Measuring the non-asymptotic convergence of sequential Monte Carlo samplers using probabilistic programming

Marco F. Cusumano-Towner and Vikash K. Mansinghka

Probabilistic modules: an interface for stochastic computations with auxiliary random choices

An existing interface for stochastic computations that enables estimation of KL divergences, but requires a log-density:

$$z \leftarrow p.\text{SIMULATE}() \text{ for } z \sim p(z)$$

$$\log p(z) \leftarrow p.\text{LOGPDF}(z)$$

The new probabilistic modules interface, a generalization for stochastic computations with uncollapsed auxiliary random choices u:

$$\left(z, \log \frac{p(u,z)}{q(u;z)}\right) \leftarrow (p,q).\text{SIMULATE}() \text{ for } u,z \sim p(u,z)$$

$$\log \frac{p(u,z)}{q(u;z)} \leftarrow (p,q).\text{REGENERATE}(z) \text{ for } u|z \sim q(u;z)$$

Implementing this interface enables estimation of upper bounds on KL divergences associated with the computation's distribution on outputs

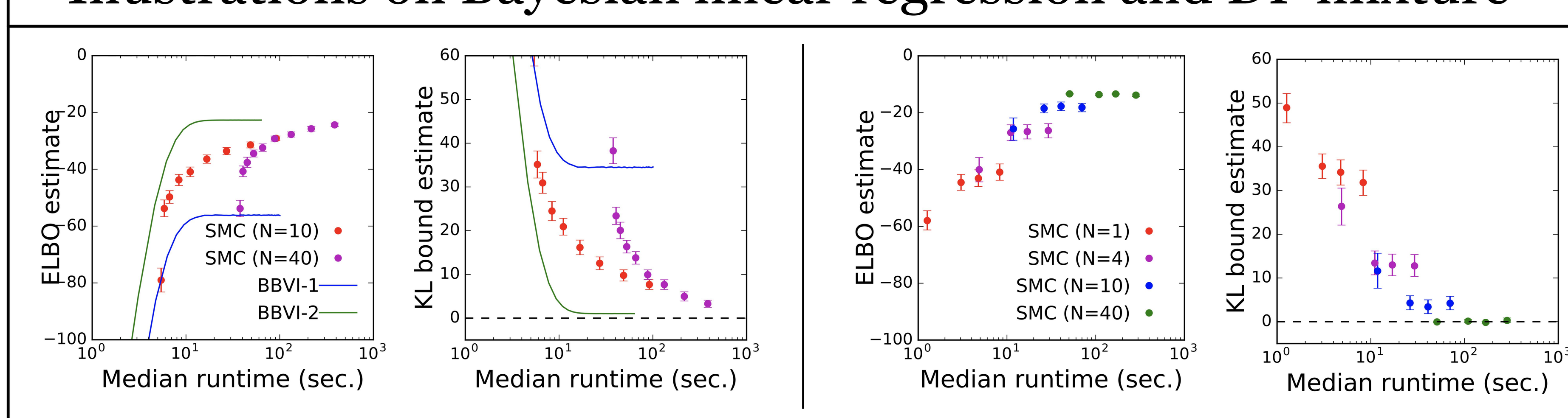
Implementing the probabilistic modules interface for the generic SMC samplers of Del Moral et al., 2006

procedure SIMULATE( )
  for $i \leftarrow 1 \ldots N$ **do**
    $x_1^i \sim k_1(\cdot)$
    $w_1^i \leftarrow w_1(x_1^i)$
  end for
  for $t \leftarrow 2 \ldots T$ **do**
    for $i \leftarrow 1 \ldots N$ **do**
      $a_{t-1}^i \sim \text{Categorical}(\text{NORMALIZE}(\mathbf{w}_{t-1}))$
      $x_t^i \sim k_t(\cdot; x_{t-1}^{a_{t-1}^i})$
      $w_t^i \leftarrow w_t(x_{t-1}^{a_{t-1}^i}, x_t^i)$
    end for
  end for
  $I_T \sim \text{Categorical}(\text{NORMALIZE}(\mathbf{w}_T))$
  $z \sim k_{T+1}(\cdot; x_T^{I_T})$
  $w_{T+1}^1 \leftarrow w_{T+1}(x_T^{I_T}, z)$
  return $\left(z, -\log\left(w_{T+1}^1 \prod_{t=1}^T \frac{1}{N}\sum_{j=1}^N w_t^j\right)\right)$
end procedure

procedure RAND-ANCESTRY(N, T)
  for $t \leftarrow 1 \ldots T$ **do**
    $I_t \sim \text{Uniform}(1, \ldots, N)$
  end for
  return $(I_1, \ldots, I_T)$
end procedure

procedure REGENERATE(z)
  $(I_1, \ldots, I_T) \sim \text{RAND-ANCESTRY}(N, T)$
  $x_T^{I_T} \sim \ell_{T+1}(\cdot; z)$
  for $t \leftarrow T-1 \ldots 1$ **do**
    $x_t^{I_t} \sim \ell_{t+1}(\cdot; x_{t+1}^{I_{t+1}})$
  end for
  (Sample ancestral trajectory using 'backwards' kernels)
  for $i \leftarrow 1 \ldots N$ **do**
    if $i \neq I_1$ **then**
      $x_1^i \sim k_1(\cdot)$
    end if
    $w_1^i \leftarrow w_1(x_1^i)$
  end for
  for $t \leftarrow 2 \ldots T$ **do**
    for $i \leftarrow 1 \ldots N$ **do**
      if $i = I_t$ **then**
        $a_{t-1}^i \leftarrow I_{t-1}$
      else
        $a_{t-1}^i \sim \text{Categorical}(\text{NORMALIZE}(\mathbf{w}_{t-1}))$
        $x_t^i \sim k_t(\cdot; x_{t-1}^{a_{t-1}^i})$
      end if
      $w_t^i \leftarrow w_t(x_{t-1}^{a_{t-1}^i}, x_t^i)$
    end for
  end for
  (Conditional SMC)
  $w_{T+1}^1 \leftarrow w_{T+1}(x_T^{I_T}, z)$
  return $-\log\left(w_{T+1}^1 \prod_{t=1}^T \frac{1}{N}\sum_{j=1}^N w_t^j\right)$
end procedure

Illustrations on Bayesian linear regression and DP mixture



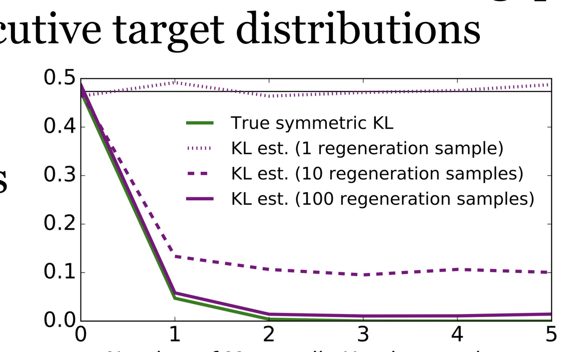
Estimated lower bounds on ELBO, and upper bounds on KL divergence to the posterior for SMC samplers applied to Bayesian linear regression (left) and Dirichlet process mixture modeling (right). SMC samplers use MCMC kernels, and are parameterized by number of particles (N, color), and number of applications of MCMC kernels between observations (different estimates in the same color). An approximate 'reference' sampler was used in place of exact posterior samples. On left, ELBO and divergence for BBVI-1 and BBVI-2, two black box variational samplers. BBVI-2 optimizes over a variational family that includes the posterior.

The accuracy of regenerate determines the bound gap

$$\mathrm{D}(p(u,z)||\pi(z)q(u;z)) = \mathrm{D}(p(z)||\pi(z)) + \underbrace{\boxed{\mathrm{E}_{z\sim p}[\mathrm{D}(p(u|z)||q(u;z))]}}_{\text{gap}}$$

For AIS Markov chains with converged detailed-balance kernels the gap is the sum of KL-divergences between consecutive target distributions $\sum_{t=0}^{T-1} \mathrm{D}(\pi_t(z)||\pi_{t+1}(z))$

For non-sequential MCMC, the generic regenerator is inaccurate, and the bound is trivial. The bound can be tightened using multiple regeneration samples.



Using subjective divergence to measure the error of approximate inference samplers

Given: (1) posterior $\pi(z)$ with unnormalized density $\tilde{\pi}(z) = \pi(z)Z_{\tilde{\pi}}$
(2) inference sampler $p(u,z)$ with output z and auxiliary random choices u
(3) regeneration sampler $q(u;z)$ proposing auxiliary choices u for z
(4) reference sampler $r(z)$

The subjective divergence estimate is:

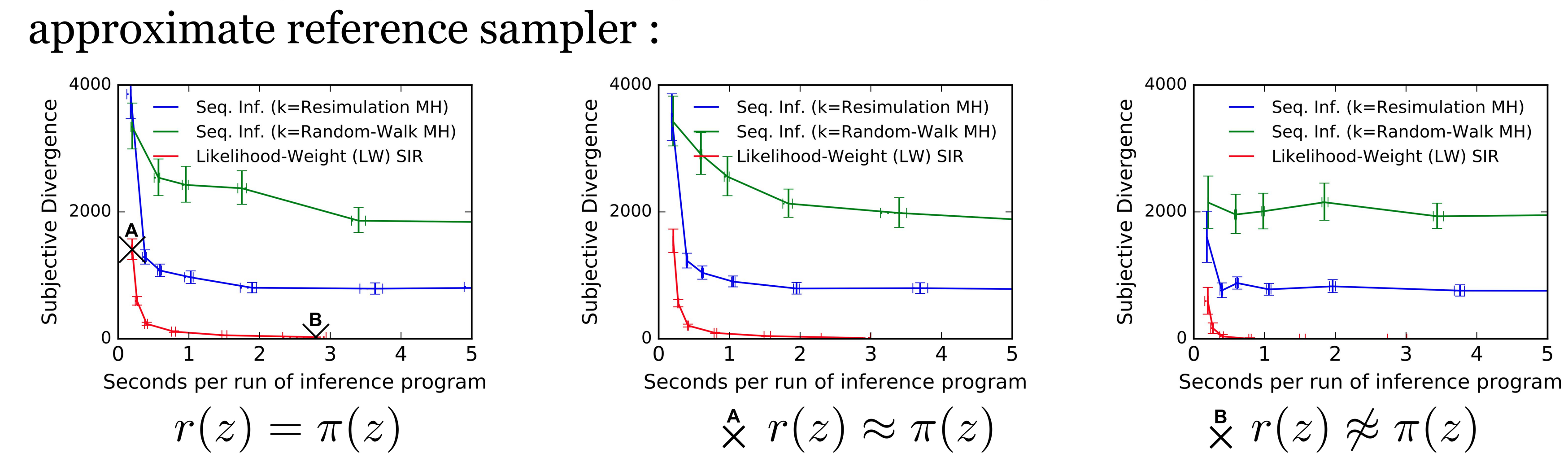
$$\hat{D} = \frac{1}{N}\sum_{i=1}^N \log \frac{p(u_1^i, z_1^i)}{\tilde{\pi}(z_1^i)q(u_1^i; z_1^i)} - \frac{1}{M}\sum_{j=1}^M \log \frac{p(u_2^j, z_2^j)}{\tilde{\pi}(z_2^j)q(u_2^j; z_2^j)} \quad \begin{array}{l} u_1^i, z_1^i \sim p(u,z) \quad i = 1 \ldots N \\ u_2^j, z_2^j \sim \pi(z)q(u;z) \quad j = 1 \ldots M \end{array}$$

Subjective divergence upper bounds the error of the sampler as quantified by KL divergence, subject to assumptions about the accuracy of the reference sampler:

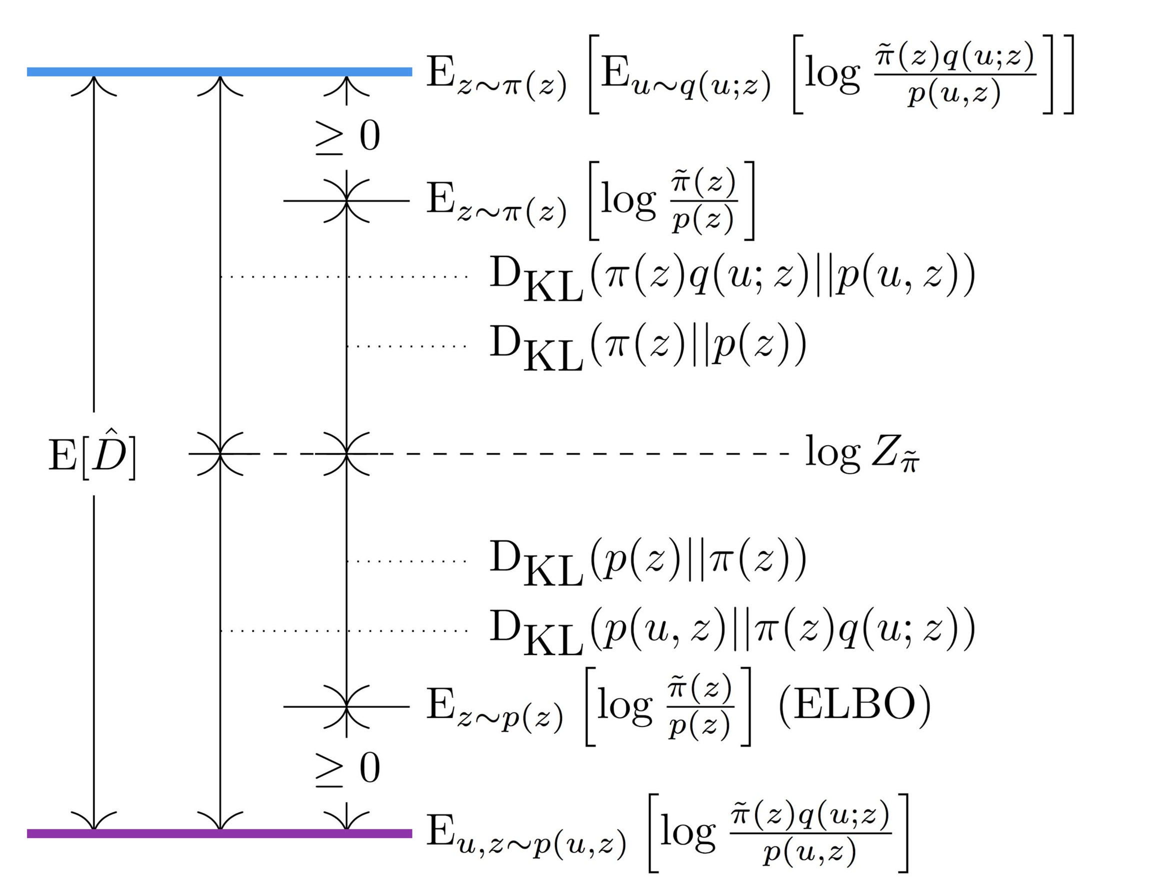
$$\mathrm{E}[\hat{D}] \geq \mathrm{D}(p(z)||\pi(z)) + \mathrm{D}(\pi(z)||p(z)) \text{ for } \mathrm{D}(r(z)||\pi(z)) = 0$$

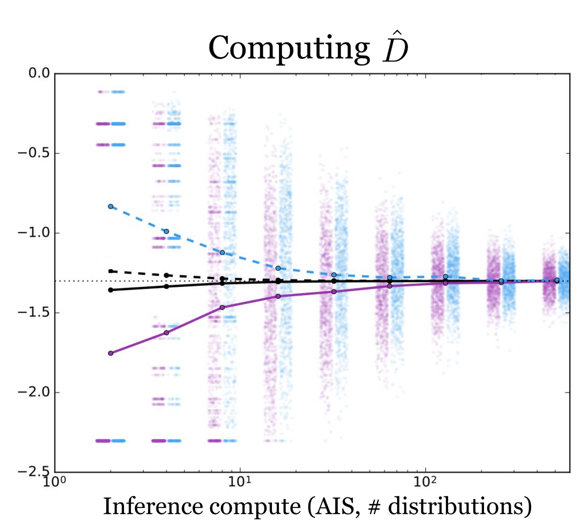
$$\mathrm{E}[\hat{D}] \geq \mathrm{D}(p(z)||\pi(z)) \text{ for } \mathrm{D}(r(z)||\pi(z)) \leq \mathrm{D}(r(z)||p(z))$$

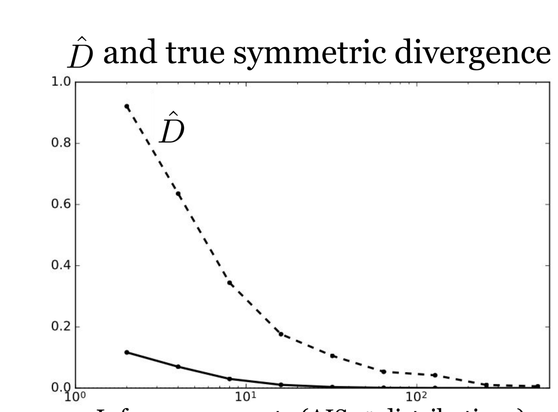
The reliability of subjective divergence can degrade gracefully with use of an approximate reference sampler :



Relationships between subjective divergence $\mathrm{E}[\hat{D}]$, the log-evidence $Z_{\tilde{\pi}}$, and key KL divergences for $r(z) = \pi(z)$:







Estimating subjective divergence for approximate inference samplers that implement the probabilistic modules interface

Require: Probabilistic module (p,q) implementing SIMULATE and REGENERATE;
posterior sampler $z \sim \pi(z)$ or reference program $z \sim r(z)$;
unnormalized posterior probability function $\tilde{\pi}(z)$.
procedure ESTIMATE-KL-BOUND$((p,q), \pi, \tilde{\pi})$
  for $i \leftarrow 1 \ldots N$ **do**
    $(z_1^i, \ell_1^i) \leftarrow (p,q).\text{SIMULATE}()$
  end for
  for $j \leftarrow 1 \ldots M$ **do**
    $z_2^j \sim \pi(z)$ ▷ Replace with sample from reference program $z_2^j \sim r(z)$ if exact posterior sampler unavailable
    $\ell_2^j \leftarrow (p,q).\text{REGENERATE}(z_2^j)$
  end for
  return $\frac{1}{N}\sum_{i=1}^N(\ell_1^i - \log\tilde{\pi}(z_1^i)) - \frac{1}{M}\sum_{j=1}^M(\ell_2^j - \log\tilde{\pi}(z_2^j))$
end procedure

Acknowledgements

The authors would like to thank Roger Grosse for suggesting application of his earlier work on bidirectional Monte Carlo to measuring KL divergences for MCMC in probabilistic programming. The authors would also like to thank Ulrich Schaechtle and Anthony Lu for testing the technique, and David Wingate, Alexey Radul, Feras Saad, and Taylor Campbell for helpful feedback and discussions. This research was supported by DARPA (PPAML program, contract number FA8750-14-2-0004), IARPA (under research contract 2015-15061000003), the Office of Naval Research (under research contract N000141310333), the Army Research Office (under agreement number W911NF-13-1-0212), and gifts from Analog Devices and Google. Marco Cusumano-Towner is supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.