Learning priors, likelihoods, or posteriors

Iain Murray
School of Informatics, University of Edinburgh

"Within the field of approximate Bayesian inference, **variational** *and* **Monte Carlo** *methods are currently the mainstay techniques."*

— `http://approximateinference.org/`

The Statistician (1987) 36, pp. 247–249

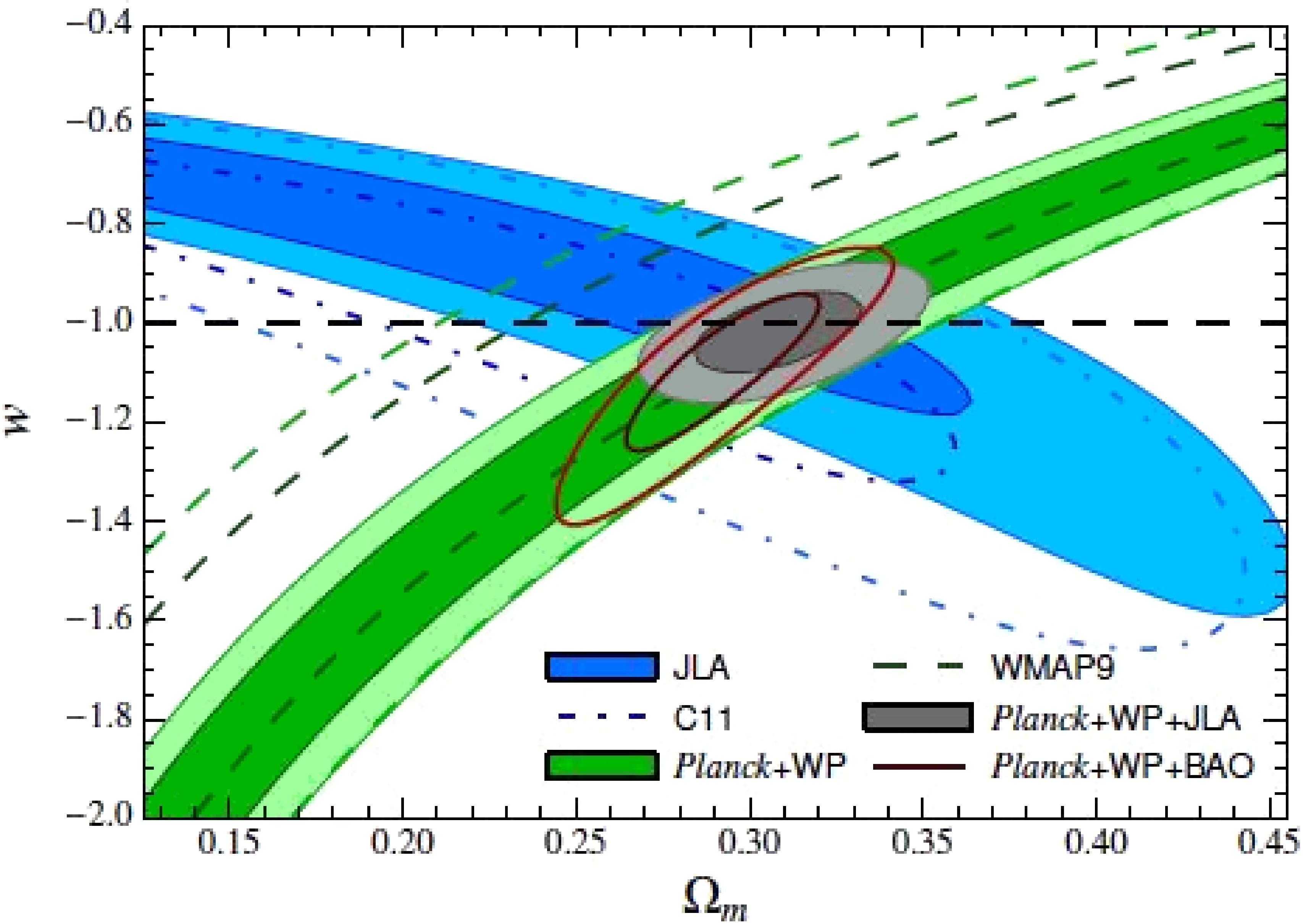
Monte Carlo is fundamentally unsound

A. O'HAGAN

Department of Statistics, University of Warwick, Coventry CV4 7AL, U.K.

Abstract. We present some fundamental objections to the Monte Carlo method of numerical integration.

Posteriors in Cosmology



https://bitbucket.org/joezuntz/cosmosis/

Recognition networks

$$\theta^{(s)} \sim p(\theta)$$

$$\mathbf{x}^{(s)} \sim p\left(\mathbf{x} \,|\, \theta^{(s)}\right)$$

Training set: $\left\{\theta^{(s)}, \mathbf{x}^{(s)}\right\}_{s=1}^{S}$

Some of the relevant work

Hinton et al. (1995, Science) — Wake Sleep, Helmholtz machine

Morris (2001, UAI) — Recognition Networks

Blum & Francois (2010, S&C) — Conditional Gaussian, neural nets

Fan, Nott, Sisson (2012, Stat) — Mixture of experts

Mitrović, Dino Sejdinović, Teh (2016, ICML) — Kernel regression

...

Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation

Papamakarios and Murray (NIPS, 2016)

Lueckmann et al. (NIPS, 2017)

— Fit $\hat{p}(\theta \,|\, \mathbf{x})$ maximize $\sum_s \log \hat{p}(\theta^{(s)} \,|\, \mathbf{x}^{(s)})$

Mixture Density Networks (Bishop, 1994)

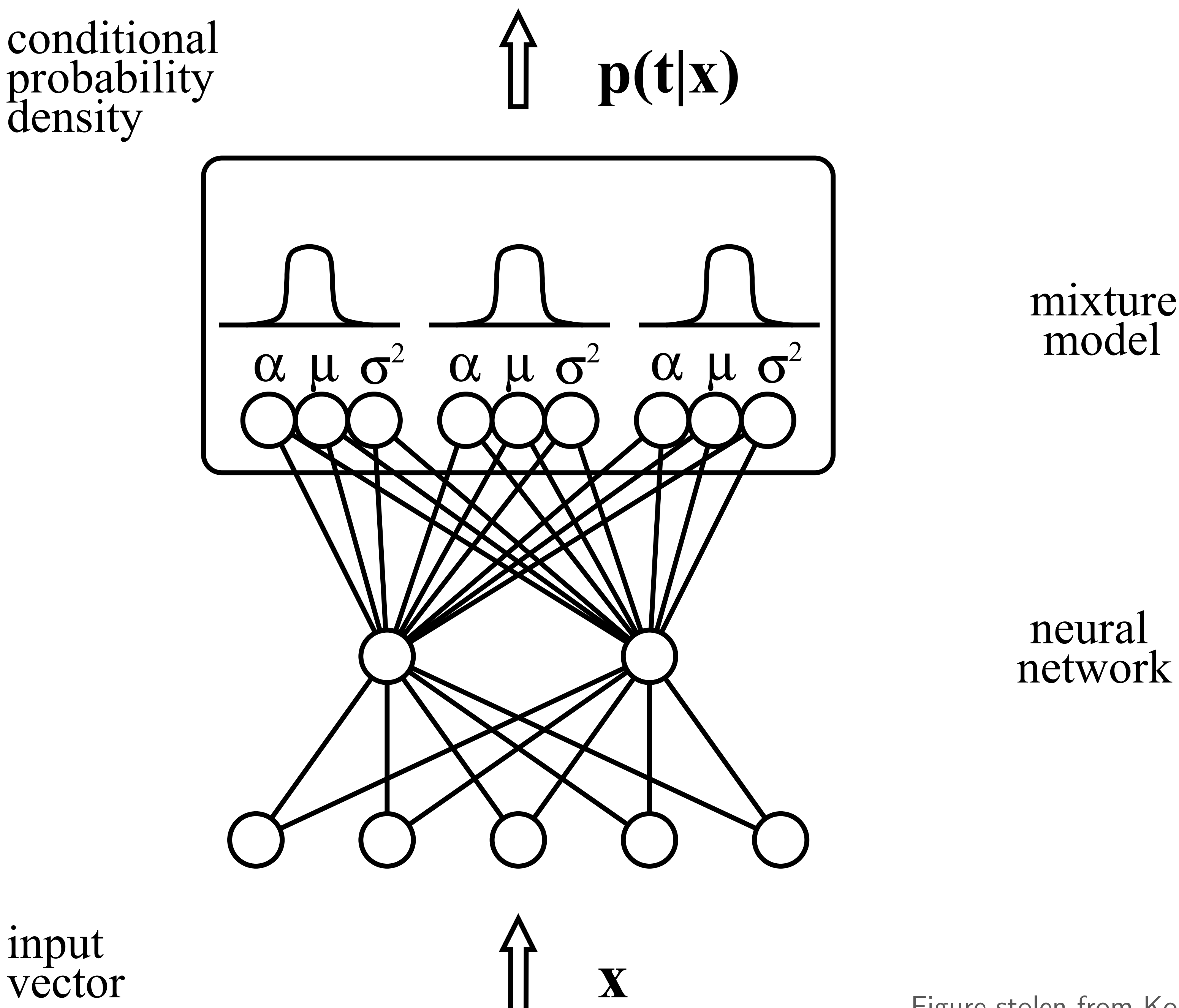


Figure stolen from Korin Richmond

Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation

Papamakarios and Murray (NIPS, 2016)
Lueckmann et al. (NIPS, 2017)

— Fit $\hat{p}(\theta \mid \mathbf{x})$ maximize $\sum_s \log \hat{p}(\theta^{(s)} \mid \mathbf{x}^{(s)})$

— $\hat{p}(\theta \mid \mathbf{x}_{\text{observed}}) \rightarrow$ approx posterior

Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation

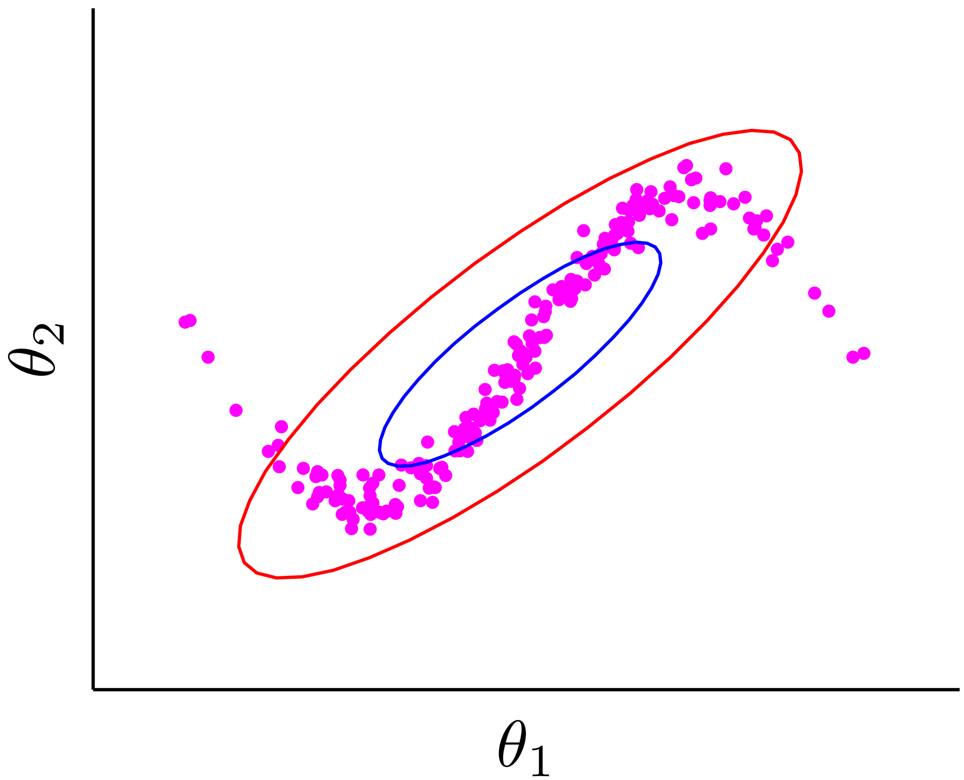
Papamakarios and Murray (NIPS, 2016)
Lueckmann et al. (NIPS, 2017)

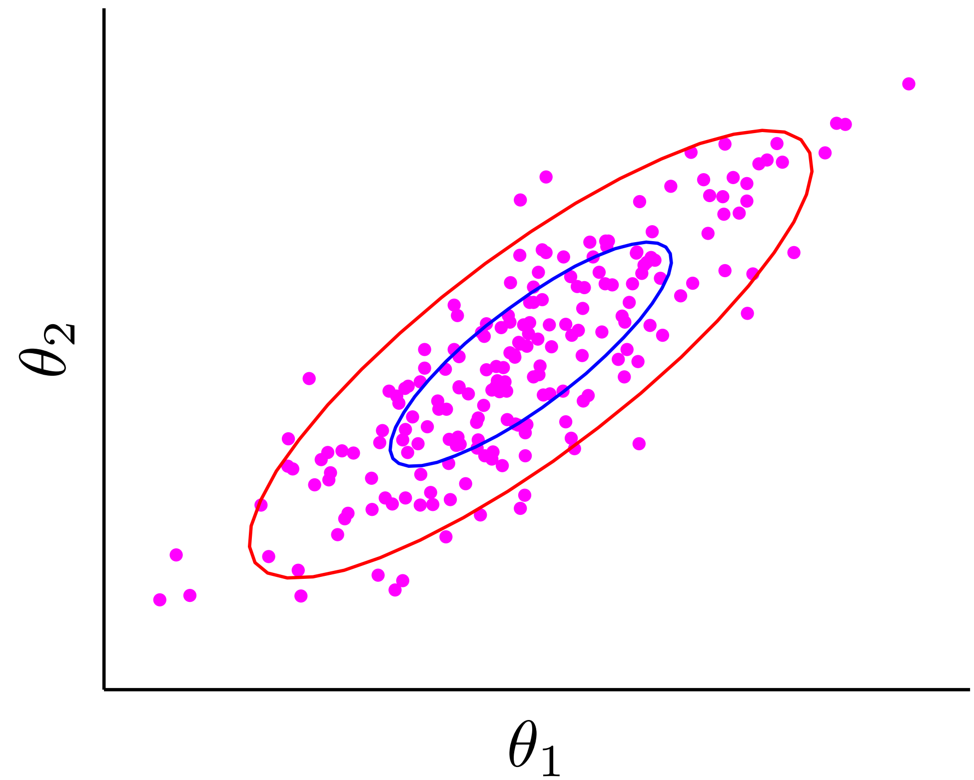
— Fit $\hat{p}(\theta \,|\, \mathbf{x})$ maximize $\sum_s \log \hat{p}(\theta^{(s)} \,|\, \mathbf{x}^{(s)})$

— $\hat{p}(\theta \,|\, \mathbf{x}_{\text{observed}}) \rightarrow$ approx posterior

— Refine fit: more simulations

Underfitting





True posterior samples

samples from Gaussian fit

— Modeling posteriors

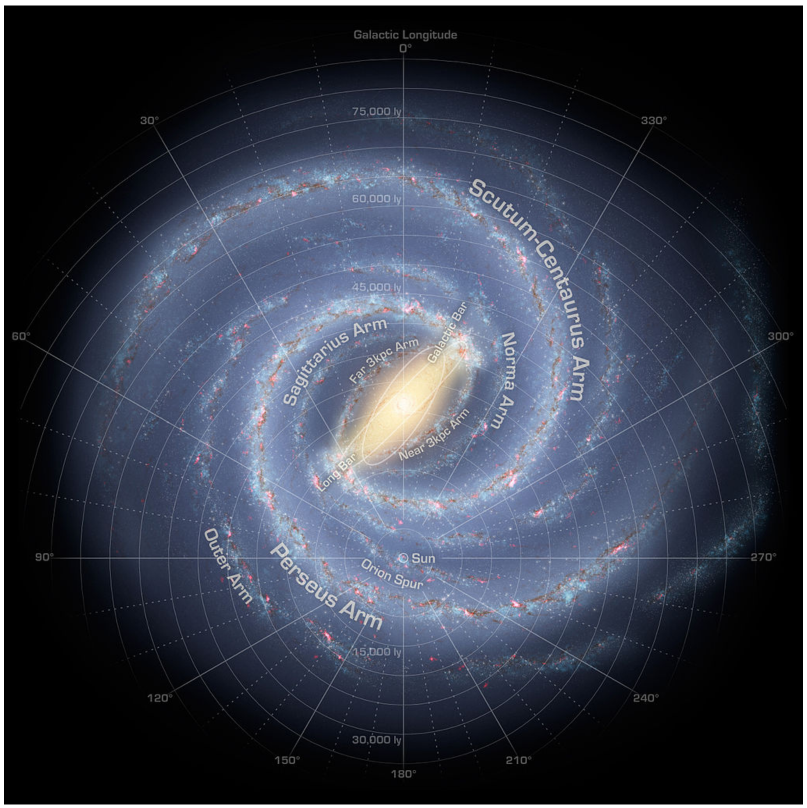
— **Modeling priors**

— Modeling likelihoods

Weighing the Milky Way

Busha, Marshall, Wechsler, Klypin and Primack (2011)

APJ 743:40



Milky Way diagram, NASA

Magellanic Clouds, ESO/S. Brunier

http://en.wikipedia.org/wiki/File:236084main_MilkyWay-full-annotated.jpg
http://www.eso.org/public/images/b01/

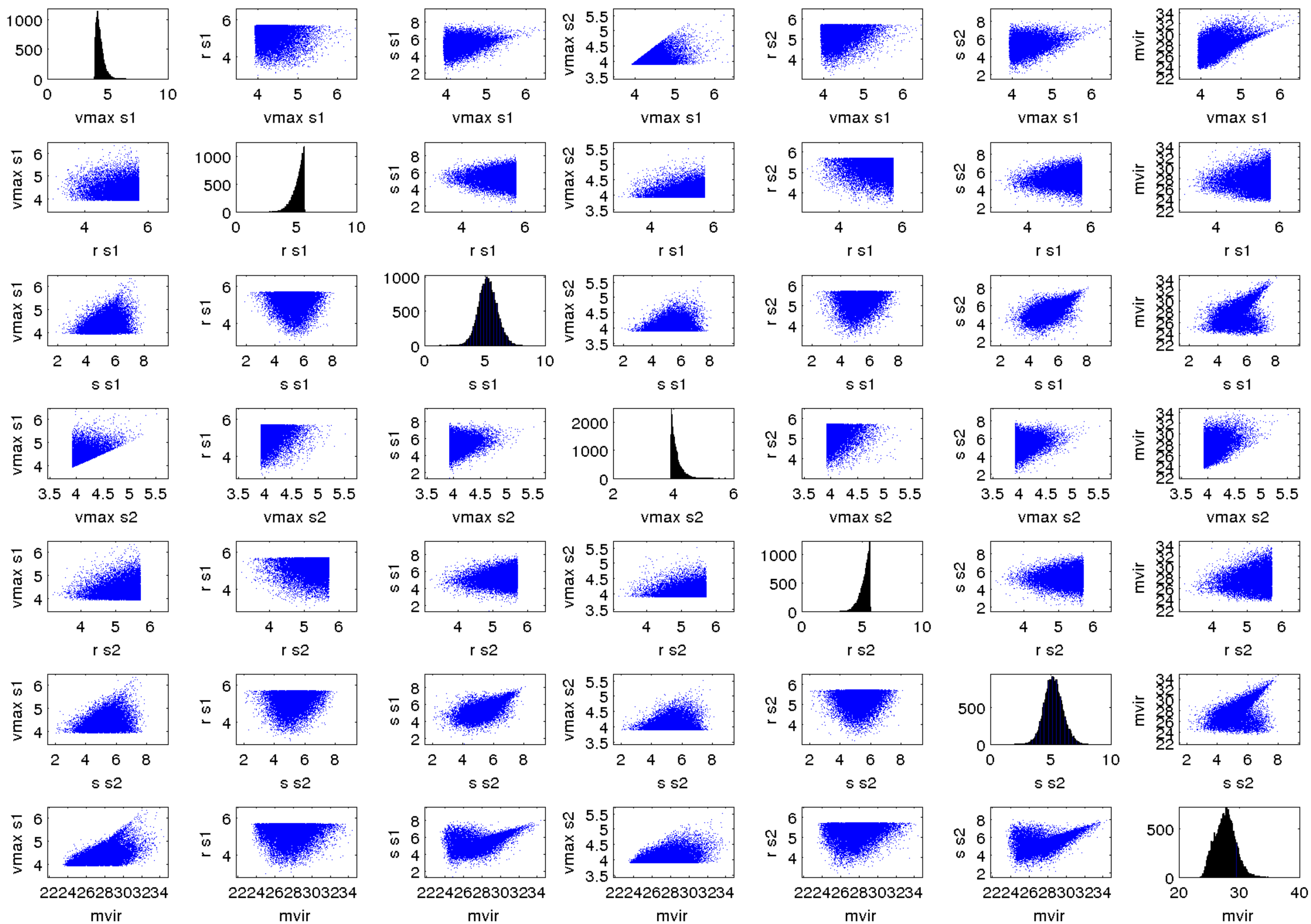
Bayesian Inference

$$p(\mathbf{x} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{x})$$

$\mathbf{x} = [r, v, m]$, vector of galaxy properties

$\mathbf{y} = [\hat{r}, \hat{v}]$, noisily observe part of $\mathbf{x}$

The prior: simulation samples

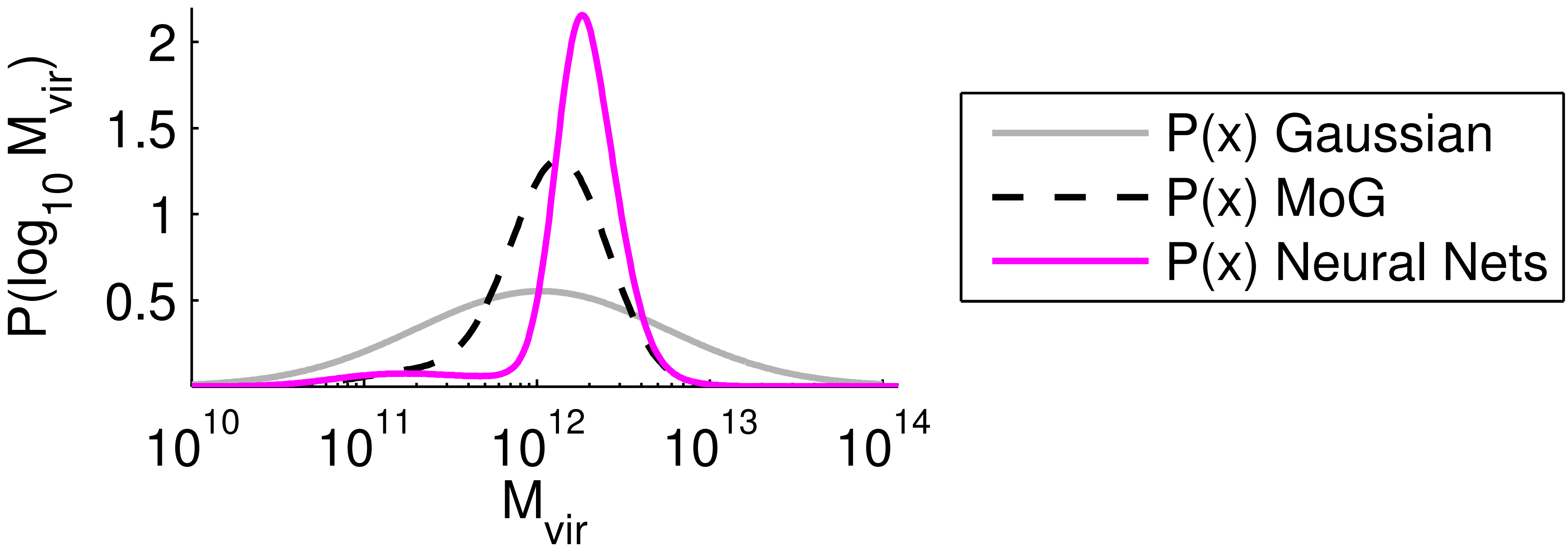


Milky Way mass

$p(\mathbf{x})$ theory: simulated galaxy properties

$p(\mathbf{y} \mid \mathbf{x})$ observations of Milky Way

$p(\mathbf{x} \mid \mathbf{y}) \rightarrow p(x_1 \mid \mathbf{y})$, posterior of mass

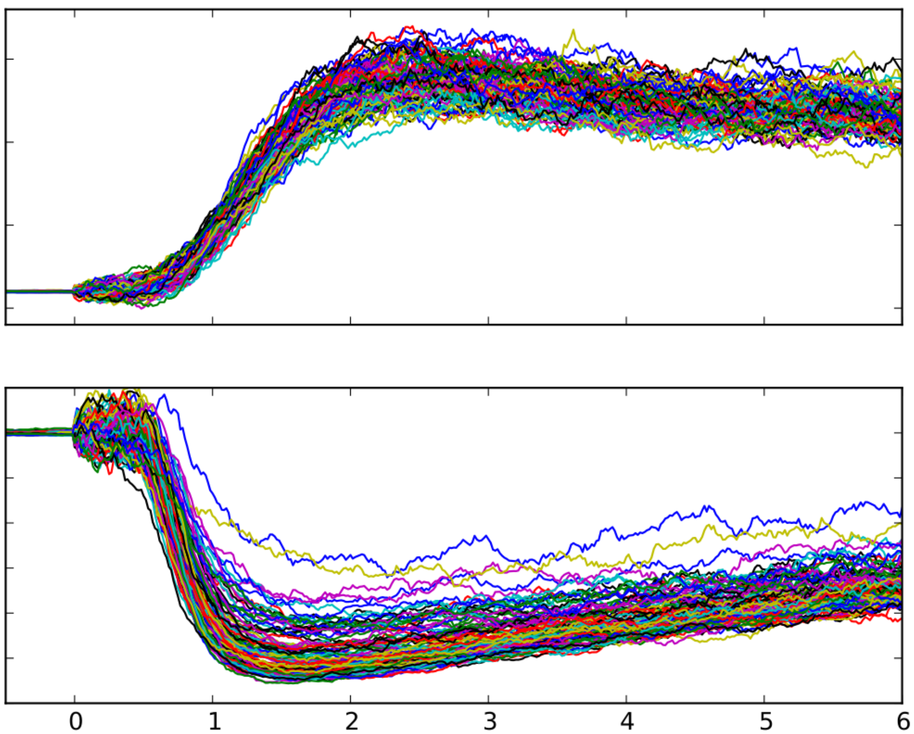


— Modeling posteriors

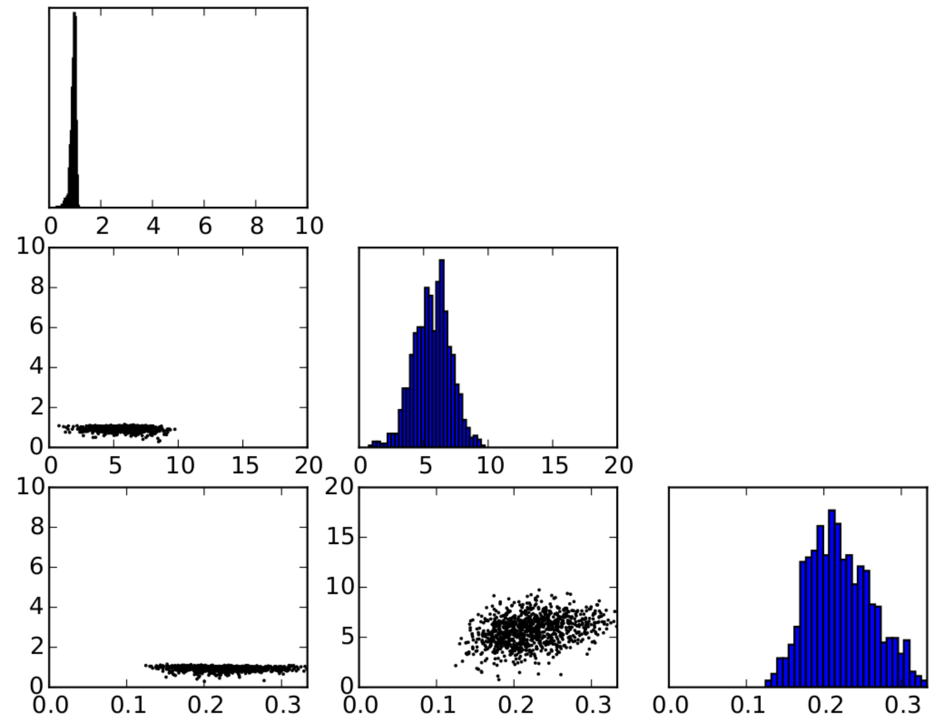
— Modeling priors

— **Modeling likelihoods**

Surrogate modeling / emulation



data $\mathbf{x}$



parameters $\boldsymbol{\theta}$

$$p(\theta \,|\, \mathcal{D}) \;\propto\; p(\theta) \prod_n p(\mathbf{x}^{(n)} \,|\, \theta)$$

Cf Cranmer, Pavez, Louppe (2016) arXiv:1506.02169

Thanks!

`http://iainmurray.net`

NADE variants, MADE, and MAF
ϵ-free ABC, pseudo-marginal slice sampling