LSB: Local Self-Balancing MCMC in Discrete Spaces

Emanuele Sansone

KU LEUVEN

LSB: Local Self-Balancing MCMC in Discrete Spaces

Emanuele Sansone

KU LEUVEN

Discrete Distributions

$$p(x) = \frac{\tilde{p}(x)}{P} \qquad x \in \{0,1\}^d$$

Discrete Distributions

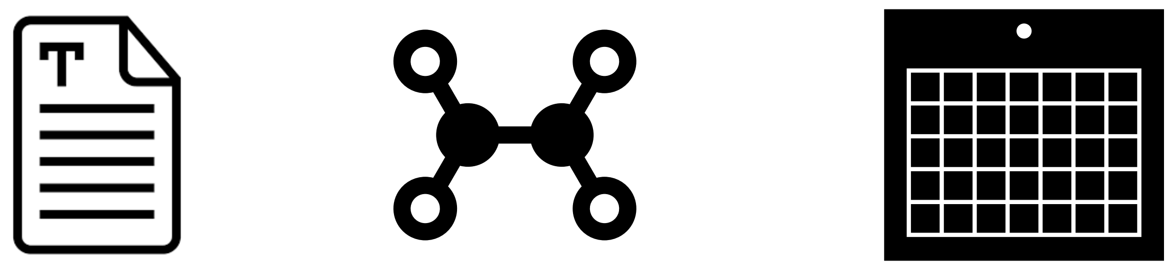
$$p(x) = \frac{\tilde{p}(x)}{P} \qquad x \in \{0,1\}^d$$

Models

Energy-based models $\tilde{p}(x) = e^{-E(x)}$

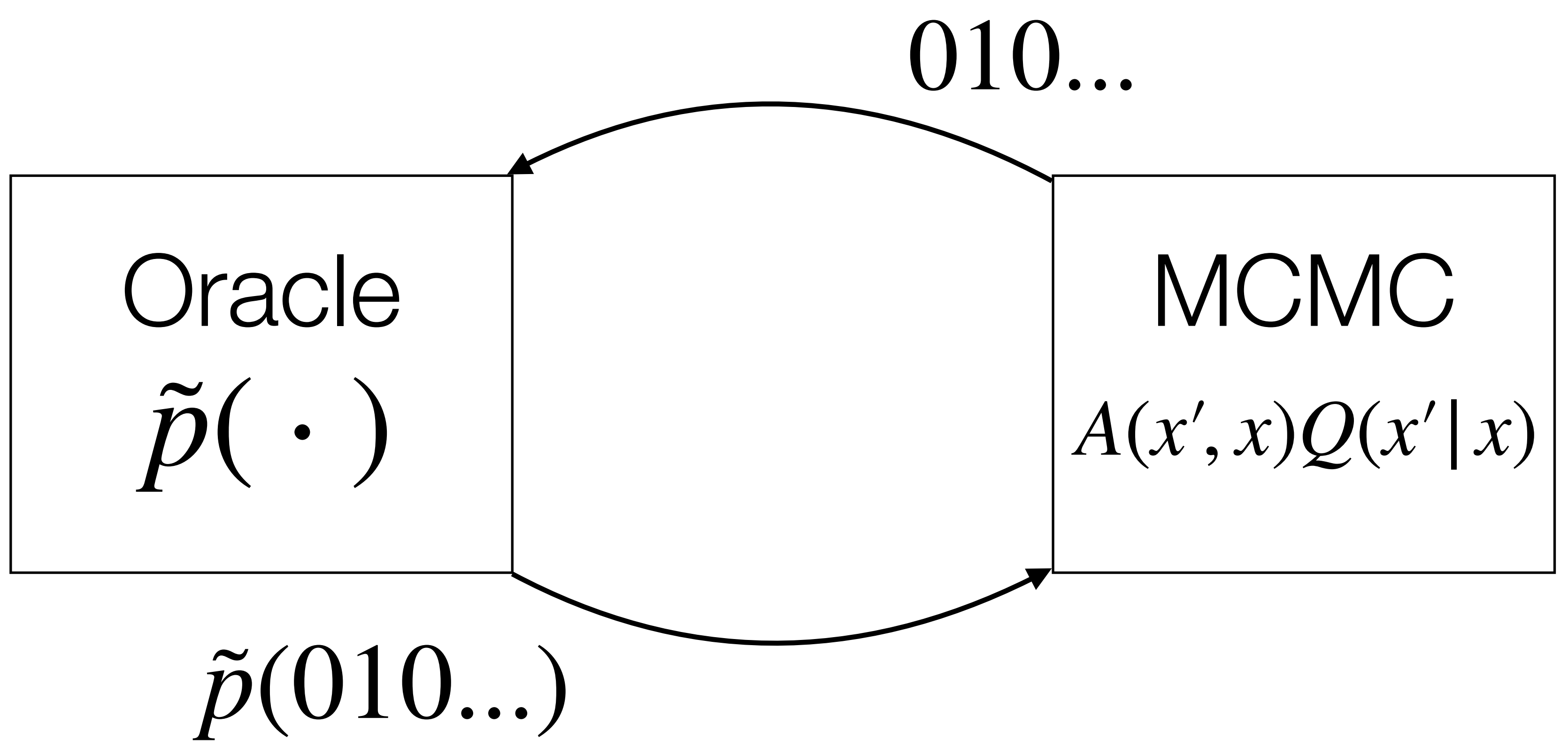
Probabilistic graphical models $\tilde{p}(x) = \prod_k \phi(x_{\{k\}})$

Data



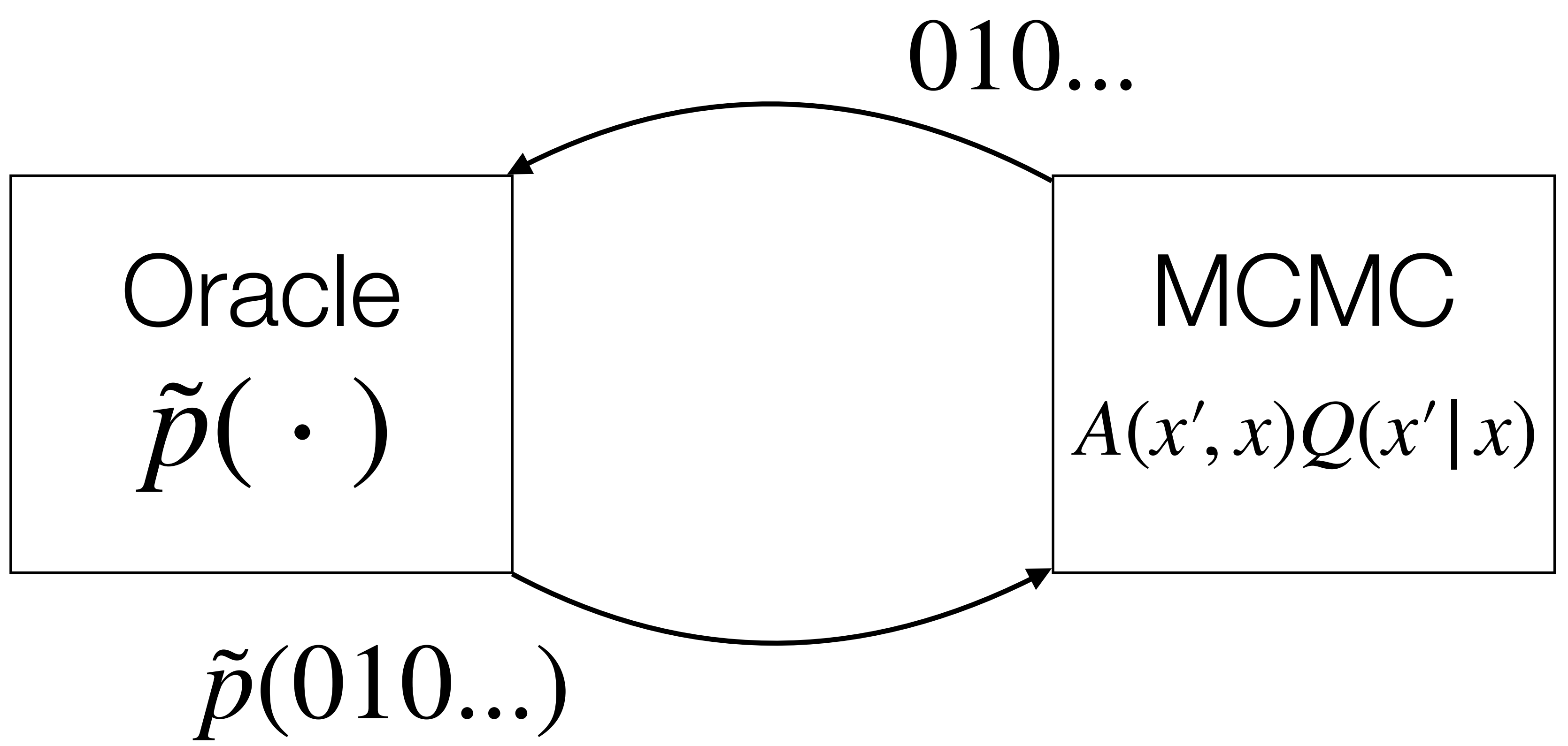
Markov Chain Monte Carlo - MCMC

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Markov Chain Monte Carlo - MCMC

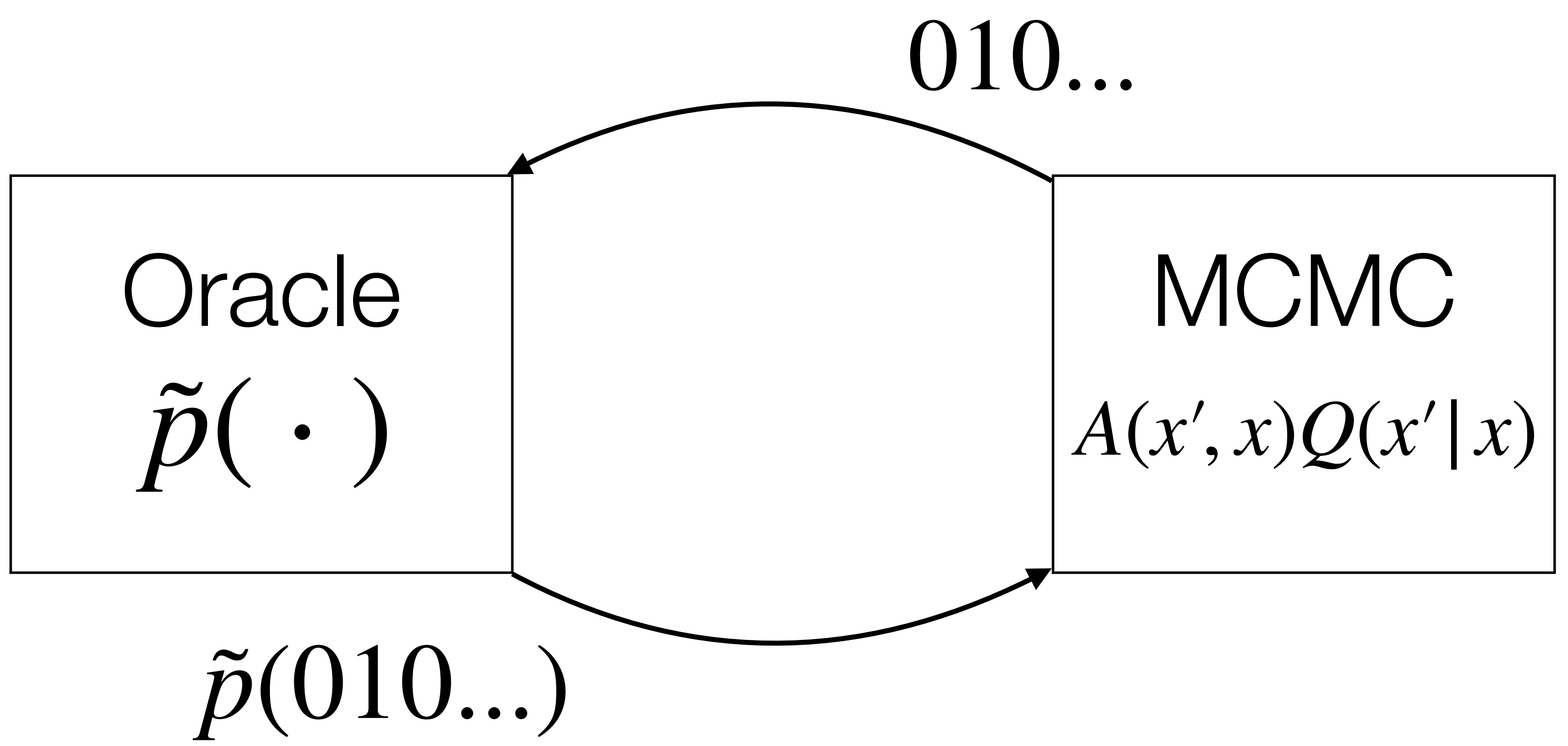
$$p(x) = \frac{\tilde{p}(x)}{P}$$



Cost of evaluation

Markov Chain Monte Carlo - MCMC

$$p(x) = \frac{\tilde{p}(x)}{P}$$

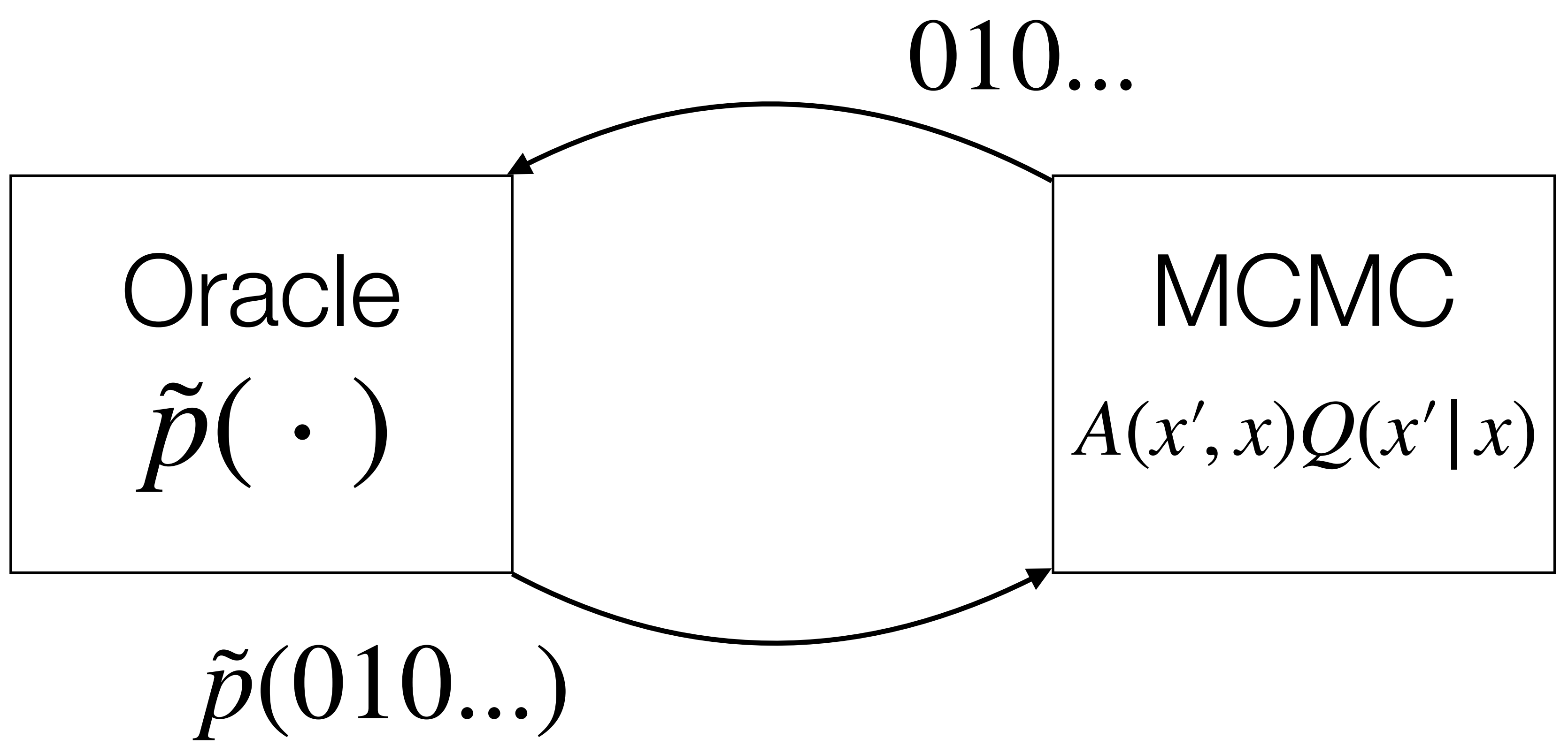


Cost of evaluation

How to learn the proposal to reduce the number of oracle evaluations?

Markov Chain Monte Carlo - MCMC

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Cost of evaluation

How to learn the proposal to reduce the number of oracle evaluations?

- Density estimation [Jaini et al. AISTATS 2021]
- Correlation-based criteria [Levy et al. ICLR 2018]

Key Contributions

1. Use of **mutual information** to assess statistical dependence between consecutive samples

Key Contributions

1. Use of **mutual information** to assess statistical dependence between consecutive samples

2. Two **parametrisations** of recent proposal distribution [Zanella, JASA 2020]

Key Contributions

1. Use of **mutual information** to assess statistical dependence between consecutive samples

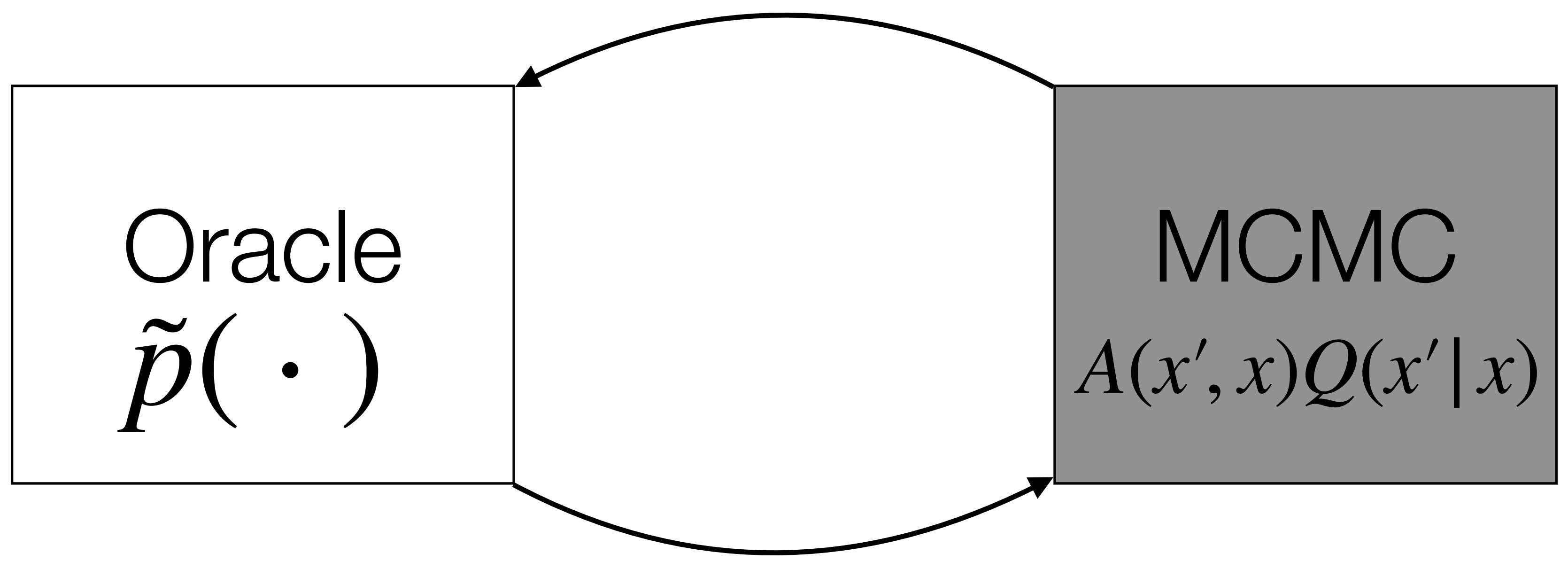
2. Two **parametrisations** of recent proposal distribution [Zanella, JASA 2020]

3. **Gradient-based procedure** to update the proposal parameters by minimising mutual information

Locally Balanced Proposals

[Zanella, JASA 2020]

$$p(x) = \frac{\tilde{p}(x)}{P}$$

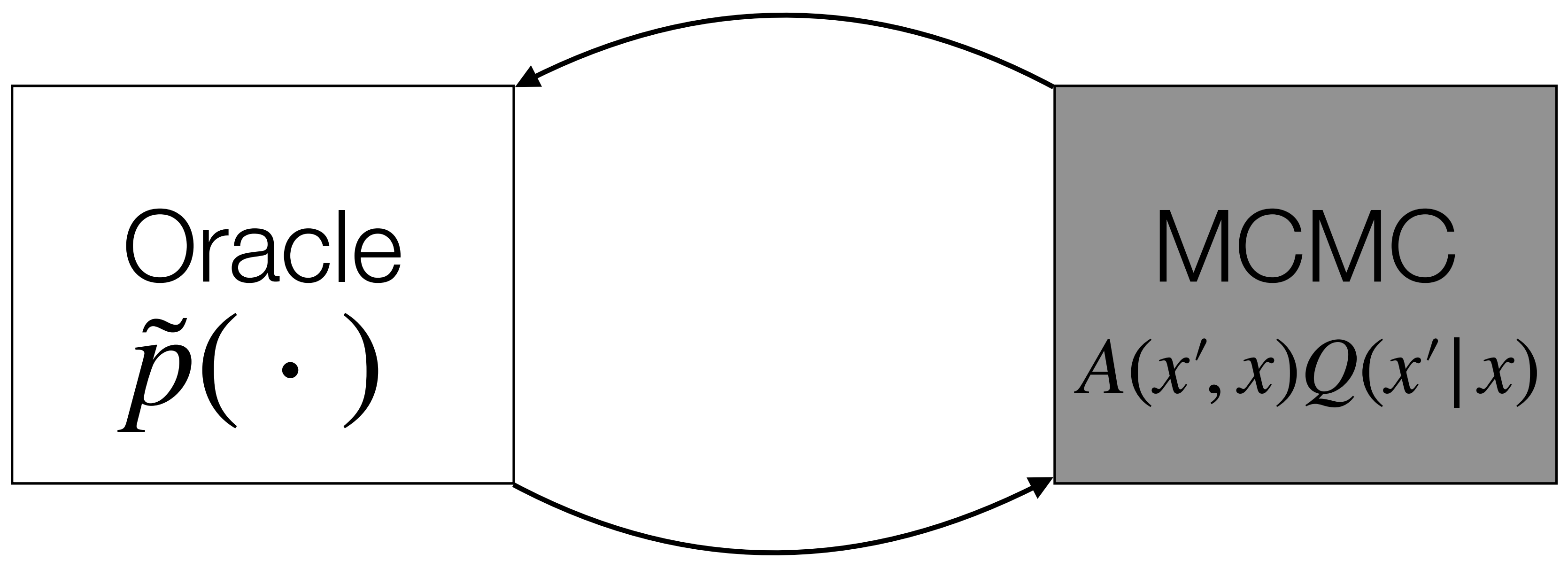


$$Q(x' \mid x) = \frac{g\left(\frac{\tilde{p}(x')}{\tilde{p}(x)}\right) 1\left[x' \in N(x)\right]}{Z(x)}$$

Locally Balanced Proposals

[Zanella, JASA 2020]

$$p(x) = \frac{\tilde{p}(x)}{P}$$



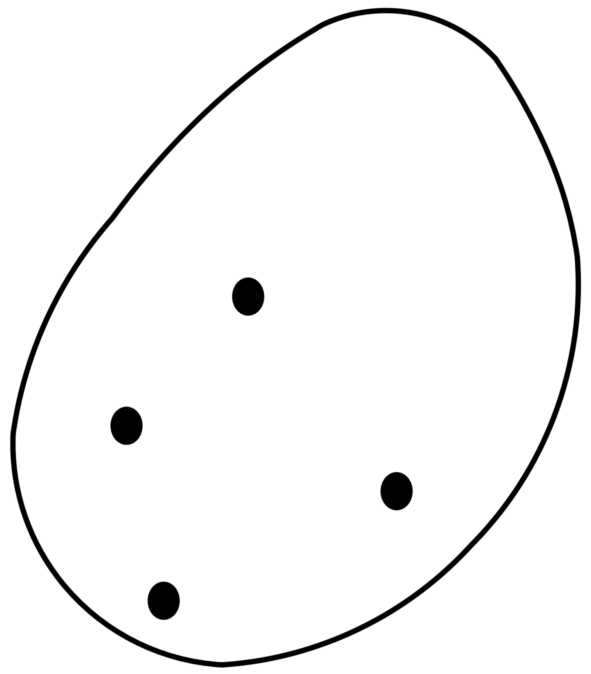
$$Q(x' \mid x) = \frac{g\left(\frac{\tilde{p}(x')}{\tilde{p}(x)}\right) 1\left[x' \in N(x)\right]}{Z(x)}$$

Balancing property $g(t) = t g\left(\frac{1}{t}\right)$

Parametrizations of g

$$g(t) = tg\left(\frac{1}{t}\right)$$

[Zanella, JASA 2020]



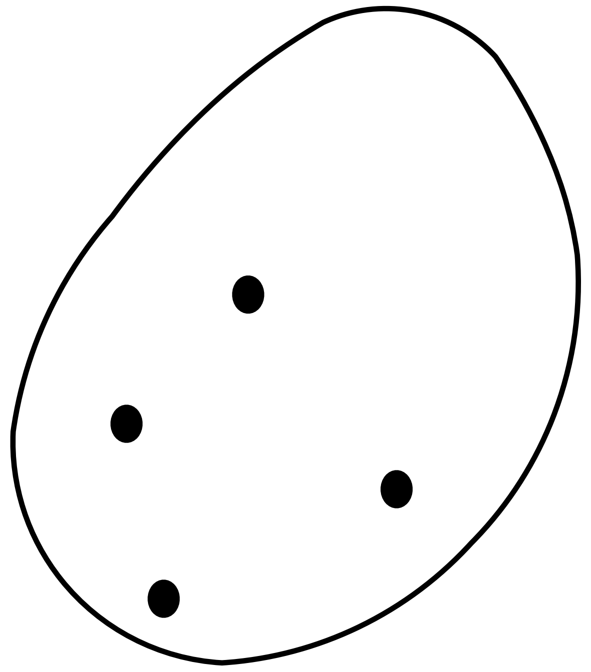
$$g_1(t) = \sqrt{t}$$
$$g_2(t) = \frac{t}{1+t}$$
$$g_3(t) = \min\{1, t\}$$
$$g_4(t) = \max\{1, t\}$$

EXPRESSIVENESS

Parametrizations of g

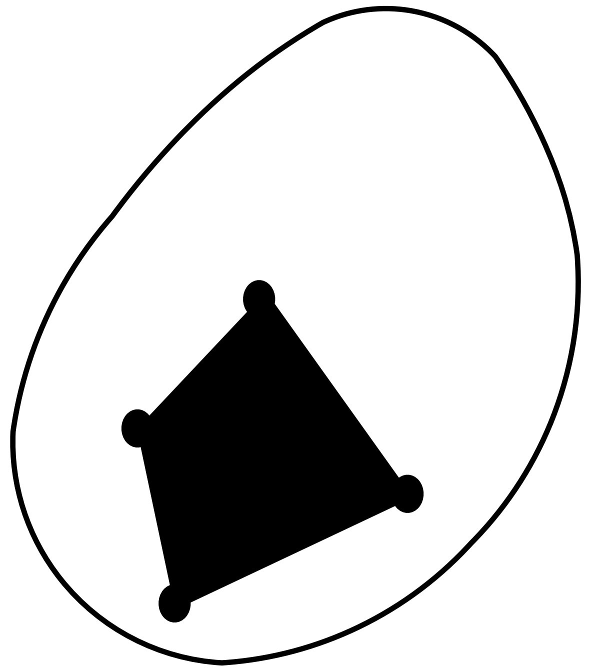
$$g(t) = t g\left(\frac{1}{t}\right)$$

[Zanella, JASA 2020]



$$g_1(t) = \sqrt{t}$$
$$g_2(t) = \frac{t}{1+t}$$
$$g_3(t) = \min\{1, t\}$$
$$g_4(t) = \max\{1, t\}$$

LSB 1



$$g_\theta(t) = \sum_i \theta_i g_i(t)$$

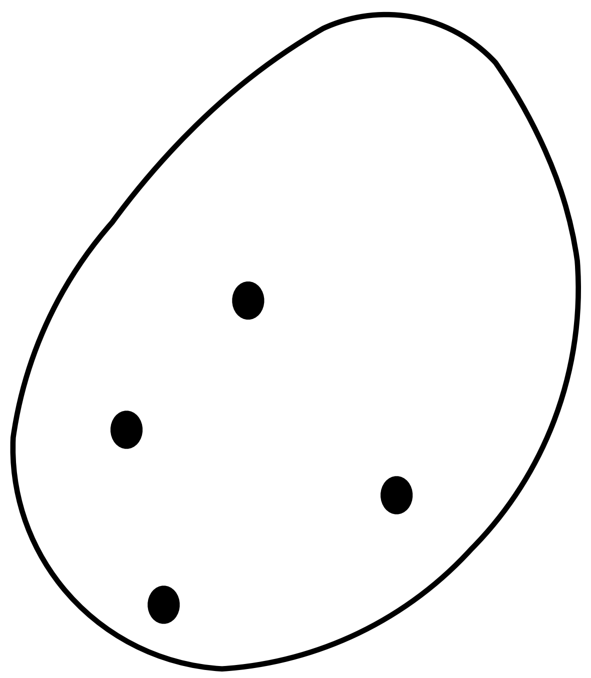
Convex set

EXPRESSIVENESS

Parametrizations of g

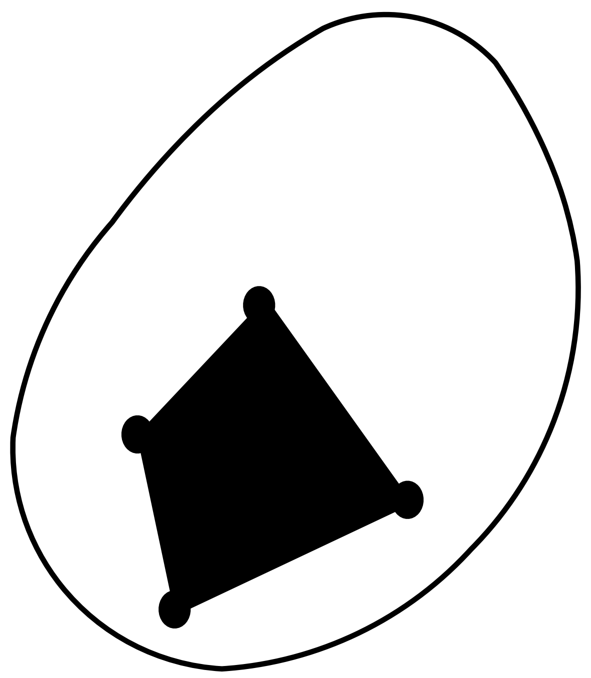
$$g(t) = tg\left(\frac{1}{t}\right)$$

[Zanella, JASA 2020]



$$g_1(t) = \sqrt{t}$$
$$g_2(t) = \frac{t}{1+t}$$
$$g_3(t) = \min\{1, t\}$$
$$g_4(t) = \max\{1, t\}$$

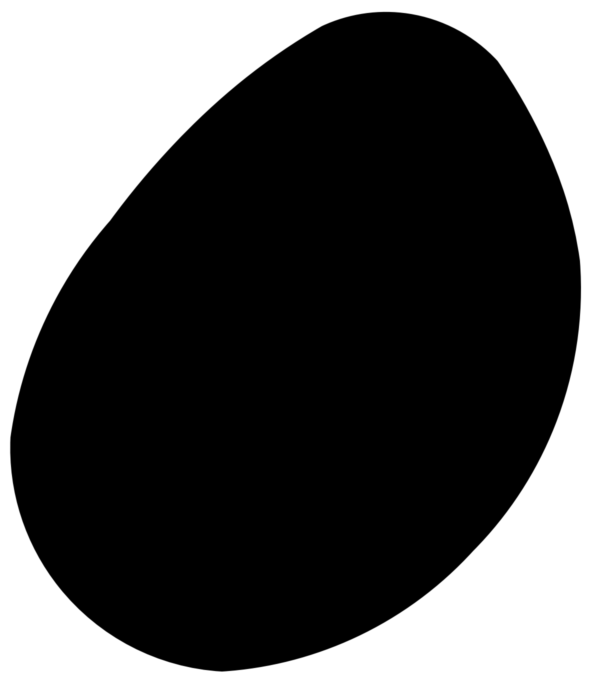
LSB 1



$$g_\theta(t) = \sum_i \theta_i g_i(t)$$

Convex set

LSB 2



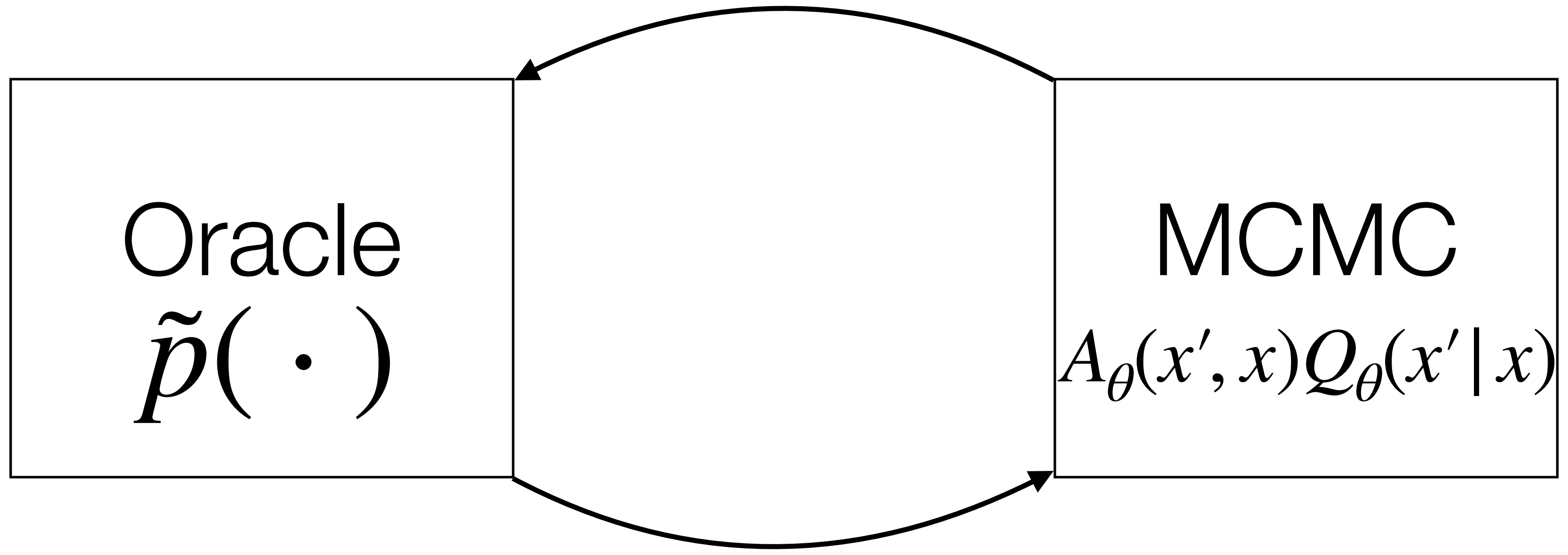
$$g_\theta(t) = \frac{h_\theta}{2} + \frac{t h_\theta(1/t)}{2}$$

h_θ Neural net

EXPRESSIVENESS

Mutual Information

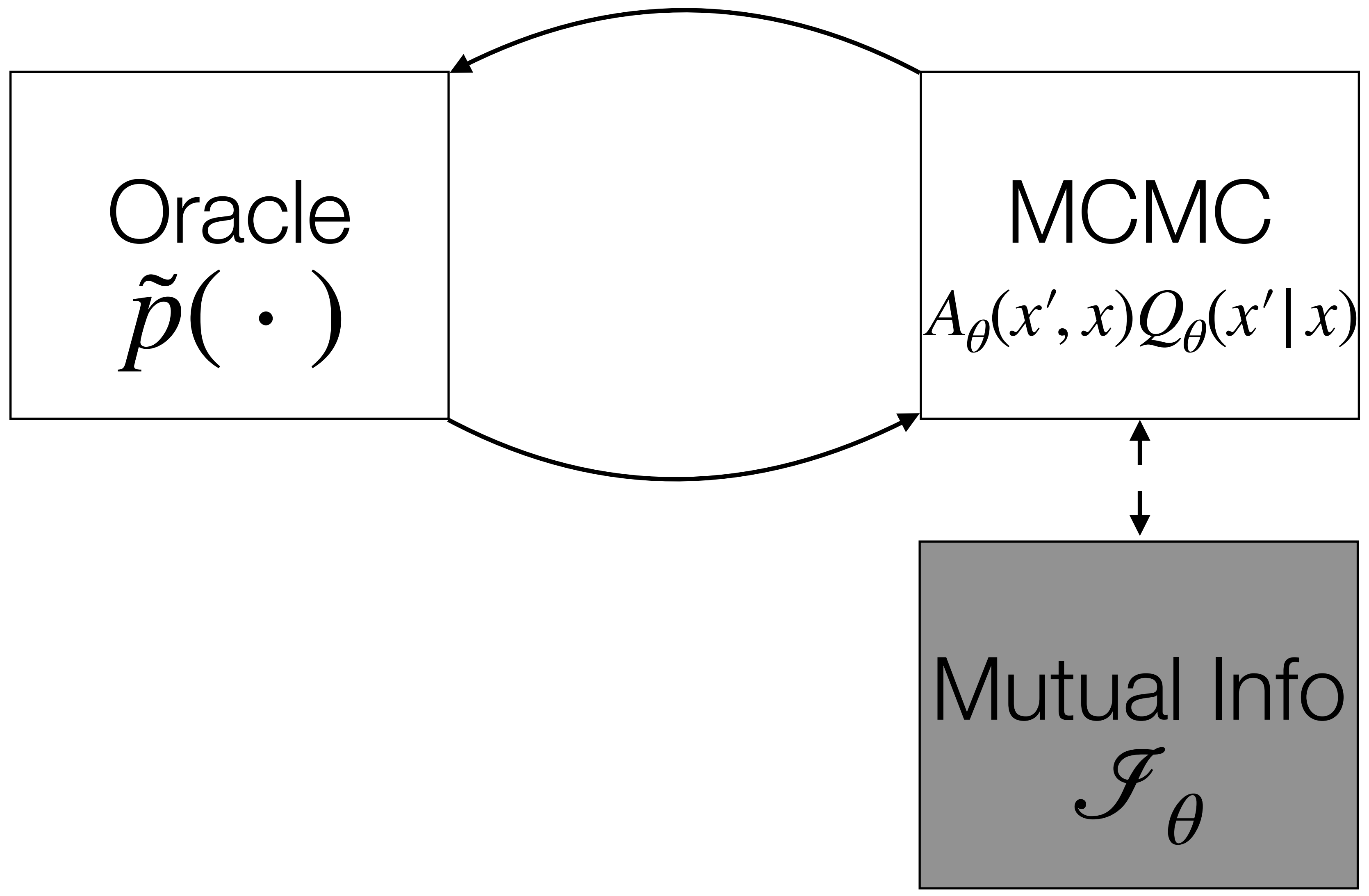
$$p(x) = \frac{\tilde{p}(x)}{P}$$



$$Q_\theta(x'|x) = \frac{g_\theta\left(\frac{\tilde{p}(x')}{\tilde{p}(x)}\right) 1\left[x' \in N(x)\right]}{Z(x)}$$

Mutual Information

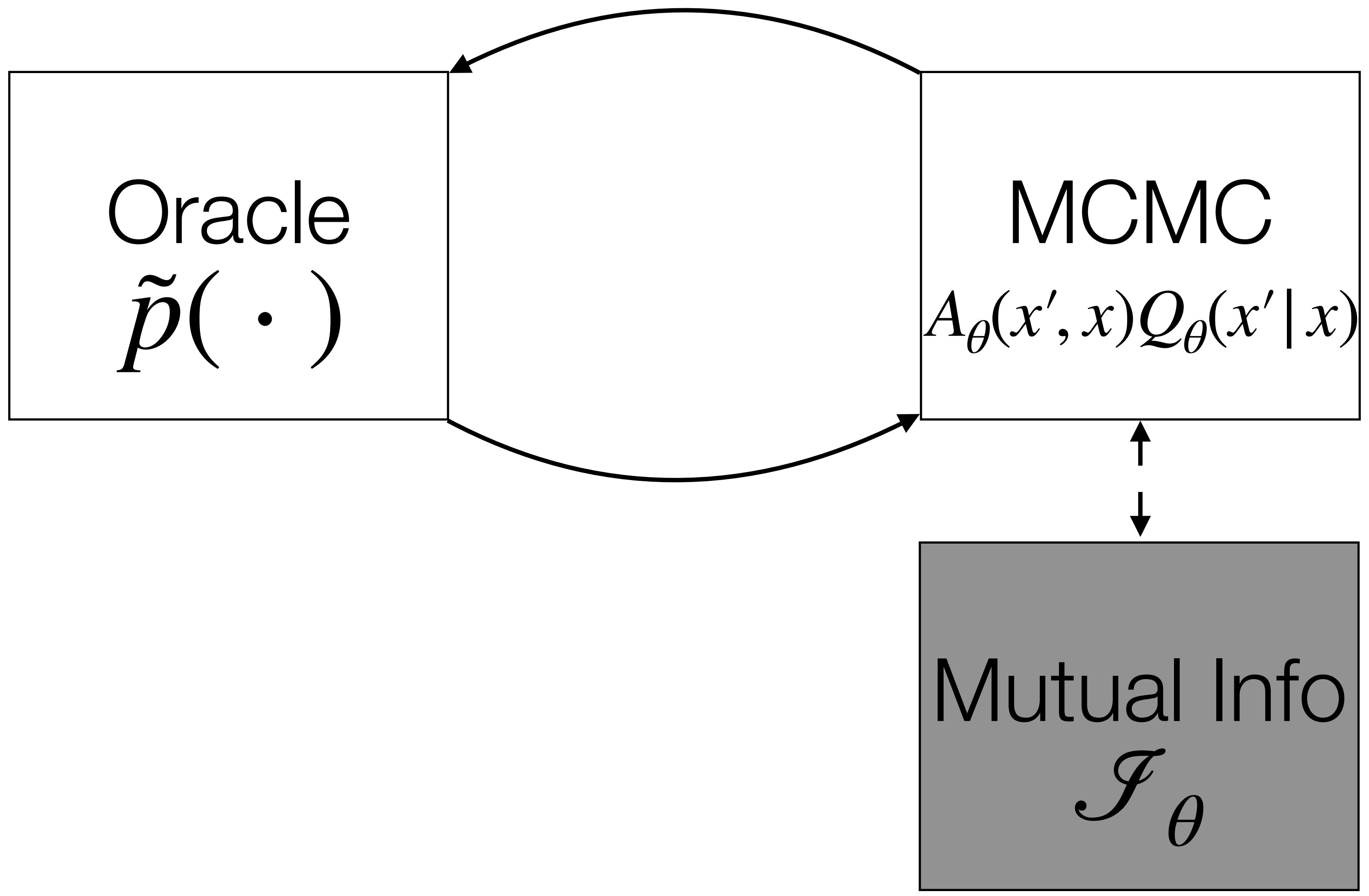
$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathscr{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

Mutual Information

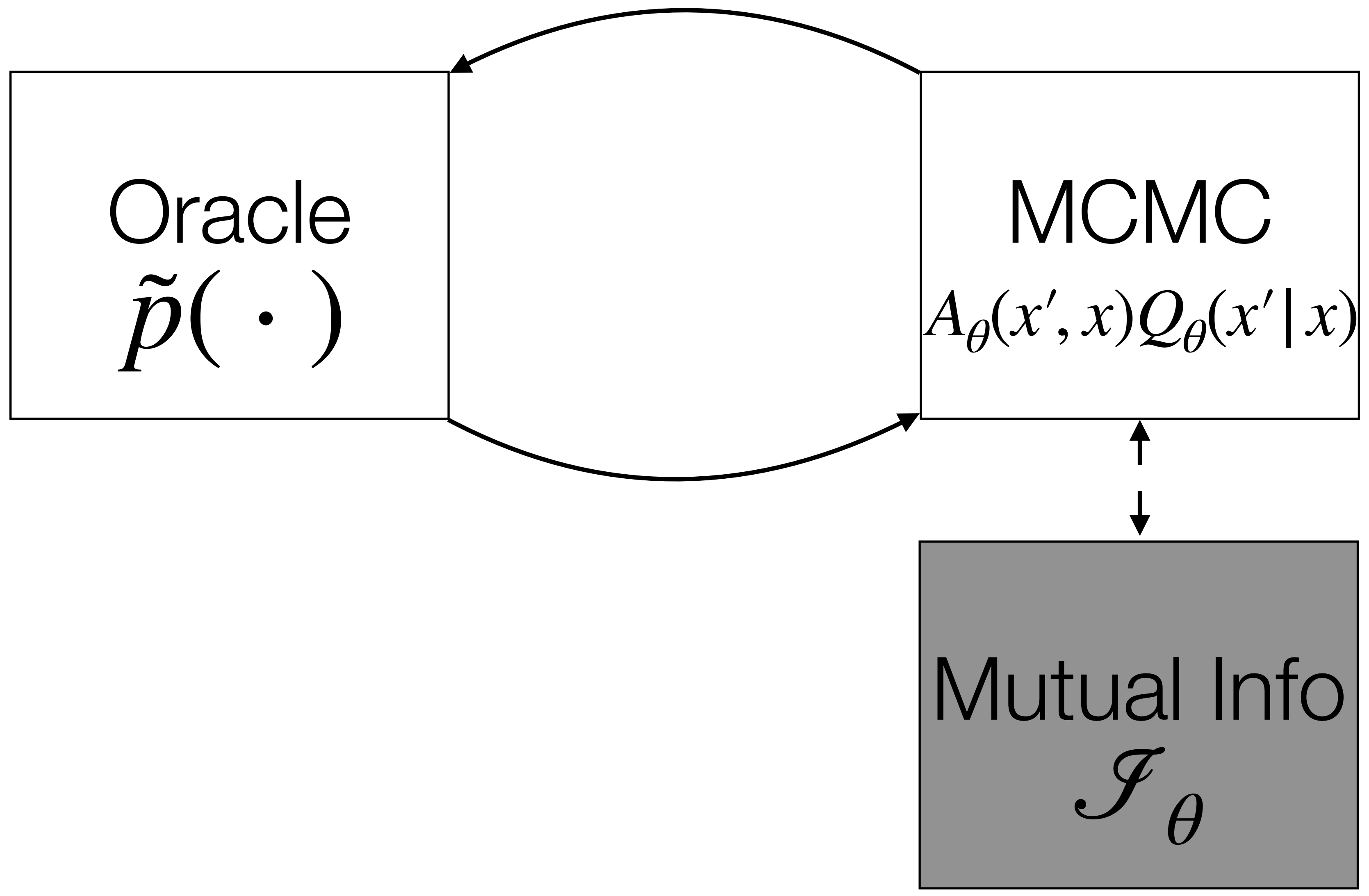
$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step
Upper bound estimation requires $O(d)$ oracle evaluations per step

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



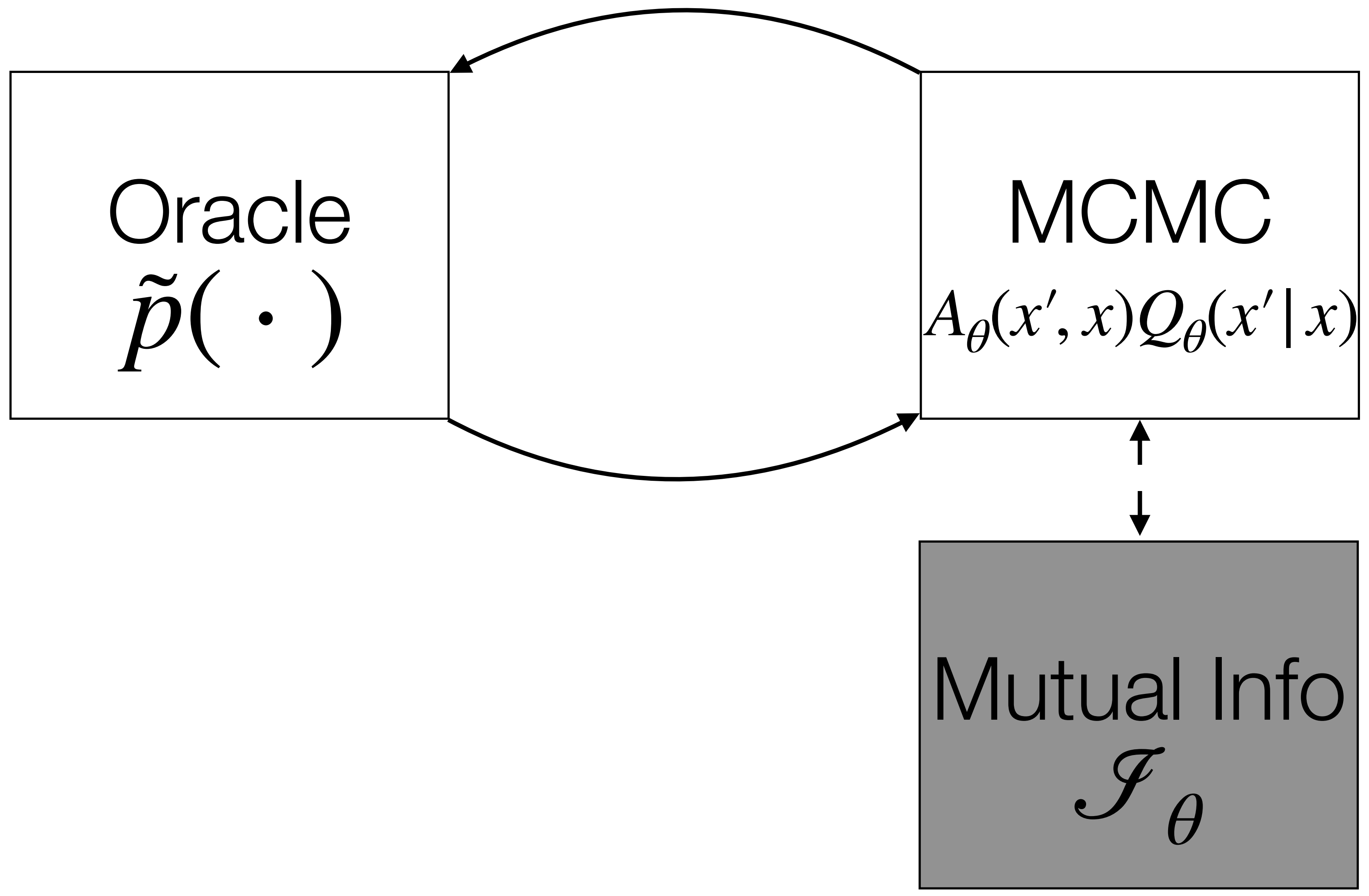
Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step
Upper bound estimation requires $O(d)$ oracle evaluations per step

$$\mathcal{I}_\theta \leq_P E_{x,x' \sim Q_1,Q_2}\left\{\frac{\tilde{p}(x)A(x',x)Q_\theta(x'|x)}{Q_1(x)Q_2(x')}\log\frac{A_\theta(x',x)Q_\theta(x'|x)}{\tilde{p}(x')}\right\} + E_{x\sim Q_1}\left\{\frac{1-A_\theta(x*,x)Q_\theta(x*|x)}{Q_1(x)}\left[\eta - \eta A - \theta(x*,x)Q_\theta(x*|x) - \tilde{p}(x)(\log\eta + 1)\right]\right\}$$

$$Q_1 = \pi U + (1-\pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

Upper bound estimation requires $O(d)$ oracle evaluations per step

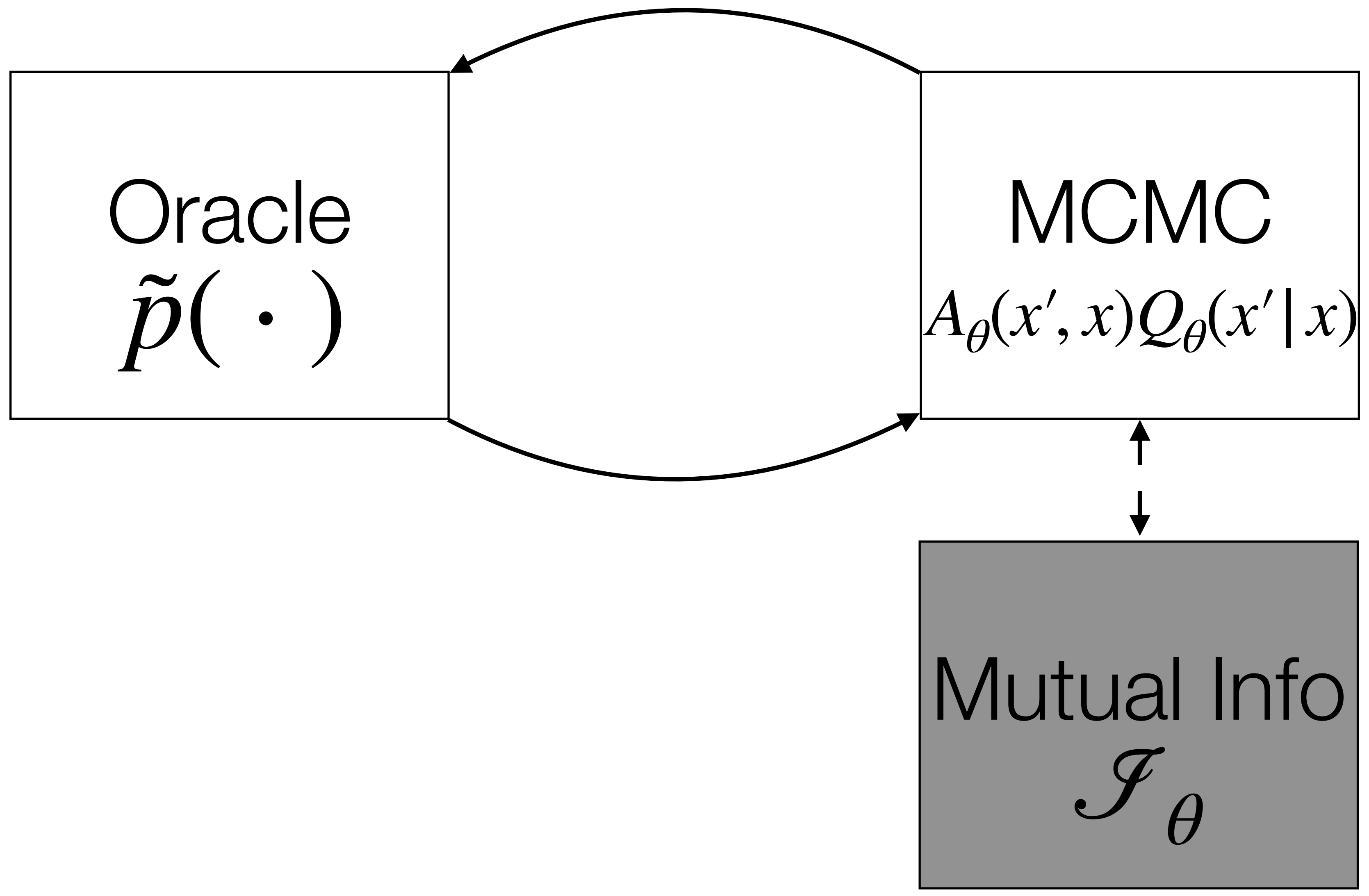
$$\mathcal{I}_\theta \leq_P E_{x,x'\sim Q_1,Q_2}\left\{\frac{\tilde{p}(x)A(x',x)Q_\theta(x'|x)}{Q_1(x)Q_2(x')}\log\frac{A_\theta(x',x)Q_\theta(x'|x)}{\tilde{p}(x')}\right\} + E_{x\sim Q_1}\left\{\frac{1-A_\theta(x^*,x)Q_\theta(x^*|x)}{Q_1(x)}\left[\eta-\eta A-\theta(x^*,x)Q_\theta(x^*|x)-\tilde{p}(x)(\log\eta+1)\right]\right\}$$

Contribution related to acceptance | **Contribution related to rejection**

$$Q_1 = \pi U + (1-\pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x^* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step
Upper bound estimation requires $O(d)$ oracle evaluations per step

$\tilde{p}$ analytical and differentiable [Grathwohl et al. 2021]

$$\tilde{p}(x') \approx \tilde{p}(x) + \nabla_x \tilde{p}(x)^T \left[x' - x\right]$$

Substitute approximation in LSB - $O(1)$ oracle evaluations per step

Fast approximation

Recap

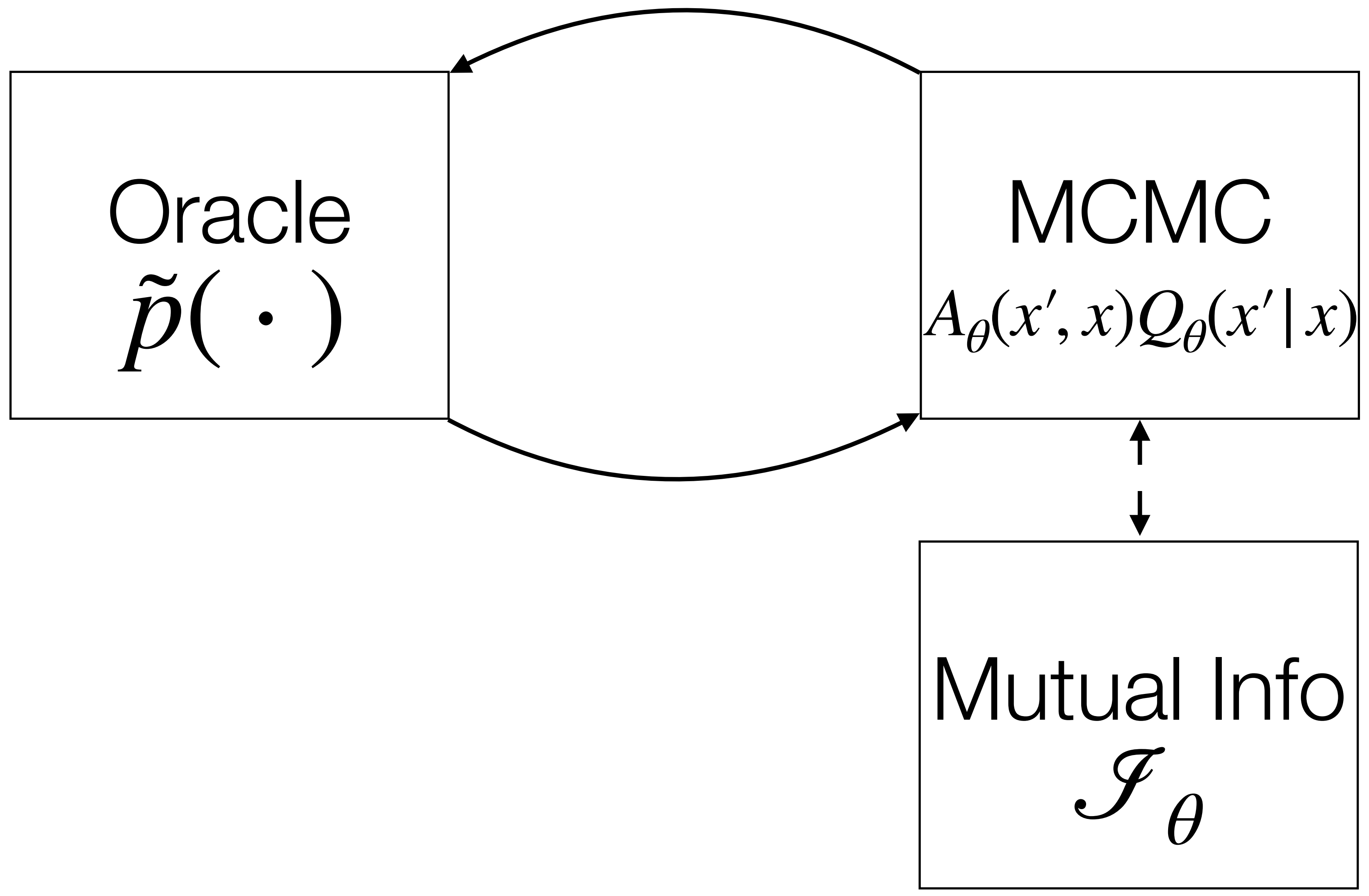
EXPRESSIVENESS →

ESTIMATORS
evaluations per step ↓

	Linear	Nonlinear
$O(d)$	LSB 1	LSB 2
$O(1)$	FLSB 1	FLSB 2

Gradient-Descent Training

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Minimization

$$Q_\theta(x' | x) = \frac{g_\theta\left(\frac{\tilde{p}(x')}{\tilde{p}(x)}\right) 1\left[x' \in N(x)\right]}{Z(x)}$$

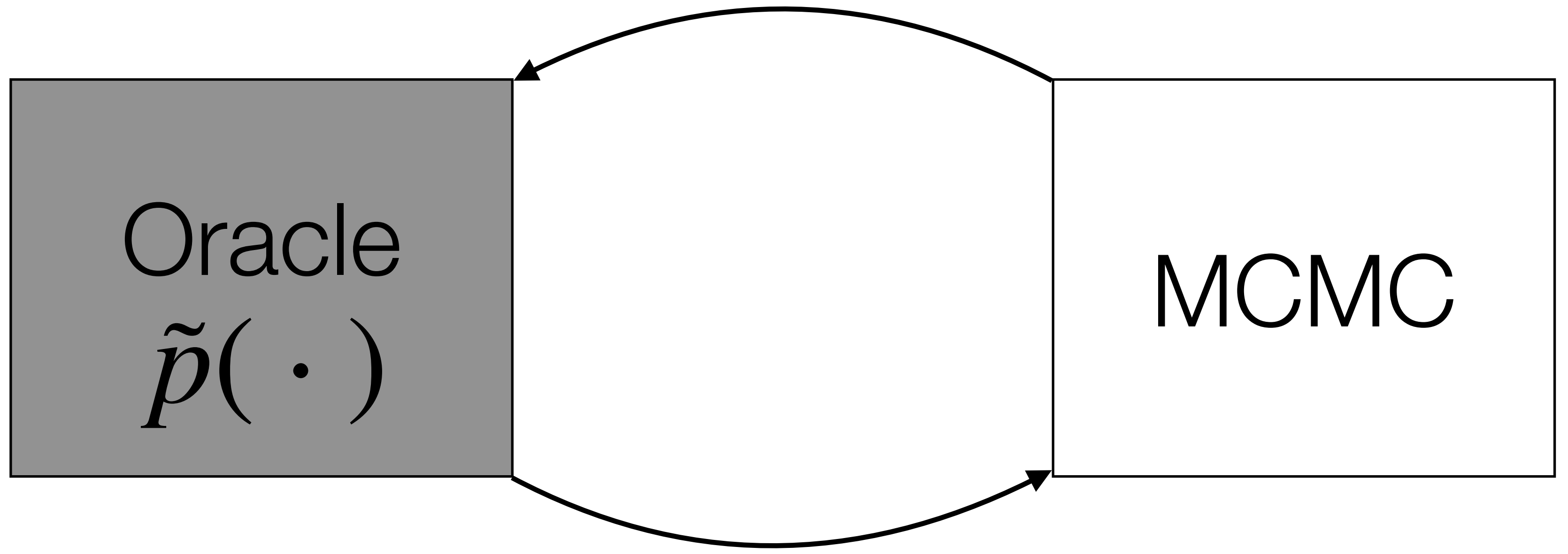
EXPRESSIVENESS

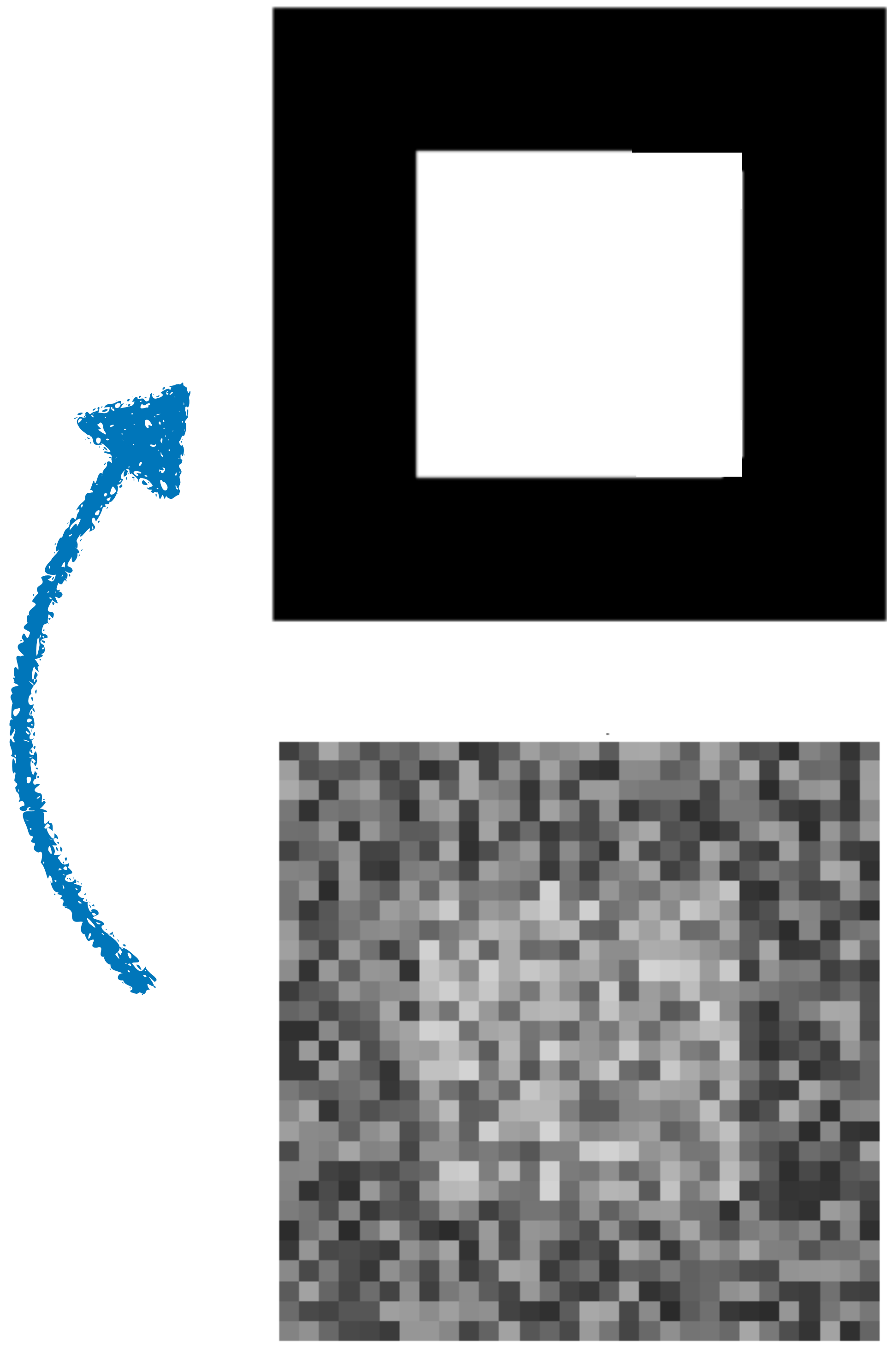
ESTIMATORS
evaluations per step

	Linear	Nonlinear
$O(d)$	LSB 1	LSB 2
$O(1)$	FLSB 1	FLSB 2

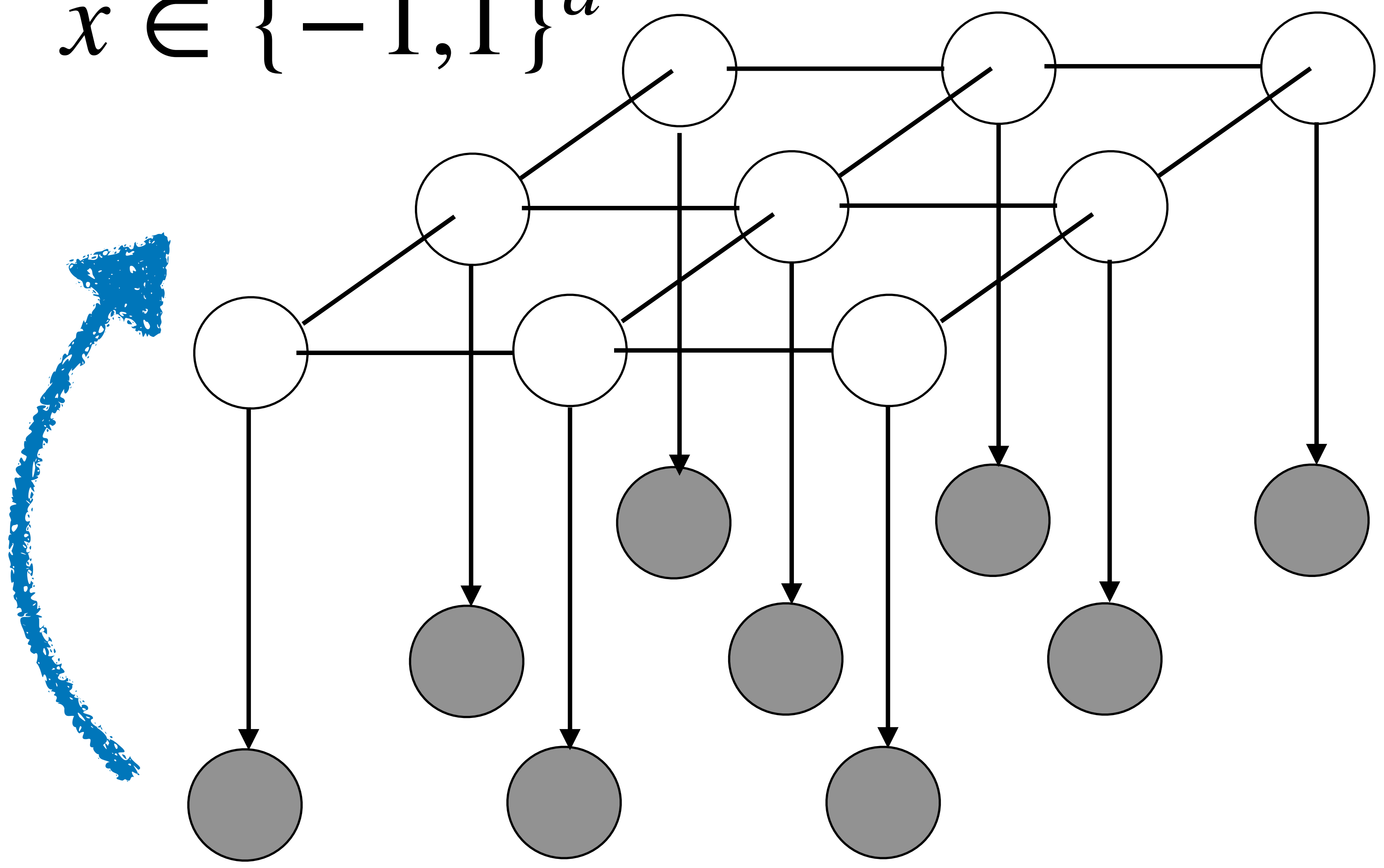
Experiments on Ising Model

$$\tilde{p}(x) = e^{\sum_{i \in V} \alpha_i x_i + \lambda \sum_{(i,j) \in E} x_i x_j}$$



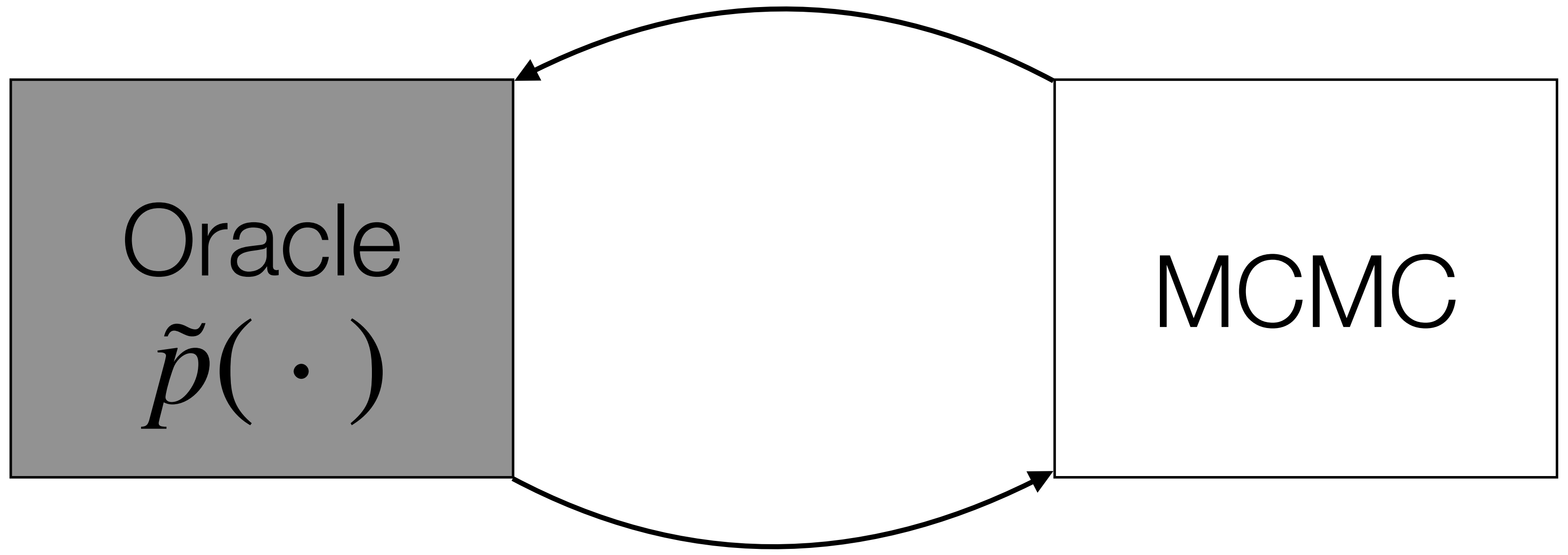


$x \in \{-1,1\}^d$

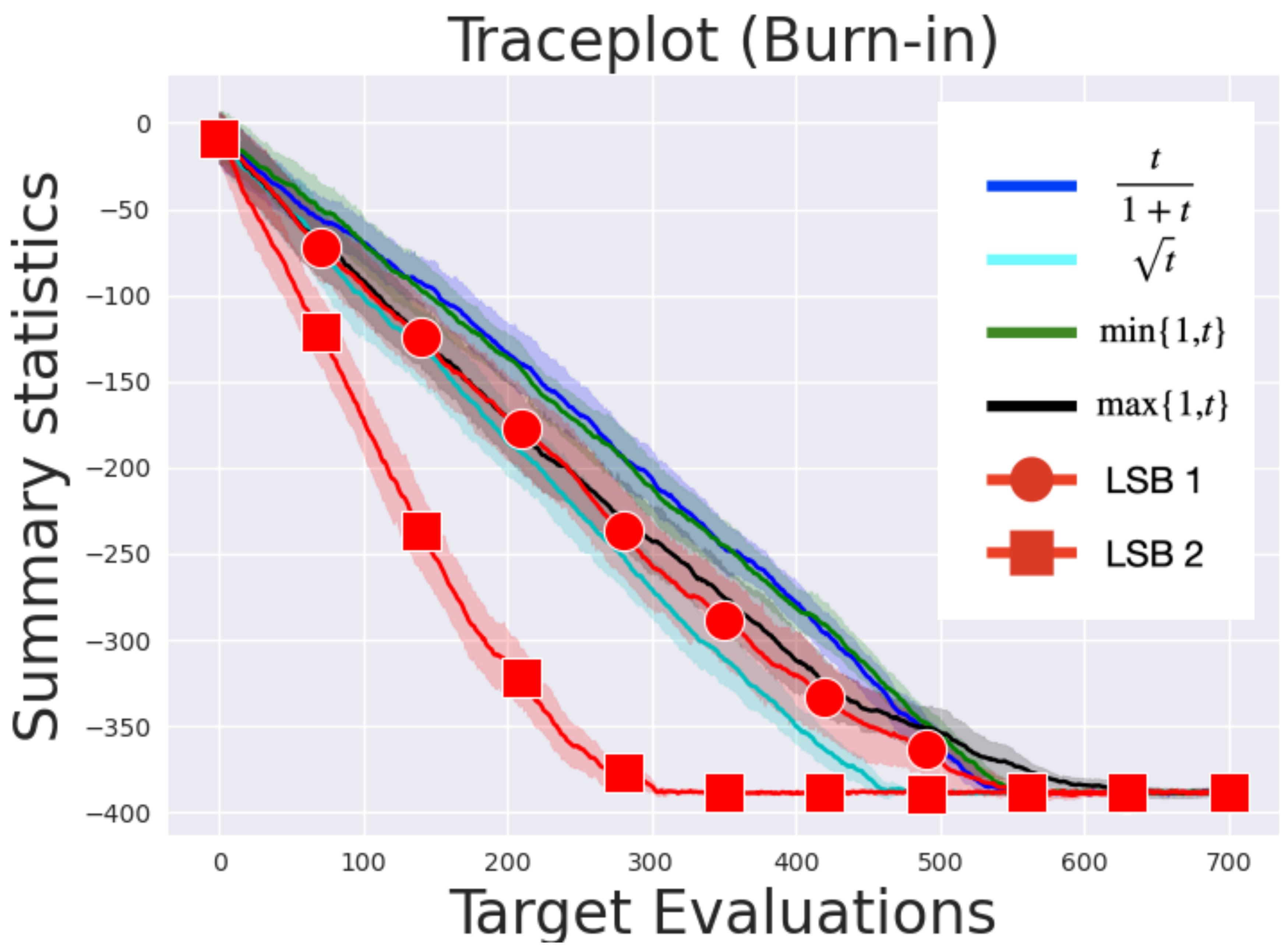


Experiments on Ising Model

$$\tilde{p}(x) = e^{\sum_{i\in V} \alpha_i x_i + \lambda \sum_{(i,j)\in E} x_i x_j}$$

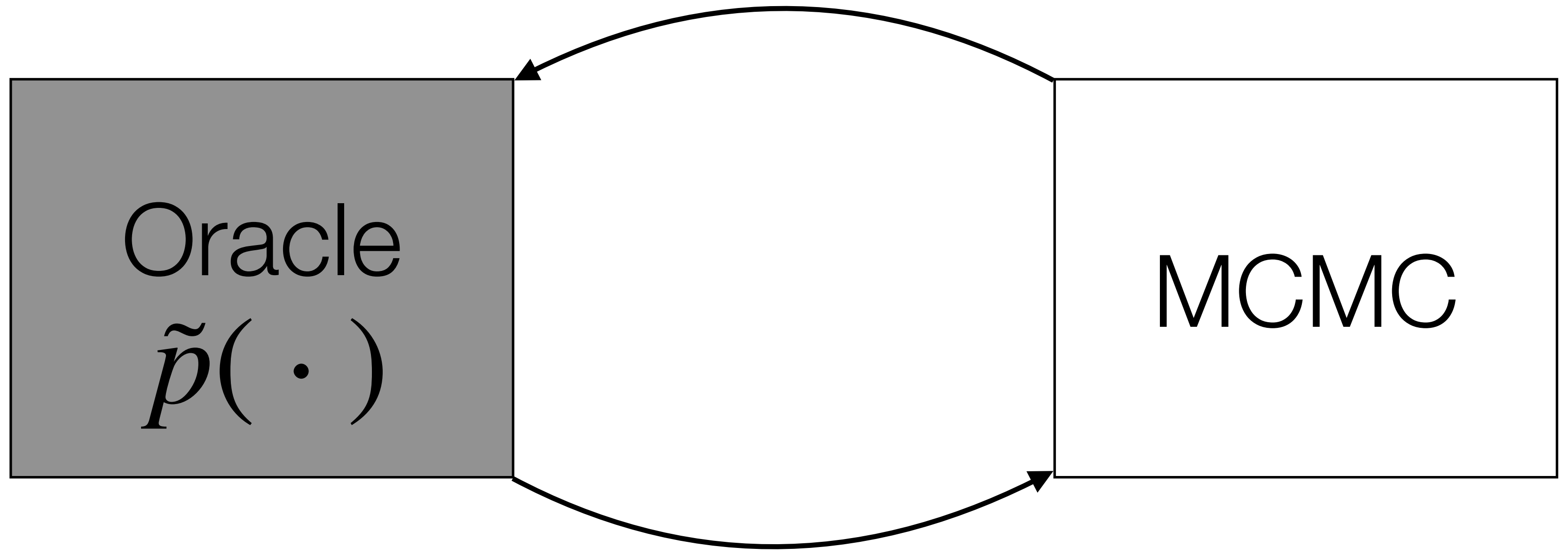


Comparison with [Zanella, JASA 2020]

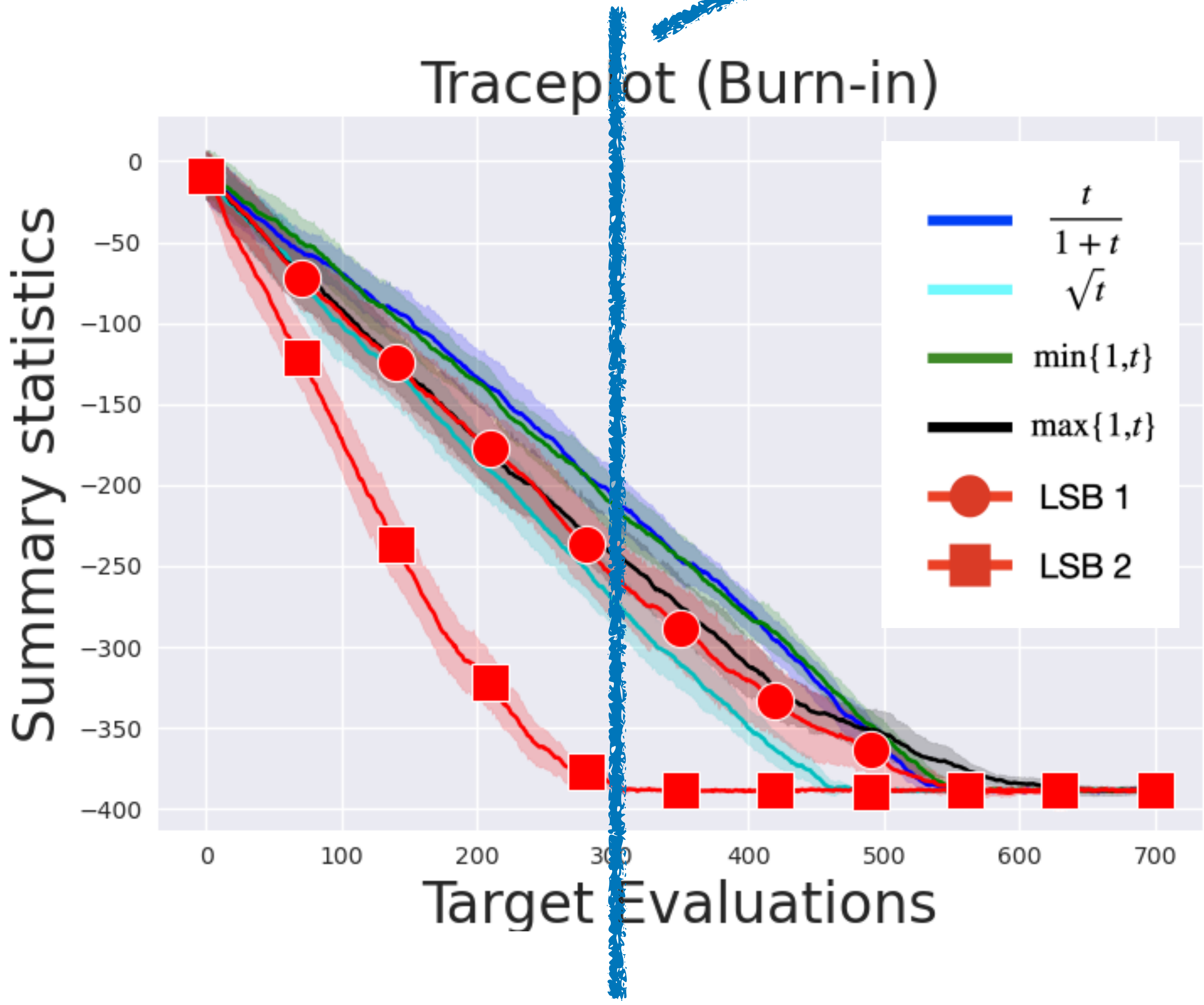


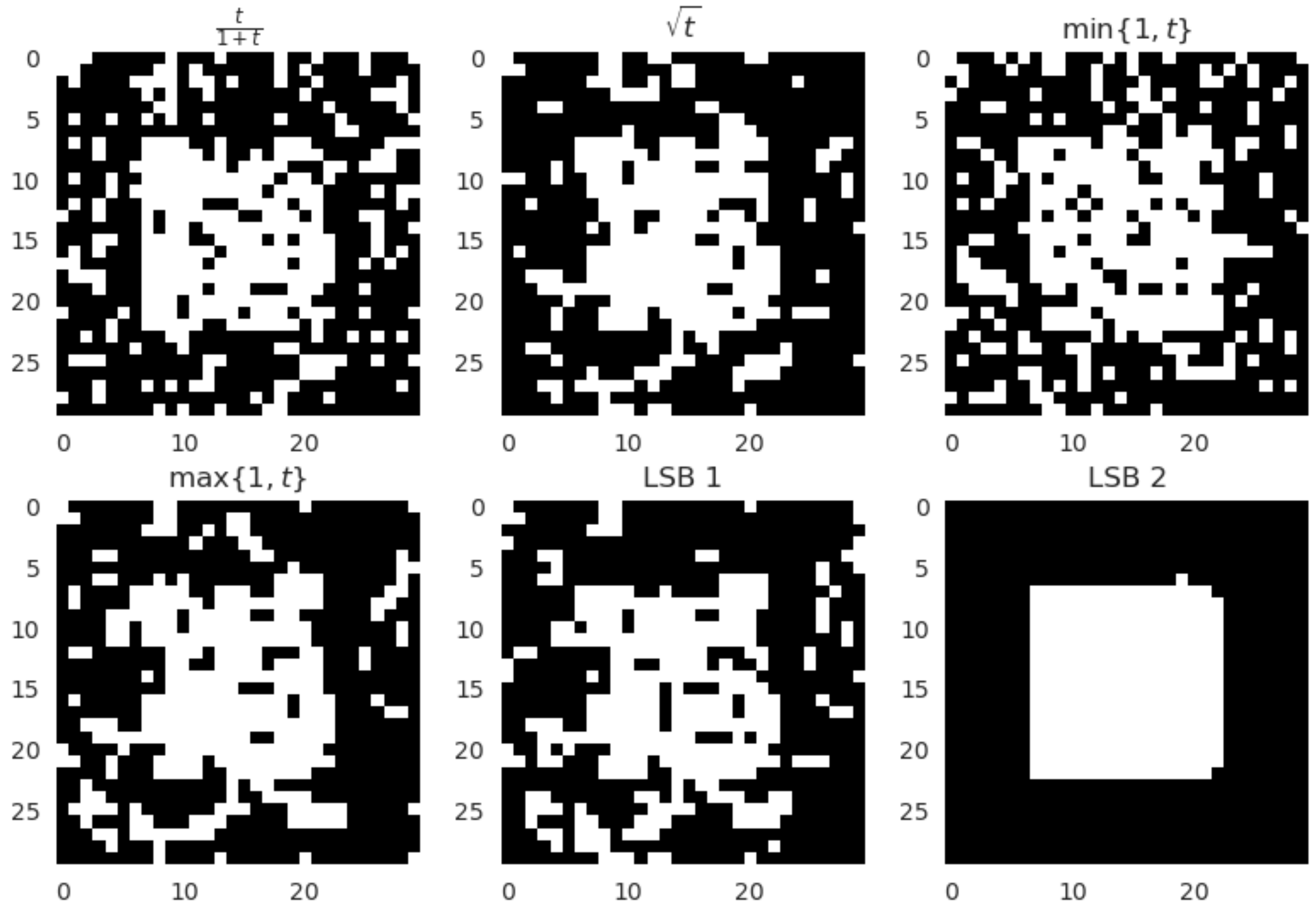
Experiments on Ising Model

$$\tilde{p}(x) = e^{\sum_{i \in V} \alpha_i x_i + \lambda \sum_{(i,j) \in E} x_i x_j}$$



Comparison with [Zanella, JASA 2020]





Thanks

Code available

You can find me at

- @skiera87
- emanuele.sansone@kuleuven.be

Special thank to:

Luc de Raedt

TAILOR

Funding

Thanks

Code available

You can find me at

- @skiera87
- emanuele.sansone@kuleuven.be

Special thank to:

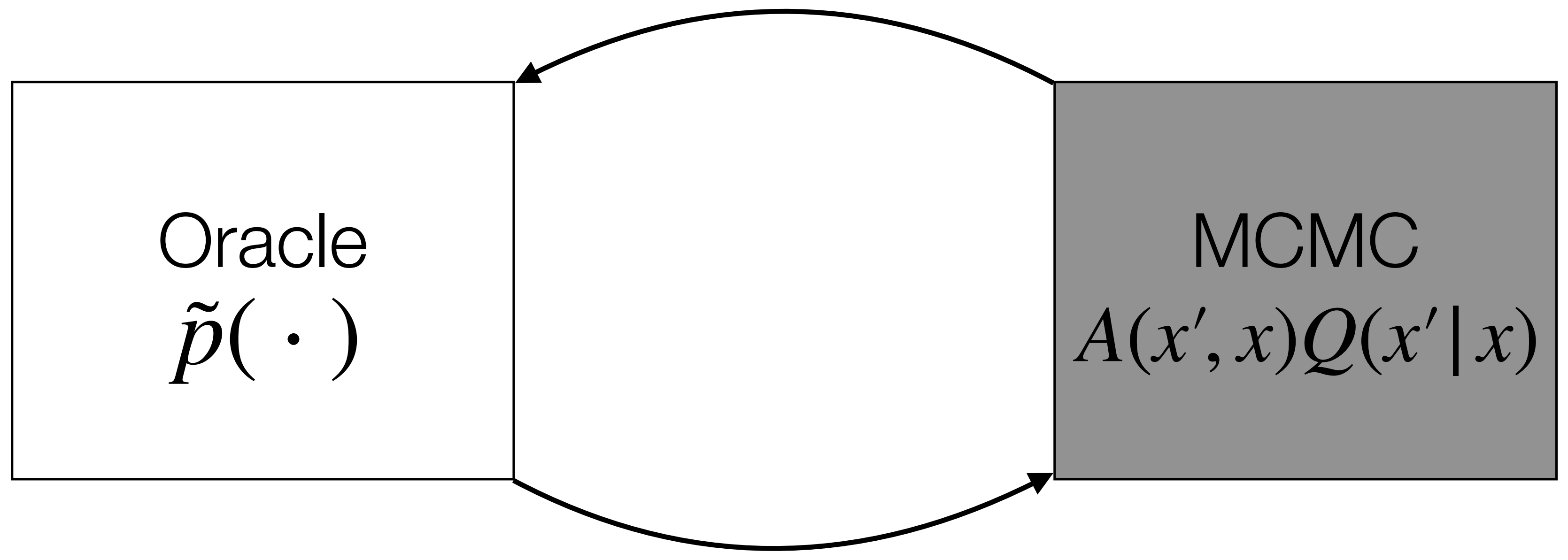
Luc de Raedt

TAILOR

Funding

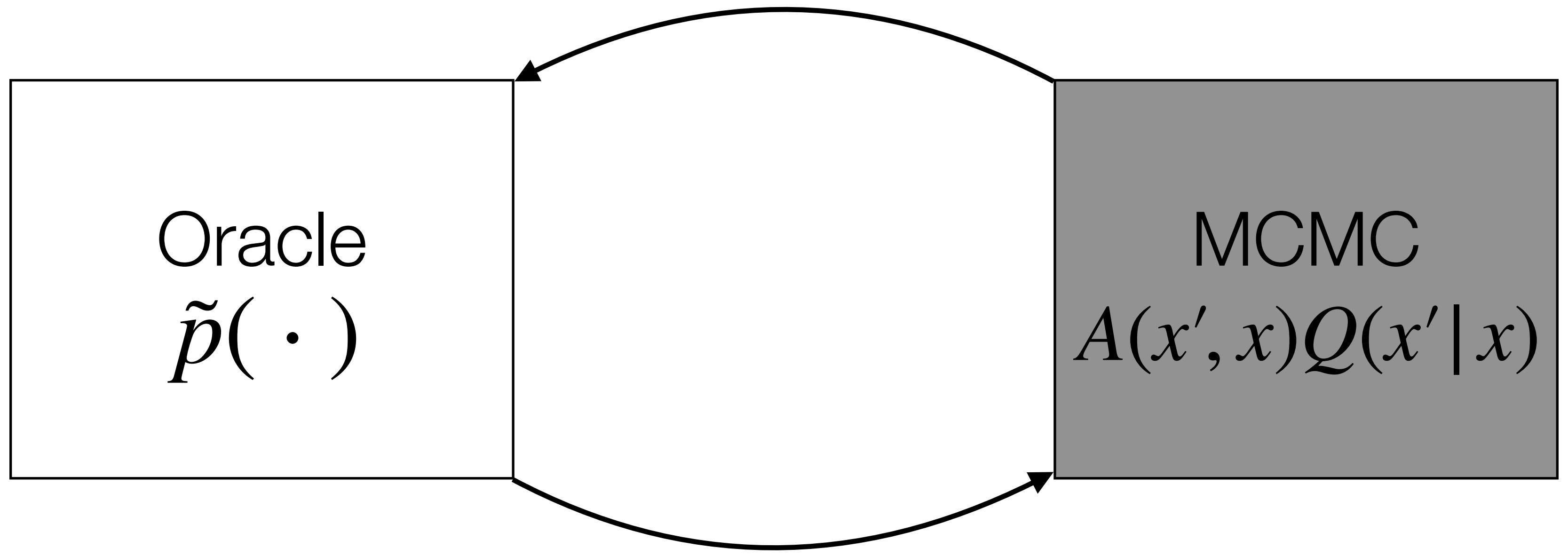
Locally Balanced Proposals

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Locally Balanced Proposals

$$p(x) = \frac{\tilde{p}(x)}{P}$$

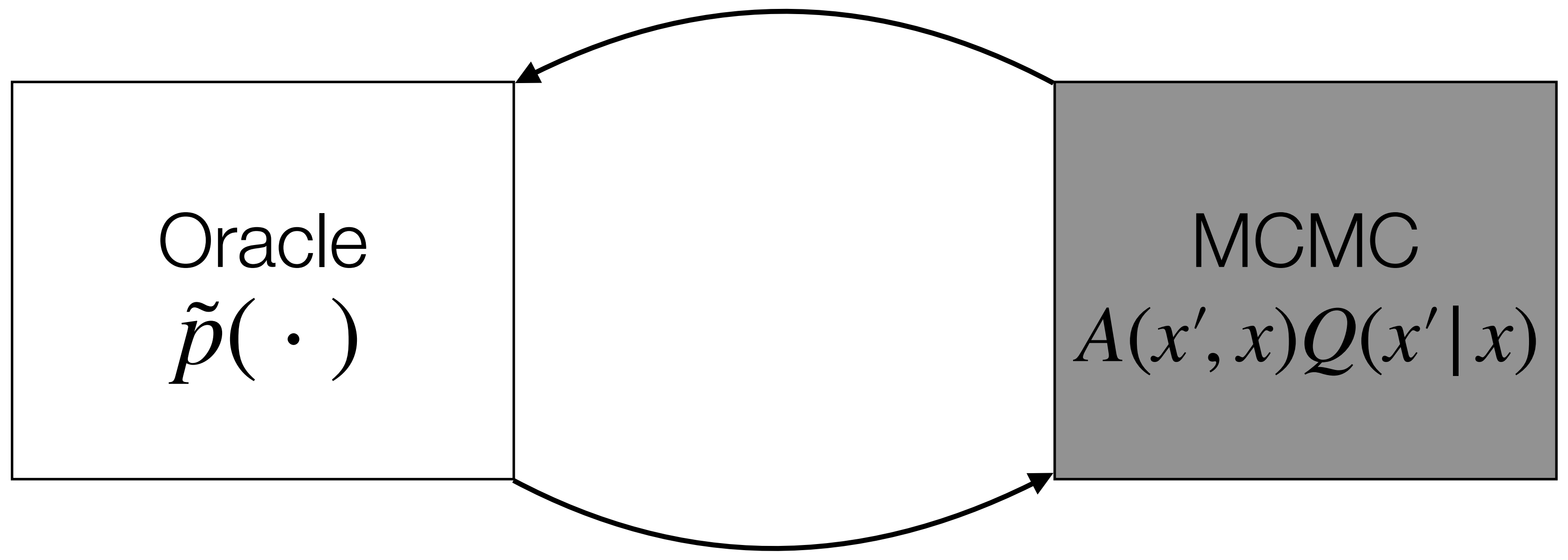


[Zanella, JASA 2020]

$$Q(x' \mid x) = \frac{g\left(\frac{\tilde{p}(x')}{\tilde{p}(x)}\right) 1\left[x' \in N(x)\right]}{Z(x)}$$

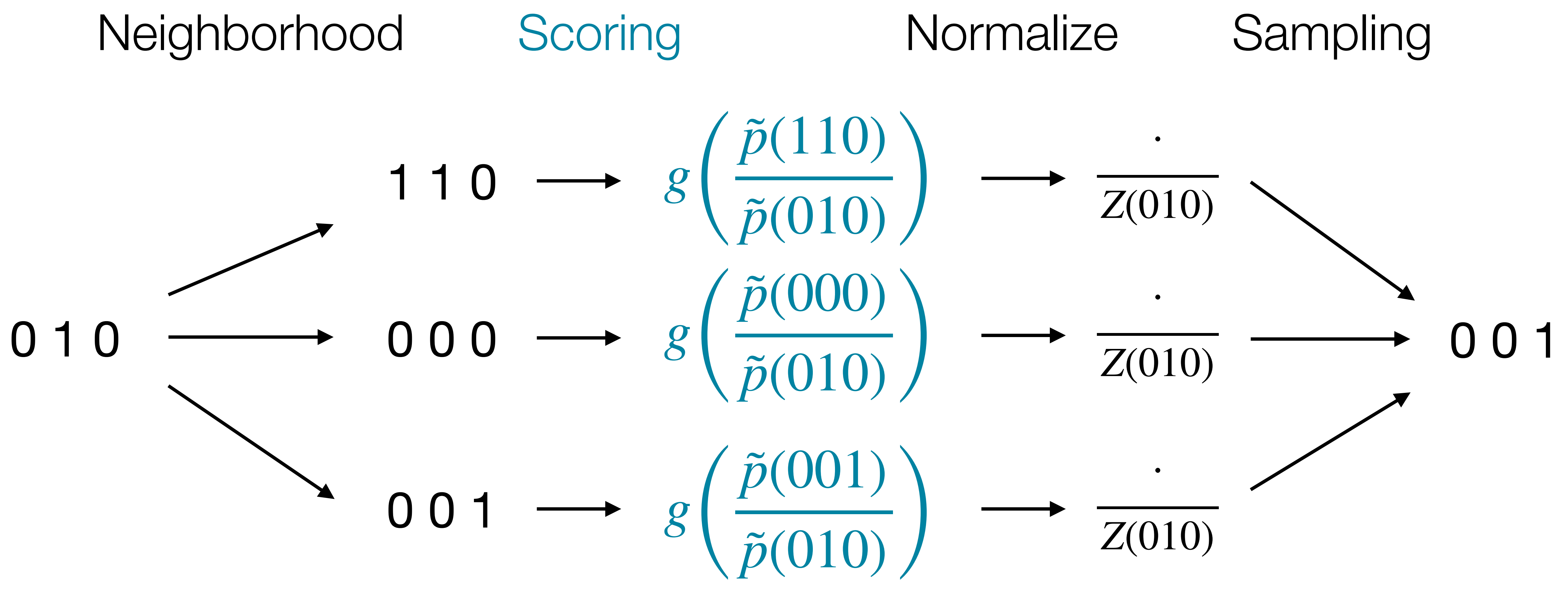
Locally Balanced Proposals

$$p(x) = \frac{\tilde{p}(x)}{P}$$



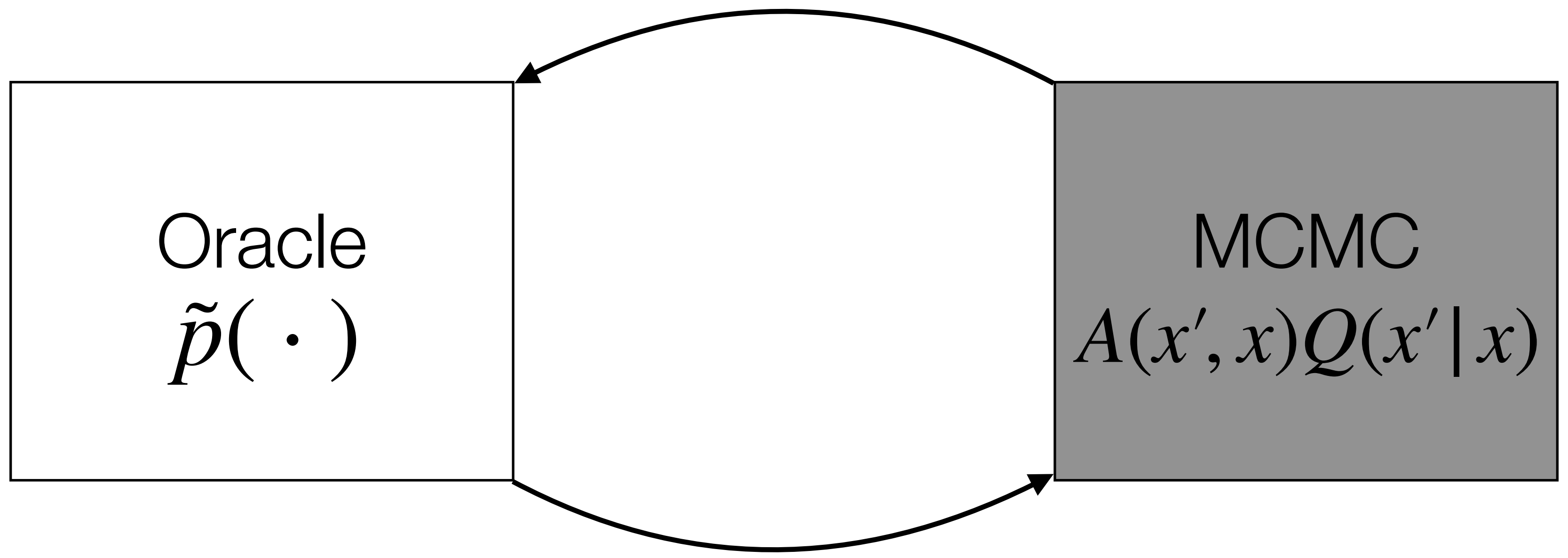
[Zanella, JASA 2020]

$$Q(x'|x) = \frac{g\left(\frac{\tilde{p}(x')}{\tilde{p}(x)}\right) 1\left[x' \in N(x)\right]}{Z(x)}$$



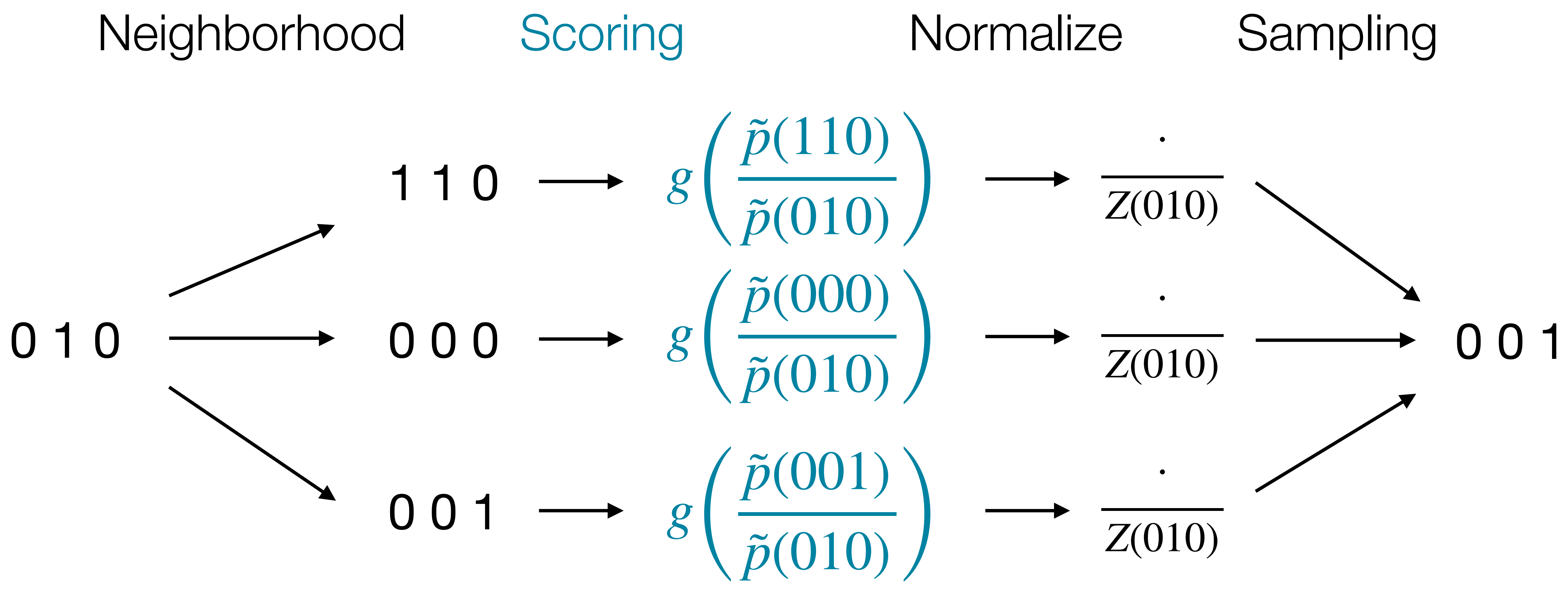
Locally Balanced Proposals

$$p(x) = \frac{\tilde{p}(x)}{P}$$



[Zanella, JASA 2020]

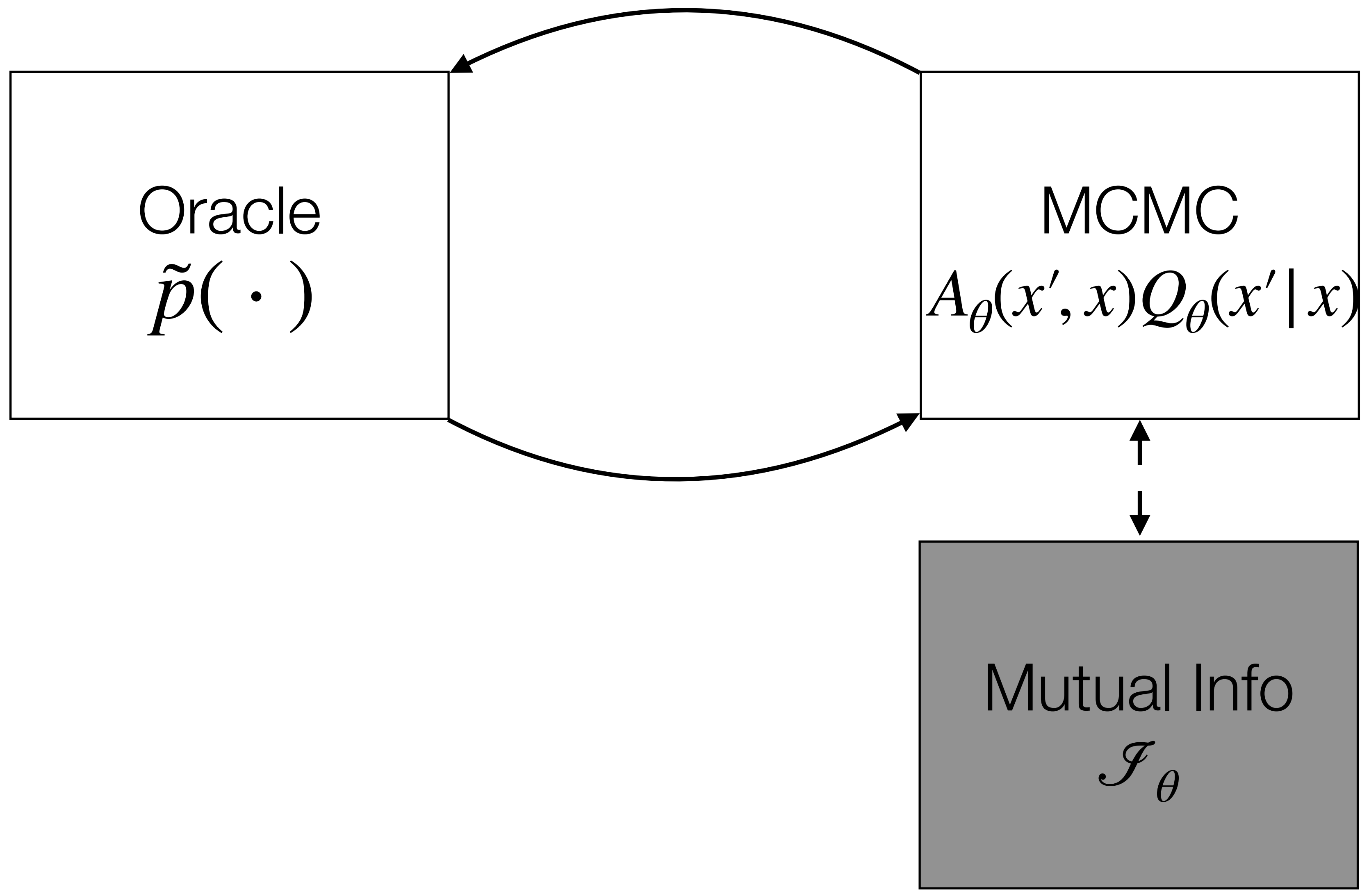
$$Q(x'|x) = \frac{g\left(\frac{\tilde{p}(x')}{\tilde{p}(x)}\right) 1\left[x' \in N(x)\right]}{Z(x)}$$



Balancing property $g(t) = t g\left(\frac{1}{t}\right)$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

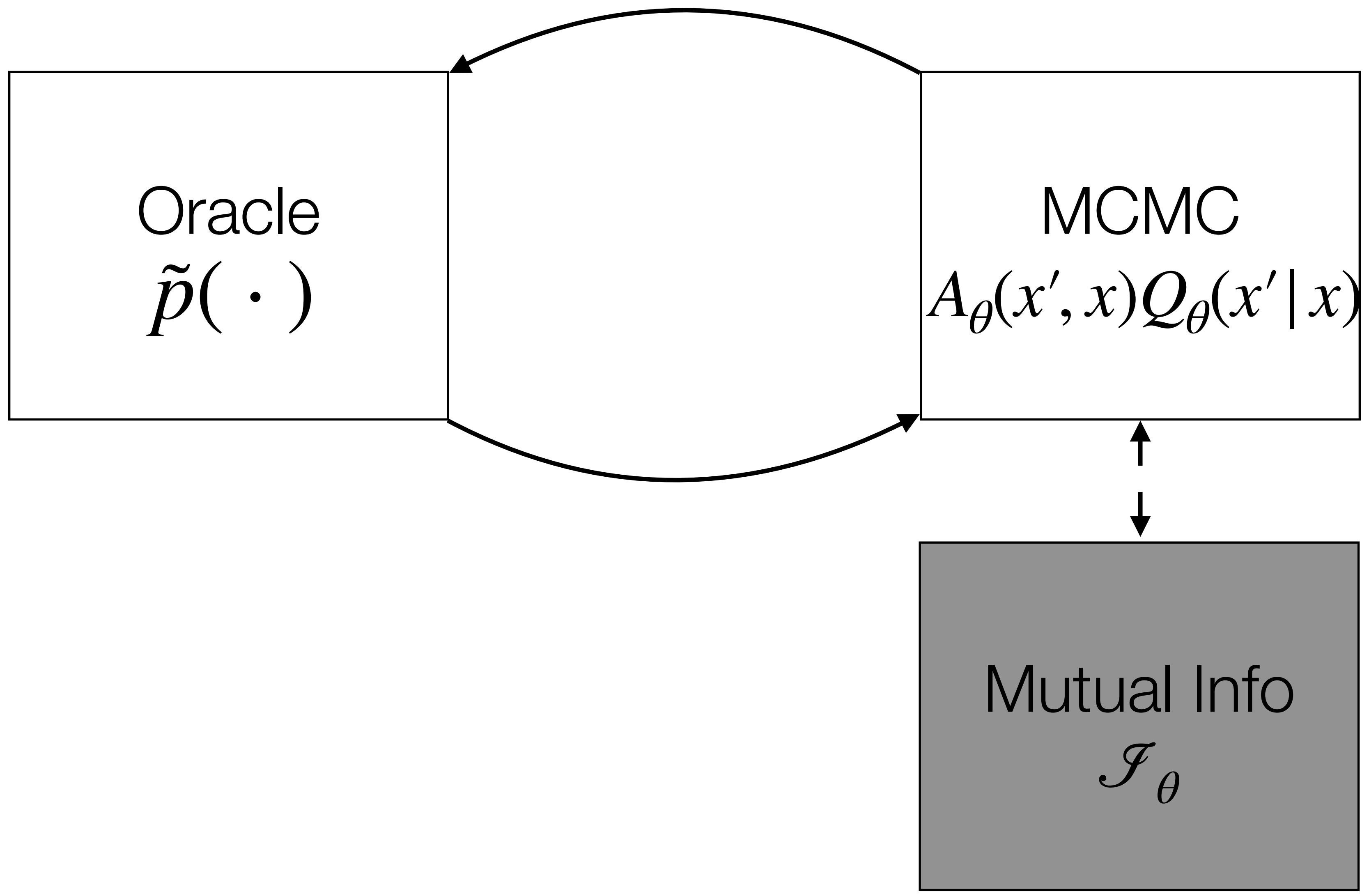
Upper bound estimation requires $O(d)$ oracle evaluations per step

$$\mathcal{I}_\theta \leq_P E_{x,x' \sim Q_1,Q_2}\left\{ \frac{\tilde{p}(x)A(x',x)Q_\theta(x'|x)}{Q_1(x)Q_2(x')} \log \frac{A_\theta(x',x)Q_\theta(x'|x)}{\tilde{p}(x')} \right\} + E_{x\sim Q_1}\left\{ \frac{1 - A_\theta(x^*,x)Q_\theta(x^*|x)}{Q_1(x)} \left[\eta - \eta A - \theta(x^*,x)Q_\theta(x^*|x) - \tilde{p}(x)(\log \eta + 1) \right] \right\}$$

$$Q_1 = \pi U + (1-\pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x^* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

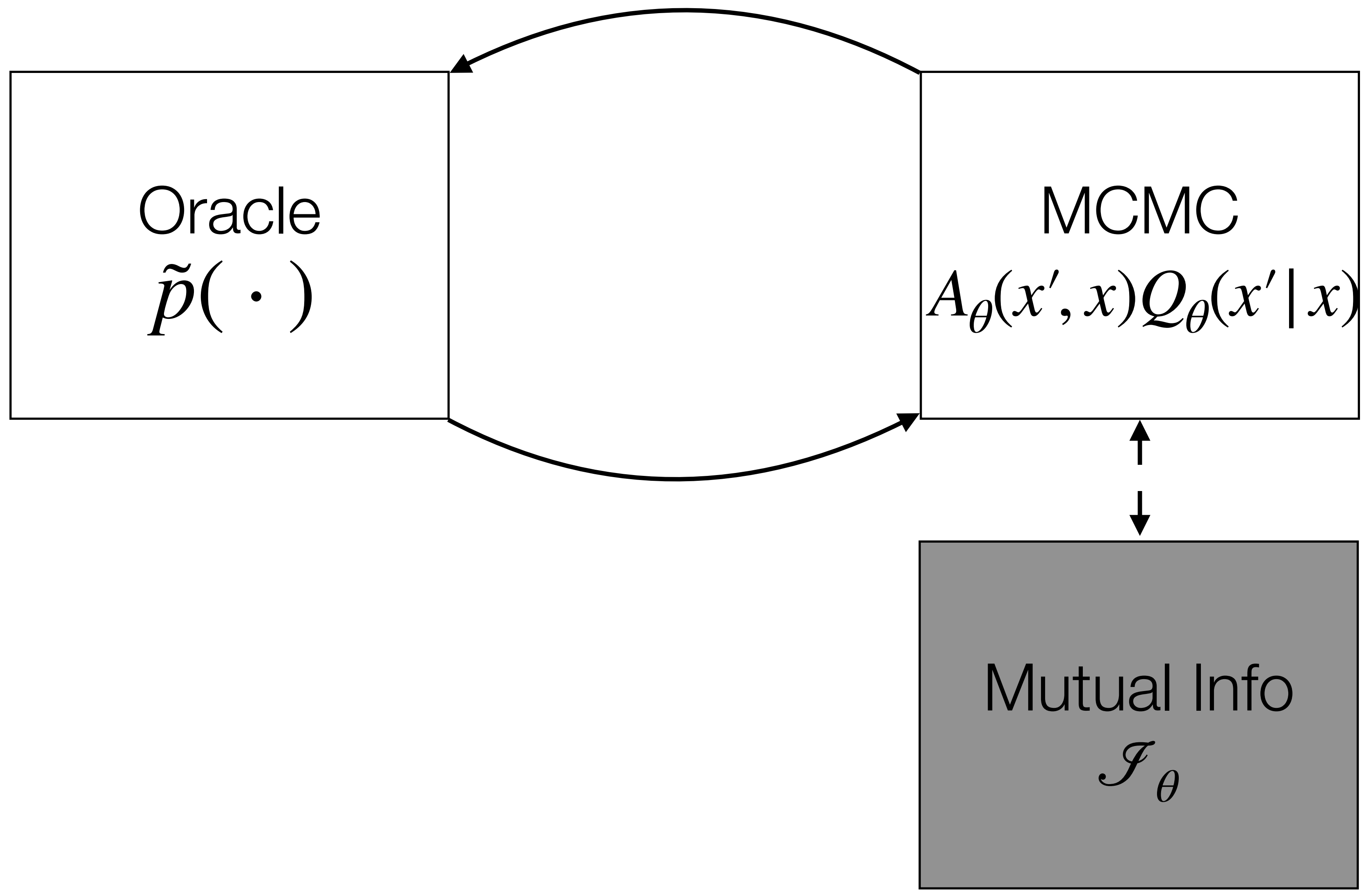
Upper bound estimation requires $O(d)$ oracle evaluations per step

$$\mathcal{I}_\theta \leq_P E_{x,x' \sim Q_1,Q_2} \left\{ \frac{\tilde{p}(x) A(x', x) Q_\theta(x' | x)}{Q_1(x) Q_2(x')} \log \frac{A_\theta(x', x) Q_\theta(x' | x)}{\tilde{p}(x')} \right\} + E_{x \sim Q_1} \left\{ \frac{1 - A_\theta(x^*, x) Q_\theta(x^* | x)}{Q_1(x)} \left[\eta - \eta A - \theta(x^*, x) Q_\theta(x^* | x) - \tilde{p}(x)(\log \eta + 1) \right] \right\}$$

$$Q_1 = \pi U + (1 - \pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x^* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

Upper bound estimation requires $O(d)$ oracle evaluations per step

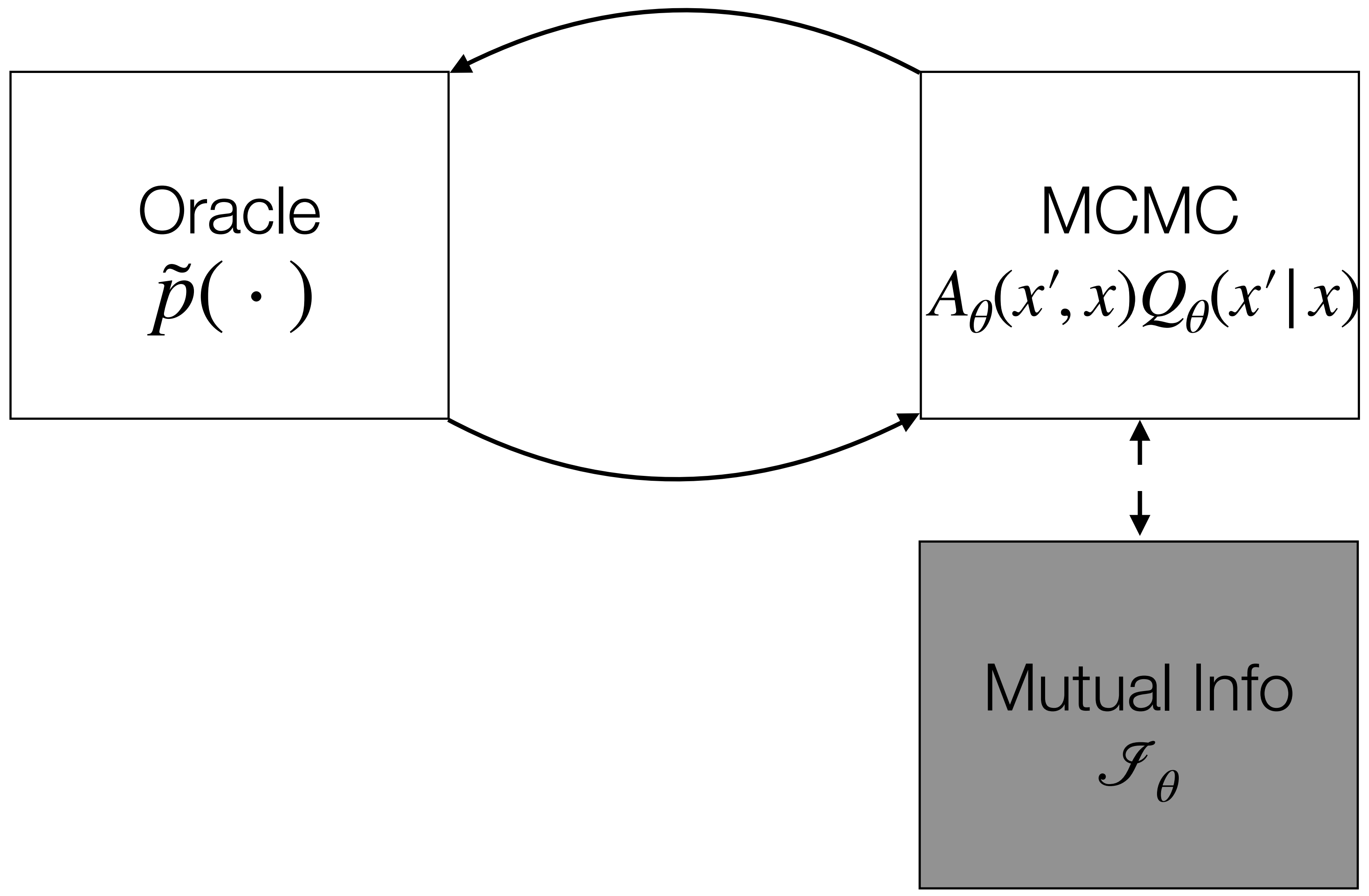
$$\mathcal{I}_\theta \leq_P E_{x,x' \sim Q_1,Q_2}\left\{ \frac{\tilde{p}(x)A(x',x)Q_\theta(x'|x)}{Q_1(x)Q_2(x')} \log \frac{A_\theta(x',x)Q_\theta(x'|x)}{\tilde{p}(x')} \right\} + E_{x \sim Q_1}\left\{ \frac{1 - A_\theta(x*,x)Q_\theta(x*|x)}{Q_1(x)} \left[\eta - \eta A - \theta(x*,x)Q_\theta(x*|x) - \tilde{p}(x)(\log \eta + 1) \right] \right\}$$

Contribution related to acceptance **Contribution related to rejection**

$$Q_1 = \pi U + (1 - \pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

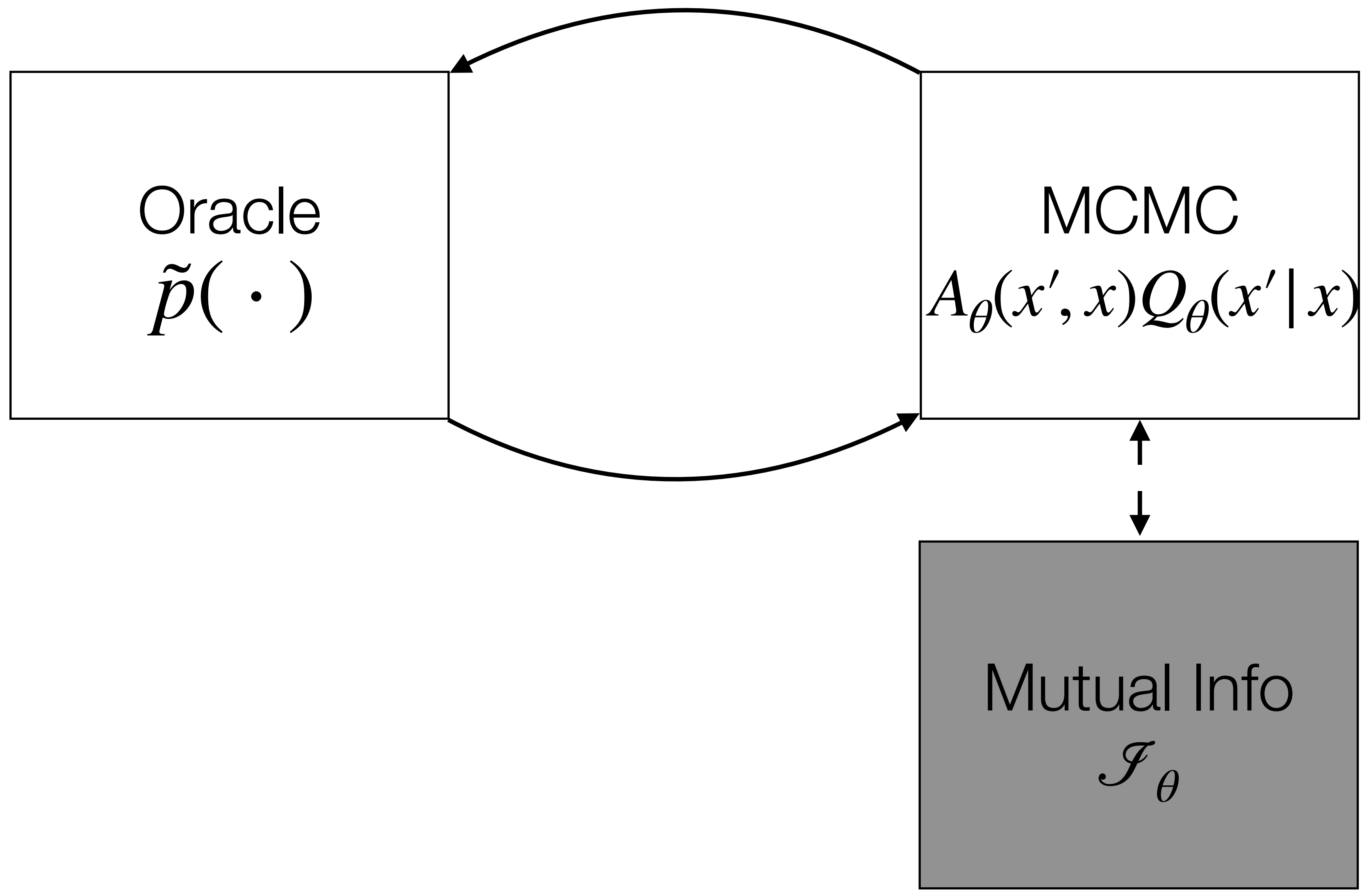
Upper bound estimation requires $O(d)$ oracle evaluations per step

$$\mathcal{I}_\theta \leq_P E_{x,x'\sim Q_1,Q_2}\left\{\frac{\tilde{p}(x)A(x',x)Q_\theta(x'|x)}{Q_1(x)Q_2(x')}\log\frac{A_\theta(x',x)Q_\theta(x'|x)}{\tilde{p}(x')}\right\} + E_{x\sim Q_1}\left\{\frac{1-A_\theta(x^*,x)Q_\theta(x^*|x)}{Q_1(x)}\left[\eta - \eta A - \theta(x^*,x)Q_\theta(x^*|x) - \tilde{p}(x)(\log\eta + 1)\right]\right\}$$

$$Q_1 = \pi U + (1-\pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x^* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

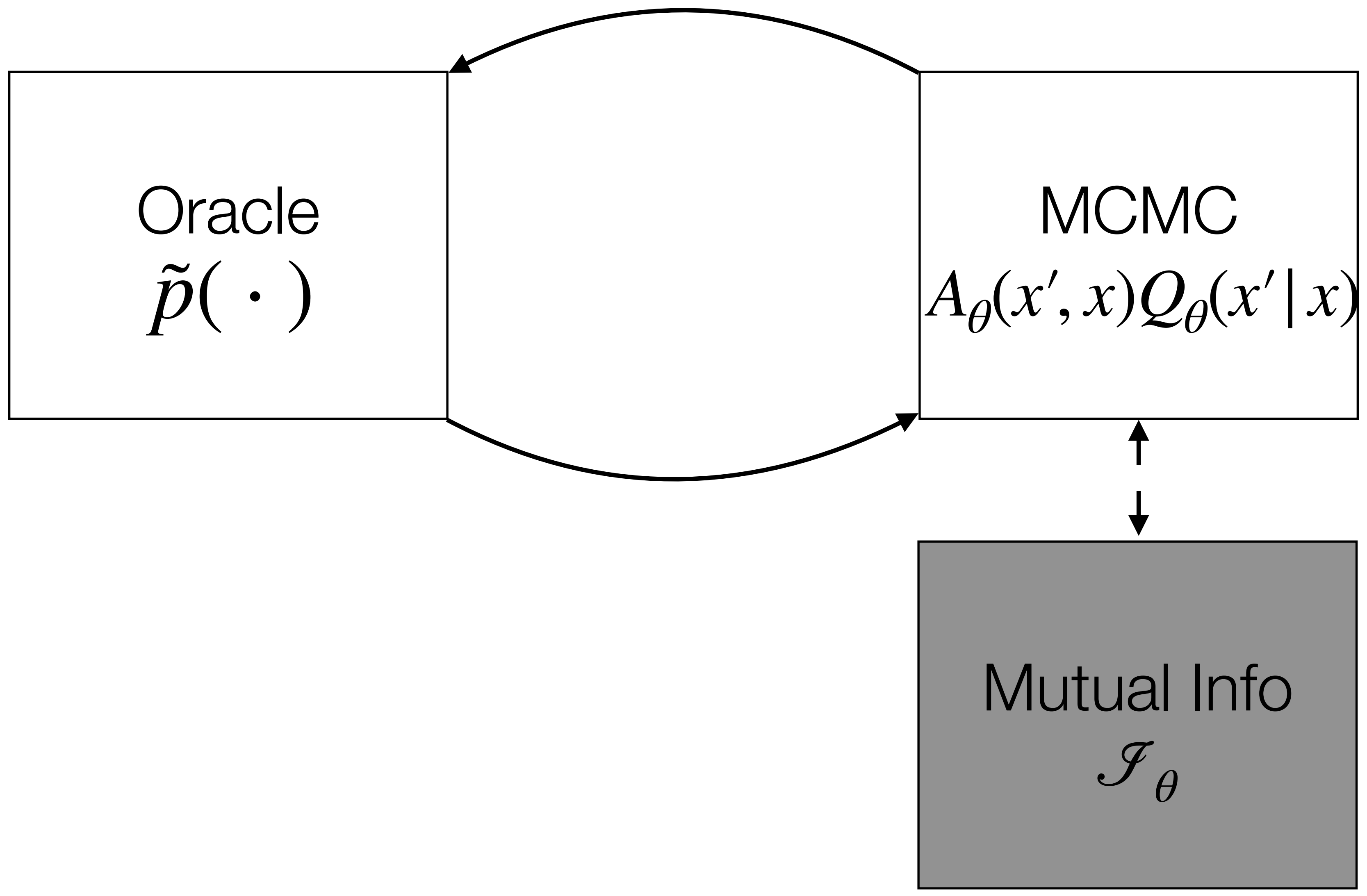
Upper bound estimation requires $O(d)$ oracle evaluations per step

$$\mathcal{I}_\theta \leq_P E_{x,x' \sim Q_1, Q_2} \left\{ \frac{\tilde{p}(x)A(x', x)Q_\theta(x' | x)}{Q_1(x)Q_2(x')} \log \frac{A_\theta(x', x)Q_\theta(x' | x)}{\tilde{p}(x')} \right\} + E_{x \sim Q_1} \left\{ \frac{1 - A_\theta(x^*, x)Q_\theta(x^* | x)}{Q_1(x)} \left[\eta - \eta A - \theta(x^*, x)Q_\theta(x^* | x) - \tilde{p}(x)(\log \eta + 1) \right] \right\}$$

$$Q_1 = \pi U + (1 - \pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x^* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

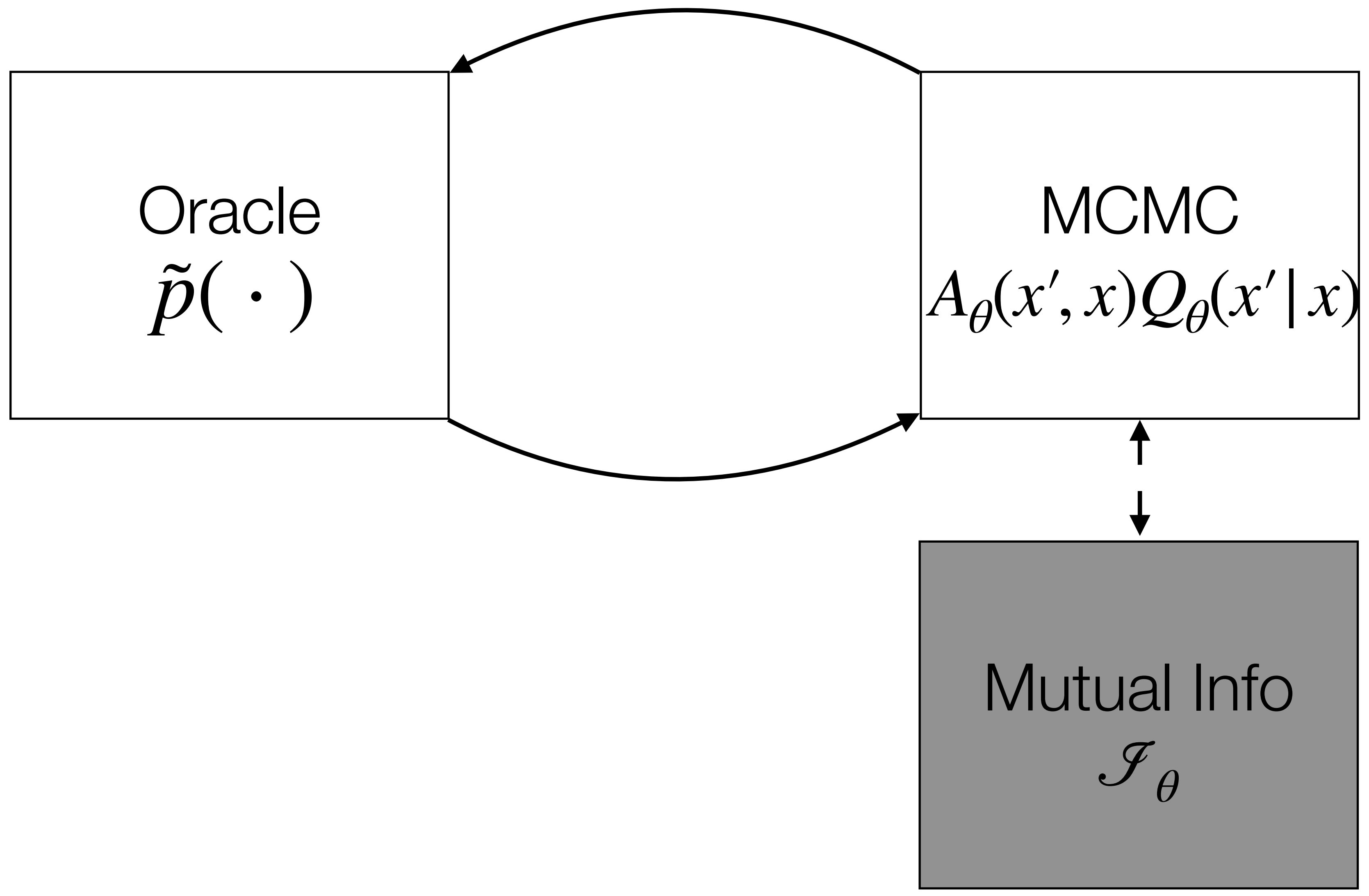
Upper bound estimation requires $O(d)$ oracle evaluations per step

$$\mathcal{I}_\theta \leq_P E_{x,x'\sim Q_1,Q_2}\left\{\frac{\tilde{p}(x)A(x',x)Q_\theta(x'|x)}{Q_1(x)Q_2(x')}\log\frac{A_\theta(x',x)Q_\theta(x'|x)}{\tilde{p}(x')}\right\} + E_{x\sim Q_1}\left\{\frac{1-A_\theta(x*,x)Q_\theta(x*|x)}{Q_1(x)}\left[\eta-\eta A-\theta(x*,x)Q_\theta(x*|x)-\tilde{p}(x)(\log\eta+1)\right]\right\}$$

$$Q_1 = \pi U + (1-\pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

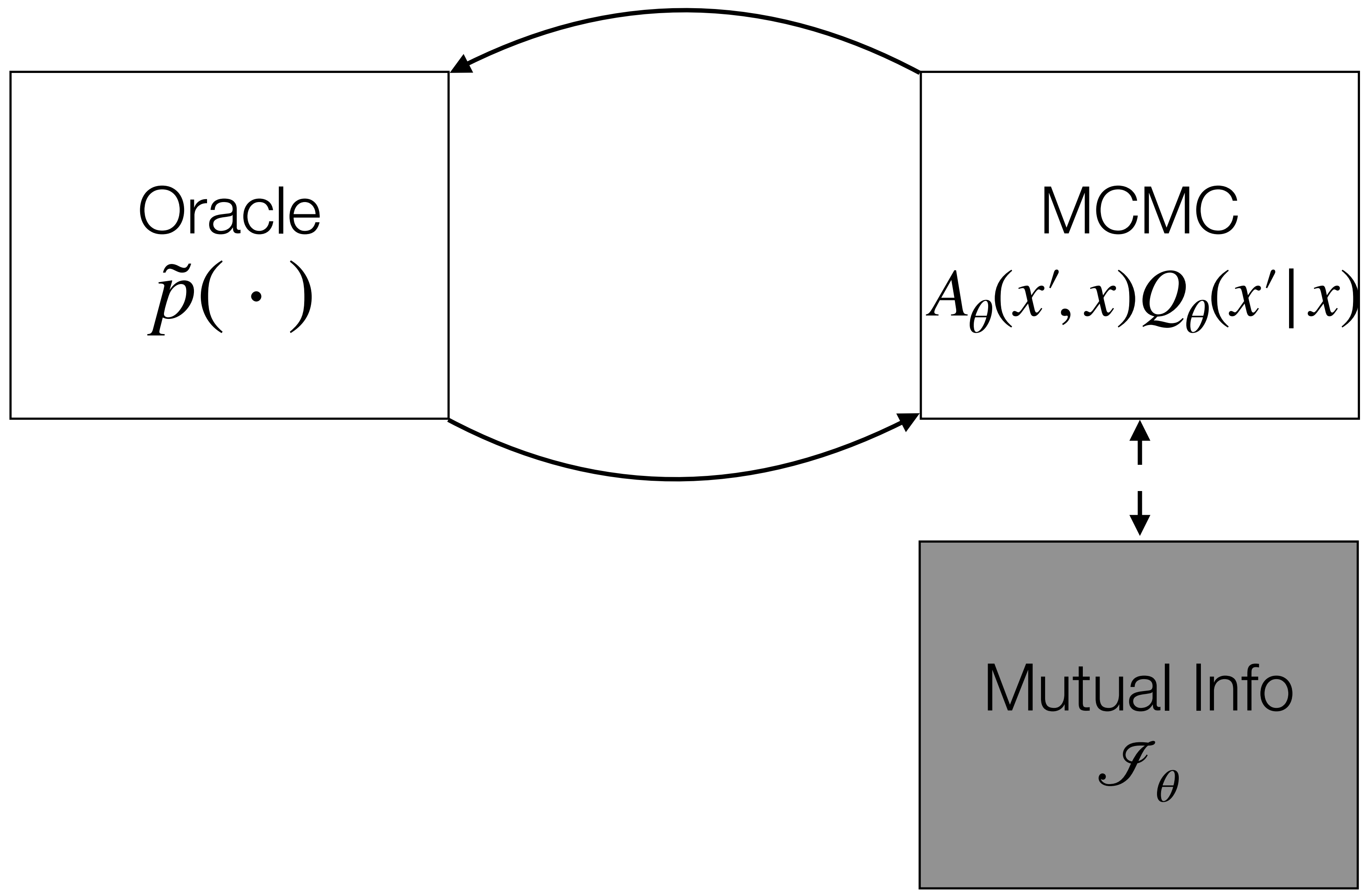
Upper bound estimation requires $O(d)$ oracle evaluations per step

$$\mathcal{I}_\theta \leq_P E_{x,x' \sim Q_1,Q_2} \left\{ \frac{\tilde{p}(x)A(x',x)Q_\theta(x'|x)}{Q_1(x)Q_2(x')} \log \frac{A_\theta(x',x)Q_\theta(x'|x)}{\tilde{p}(x')} \right\} + E_{x \sim Q_1} \left\{ \frac{1 - A_\theta(x^*,x)Q_\theta(x^*|x)}{Q_1(x)} \left[\eta - \eta A - \theta(x^*,x)Q_\theta(x^*|x) - \tilde{p}(x)(\log \eta + 1) \right] \right\}$$

$$Q_1 = \pi U + (1-\pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x^* \sim U_{N(x)}$$

Mutual Information

$$p(x) = \frac{\tilde{p}(x)}{P}$$



Theorem. Exact estimation of $\mathcal{I}_\theta$ requires $O(d^2)$ oracle evaluations per step

Upper bound estimation requires $O(d)$ oracle evaluations per step

$$\mathcal{I}_\theta \leq_P E_{x,x' \sim Q_1, Q_2}\left\{ \frac{\tilde{p}(x)A(x', x)Q_\theta(x' | x)}{Q_1(x)Q_2(x')} \log \frac{A_\theta(x', x)Q_\theta(x' | x)}{\tilde{p}(x')} \right\} + E_{x \sim Q_1}\left\{ \frac{1 - A_\theta(x^*, x)Q_\theta(x^* | x)}{Q_1(x)} \left[\eta - \eta A - \theta(x^*, x)Q_\theta(x^* | x) - \tilde{p}(x)(\log \eta + 1) \right] \right\}$$

$$Q_1 = \pi U + (1 - \pi)\delta_x \qquad Q_2 = Q_{stop(\theta)} \qquad x^* \sim U_{N(x)}$$

Experiments on Restricted Boltzmann Machines (II)

$$\tilde{p}(x) = e^{-b-\sum_i \alpha_i x_i - \sum_{(i,j)} W_{ij} x_i x_j}$$

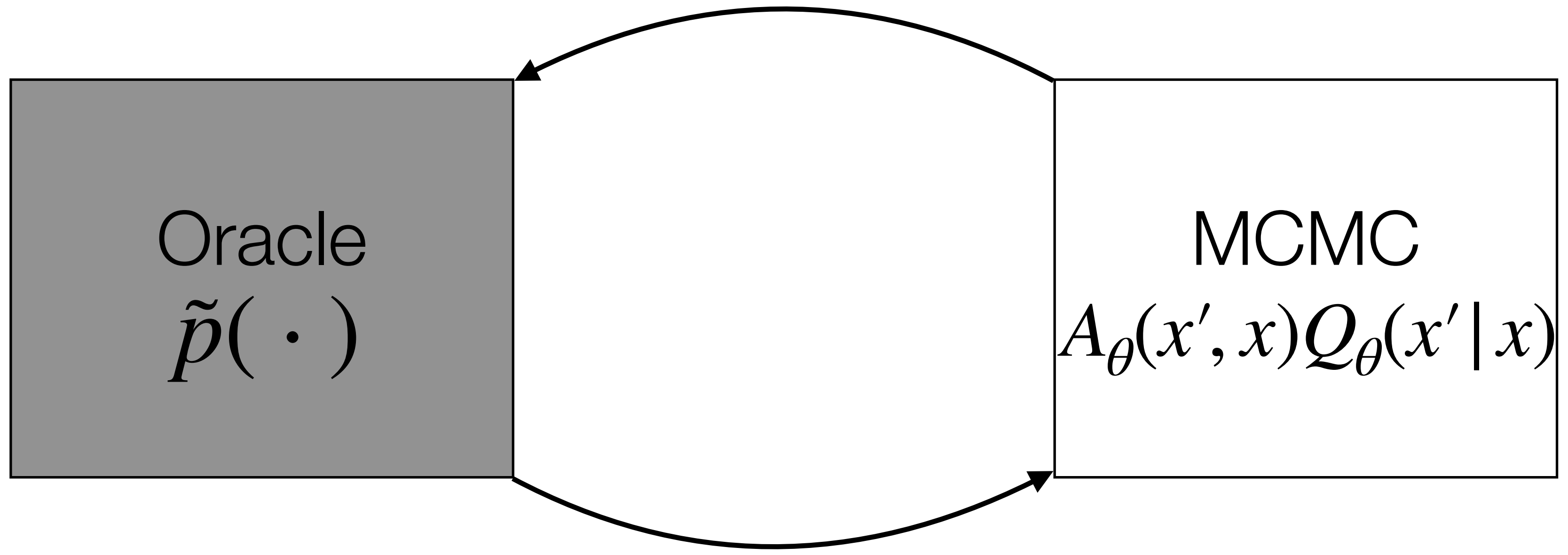
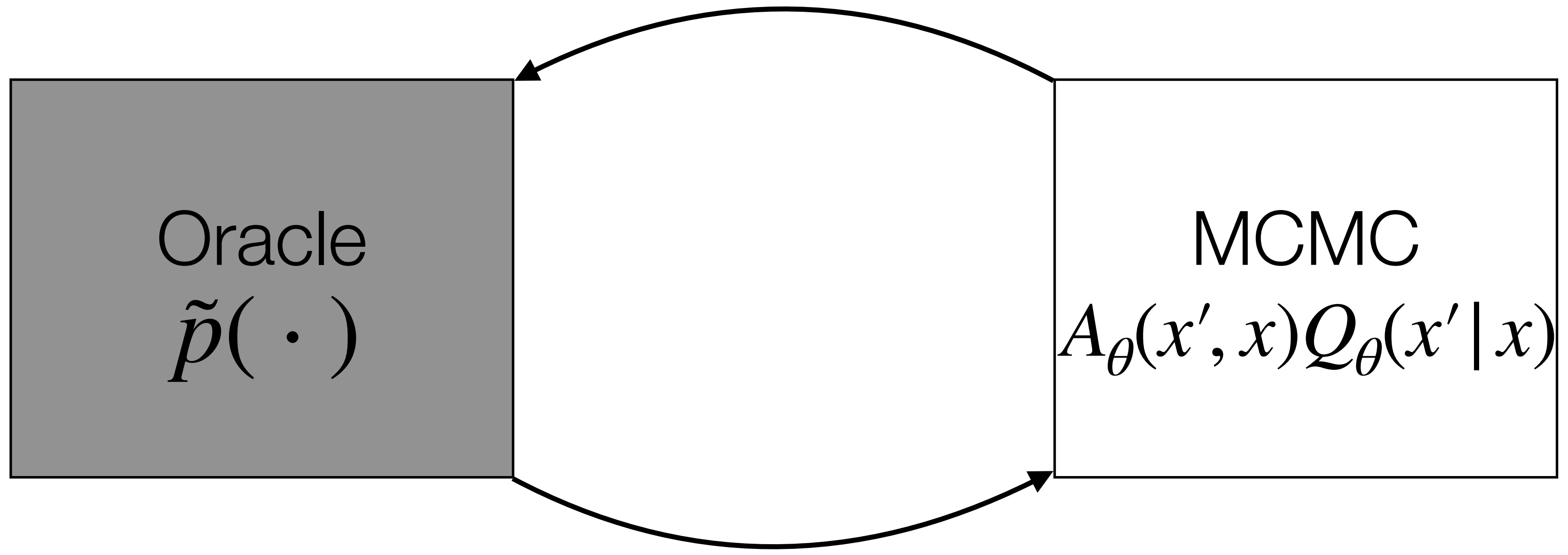


Image data

Experiments on Restricted Boltzmann Machines (II)

$$\tilde{p}(x) = e^{-b - \sum_i \alpha_i x_i - \sum_{(i,j)} W_{ij} x_i x_j}$$



Comparison with Gibbs sampling, Hamming Ball sampler and Gibbs-With-Gradients [Grathwohl et al., ICML 2021]

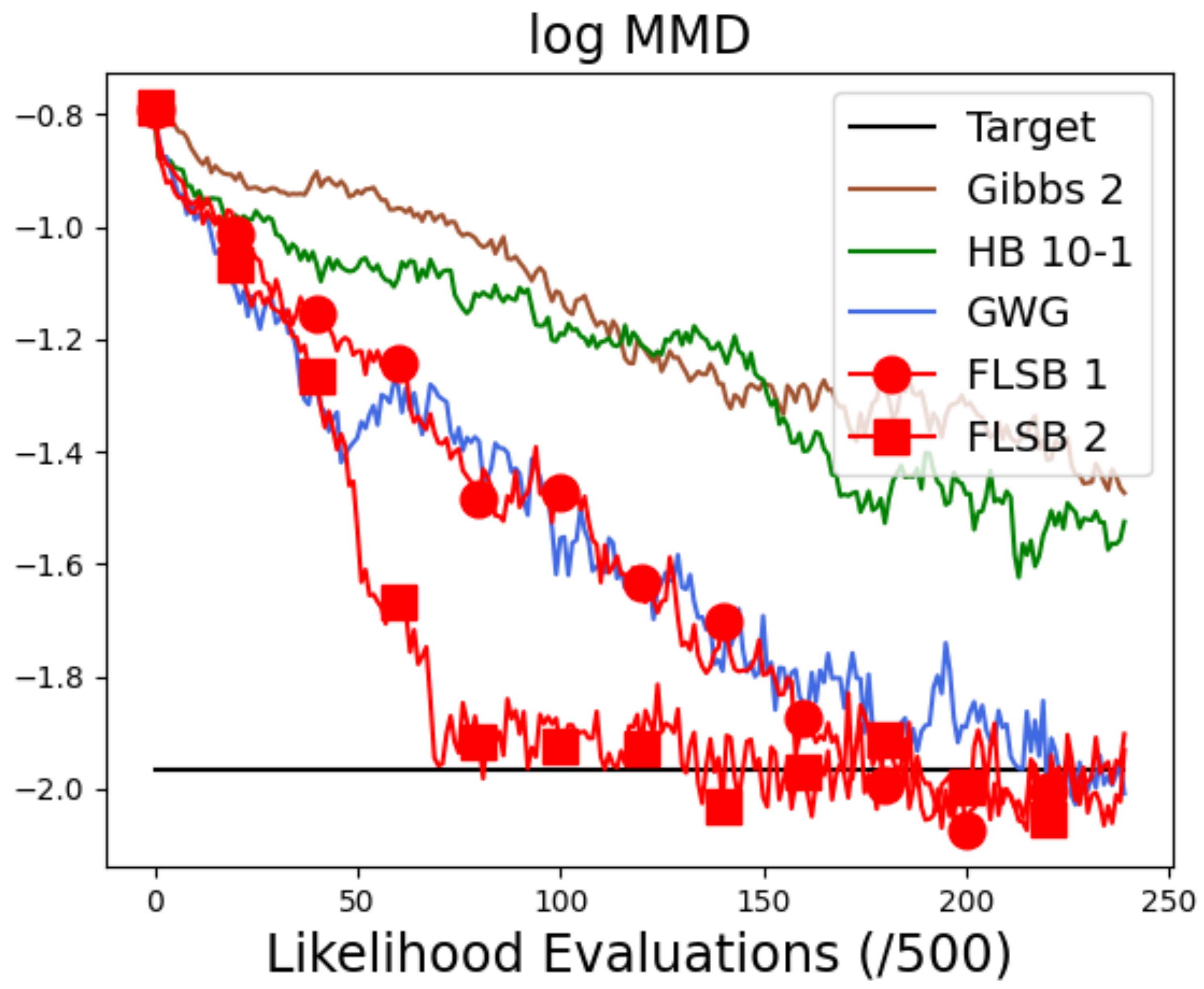
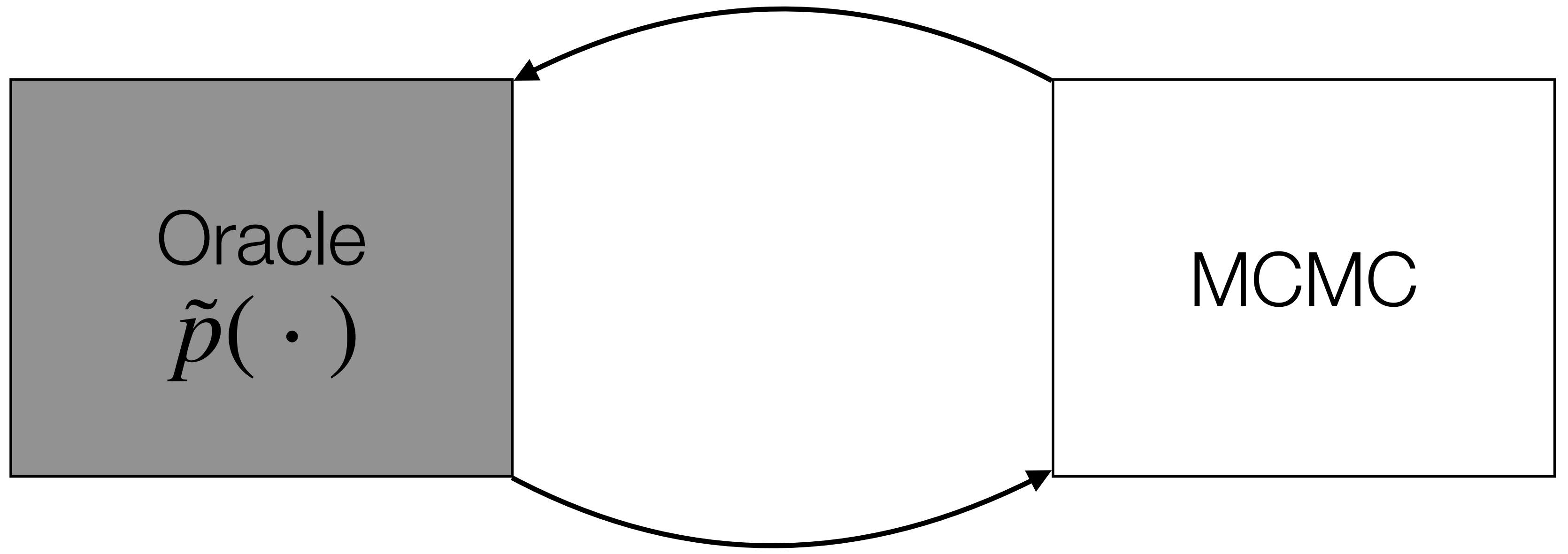


Image data

Experiments on Restricted Boltzmann Machines (II)

$$\tilde{p}(x) = e^{-b - \sum_i \alpha_i x_i - \sum_{(i,j)} W_{ij} x_i x_j}$$



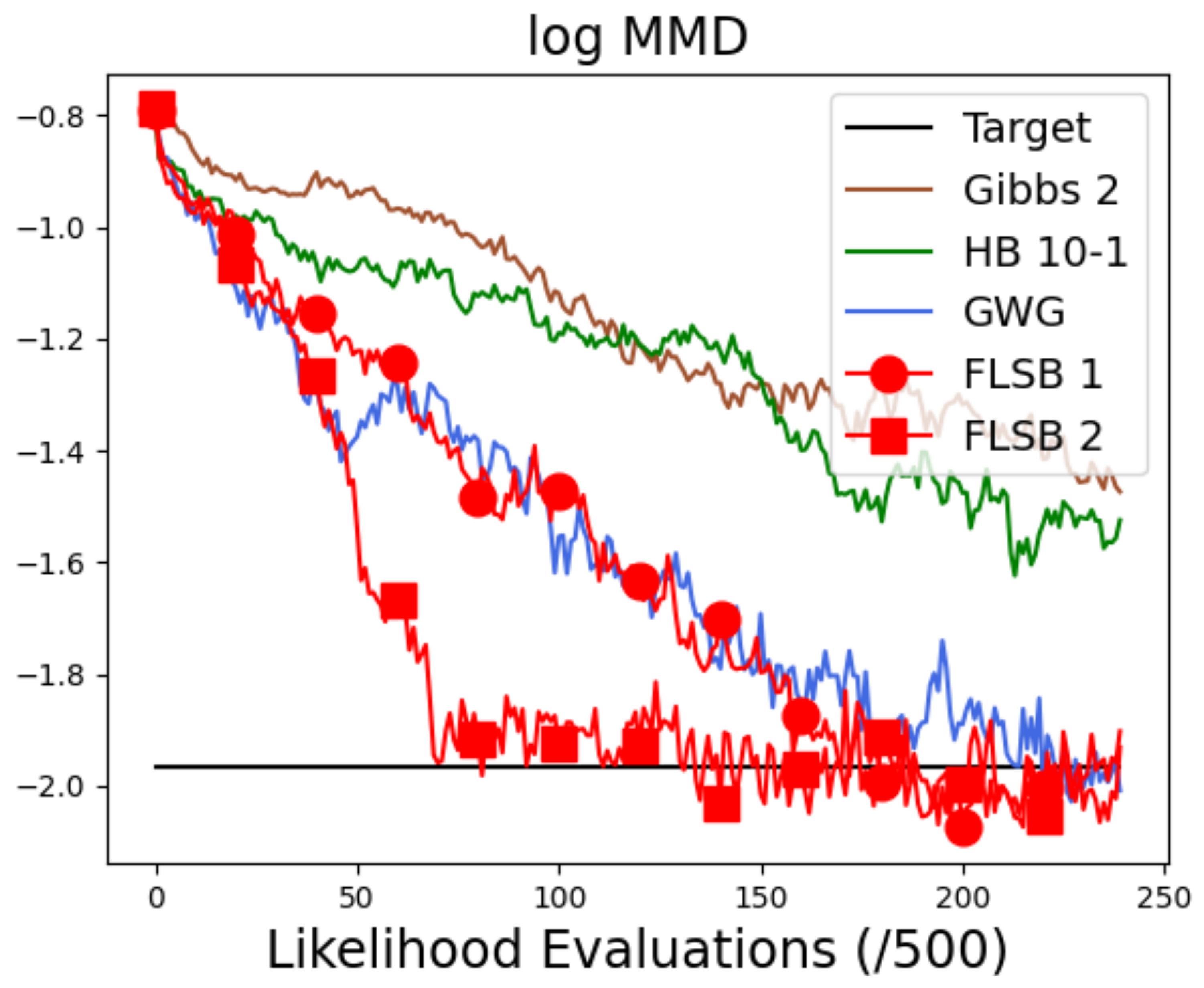
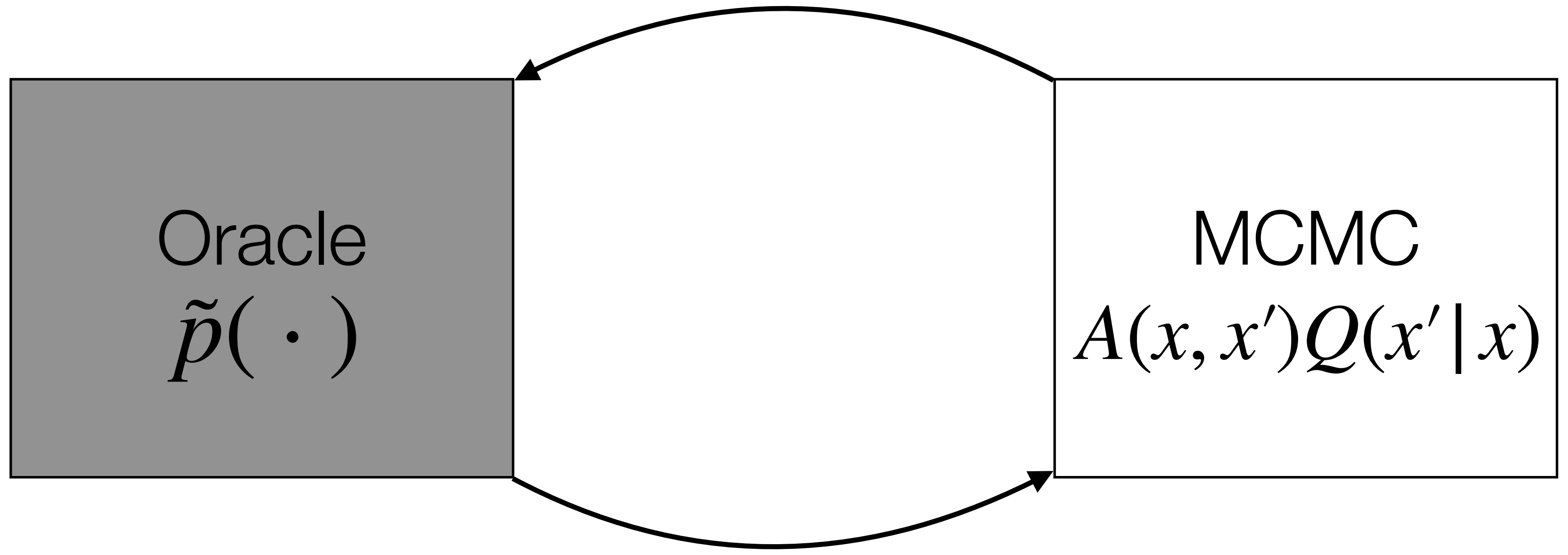
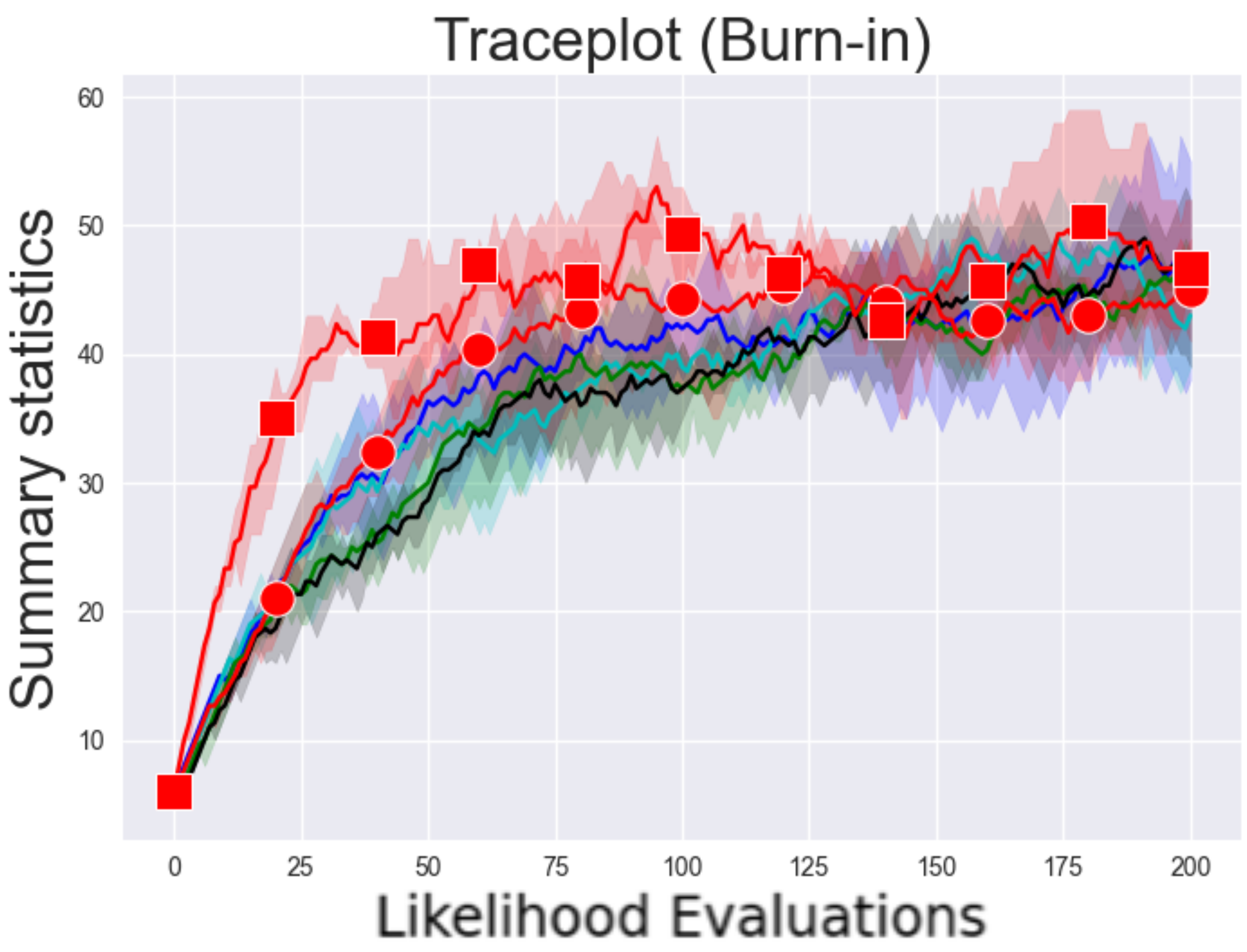


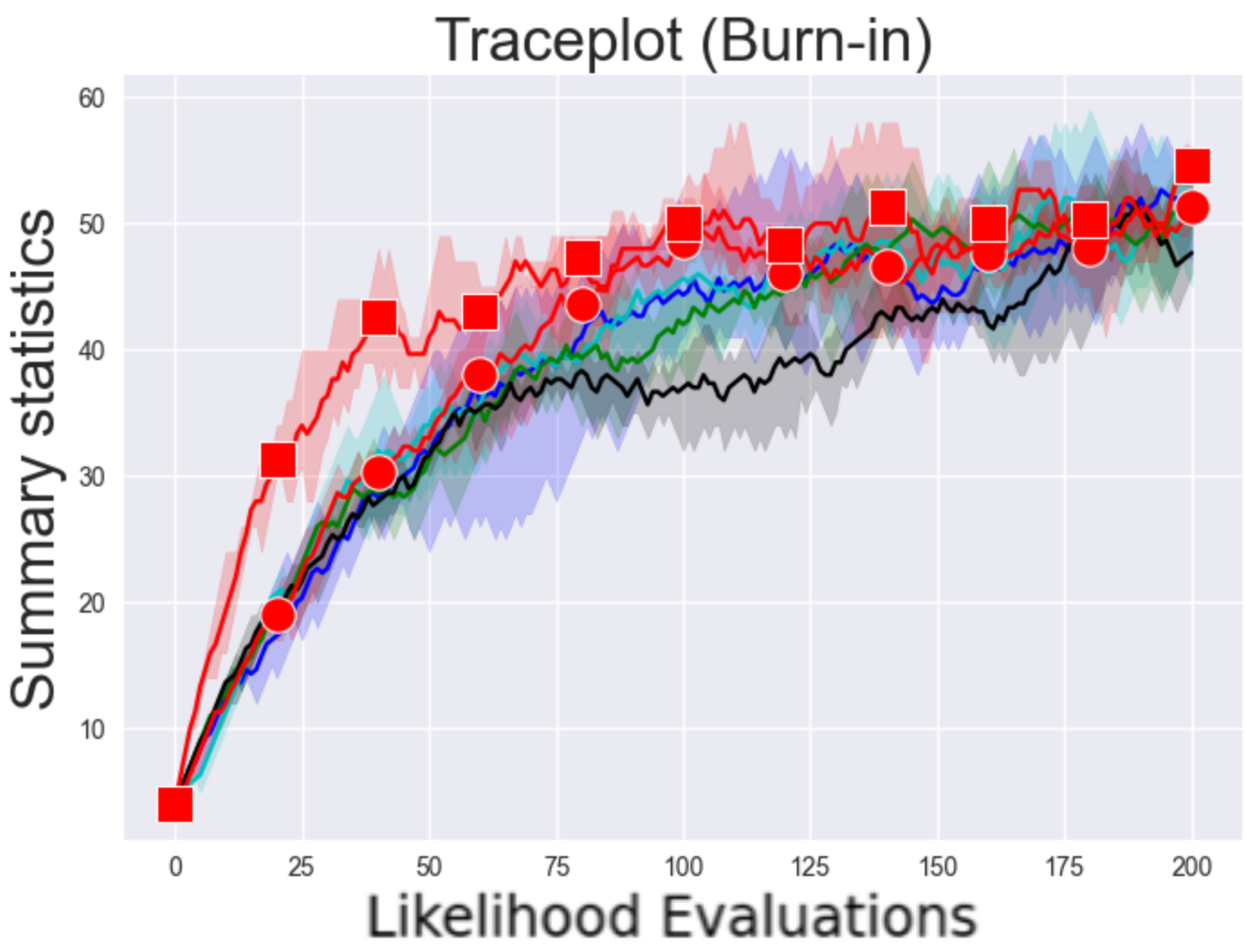
Image data

Experiments on Markov Networks (III)

$$\tilde{p}(x) = \prod_k \phi(x_{\{k\}})$$







Graph data

Comparison with [Zanella, JASA 2020]

Additional Experiments

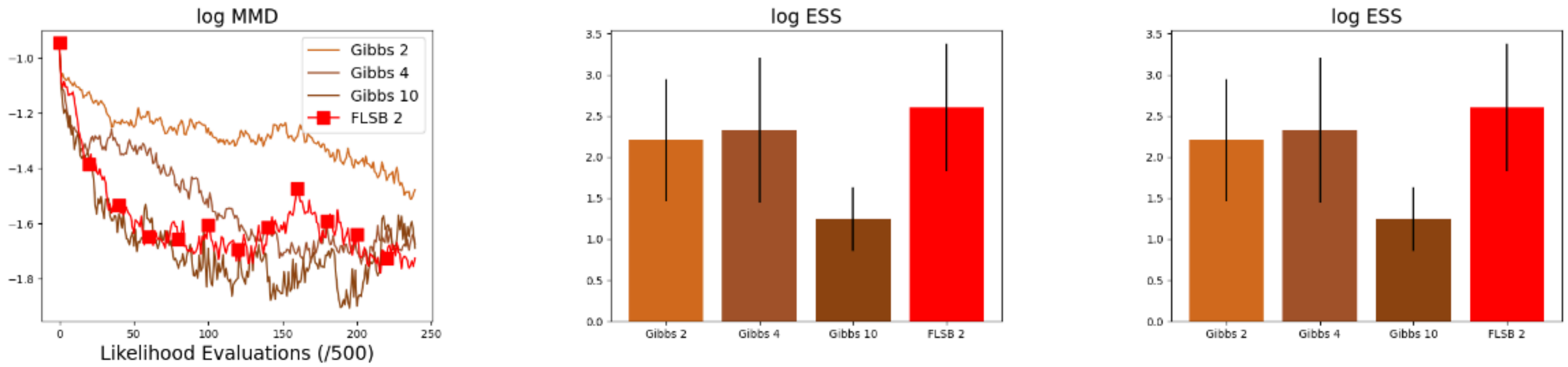


Figure 8. Comparison of FLSB 2 against block Gibbs sampling with block size of 2, 4 and 10 variables on RBMs.

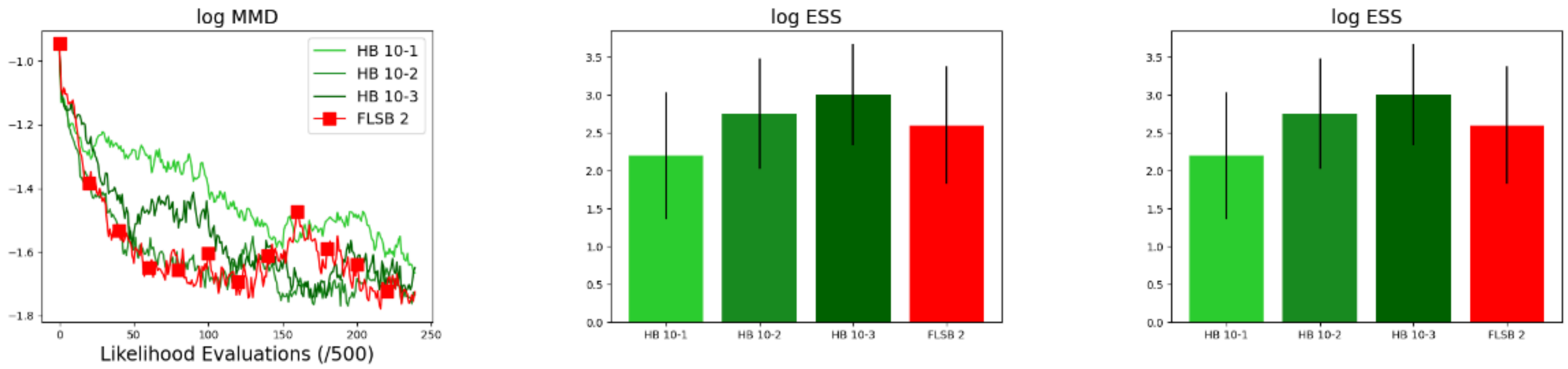


Figure 9. Comparison of FLSB 2 against the Hamming Ball sampler using 10 variables per block and updating 1, 2 and 3 variables per step.

In Numbers

Table 4. Summary of the properties of different approaches.

Method	Target likelihood evaluations per sampling step	Number of variables modified per sampling step
Gibbs 2	4	2
Gibbs 4	16	4
Gibbs 10	1024	10
HB-10-1	20	1
HB-10-2	180	2
HB-10-3	960	3
FLSB 2	**1**	1