Fast Variational Estimation of Mutual Information for Implicit and Explicit Likelihood Models

Caleb Dahlke[t], Sue Zheng[§], Jason Pacheco[t]

University of Arizona[t], Analog Devices[§]

Link to Paper

Contributions

- Construct a **moment matching solution** for a widely used class of **variational mutual information** (MI) estimators that is equivalent to costly numerical optimization methods.
- Demonstrate multiple **orders of magnitude computational speedup** over the standard optimization-based solutions.
- Show the **efficient application** of the moment matching solution to variational estimates for **"implicit" models** that lack a closed form likelihood function.

Estimating Mutual Information

Mutual Information

$$I(X,Y) = \mathbb{E}_{p(x,y)}[\log p(x|y) - \log p(x)]$$

Variational Marginal (upper bound)

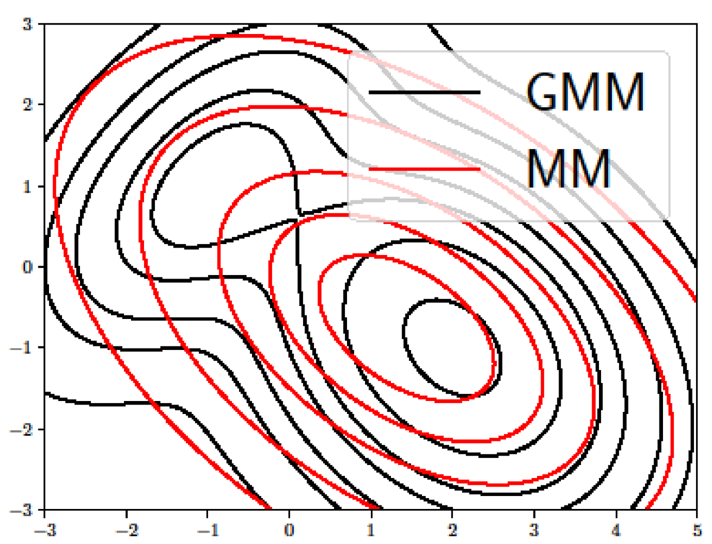
$$I(X,Y) \leq \mathbb{E}_{p(x,y)}[\log p(x|y) - \log q_m(x)] = I_{marg}(X,Y)$$

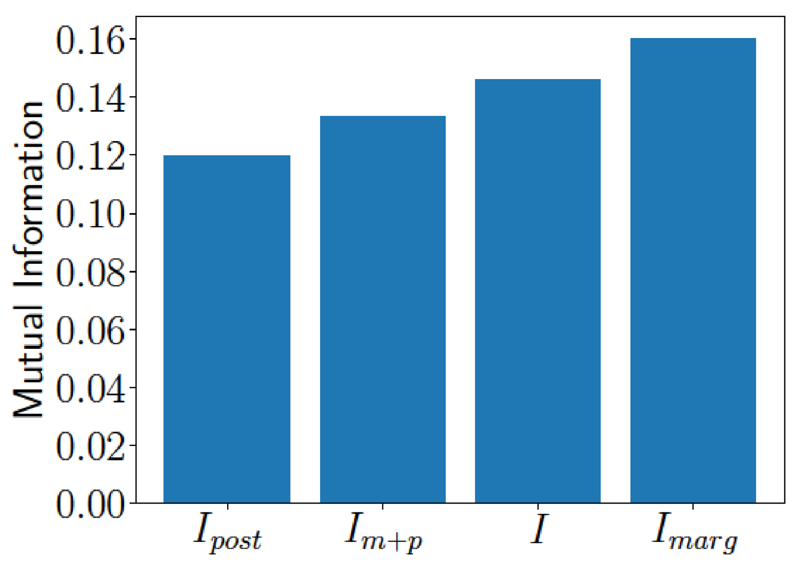
Variational Posterior (lower bound)

$$I(X,Y) \geq \mathbb{E}_{p(x,y)}\left[\log q_p(x|y) - \log p(x)\right] = I_{post}(X,Y)$$

Variational Marginal + Posterior (approximation)

$$I(X,Y) \approx \mathbb{E}_{p(x,y)}\left[\log q_p(x|y) - \log q_m(x)\right] = I_{m+p}(X,Y)$$





Moment Matching = Optimization

Let $q_m(x)$ and $q(x,y)$ be in the exponential family

$$p(X,Y) = \underbrace{h(X,Y)}_{\text{Base Measure}} \exp[\underbrace{\eta}_{\text{Natural Parameters}}{}^T \underbrace{T(X,Y)}_{\text{Sufficient Statistics}} - \underbrace{A(\eta)}_{\text{log Partition}}]$$

Further, let $q(x,y)$ satisfy the linear conditional expectations property:

$$\mathbb{E}_{q_p(x,y)}[T(X,Y)] = \sum_i^k g_i(\eta)\tau_i(Y)$$

where $\tau_i(Y)$ is the i^{th} sufficient statistic dependent only on Y, and $g_i(\eta)$ are any functions only on the natural parameters. Then, **moment matching (MM) the joint $\mathbf{q(x,y)}$ and marginal $\boldsymbol{q_m(x)}$**

$$\mathbb{E}_{q(x,y)}[T(X,Y)] = \mathbb{E}_{p(x,y)}[T(X,Y)]$$

$$\mathbb{E}_{q_m(x)}[T(X)] = \mathbb{E}_{p(x)}[T(X)]$$

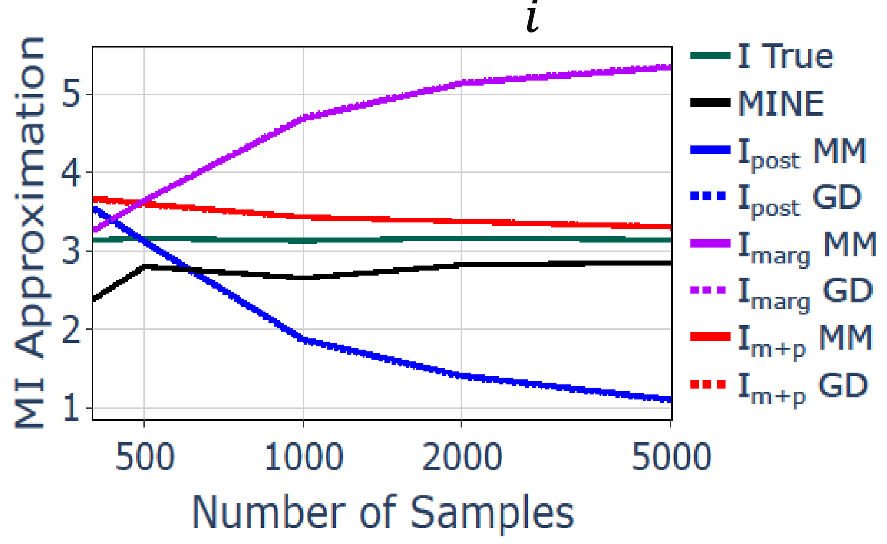
yield optimal $\boldsymbol{q_m(x)}$ for $I_{marg}(X,Y)$, $\boldsymbol{q_p(x,y)} \propto q(x,y)$ for $I_{post}(X,Y)$, and together minimize the bound [1] on $I_{m+p}(X,Y)$.

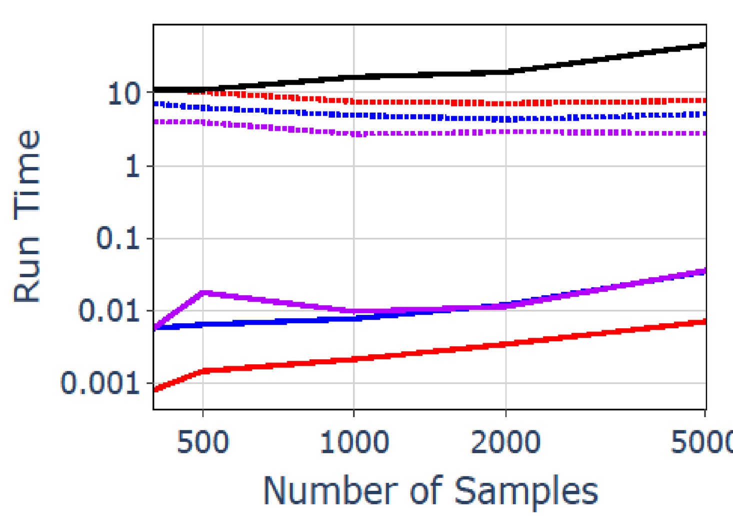
Multivariate Gaussian Mixture Model

Explicit Likelihood:

High dimensional closed form likelihood of a Gaussian mixture model.

$$p(x,y) = \sum_i^5 w_i N(\mu_i, \Sigma_i) \qquad X \in \mathbb{R}^{60} \quad Y \in \mathbb{R}^5$$

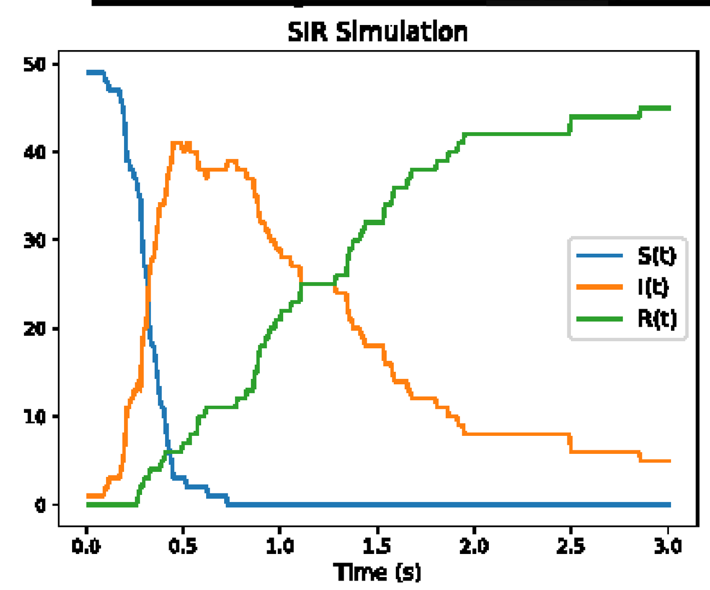




$I(X,Y)$, $I_{marg}(X,Y)$, $I_{post}(X,Y)$, $I_{m+p}(X,Y)$, and $I_{MINE}(X,Y)$ [2] plotted versus various sample sizes, showing a drastic computation speed up of MM.

SIR Epidemiology Model

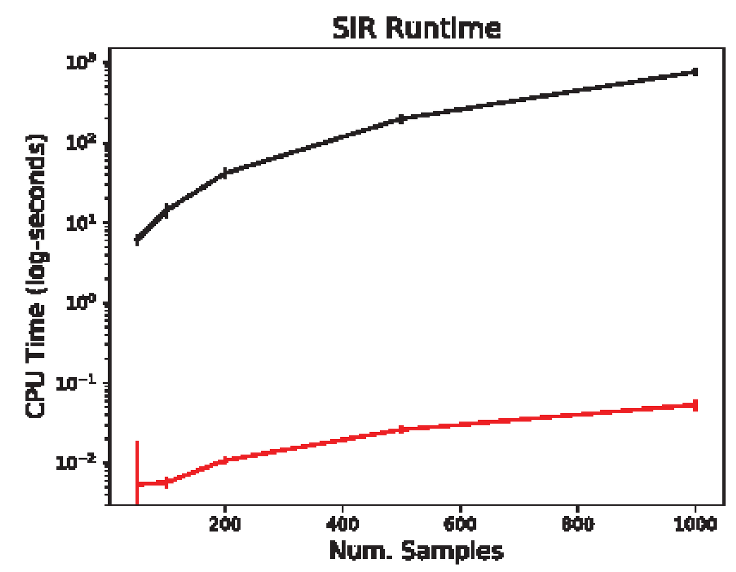
Susceptible $S(t)$, Infected $I(t)$, Recovered $R(t)$:

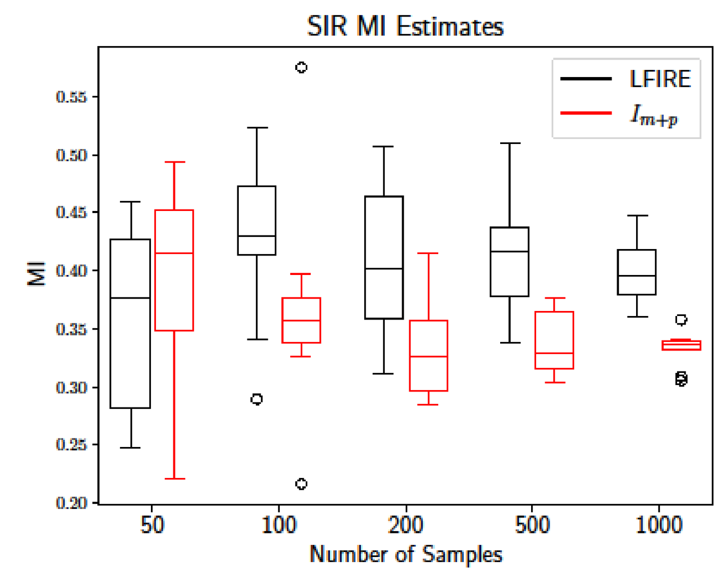


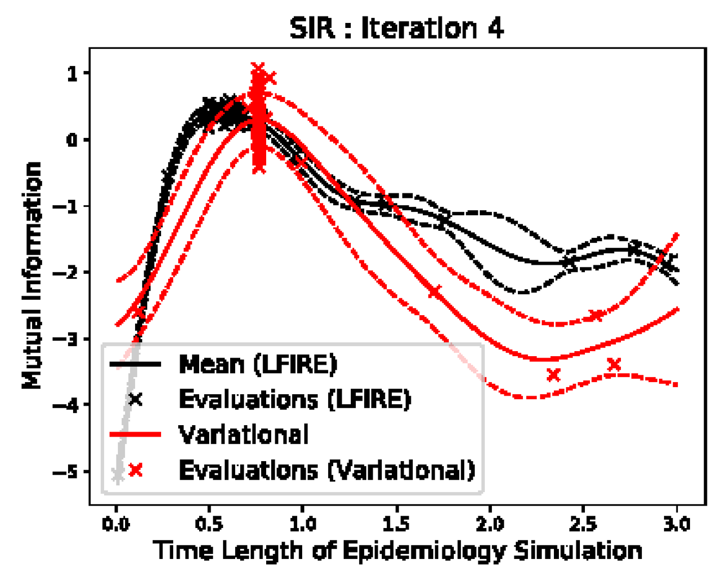
$$S(t+\Delta t) = S(t) - \Delta I(t)$$
$$I(t+\Delta t) = I(t) + \Delta I(t) - \Delta R(t)$$
$$R(t+\Delta t) = R(t) + \Delta R(t)$$
$$\Delta I(t) \sim Binomial\left(S(t), \frac{\beta I(t)}{N}\right)$$
$$\Delta R(t) \sim Binomial(I(t), \gamma)$$

Implicit Likelihood:

No closed form distribution due to nuisance variables or simulation dependent model.







Sequential design for optimal observation time, t_K^*, given the history, $\mathcal{H}_K$. Bayesian optimization using a Gaussian process is used to determine the max MI.

$$t_K^* = argmax\, I(\{\beta,\gamma\},\{S(t),I(t)\}|\mathcal{H}_K)$$

I_{m+p} and I_{LFIRE} [3] found similar maxima and I_{m+p} provides multiple orders of magnitude speed up.

References

[1] A. Foster, M. Jankowiak, E. Bingham, P. Horsfall, Y. W. Teh, T. Rainforth, and N. Goodman. "Variational Bayesian optimal experimental design". In *Advances in Neural Information Processing Systems* 32, pages 14036–14047. 2019.

[2] M. I. Belghazi, A. Baratin, S. Rajeswar, S. Ozair, Y. Bengio, A. Courville, and R. D. Hjelm. "Mine: Mutual information neural estimation". 2018.

[3] S. Kleinegesse, C. Drovandi, and M. U. Gutmann. "Sequential bayesian experimental design for implicit models via mutual information". *Bayesian Analysis*, 16(3):773–802, 2021.