Three Quantities and Their Relations

Zi Yin

November 29, 2014

1 Three Quantities

We adopt some notations from Chandra Nair by defining:

(a) 
\[ r^*(X; Y) = \lim_{p \to \infty} r^*_p(X; Y), \]

where \( r^*_p(\cdot, \cdot) \) is the hypercontractivity ribbon:

\[ r^*_p(X; Y) = \frac{1}{p} \inf\{ q : ||E[g(Y)|X]||_p \leq ||g(Y)||_q, \forall g \text{ Borel} \}; \]

(b) 
\[ s^*(X; Y) = \sup_{r(x) \neq p(x)} \frac{D(r(y)||p(y))}{D(r(x)||p(x))} \]

where \( r(y), p(y) \) are distributions on \( Y \) induced by \( r(x), p(x) \) through the same channel \( w(y|x) \);

(c) 
\[ \lambda^*(X; Y) = \sup_{U:U \rightarrow X \rightarrow Y} \frac{I(U; Y)}{I(U; X)} \]

We now investigate some relations between the three quantities.
2 Relationship between $s^*$ and $\lambda^*$

1. $s^* \leq \lambda^*$. We can show this, by fixing a sequence of distributions $r_n$ on $\mathcal{X}$ with divergence ratio $s_n = \frac{D(r_n(y)||r(x))}{D(r_n(x)||p(x))}$ asymptotically achieving $s^*$. Define binary random variable $U_\epsilon \sim \text{Bern}(\epsilon)$, and define $X_{n,\epsilon}$:

$$L\{X_{n,\epsilon}|U_\epsilon = 0\} \sim \frac{1}{1-\epsilon}p(x) - \frac{\epsilon}{1-\epsilon}r_n(x),$$

$$L\{X_{n,\epsilon}|U_\epsilon = 1\} \sim r_n(x).$$

It’s easy to check that $X_{n,\epsilon} \sim p(x)$. Now define $g(\epsilon) = \frac{I(U_\epsilon; Y)}{I(U_\epsilon; X_{n,\epsilon})}$. It’s again easy to show (by L’Hospital’s rule) that

$$g(0^+) = \frac{D(r_n(y)||p(y))}{D(r_n(x)||p(x))}$$

So we know $\lambda^* \geq s_n$. $n \to \infty$ leads to $\lambda^* \geq s^*$.

2. $s^* \geq \lambda^*$. Fix a sequence $\{U_n : U_n - X - Y\}$ with information ratio $\lambda_n = \frac{I(U_n; Y)}{I(U_n; X)}$ achieving $\lambda^*$. We can expand $\lambda_n$ and note $\frac{a+b}{c+d} \leq \max\{\frac{a}{b}, \frac{c}{d}\}$:

$$\lambda_n = \frac{\sum p_U(u)D(Y_u||p(y))}{\sum p_U(u)D(X_u||p(x))} \leq s^*$$

Where $X_u$ (or $Y_u$) are the conditional distributions of $X|U = u$ (resp. $Y|U = u$). Let $n \to \infty$ we get $\lambda^* \leq s^*$. 

\[\Box\]
3 Relationship between $s^*$ and $r^*$

1. $r^* \leq s^*$. By definition of $r_p$, we have

$$||E[g(Y)|X]||_p \leq ||g(Y)||_{pr_p}, \forall g \in mB^+$$

In particular, we have

$$||E[g(Y)|X]||_{pr_p}^p \leq ||g(Y)||_{pr_p}^p, \forall g \in mB^+$$ \tag{1}

Note the left hand side of (1) equals:

$$LHS = ||E[g(Y)|X]||_{pr_p}^p$$

$$= \{\sum_x p(x)\left[\sum_y w(y|x)g(y)^p\right]^{r_p}\}^{r_p}$$

$$= \{\sum_x p(x)\left[\sum_y w(y|x)h(y)^{\frac{1}{pr_p}}\right]^{r_p}\}^{r_p}$$

where we set $h(y) = g(y)^{pr_p}$ (note we now fix $h(y)$). Consider the summand $[\sum_y w(y|x)h(y)^{\frac{1}{pr_p}}]^p$ in the limiting case where $p \to \infty$:

$$[\sum_y w(y|x)h(y)^{\frac{1}{pr_p}}]^p = [\sum_y w(y|x)(h(y)^{\frac{1}{pr_p}} - 1) + 1]^p$$

$$= \{\sum_y w(y|x)(h(y)^{\frac{1}{pr_p}} - 1) + 1\}^p\sum_y w(y|x)(h(y)^{\frac{1}{pr_p}} - 1)$$

$$\xrightarrow{\text{(a)}} \exp\{\sum_y \frac{1}{r^*} w(y|x) \log h(y)\}$$

$$= \prod_y h(y)^{\frac{w(y|x)}{r^*}}$$

where in (a) we observe

$$p(h(y)^{\frac{1}{pr_p}} - 1) \to \frac{1}{r^*} \log h(y)$$

This simplified proof is due to Chandra Nair based on Ahlswede’s proof.
since $h(y)^{\frac{1}{p_r}} = 1 + \frac{1}{p_r} \log h(y) + o(p^{-2})$. Further the right hand side of the hypercontractivity inequality (1):

$$RHS = ||g(Y)||^{p_{rp}}_{prp} = \mathbb{E}[h(Y)]$$

As a result, we finally arrive at

$$\left( \sum_x p(x) \prod_y h(y)^{\frac{w(y|x)}{r^*}} \right)^{r^*} \leq \mathbb{E}[h(Y)] \quad (2)$$

By the definition of $r_p$, for any $\epsilon > 0$ there exists a $g_\epsilon$ (hence $h_\epsilon$) which reverses the direction of the hypercontractivity inequality with $r^*$ replaced by $r^* - \epsilon$. So in this case

$$\left( \sum_x p(x) \prod_y h_\epsilon(y)^{\frac{w(y|x)}{r^*}} \right)^{r^* - \epsilon} > \mathbb{E}[h_\epsilon(Y)] \quad (3)$$

Now we define a new probability on $\mathcal{X}$ by

$$q(x) = C p(x) \prod_y h_\epsilon(y)^{\frac{w(y|x)}{r^* - \epsilon}}$$

where $C$ is a normalization constant. Consider $D(q(x)||p(x))$:

$$D(q(x)||p(x)) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$

$$= \sum_x q(x) \log \frac{C p(x) \prod_y h_\epsilon(y)^{\frac{w(y|x)}{r^* - \epsilon}}}{p(x)}$$

$$= \log C + \sum_y \frac{q(y)}{r^* - \epsilon} \log h_\epsilon(y)$$

$$= \log C + \frac{1}{r^* - \epsilon} \left( \sum_y q(y) \log \frac{q(y)}{p(y)} + \sum_y q(y) \log \frac{p(y) h_\epsilon(y)}{q(y)} \right)$$

$$\leq \log C + \frac{1}{r^* - \epsilon} \left( \sum_y q(y) \log \frac{q(y)}{p(y)} + \log \left( \sum_y q(y) \frac{p(y) h_\epsilon(y)}{q(y)} \right) \right)$$

$$= \log C + \frac{1}{r^* - \epsilon} \sum_y q(y) \log \frac{q(y)}{p(y)} + \frac{1}{r^* - \epsilon} \log(\mathbb{E}[h_\epsilon])$$

$$\leq \frac{1}{r^* - \epsilon} D(q(y)||p(y)) \quad (4)$$
where (a) is by Jensen and (b) is due to the fact that \( \log C + \frac{1}{r^* - \epsilon} \log \mathbb{E}[h_r(Y)] < 0 \) as by (3)

\[
1 = \sum q(x) = C \sum_x p(x) \prod_y h_c(y) \frac{w(y|x)}{r^* - \epsilon} > C \mathbb{E}[h_c(y)] \frac{1}{r^* - \epsilon}.
\]

As a result, \( r^* - \epsilon \leq s^* \) for all \( \epsilon > 0 \). So \( r^* \leq s^* \). \( \square \)

2. \( s^* \leq r^* \). Since we know already by (2) that

\[
\mathbb{E}\left[ \prod h(y)^{w(y|x)} \right]^{r^*} \leq \mathbb{E}[h(Y)], \forall h \in mY
\]

For any distribution \( r(x) \), consider \( h_r(y) = \frac{r(y)}{p(y)} \), where \( r(y), p(y) \) are induced by the channel \( w(y|x) \). We know that \( \mathbb{E}[h_r(Y)] = 1 \). Now, consider

\[
1 = \mathbb{E}[h_r(Y)]^{\frac{1}{r^*}}
\geq \mathbb{E}\left[ \prod h_r(y)^{w(y|x)} \right]
= \mathbb{E}\left[ \exp\left\{ \frac{1}{r^*} \sum w(y|x) \log h_r(y) \right\} \right]
= \sum r(x) \exp\left\{ \frac{1}{r^*} \sum w(y|x) \log h_r(y) - \log \frac{r(x)}{p(x)} \right\}
\geq \exp\left\{ \sum r(x)\left( \frac{1}{r^*} \sum w(y|x) \log h_r(y) - \log \frac{r(x)}{p(x)} \right) \right\}
= \exp\left\{ \frac{1}{r^*} \sum r(y) \log h_r(y) - \sum r(x) \log \frac{r(x)}{p(x)} \right\}
= \exp\left\{ \frac{1}{r^*} D(r(y)||p(y)) - D(r(x)||p(x)) \right\}
\]

So we have

\[
\frac{D(r(y)||p(y))}{D(r(x)||p(x))} \leq r^*, \forall r
\]

As a result, \( s^* \leq r^* \). \( \square \)
4 Relationship between $\lambda^*$ and $r^*$

1. $\lambda^* \leq r^*$. Consider any $U - X - Y$, pick $u^n \in \mathcal{T}_e^{(n)}(U)$. For $\epsilon'' > \epsilon'$, define the conditional typicality sets $\mathcal{T}_1^{(n)} = \mathcal{T}_{e'}^{(n)}(X|u^n)$ and $\mathcal{T}_2^{(n)} = \mathcal{T}_{e''}^{(n)}(Y|u^n)$. We slightly abuse notations here, substituting $\mathbb{P}_{X^n}$ (or $\mathbb{P}_{Y^n}$) by $\mathbb{P}$ when there’s no ambiguity. Consider

$$\text{Pr}(X^n \in \mathcal{T}_1^{(n)}, Y^n \in \mathcal{T}_2^{(n)}) = \mathbb{E}[1_{\mathcal{T}_1^{(n)}} \mathbb{E}[1_{\mathcal{T}_2^{(n)}} | X^n]]$$

$$\leq ||1_{\mathcal{T}_1^{(n)}}||_q \mathbb{E}[1_{\mathcal{T}_2^{(n)}} | X^n]||_p$$

$$\leq \mathbb{P}(\mathcal{T}_1^{(n)})^{\frac{1}{q}} ||1_{\mathcal{T}_2^{(n)}}||_{prp}$$

$$= \mathbb{P}(\mathcal{T}_1^{(n)})^{\frac{1}{q}} \mathbb{P}(\mathcal{T}_2^{(n)})^{\frac{1}{prp}}$$

where for the (a) we used Hölder and (b) hypercontractivity inequality together with tensorization of $r_p(X;Y)$. Taking log on both sides of (5), we have

$$\frac{1}{n} \log \mathbb{P}(\mathcal{T}_1^{(n)}) \text{Pr}(Y^n \in \mathcal{T}_2^{(n)} | X^n \in \mathcal{T}_1^{(n)}) \leq \frac{1}{q} \frac{1}{n} \log \mathbb{P}(\mathcal{T}_1^{(n)}) + \frac{1}{pr_p} \frac{1}{n} \log \mathbb{P}(\mathcal{T}_2^{(n)})$$

By typicality lemmas,

$$\frac{1}{n} \log \mathbb{P}(\mathcal{T}_1^{(n)}) \rightarrow -I(U;X),$$

$$\frac{1}{n} \log \mathbb{P}(\mathcal{T}_2^{(n)}) \rightarrow -I(U;Y), \text{ and}$$

$$\text{Pr}(Y^n \in \mathcal{T}_2^{(n)} | X^n \in \mathcal{T}_1^{(n)}) \rightarrow 1.$$

Letting $n \rightarrow \infty$ in (6), we get

$$r_p \geq \frac{I(U;Y)}{I(U,X)}, \text{ for all } U - X - Y$$

As a result, $r_p \geq \lambda^*$ for all $p > 1$. Let $p \rightarrow \infty$ we have $r^* \geq \lambda^*$.

\[\square\]

\[\text{2 Again, This is simplified by Chandra based on Ahlswede's proof.}\]
2. $r^* \leq \lambda^*$. For any $h \in mY^+$, define

$$q(x) = C p(x) \prod_y h(y) \frac{w(y|x)}{\lambda^*}$$

where $C = \mathbb{E}[\prod h(y) \frac{w(y|x)}{\lambda^*}]^{-1}$ is a normalization constant. Define binary random variable $U_\delta \sim Bern(\delta)$, and define $X_{\epsilon,\delta}$:

$$\mathcal{L}\{X_\delta|U_\delta = 0\} \sim \frac{1}{1-\delta} p(x) - \frac{\delta}{1-\delta} q(x),$$

$$\mathcal{L}\{X_\delta|U_\delta = 1\} \sim q(x).$$

when $\delta$ is small, $\mathcal{L}\{X_\delta|U_\delta = 0\}$ is a valid distribution, and $X \sim p(x)$. As usual, define $g_\epsilon(\delta) = \frac{I(U_\delta; Y)}{I(U_\delta; X_\delta)}$. $\delta \rightarrow 0$ leads to (as previously in section 2)

$$\lambda^* \geq g_\epsilon(0^+) = \frac{D(q(y)||p(y))}{D(q(x)||p(x))}$$

(7)

Note that

$$D(q(x)||p(x)) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$

$$= \log C + \sum_y q(y) \frac{\log \frac{q(y)}{\lambda^*}}{\log h(y)}$$

$$= \log C + \frac{1}{\lambda^*} \left( \sum_y q(y) \log \frac{q(y)}{p(y)} + \sum_y q(y) \log \frac{p(y)h(y)}{\frac{q(y)}{\lambda^*}} \right)$$

$$\leq \log C + \frac{1}{\lambda^*} \left( \sum_y q(y) \log \frac{q(y)}{p(y)} + \log \left( \sum_y q(y) \frac{p(y)h(y)}{\frac{q(y)}{\lambda^*}} \right) \right)$$

$$= \log C + \frac{1}{\lambda^*} \left( D(q(y)||p(y)) + \log \mathbb{E}[h(Y)] \right)$$

(8)

Combining (7) and (8)

$$\lambda^* \geq \frac{D(q(y)||p(y))}{D(q(x)||p(x))} \geq \lambda^* - \lambda^* \log C - \log \mathbb{E}[h(Y)]$$

So

$$\mathbb{E}[h(Y)] \geq C^{-\lambda^*} = \mathbb{E}[\prod h(y) \frac{w(y|x)}{\lambda^*}]^{\lambda^*}$$

Since this holds for any $h \in mY^+$, we have $r^* \leq \lambda^*$. \qed
5 References


