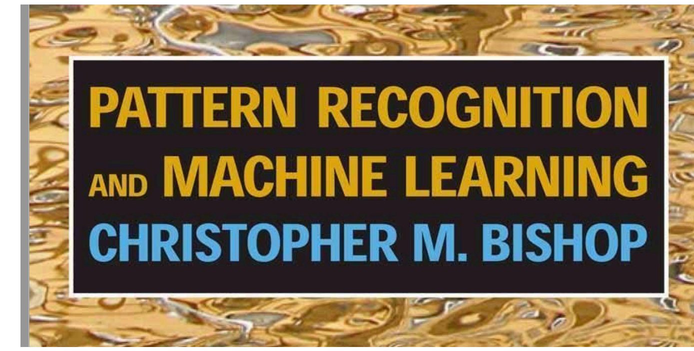
Bayesian Model Comparison

SBI workshop

Jan Boelts, September 22, 2021

Bayesian model comparison vs Inference

Definition from Bishop 2006



Bayesian inference: posterior over parameters

model M_i is fixed, infer parameters: $p(w \mid D) \propto p(w)\boxed{p(D \mid w)}$

likelihood

Model comparison: posterior over models

$$p(\mathcal{M}_i|\mathcal{D}) \propto p(\mathcal{M}_i)\boxed{p(\mathcal{D}|\mathcal{M}_i)}. \qquad (3.66)$$

Model Evidence aka **Marginal** likelihood

Bayesian model comparison
marginalises over model parameters

Model comparison: posterior over models

$$p(\mathcal{M}_i|\mathcal{D}) \propto p(\mathcal{M}_i)\boxed{p(\mathcal{D}|\mathcal{M}_i)}. \tag{3.66}$$

Model Evidence aka **Marginal** likelihood

Bayesian model comparison

marginalises over model parameters

Model comparison: posterior over models

$$p(\mathcal{M}_i|\mathcal{D}) \propto p(\mathcal{M}_i)\boxed{p(\mathcal{D}|\mathcal{M}_i)}. \tag{3.66}$$

Model Evidence aka **Marginal** likelihood

Model evidence aka Marginal likelihood

$$p(\mathcal{D}|\mathcal{M}_i) = \int \boxed{p(\mathcal{D}|\mathbf{w}, \mathcal{M}_i)}\,\boxed{p(\mathbf{w}|\mathcal{M}_i)}\,\mathrm{d}\mathbf{w}. \tag{3.68}$$

Bayesian model comparison

marginalises over model parameters

Model comparison: posterior over models

$$p(\mathcal{M}_i|\mathcal{D}) \propto p(\mathcal{M}_i)\boxed{p(\mathcal{D}|\mathcal{M}_i)}. \qquad (3.66)$$

Model Evidence aka **Marginal** likelihood

Model evidence aka Marginal likelihood

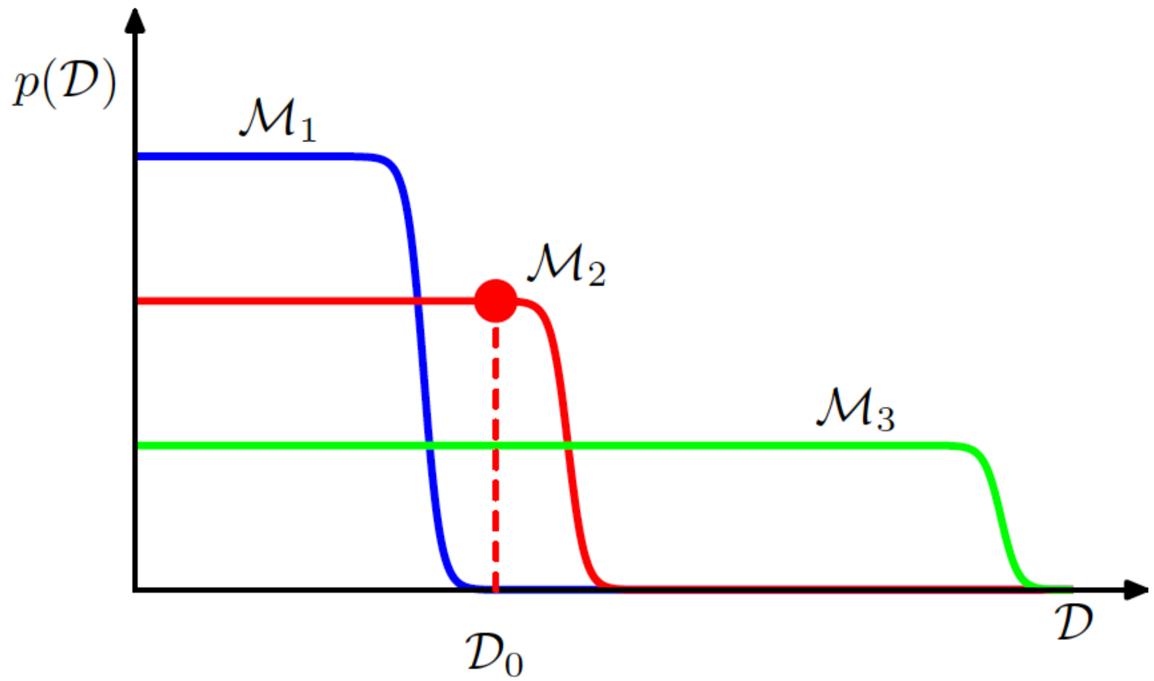
$$p(\mathcal{D}|\mathcal{M}_i) = \int \boxed{p(\mathcal{D}|\mathbf{w}, \mathcal{M}_i)}\boxed{p(\mathbf{w}|\mathcal{M}_i)}\,\mathrm{d}\mathbf{w}. \qquad (3.68)$$

likelihood of data given w

prior of w under M_i

Bayesian model comparison
penalises too complex models

Figure 3.13 Schematic illustration of the distribution of data sets for three models of different complexity, in which $\mathcal{M}_1$ is the simplest and $\mathcal{M}_3$ is the most complex. Note that the distributions are normalized. In this example, for the particular observed data set $\mathcal{D}_0$, the model $\mathcal{M}_2$ with intermediate complexity has the largest evidence.



Bishop, 2006

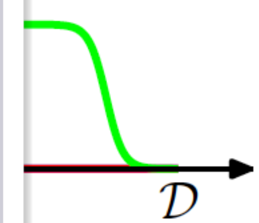
the simple model cannot fit the data

the complex model spreads its
predictive power over too many parameters

Automatic Occam's razor

penalises to

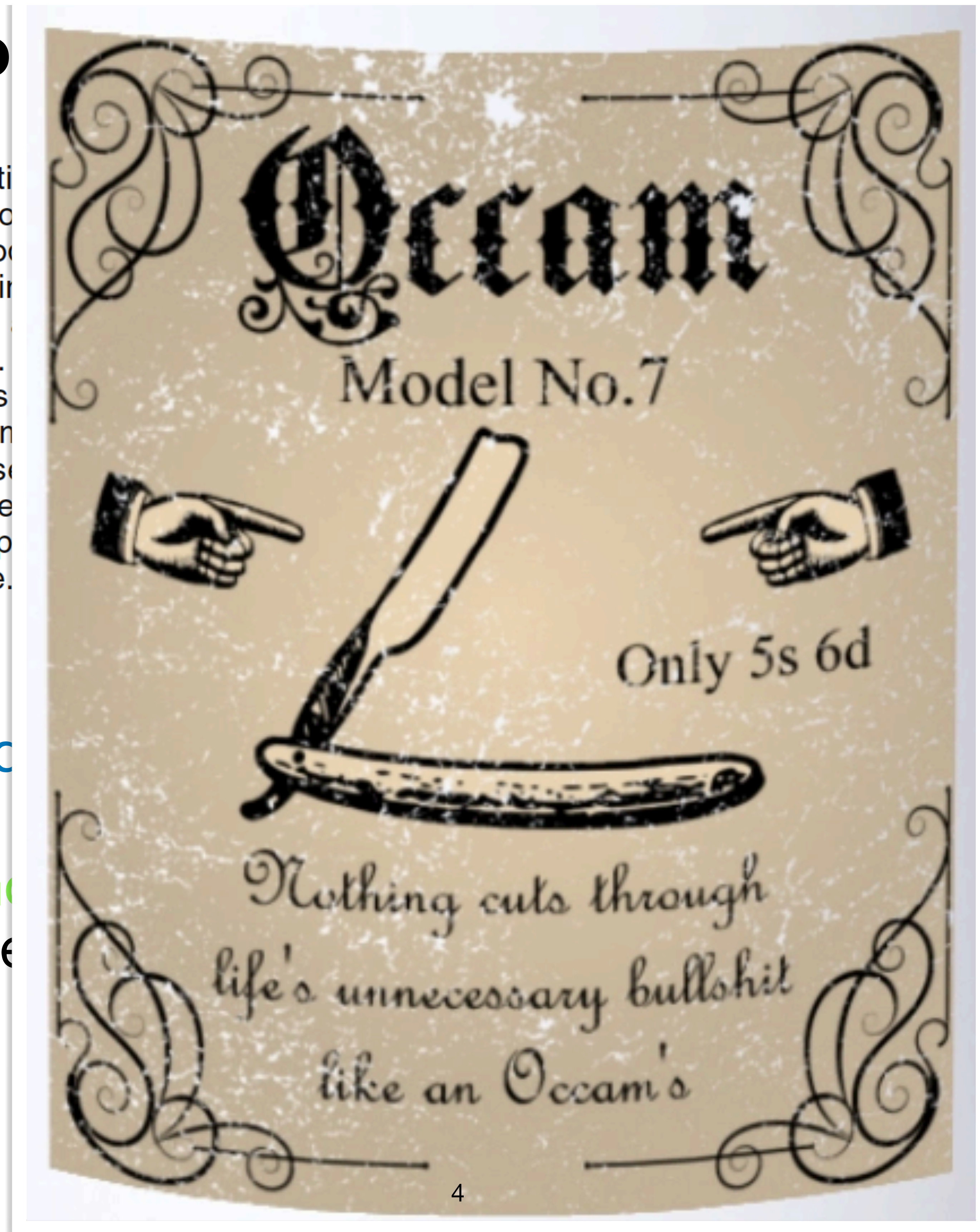
Figure 3.13 Schemati distributio three mo plexity, i simplest complex. tributions this exam ular obs the mode ate comp evidence.



Bishop, 2006

the simple mod

the complex m
predictive powe



spreadshirt.ch

Bayes factors

Bayesian hypothesis testing

$$K = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} = \frac{\Pr(M_1|D)}{\Pr(M_2|D)} \frac{\Pr(M_2)}{\Pr(M_1)}.$$

posterior odds times prior odds

https://en.wikipedia.org/wiki/Bayes_factor

Bayes factors

Bayesian hypothesis testing

$$K = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} = \frac{\Pr(M_1|D)}{\Pr(M_2|D)} \frac{\Pr(M_2)}{\Pr(M_1)}.$$

K	dHart	bits	Strength of evidence
< 10^0	< 0	< 0	Negative (supports M_2)
10^0 to $10^{1/2}$	0 to 5	0 to 1.6	Barely worth mentioning
$10^{1/2}$ to 10^1	5 to 10	1.6 to 3.3	Substantial
10^1 to $10^{3/2}$	10 to 15	3.3 to 5.0	Strong
$10^{3/2}$ to 10^2	15 to 20	5.0 to 6.6	Very strong
> 10^2	> 20	> 6.6	Decisive

https://en.wikipedia.org/wiki/Bayes_factor

BMC in practice

difficult because we don't have likelihoods

$$p(\mathcal{D}|\mathcal{M}_i) = \int \cancel{p(\mathcal{D}|\mathbf{w}, \mathcal{M}_i)} p(\mathbf{w}|\mathcal{M}_i) \, \mathrm{d}\mathbf{w}. \tag{3.68}$$

BMC in practice

difficult because we don't have likelihoods

$$p(\mathcal{D}|\mathcal{M}_i) = \int \cancel{p(\mathcal{D}|\mathbf{w}, \mathcal{M}_i)} p(\mathbf{w}|\mathcal{M}_i) \, \mathrm{d}\mathbf{w}. \tag{3.68}$$

different approaches:

- first approach: via ABC
- more recenlty: conditional density estimation / SBI

BMC with ABC

BMC with ABC

ABC approach:

- in plain rejection ABC or SMC-ABC

- add prior over models to ABC algorithm

- count how often $d(x, x_o) < \epsilon$ for each model

- estimate model posterior from frequencies

BMC with ABC

ABC approach:

- in plain rejection ABC or SMC-ABC

- add prior over models to ABC algorithm

- count how often $d(x, x_o) < \epsilon$ for each model

- estimate model posterior from frequencies

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, *6*(31), 187-202.

BMC with ABC

ABC approach:

- in plain rejection ABC or SMC-ABC
- add prior over models to ABC algorithm
- count how often $d(x, x_o) < \epsilon$ for each model
- estimate model posterior from frequencies

Algorithm:

sample model	$m_i \sim p(m)$
sample parameter	$w_i \sim p(w \mid m_i)$
simulate	$x_i \sim f(w_i)$
compare	$d(x_i, x_o) < \epsilon\,?$

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, *6*(31), 187-202.

BMC with SBI
ongoing research

idea: conditional density estimation with NNs

BMC with SBI

ongoing research

idea: conditional density estimation with NNs

but the posterior over models is discrete!

BMC with SBI

ongoing research

idea: conditional density estimation with NNs

but the posterior over models is discrete!

1) Conditional **mass** estimation for model-posterior

BMC with SBI

ongoing research

idea: conditional density estimation with NNs

but the posterior over models is discrete!

1) Conditional **mass** estimation for model-posterior

2) Density ratio estimation

BMC with SBI

ongoing research

idea: conditional density estimation with NNs

but the posterior over models is discrete!

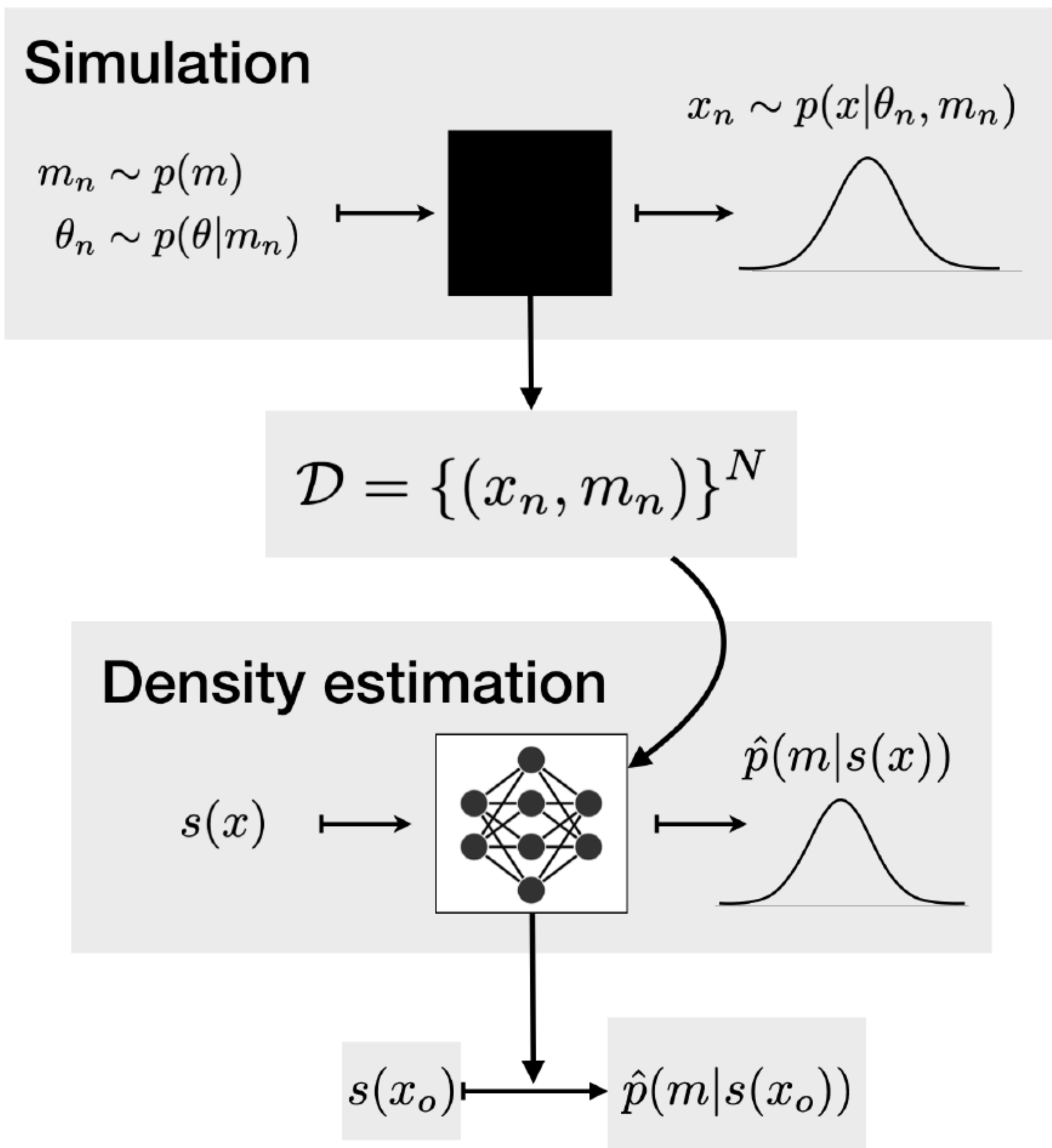
1) Conditional **mass** estimation for model-posterior

2) Density ratio estimation

3) Evidential deep learning for BMC

1) CDE for model-posterior

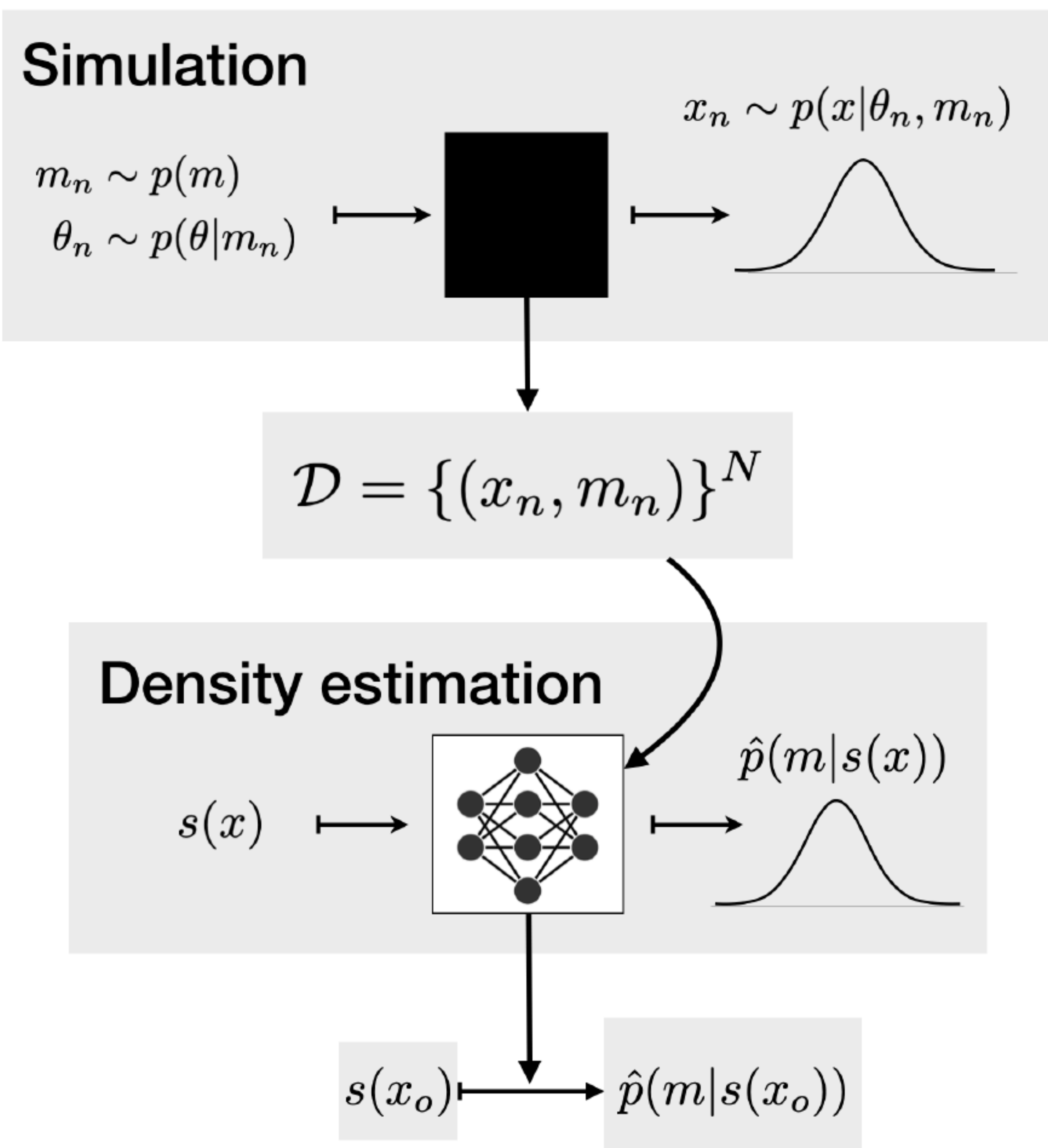
multi-class classification over models

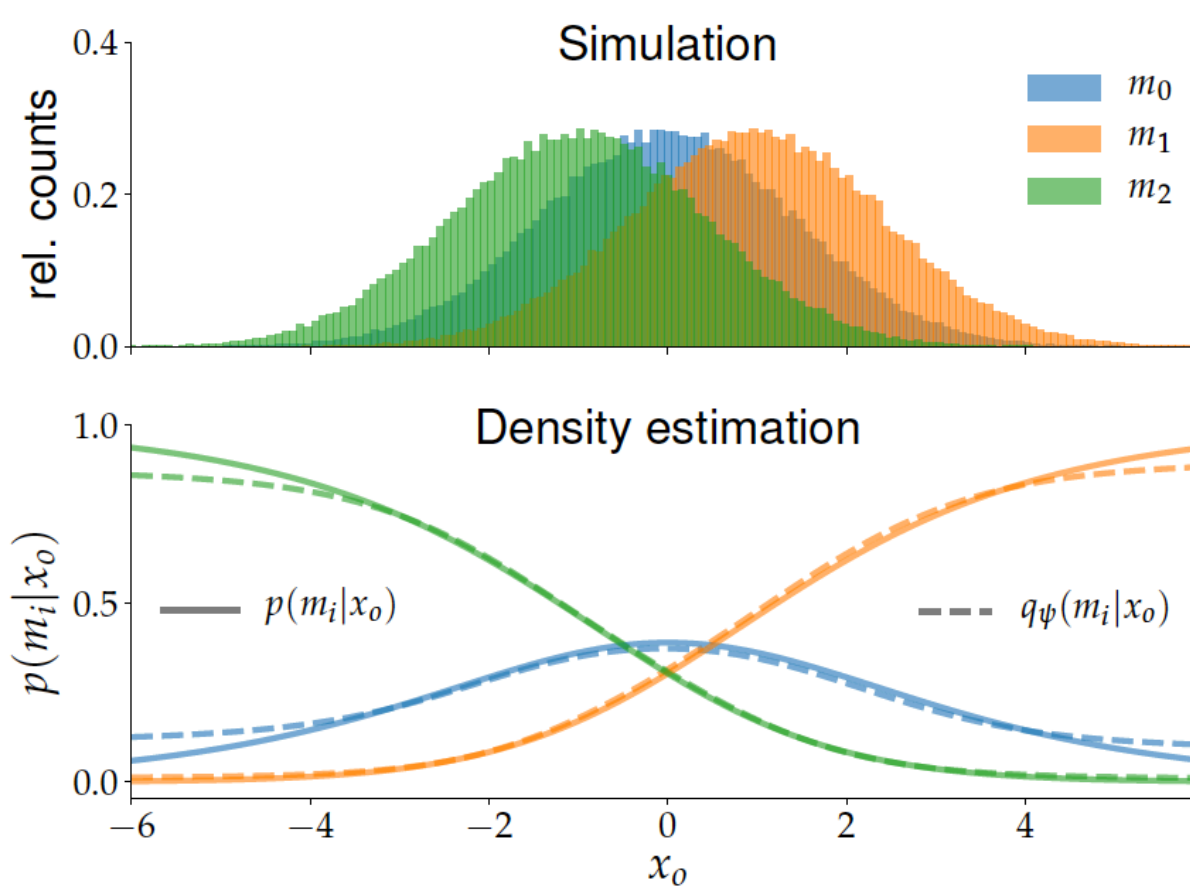


Boelts et al. 2019, Comparing Neural Simulations using Neural Density estimation

1) CDE for model-posterior

multi-class classification over models

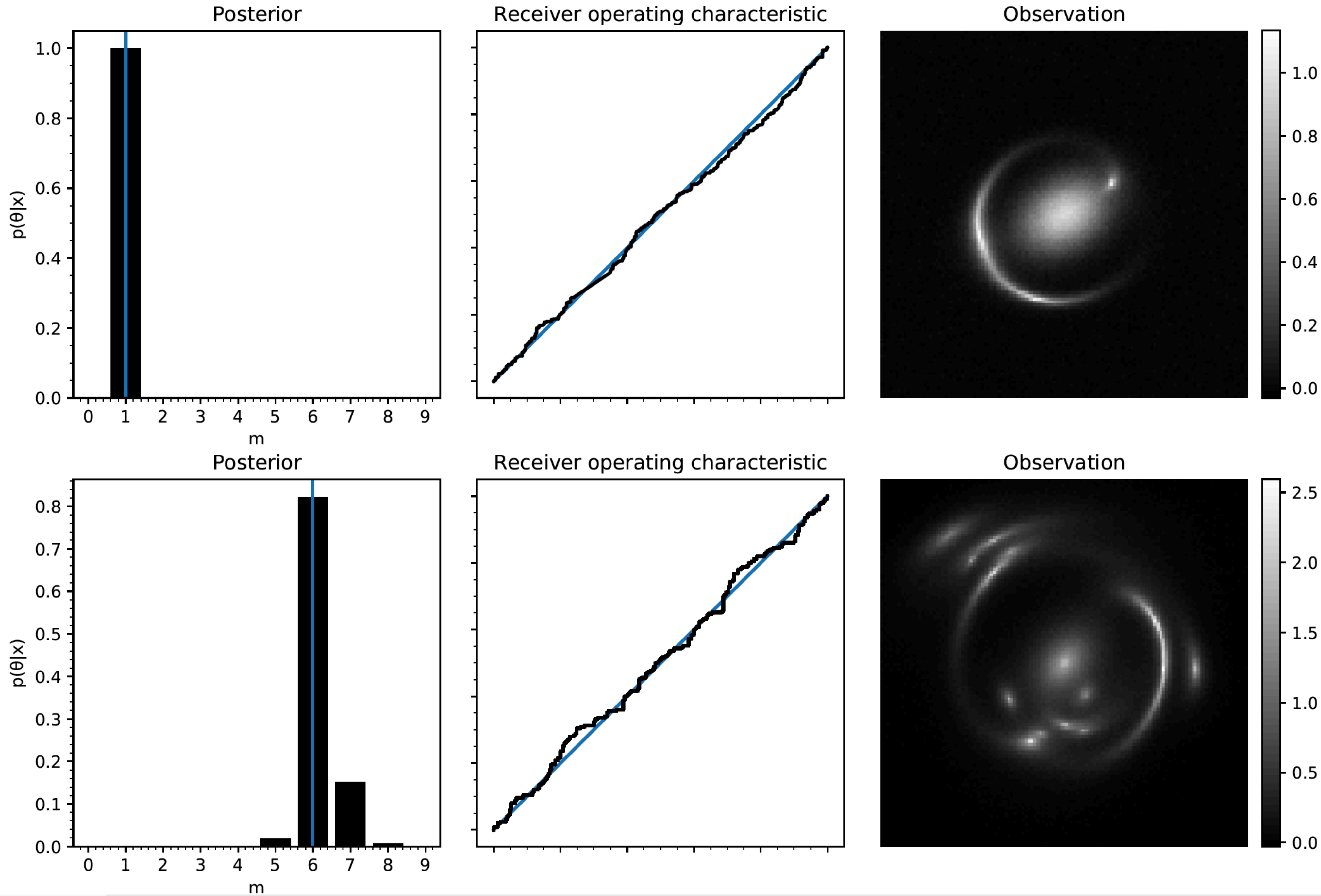




Boelts et al. 2019, Comparing Neural Simulations using Neural Density estimation

2) Neural Ratio Estimation

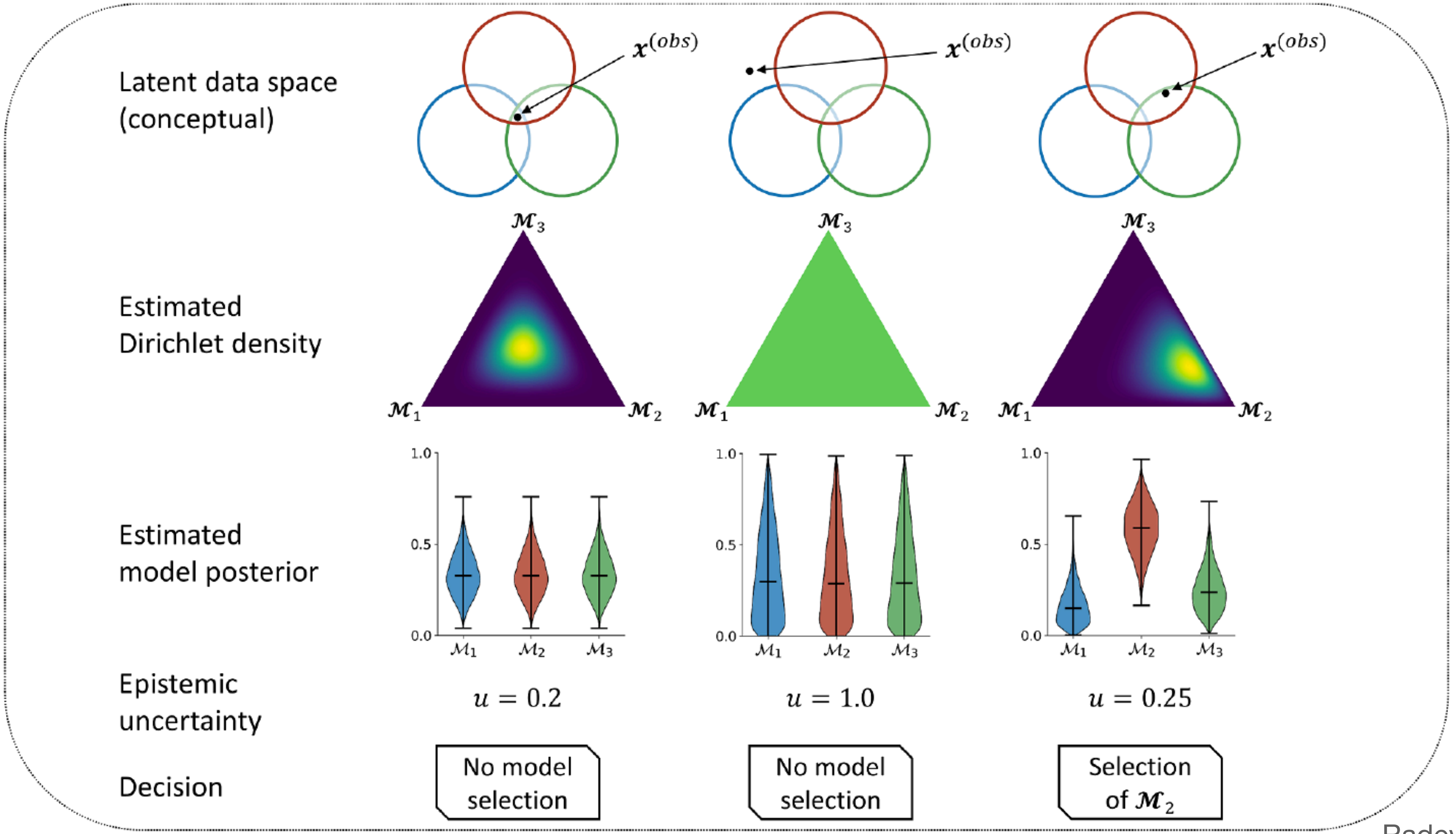
ratio estimator on discrete model prior + MCMC



Hermans et al. 2020, “Likelihood-free MCMC with Amortized Approximate Ratio Estimators”

3) Evidential neural networks

quantify classification uncertainty



Radev et al. 2020

Sensoy et a. 2018, "Evidential deep learning to quantify classification uncertainty."

Radev et al. 2020, "Amortized Bayesian Model Comparison with Evidential Deep Learning"

Discussion