Nesting Probabilistic Programs

Tom Rainforth

With help from Rob Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood

UNIVERSITY OF
OXFORD

UNIVERSITY OF OXFORD

Probabilistic programming presents and opportunity to nest queries

```
(defquery outer [D]
 (let [y (sample (beta 2 3))
       dist (conditional inner)
       z (sample (dist y D))]
  (* y z)))
```

```
(defquery inner [y D]
 (let [z (sample (gamma y 1))]
  (observe (normal y z) D)
  z))
```

Nested Expectations

$$\gamma_0 = \mathbb{E}\left[f_0\left(y, \mathbb{E}\left[f_1\left(y, z\right)|y\right]\right)\right]$$

$$\gamma_0 = \mathbb{E}[f_0(y, \mathbb{E}[f_1(y, z, \mathbb{E}[f_2(y, z, u)|y, z])|y])]$$

UNIVERSITY OF OXFORD

There are problems we cannot encode without nesting

- E.g. entropy of KL divergence of a marginal distribution:

$$H\left(p(y)\right) = -\mathbb{E}\left[\log\left(\mathbb{E}\left[p(y|\theta)|y\right]\right)\right]$$

UNIVERSITY OF OXFORD

Agents reasoning about other agents

Bad cards, should probably fold

Can I bluff him though? I doubt he has good cards

Cards are ok, but if he bets I'll fold

He might be bluffing though…

Aims

- When does nesting probabilistic programs no longer give a standard Bayesian model?
- What are the statistical implications?
- How can we construct valid inference engines?

Three Ways We Might Nest

- Nested conditioning
- Nested inference
- Estimates as first class variables

Nested Conditioning

```
(defquery outer [D]
 (let [y (sample (beta 2 3))]
  (observe (nest inner [y D] :smc 100) nil)
  y))
```

UNIVERSITY OF OXFORD

Formalizing Nested Conditioning

$$\pi_o(y) = \boxed{\psi(y)}\,\boxed{p_i(y)} \approx \psi(y)\ \hat{p}_i(y)$$

Density terms from outer

Marginal likelihood for inner

UNIVERSITY OF OXFORD

Formalizing Nested Conditioning

$$\mathbb{E}_{q(y)}\left[\frac{\psi(y)\hat{p}_i(y)}{q(y)}\right] = \mathbb{E}_{q(y)}\left[\frac{\pi_o(y)}{q(y)}\right]$$

Nested Inference

```
(defquery inner [y D]
 (let [z (sample (gamma y 1))]
  (observe (normal y z) D)
  z))


(defquery outer [D]
 (let [y (sample (beta 2 3))
       dist (conditional inner)
       z (sample (dist y D))]
  (* y z)))
```

UNIVERSITY OF OXFORD

Formalizing Nested Inference

$$\pi_o(y,z) = \psi(y,z)p_i(z|y)$$

$$p_i(z|y) = \frac{\pi_i(y,z)}{\int \pi_i(y,z')dz'}$$

$$\pi_o(y,z) = \frac{\psi(y,z)\pi_i(y,z)}{\int \pi_i(y,z')dz'}$$

UNIVERSITY OF OXFORD

Formalizing Nested Inference

$$\pi_o(y,z) \approx \psi(y,z)\hat{p}_i(z|y)$$

$$\mathbb{E}_{q(y,z)}\left[\frac{\pi_o(y,z)}{q(y,z)}\right] = \mathbb{E}_{q(y,z)}\left[\frac{\psi(y,z)\pi_i(y,z)}{q(y,z)\mathbb{E}_{z'\sim q(z|y)}\left[\pi_i(y,z')/q(z'|y)\right]}\right]$$

UNIVERSITY OF OXFORD

Estimates as First Class Variables

```
(defm prior [] (normal 0 1))
(defm lik [theta d] (normal theta d))

(defquery inner-q [y d]
 (let [theta (sample (prior))]
  (observe (lik theta d) y)))

(defn inner-E [y d M]
 (->> (doquery :importance
        inner-q [y d])
      (take M)
      log-marginal))

(with-primitive-procedures [inner-E]
```

```
 (defquery outer-q [d M]
  (let [theta (sample (prior))
        y (sample (lik theta d))
        log-lik (observe*
                  (lik theta d) y)
        log-marg (inner-E y d M)]
   (- log-lik log-marg))))

(defn outer-E [d M N]
 (->> (doquery :importance
        outer-q [d M])
      (take N)
      collect-results
      empirical-mean))
```

UNIVERSITY OF OXFORD

Estimates as First Class Variables

$$\pi_0(y) = \psi(y, z(y)) \approx \psi(y, \hat{z}(y))$$

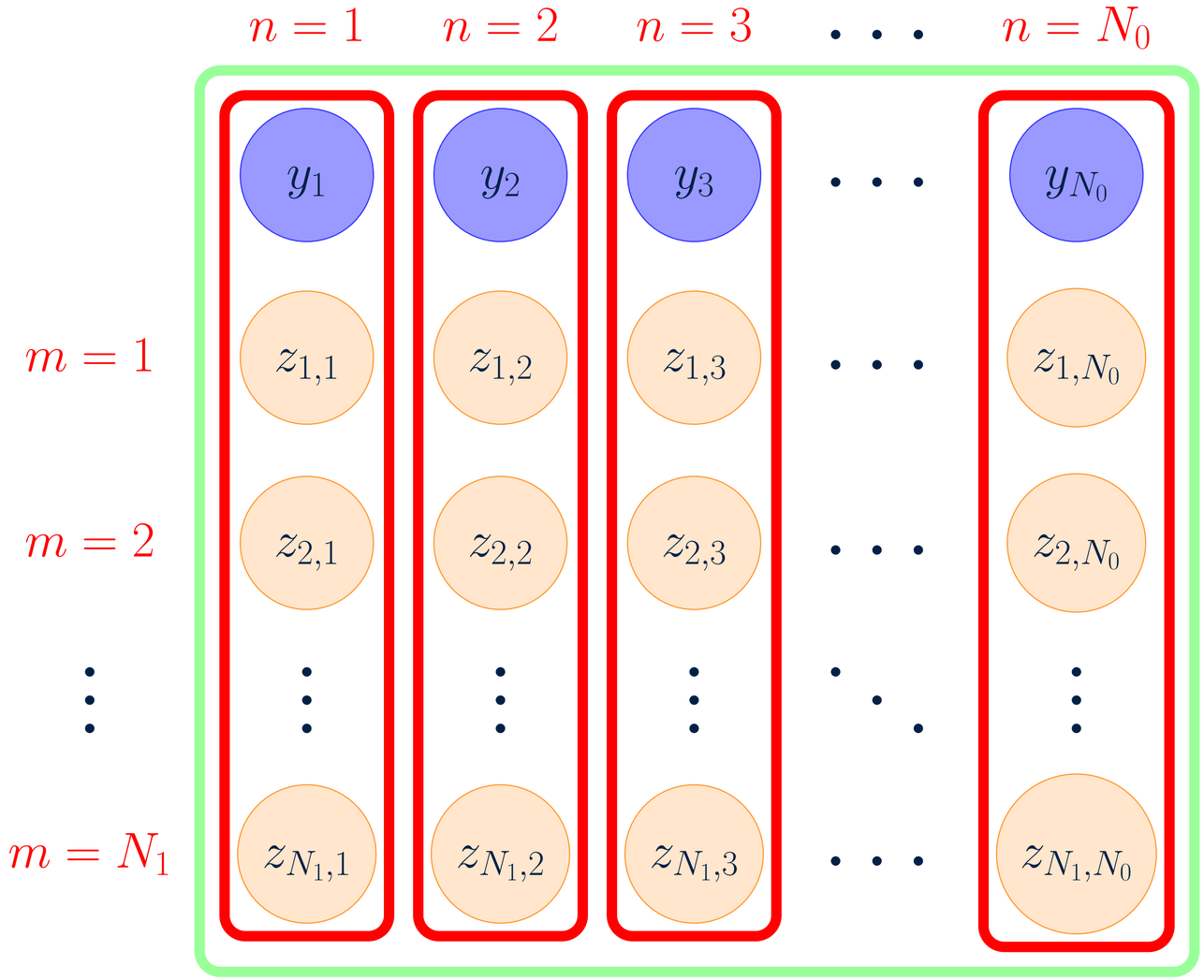
$$\mathbb{E}_{q(y)}\left[\frac{\pi_0(y)}{q(y)}\right] = \mathbb{E}_{q(y)}\left[\frac{\psi(y, \mathbb{E}\left[\hat{z}(y)|y\right])}{q(y)}\right]$$

Nested Monte Carlo

$$I_0 = \frac{1}{N_0} \sum_{n=1}^{N_0} f_0\left(y_n, I_1^n(y_n)\right) \quad \text{where} \quad y_n \sim p(y)$$

$$I_1^n(y_n) = \frac{1}{N_1} \sum_{m=1}^{N_1} f_1\left(y_n, z_{m,n}\right) \quad \text{where} \quad z_{m,n} \sim p(z|y = y_n)$$

[Rainforth, Cornish, Yang, Warrington, and Wood. On the Opportunities and Pitfalls of Nesting Monte Carlo Estimators. ICML 2018]


$n = 1$
$n = 2$
$n = 3$
$n = N_0$
y_1
y_2
y_3
y_{N_0}
$m = 1$
$m = 2$
$m = N_1$
$z_{1,1}$
$z_{1,2}$
$z_{1,3}$
z_{1,N_0}
$z_{2,1}$
$z_{2,2}$
$z_{2,3}$
z_{2,N_0}
$z_{N_1,1}$
$z_{N_1,2}$
$z_{N_1,3}$
z_{N_1,N_0}

UNIVERSITY OF OXFORD

Convergence Rate

- If each f_k is continuously differentiable, the MSE converges at a rate

$$O\left(\frac{1}{N_0} + \left(\sum_{k=1}^{D} \frac{1}{N_k}\right)^2\right)$$

Variance

Bias Squared

- We need each $N_k \to \infty$ for convergence

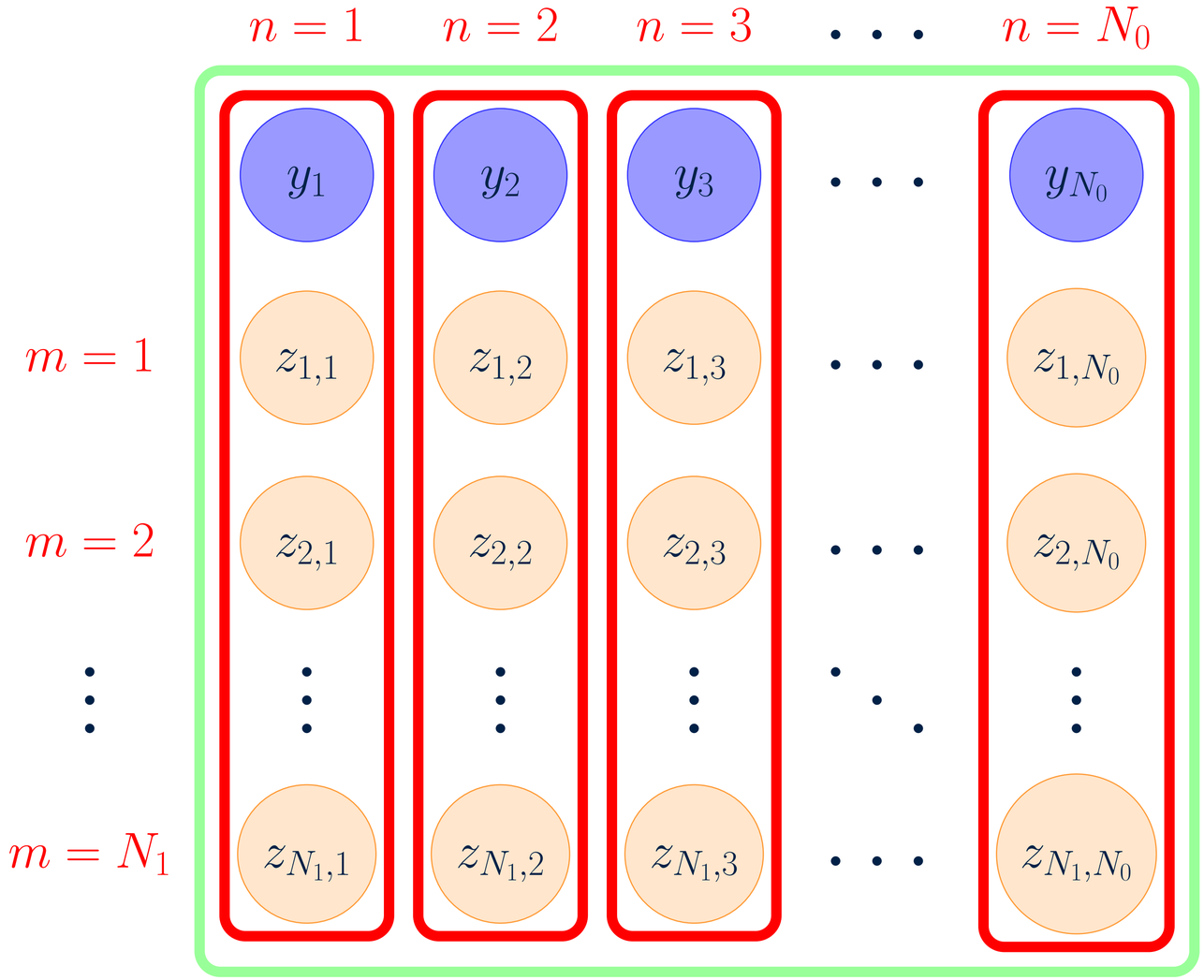
Optimal Setup

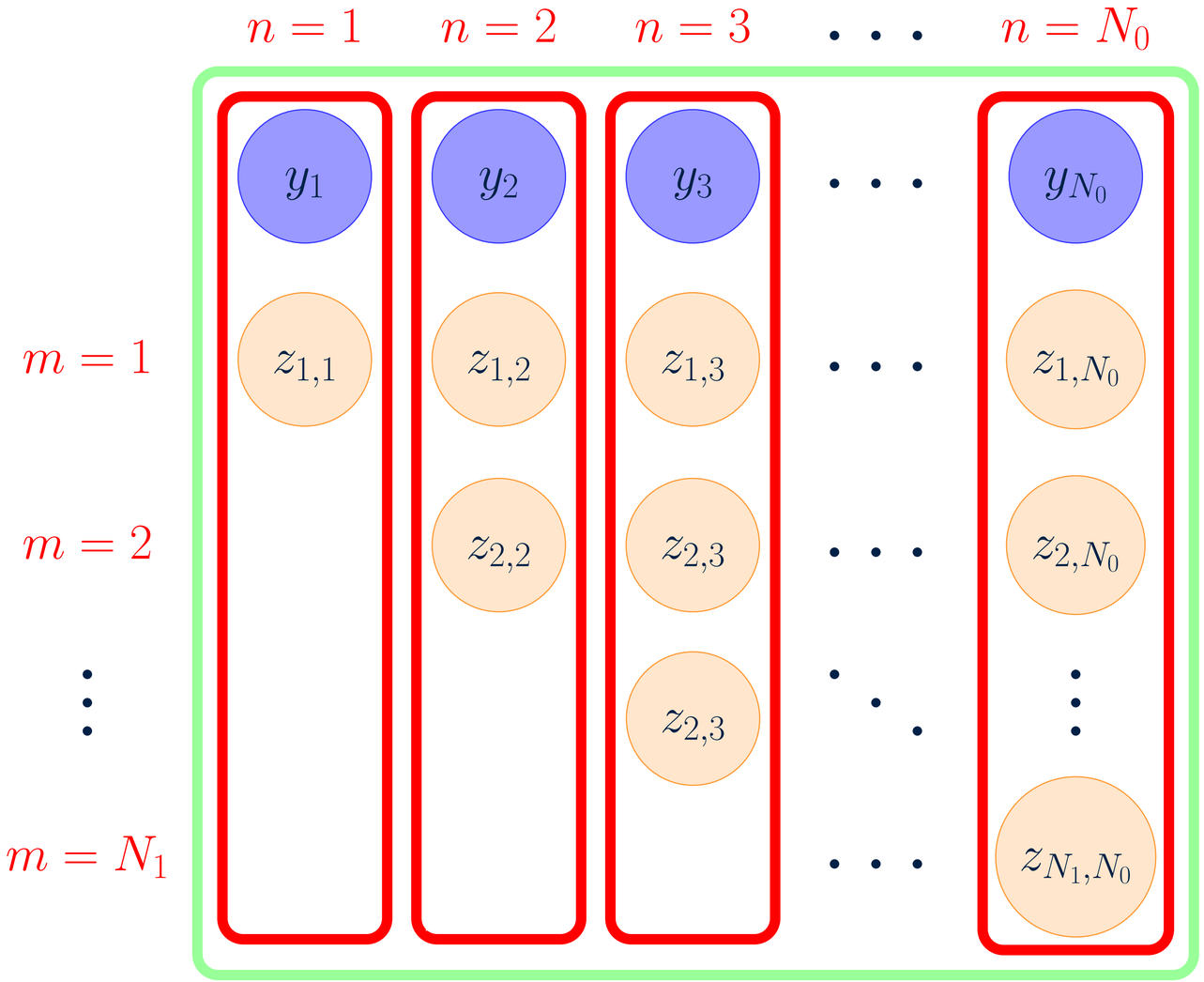
- The asymptotic mean squared error is minimised when

$$\sqrt{N_0} \propto N_1 \propto N_2 \propto \cdots \propto N_D$$

- This yields an overall convergence rate of $O\left(\frac{1}{T^{\frac{2}{2+D}}}\right)$

$$D = 0,\ O\left(\frac{1}{T}\right); \quad D = 1,\ O\left(\frac{1}{T^{\frac{2}{3}}}\right); \quad D = 2,\ O\left(\frac{1}{T^{\frac{1}{2}}}\right)$$


$n = 1$
$n = 2$
$n = 3$
$n = N_0$
y_1
y_2
y_3
y_{N_0}
$m = 1$
$m = 2$
$m = N_1$
$z_{1,1}$
$z_{1,2}$
$z_{1,3}$
z_{1,N_0}
$z_{2,1}$
$z_{2,2}$
$z_{2,3}$
z_{2,N_0}
$z_{N_1,1}$
$z_{N_1,2}$
$z_{N_1,3}$
z_{N_1,N_0}


$n = 1$
$n = 2$
$n = 3$
$n = N_0$
y_1
y_2
y_3
y_{N_0}
$m = 1$
$m = 2$
$m = N_1$
$z_{1,1}$
$z_{1,2}$
$z_{1,3}$
z_{1,N_0}
$z_{2,2}$
$z_{2,3}$
$z_{2,3}$
z_{2,N_0}
z_{N_1,N_0}

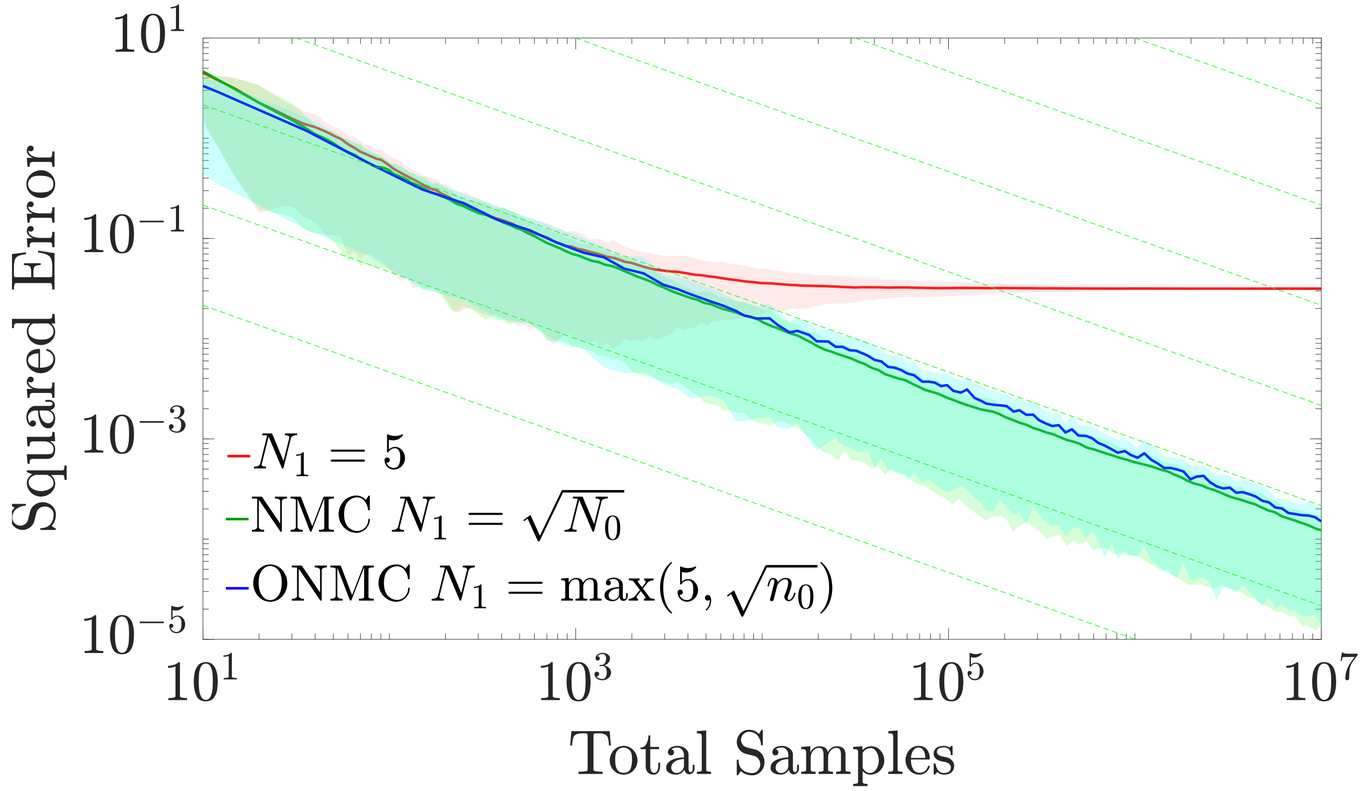
UNIVERSITY OF OXFORD

Online NMC

$$\mathrm{Var}[\mathrm{ONMC}] \to c\,\mathrm{Var}[\mathrm{NMC}] \quad \text{where} \quad c < 1$$
$$\mathrm{Bias}[\mathrm{ONMC}] \to d\,\mathrm{Bias}[\mathrm{NMC}] \quad \text{where} \quad d < 2$$

UNIVERSITY OF OXFORD

Online NMC



Take Homes

- Nesting probabilistic programs allows encoding of models we could not otherwise express
  - Allows encoding of nested expectations
- We need to be careful in the design of our inference engines
- Online NMC gives an easy way of making sure consistency is maintained

UNIVERSITY OF OXFORD

Thanks!

Rob Cornish

Hongseok Yang

Andrew Warrington

Frank Wood

[Rainforth, Cornish, Yang, Warrington, and Wood. On the Opportunities and Pitfalls of Nesting Monte Carlo Estimators. ICML 2018]

[Rainforth. Nesting Probabilistic Programs. UAI 2018]

Contact: rainforth@stats.ox.ac.uk