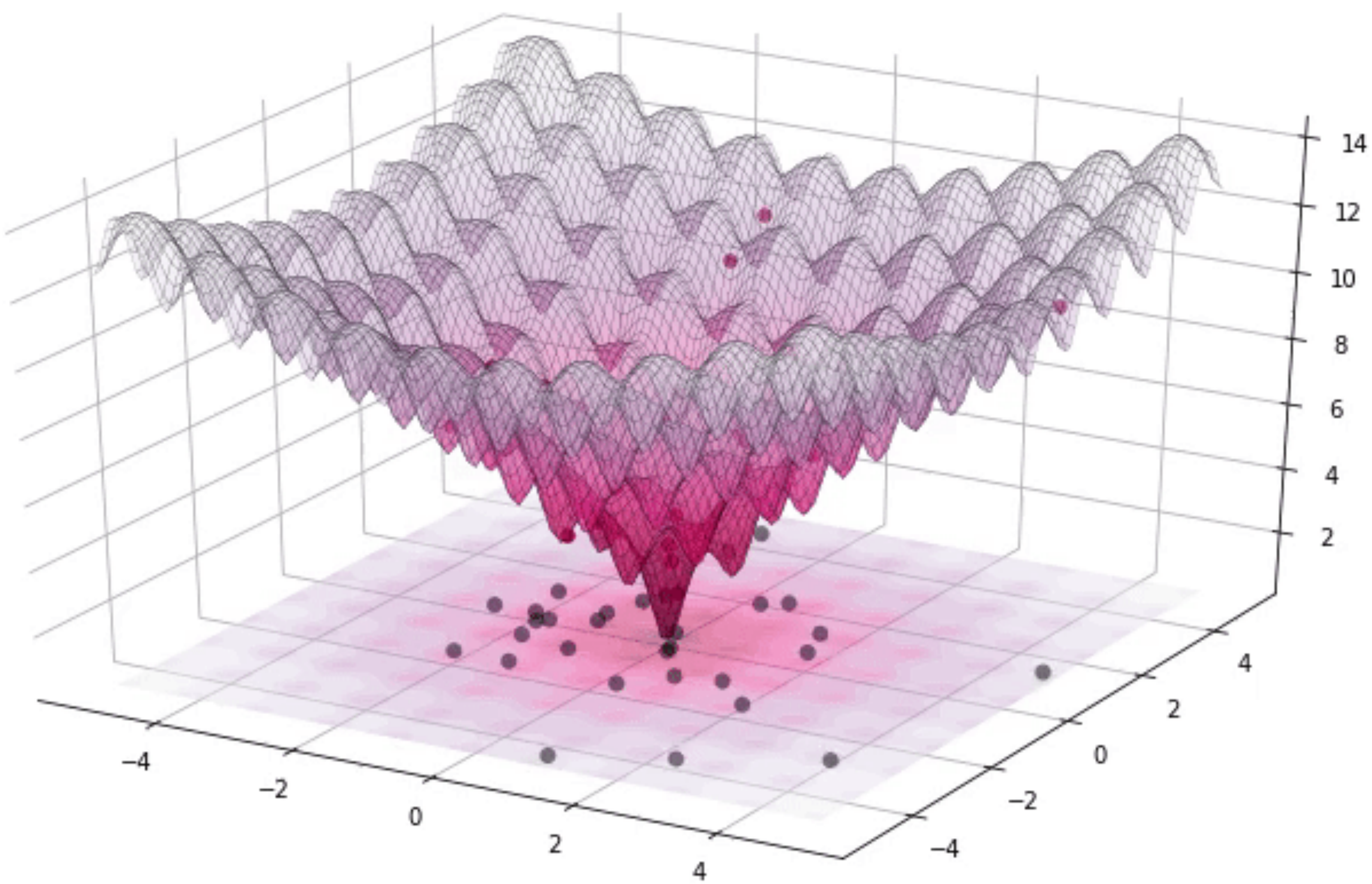


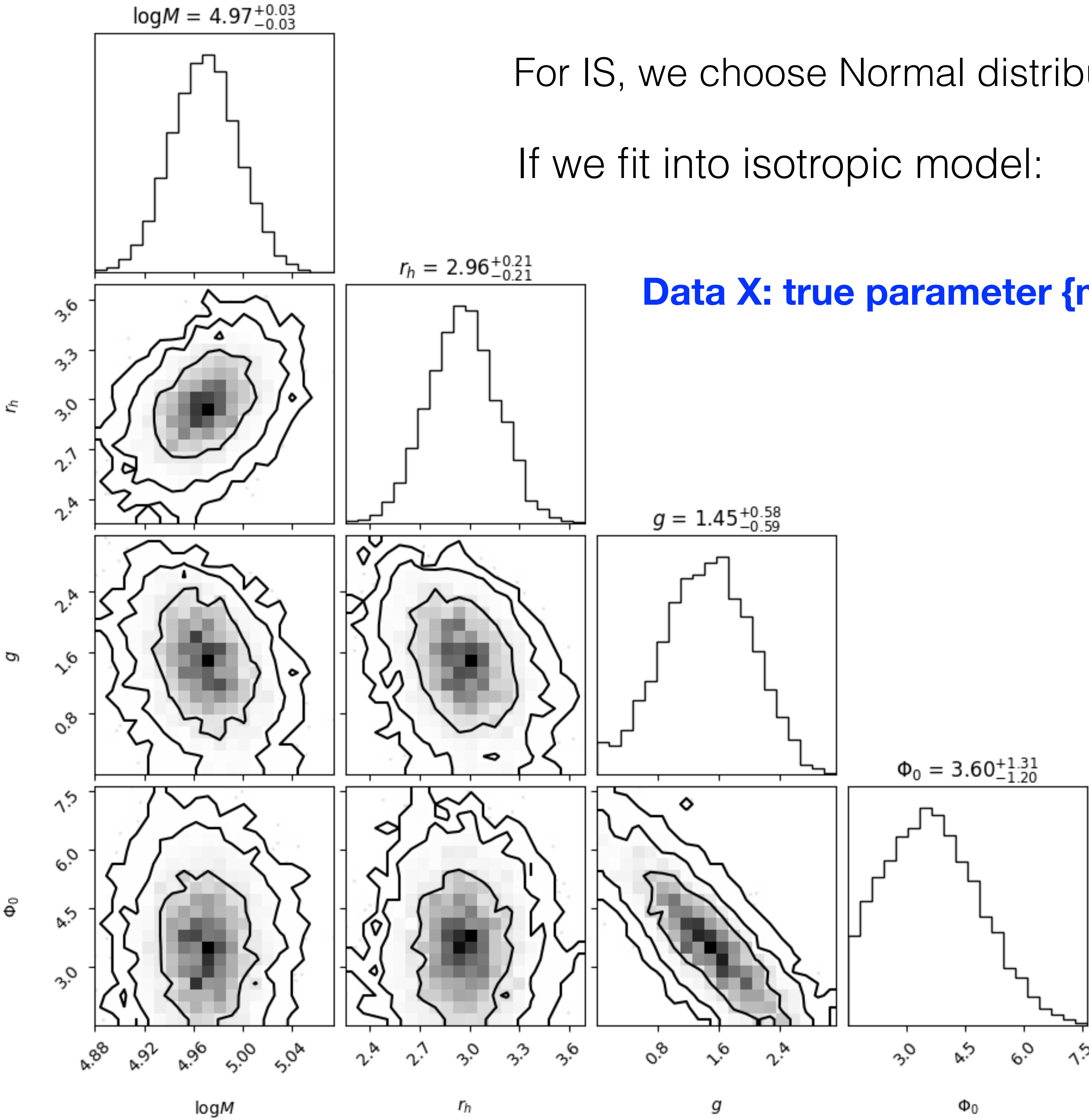
Optimization, get parameter estimates

Max posterior =

Min {Negative of the log-posterior function of the input parameters}

There are many optimization methods like gradient/coordinate descent
Here we use **Differential Evolution**
Target function not required to be differential (not even continuous)

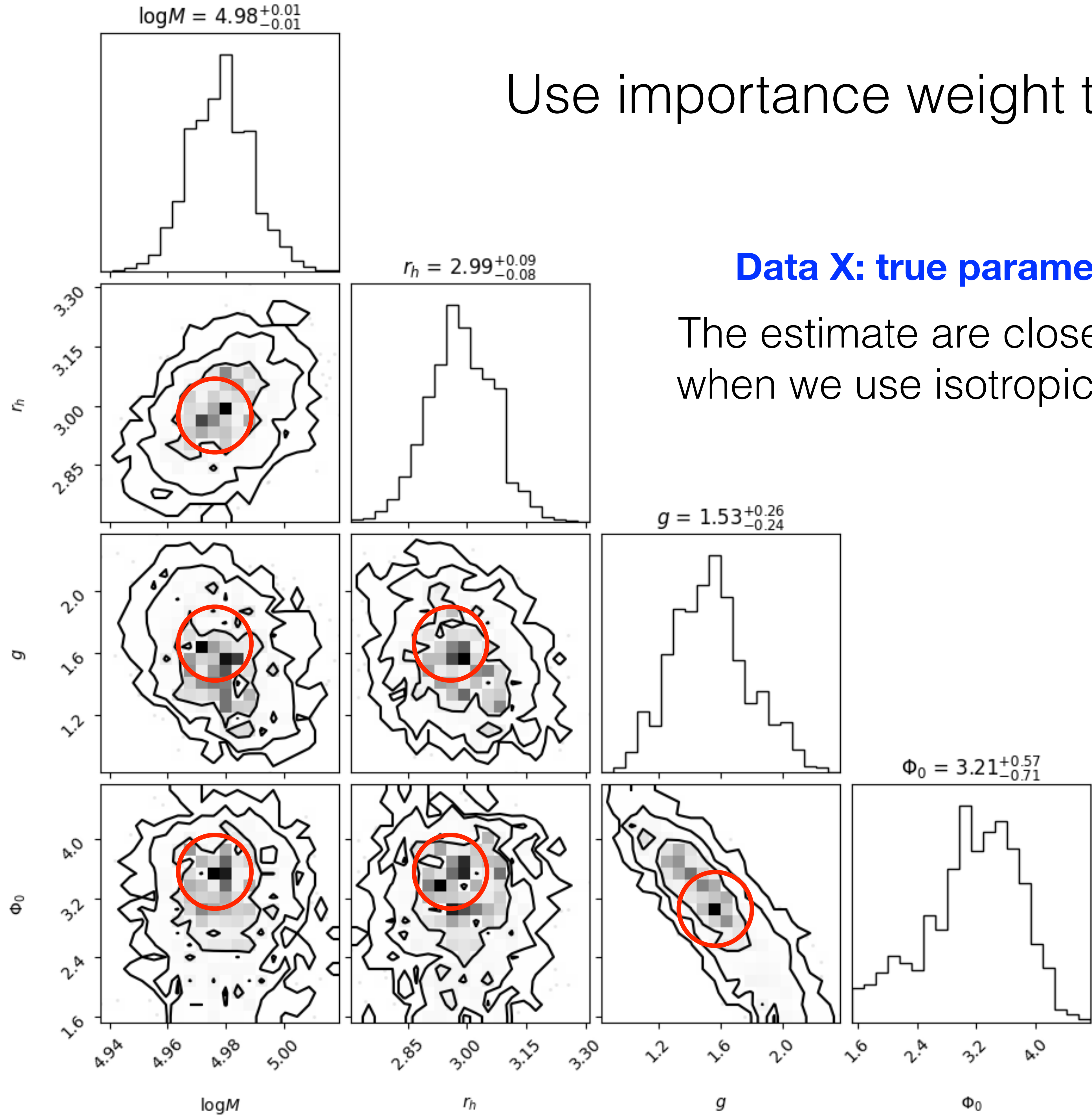




For IS, we choose Normal distribution as proposal

If we fit into isotropic model:

Data X: true parameter {m5r3g1.5phi3.0}



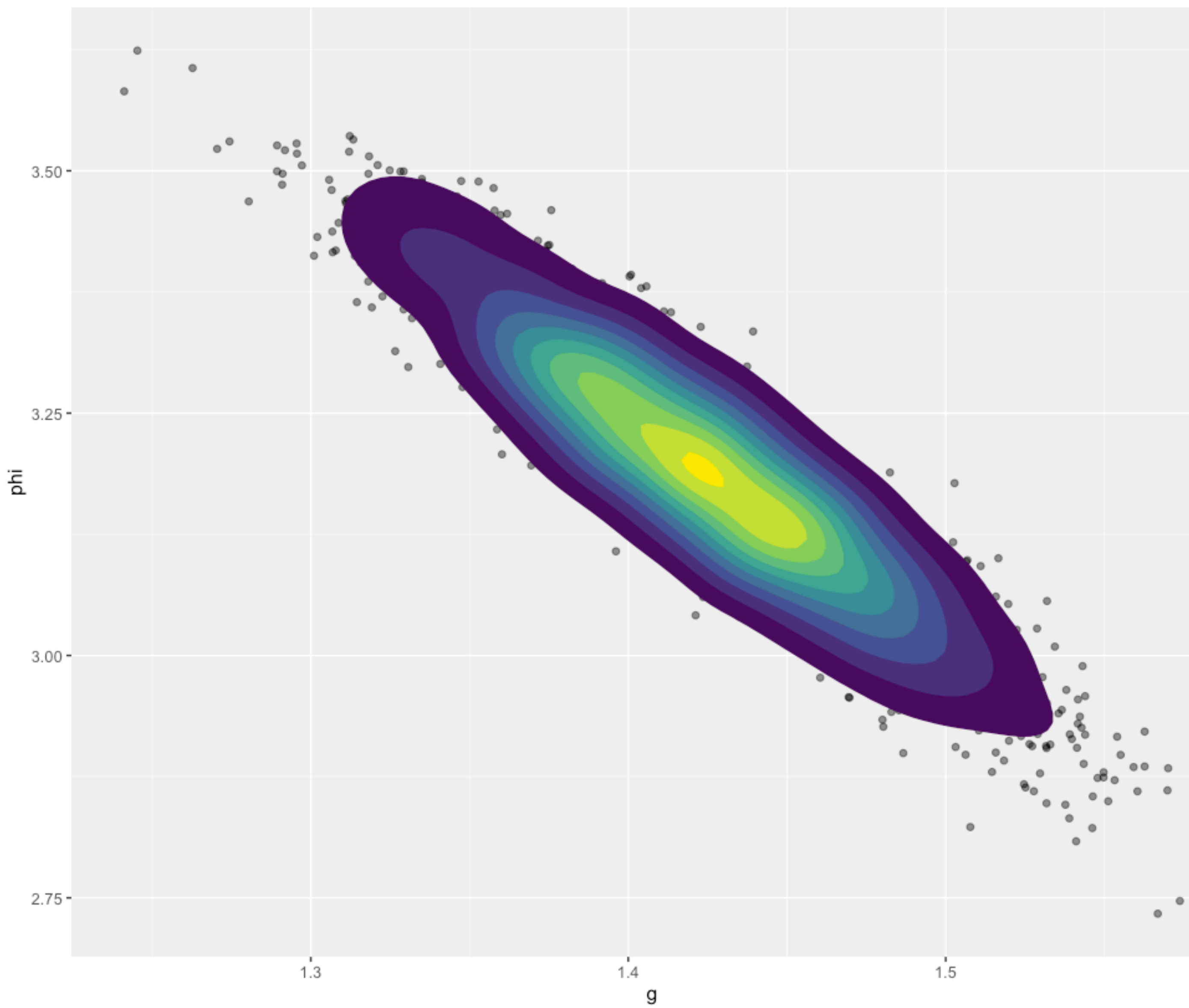
Use importance weight to resample:

Data X: true parameter {m5r3g1.5phi3.0}

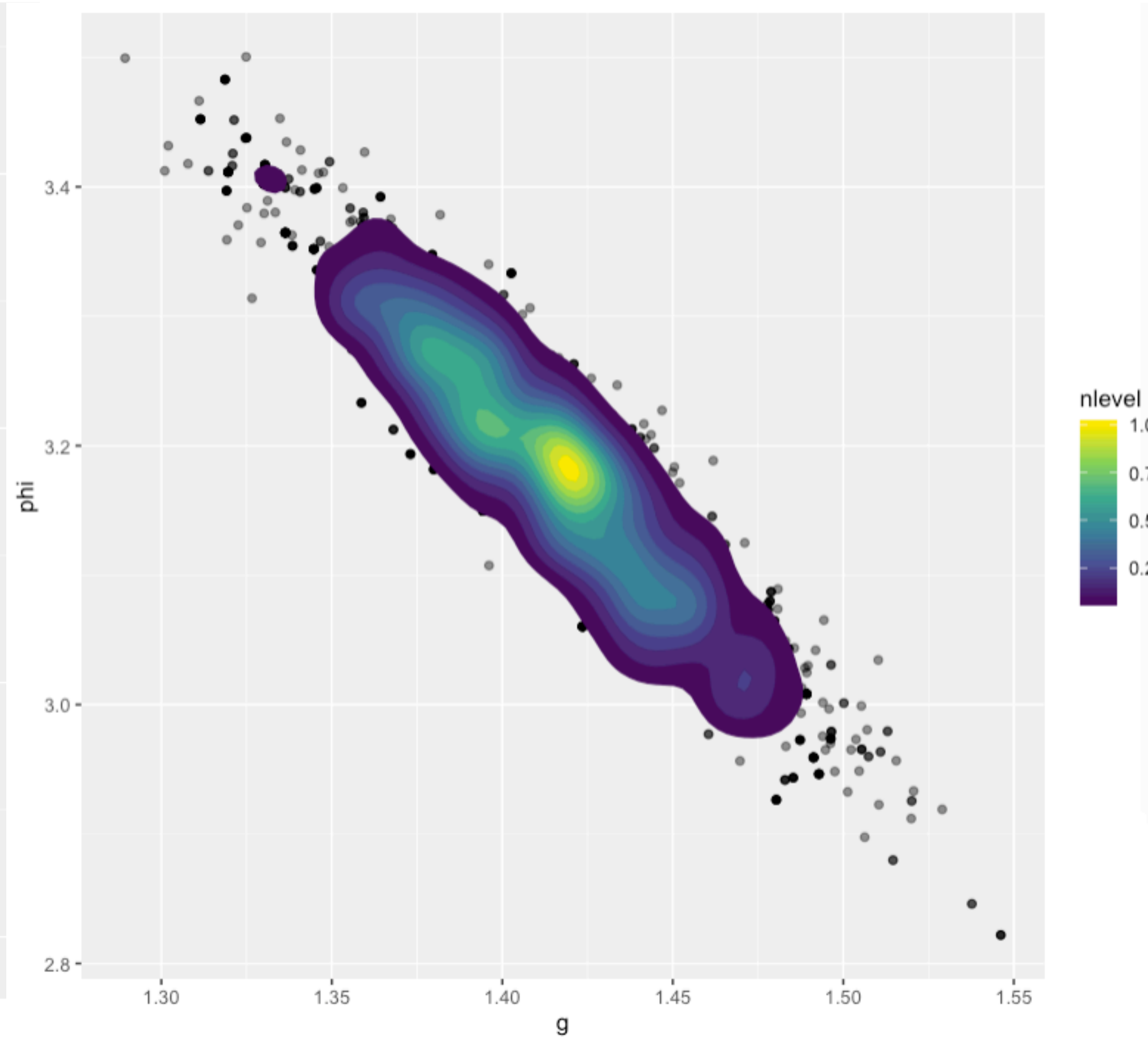
The estimate are close to true parameter when we use isotropic model

g vs phi

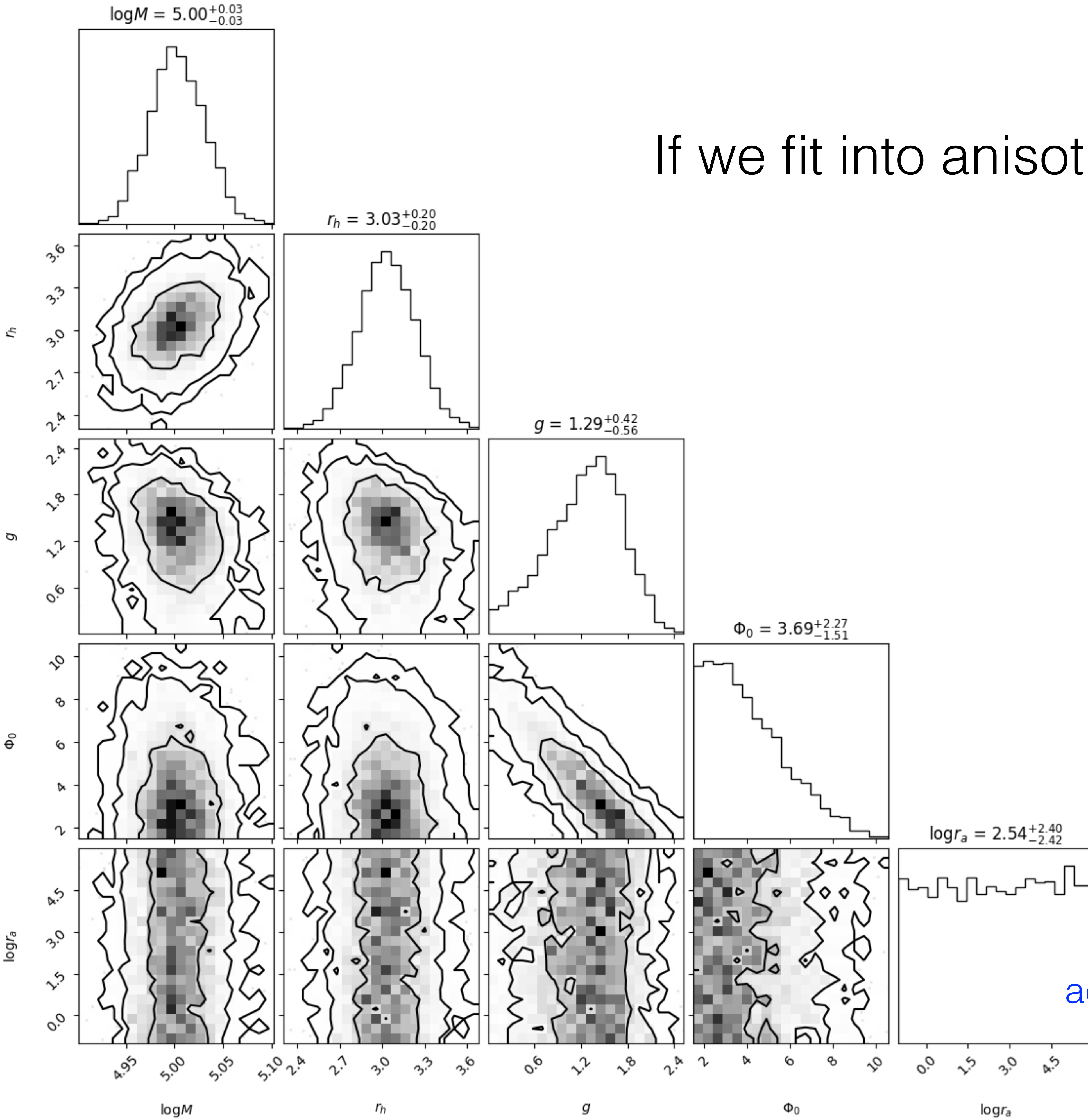
IS



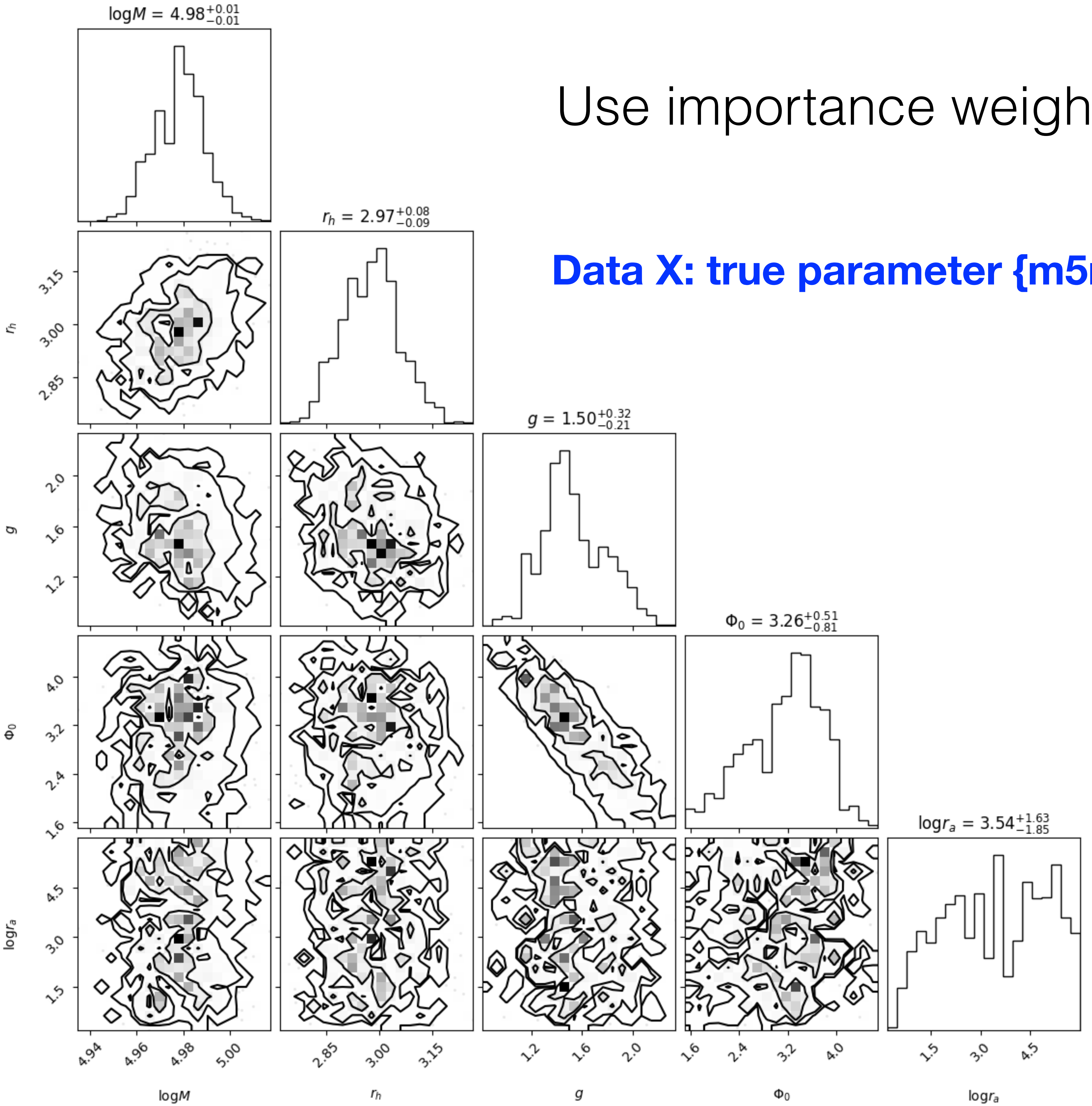
IS with weight



If we fit into anisotropic model



additional parameter r_a



Use importance weight to resample

Data X: true parameter {m5r3g1.5phi3.0}

what's the estimate for r_a?

#Effective sample size
("exchange rate"
btw dependent &independent samples)

$$n_{\text{eff}} \equiv \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} w_i^2}$$

Model 1: 305.5455 Total 2500 samples

Model 2: 113.3407 (proposal distribution not good enough)

Bayes Factor logK = $\frac{P(M1|X)}{P(M2|X)} = \frac{P(X|M1)}{P(X|M2)}\frac{P(M1)}{P(M2)}$ = 2.41

$\log_{10} K$	K	Strength of evidence
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

As expected, we prefer Model 1

Markov Chain Monte Carlo (MCMC)

Markov Chain: Time homogenous property

$$\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n)$$

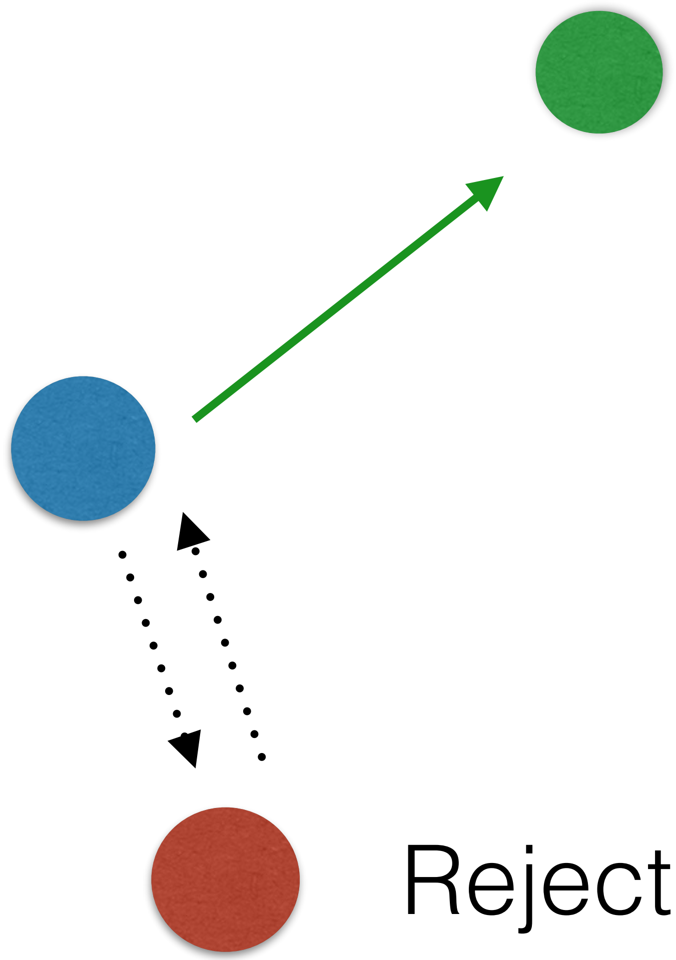
MCMC: By constructing a Markov Chain such that it will converge to its desired stationary distribution
Ex: Metropolis Hasting Algorithm

In Bayesian, the samples generated by MCMC can be used to evaluate the integral over high dimensional variable

Same Goal: Calculate the likelihood

1. Calculate difference of logs of likelihood*prior between current and new proposed points = y
2. Generate X~ Unif[0,1] independently

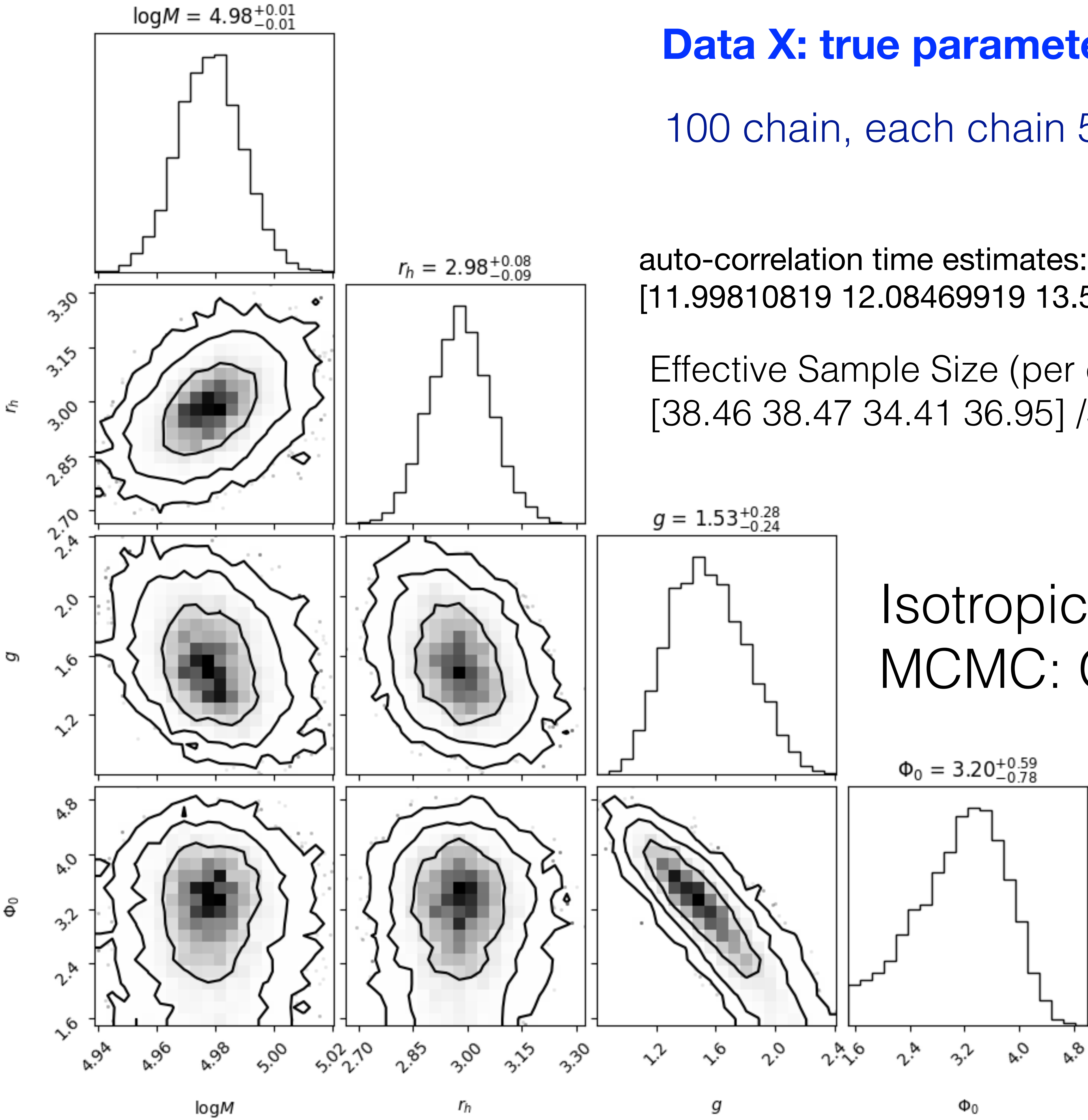
Detailed Options:
Thinning
Finite Adaptive MCMC



Accept
if exp(y) > x

Reject

if return 0 probability from prior
if return -Inf from the log likelihood
if exp(y) < x



Data X: true parameter {m5r3g1.5phi3.0}

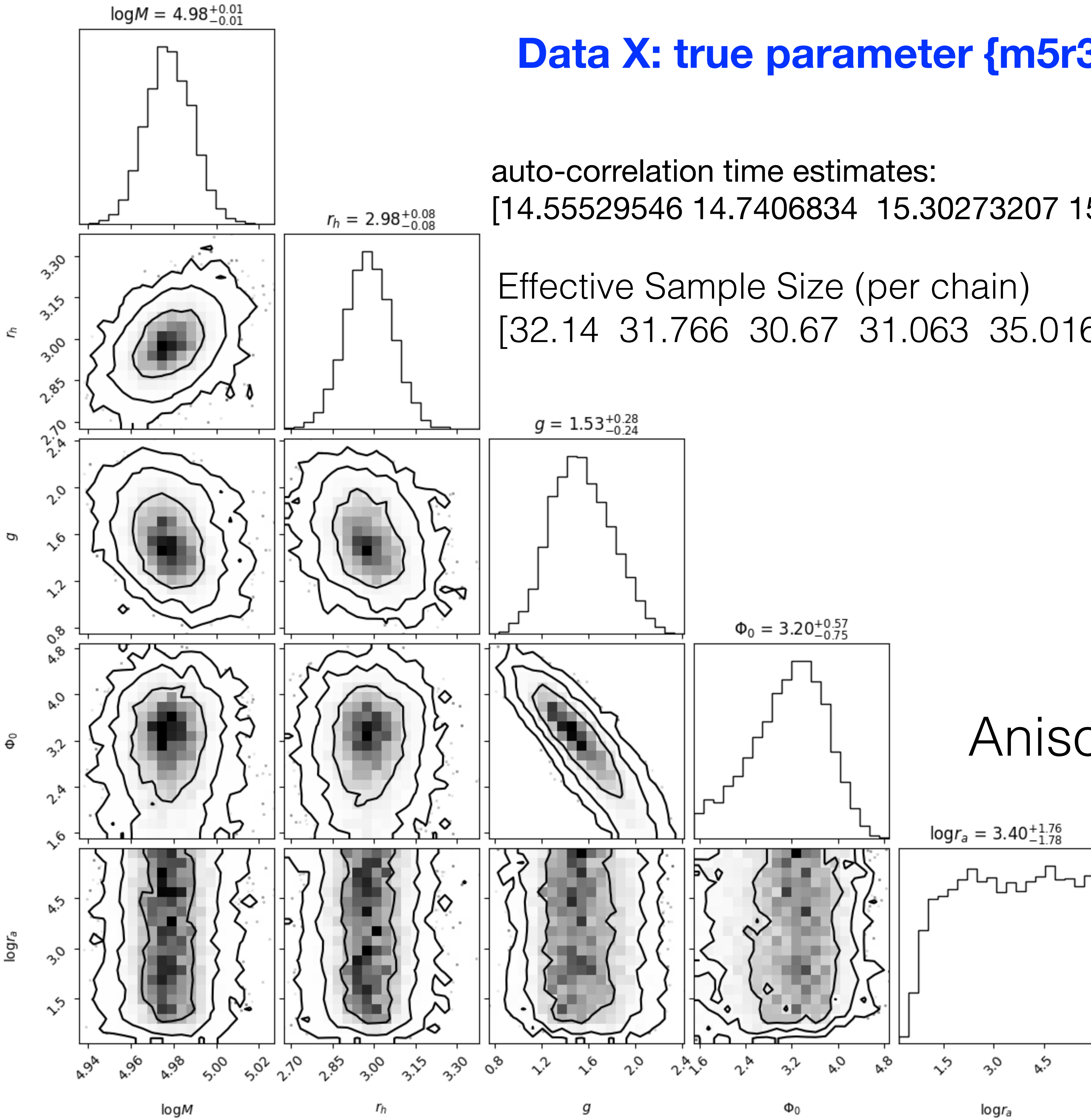
100 chain, each chain 500 samples

auto-correlation time estimates:
[11.99810819 12.08469919 13.53103676 12.53233813]

Effective Sample Size (per chain)
[38.46 38.47 34.41 36.95] /500 samples

Isotropic Model
MCMC: Good Estimate

Data X: true parameter {m5r3g1.5phi3.0}



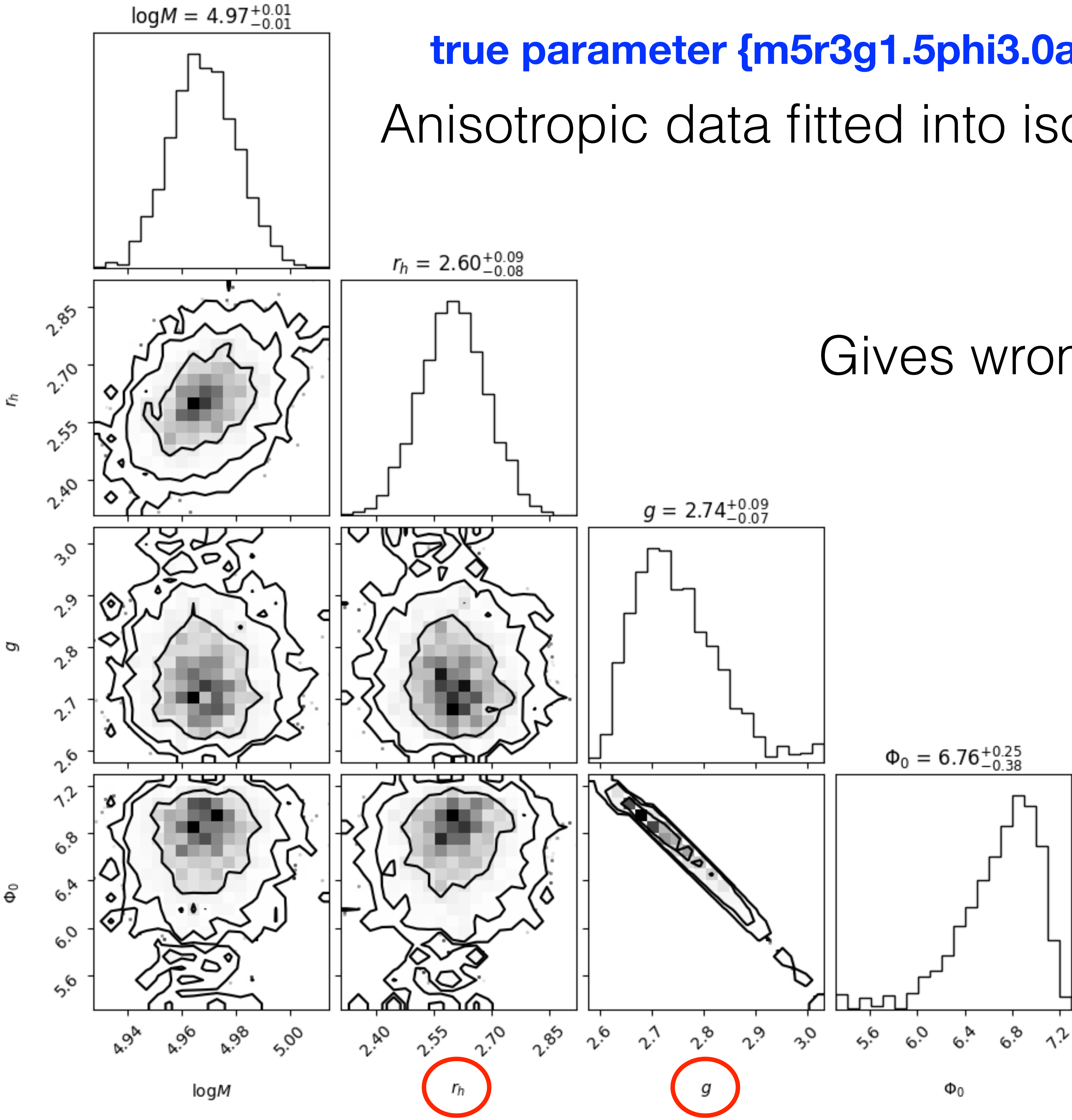
auto-correlation time estimates:
[14.55529546 14.7406834 15.30273207 15.09608124 13.27906024]

Effective Sample Size (per chain)
[32.14 31.766 30.67 31.063 35.016] /500 samples

Anisotropic Model

true parameter {m5r3g1.5phi3.0a0.8}

Anisotropic data fitted into isotropic model



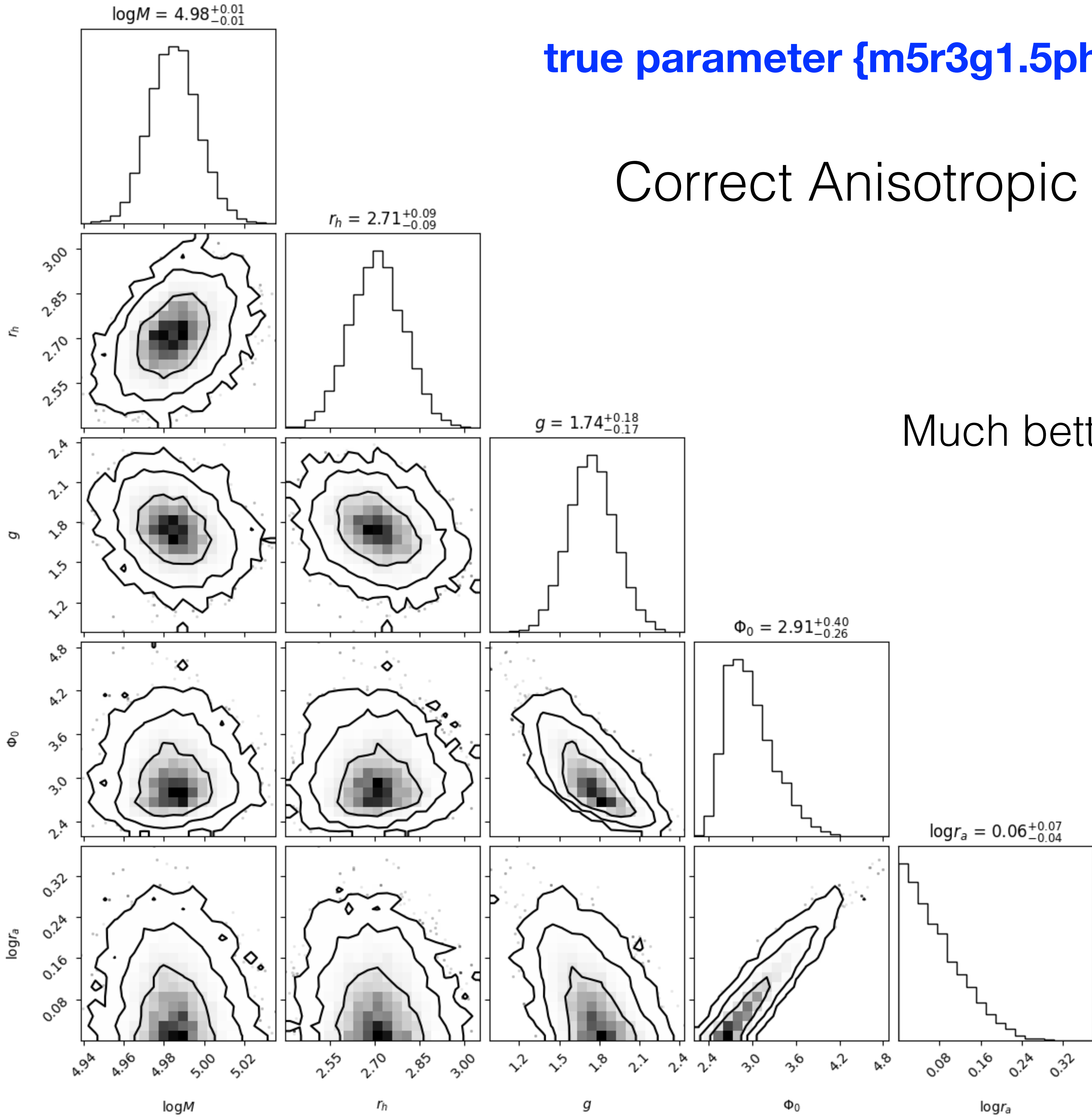
Gives wrong estimate

Missing r_a

true parameter {m5r3g1.5phi3.0a0.8}

Correct Anisotropic Model

Much better estimates



Next Step:

1. Try other sampling methods, such as Bridge Sampling, Nested Sampling
2. Consider the case when we don’t have complete dataset (Projection/ missing positions)
3. Error Measurements/ Uncertainty in the model