# On Variational Bounds of Mutual Information

Alonso Granados

April 8, 2024

# Motivation for Mutual Information

- Neural networks capacity
- Bayesian experimental design
- Computational neuroscience

# Mutual Information

$$I(X, Y) = \mathbb{E}_{p(x,y)}\left[\frac{p(x, y)}{p(x)p(y)}\right] = \mathrm{KL}(p(x, y)||p(x)p(y)) \qquad (1)$$

# Variational Upper Bound

When $p(y|x)$ is known, we can introduce a variational distribution $q(y)$ for $p(y)$:

$$\begin{aligned} I(X,Y) &= \mathbb{E}_{p(x,y)}\left[\frac{p(y|x)}{p(y)}\right] && (2) \\ &= \mathbb{E}_{p(x,y)}\left[\frac{q(y)p(y|x)}{q(y)p(y)}\right] && (3) \\ &= \mathbb{E}_{p(x,y)}\left[\frac{p(y|x)}{q(y)}\right] - \mathrm{KL}(p(y)||q(y)) && (4) \\ &\leq \mathbb{E}_{p(x)}\left[\mathrm{KL}(p(y|x)||q(y))\right] && (5) \end{aligned}$$

# Variational Lower Bound

For the lower bound, we replace the intractable $p(x|y)$ for $q(x|y)$:

$$
\begin{aligned}
I(X, Y) &= \mathbb{E}_{p(x,y)}\left[\frac{p(x|y)}{p(x)}\right] && (6)\\
&= \mathbb{E}_{p(x,y)}\left[\frac{q(x|y)p(x|y)}{q(x|y)p(x)}\right] && (7)\\
&= \mathbb{E}_{p(x,y)}\left[\frac{q(x|y)}{p(x)}\right] + \mathrm{KL}(p(x|y)||q(x|y)) && (8)\\
&\geq \mathbb{E}_{p(x,y)}\left[\log q(x|y)\right] + h(X) && (9)
\end{aligned}
$$

# Unnormalized Lower Bounds

We choose an energy-based variational family that uses a critic $f(x, y)$ and is scaled by $p(x)$:

$$q(x|y) = \frac{p(x)}{Z(y)} e^{f(x,y)}.$$

Hence, we obtain the lower bound (Unnormalized version of the Barber and Agakov bound):

$$I(X, Y) \geq \mathbb{E}_{p(x,y)} [f(x, y)] - \mathbb{E}_{p(y)}[\log Z(y)] = I_{UBA},$$

which is tight when $f(x, y) = \log p(y|x) + c(y)$.

Applying Jensen's inequality, we recover Donsker & Varadhan bound:

$$I(X,Y) \geq \mathbb{E}_{p(x,y)}[f(x,y)] - \log \mathbb{E}_{p(y)}[Z(y)] = I_{DV}.$$

We can also apply it in the other direction:

$$\log Z(y) = \log \mathbb{E}_{p(x)}[e^{f(x,y)}] \geq \mathbb{E}_{p(x)}[f(x,y)]$$

But then we have,

$$\mathbb{E}_{p(x,y)}[f(x,y)] - \mathbb{E}_{p(x,y)}[f(x,y)] \geq \mathbb{E}_{p(x,y)}[f(x,y)] - \mathbb{E}_{p(y)}[\log Z(y)].$$

Using the inequality $\log(x) \leq \frac{x}{a} + \log(a) - 1$, we obtain the Tractable version of Barber and Agakov bound:

$$\geq \mathbb{E}_{p(x,y)}\left[f(x,y)\right] - \mathbb{E}_{p(y)}\left[\frac{\mathbb{E}_{p(x)}[e^{f(x,y)}]}{a(y)} + \log(a(y)) - 1\right] = I_{TUBA}.$$

If $a(y) = e$, then we recover the Nguyen, Wainwright and Jordan bound:

$$I(X,Y) \geq \mathbb{E}_{p(x,y)}\left[f(x,y)\right] - e^{-1}\mathbb{E}_{p(y)}[Z(y)] = I_{NWJ}.$$

with optimal critic $f^*(x,y) = 1 + \log\frac{p(x|y)}{p(x)}$.

# Multi-sample unnormalized lower bounds

Assumption: We want to estimate $I(X_1, Y)$ and we have samples from $p(x_1)p(y|x_1)$ and $K-1$ additional samples $x_{2:K} \sim r^{K-1}(x_{2:K})$ (independent from $X_1$ and $Y$). Then,

$$I(X_1; Y) = I(X_1, X_{2:K}; Y.)$$

The critic can now depend on the additional samples. Hence, we consider the critic $1 + \log \frac{f(x_1,y)}{a(y;x_{1:K})}$. So we obtain the bound:

$$I(X_1; Y) \geq 1 + \mathbb{E}_{p(x_{1:K})p(y|x_1)} \left[ \log \frac{e^{f(x_1,y)}}{a(y;x_{1:K})} \right] - \mathbb{E}_{p(x_{1:K})p(y)} \left[ \frac{e^{f(x_1,y)}}{a(y;x_{1:K})} \right]$$

Now let's choose the form

$$a(y; x_{1:K}) = m(y; x_{1:K}) = \frac{1}{K} \sum_{i=1}^{K} e^{f(x_i, y)}.$$

Then:

$$\mathbb{E}_{p(x_{1:K})p(y)} \left[ \frac{e^{f(x_1, y)}}{m(y; x_{1:K})} \right] = \frac{1}{K} \sum_{i=1}^{K} \mathbb{E} \left[ \frac{e^{f(x_i, y)}}{m(y; x_{1:K})} \right] = 1$$

when $x_{1:K} \sim \prod_{i=1}^{K} p(x_i)$.

Hence, we recover the lower bound proposed by van der Oord:

$$I(X,Y) \geq \mathbb{E}\left[\frac{1}{K}\sum_{i=1}^{K}\log\frac{e^{f(x_i,y_i)}}{\frac{1}{K}\sum_{i=1}^{K}e^{f(x_i,y_i)}}\right] = I_{NCE}$$

In particular, $I_{NCE} \leq \log K$, meaning that this bound is loose when $I(X,Y) > \log K$.

# Nonlinearly interpolated lower bounds

Now let's set the critic to $1 + \log \frac{e^{f(x_1,y)}}{\alpha m(y;x_{1:K})+(1-\alpha)q(y)}$ where $\alpha \in [0, 1]$:

$$I_\alpha = 1 + \mathbb{E}_{p(x_{1:K})p(y|x_1)}\left[\log \frac{e^{f(x_1,y)}}{\alpha m(y;x_{1:K}) + (1-\alpha)q(y)}\right] \quad (10)$$

$$-\mathbb{E}_{p(x_{1:K})p(y)}\left[\frac{e^{f(x_1,y)}}{\alpha m(y;x_{1:K}) + (1-\alpha)q(y)}\right] \quad (11)$$

**Conjecture**: Optimal critic is $f(x, y) = \log p(y|x)$ and $q(y) = p(y)$.

# Special cases:

When $p(y|x)$ is known, we can use it as our critic for $I_{NCE}$:

$$I(X;Y) \geq \mathbb{E}\left[\frac{1}{K}\sum_{i=1}^{K}\log\frac{p(y_i|x_i)}{\frac{1}{K}\sum_{i=1}^{K}p(y_i|x_i)}\right]$$

We can approximate $p(y) \approx \frac{1}{K}\sum_i p(y|x_i)$:

$$I(X;Y) \leq \mathbb{E}\left[\frac{1}{K}\sum_{i=1}^{K}\log\frac{p(y_i|x_i)}{\frac{1}{K-1}\sum_{i\neq j}p(y_i|x_i)}\right]$$

For $I_{NWJ}$, the optimal critic is given by $1 + \log \frac{p(y|x)}{p(y)}$. Hence, we can replace $p(y)$ with $q(y)$ and optimize w.r.t. $q$:

$$I \geq \mathbb{E}_{p(x,y)}\left[\log \frac{p(y|x)}{q(y)}\right] - \mathbb{E}_{p(y)}\left[\frac{\mathbb{E}_{p(x)} p(y|x)}{q(y)}\right] + 1$$

# Experiments