Appendix of "Hypothesis Testing Interpretations and Rényi Differential Privacy"

Borja Balle Gilles Barthe Marco Gaboardi
Justin Hsu Tetsuya Sato

February 29, 2020

A Weak version of Birkhoff-von Neumann Theorem (Theorem 11)

Theorem 1 (Weak Birkhoff-von Neumann theorem)**.** *Let $k, l \in \mathbb{N}$ and $k > l$. For any $\gamma\colon k \to \mathrm{Prob}(l)$, there are $\gamma_1, \gamma_2, \ldots, \gamma_N\colon k \to l$ and $0 \le a_1, a_2, \ldots, a_N \le 1$ such that $\sum_{m=1}^{N} a_m = 1$ and $\gamma(i) = \sum_{m=1}^{N} a_m \mathbf{d}_{\gamma_m(i)}$ for any $1 \le i \le k$.*

The cardinal k can be relaxed to countable infinite cardinal ω, and then the families $\{\gamma_j\}_j$ and $\{a_j\}_j$ may be countable infinite.

Proof. Consider the following matrix representation f of γ:

$$f = \begin{pmatrix} f_{1,1} & \cdots & f_{l,1} \\ \vdots & & \vdots \\ f_{1,k} & \cdots & f_{l,k} \end{pmatrix}.$$

where $f_{i,j} = \gamma(i)(j)$ and $\sum_{j=1}^{N} f_{i,j} = 1$ for any $1 \le i \le l$.

For any $h\colon k \to l$, the matrix representation g of $(\{x \mapsto \mathbf{d}_x\} \circ h)$ is

$$g = \begin{pmatrix} g_{1,1} & \cdots & g_{l,1} \\ \vdots & & \vdots \\ g_{1,k} & \cdots & g_{l,k} \end{pmatrix}$$

satisfying that for any $1 \le i \le l$, there is exactly $1 \le j \le k$ such that $g_{i,j} = 1$ and $g_{i,s} = 0$ for $s \ne j$. Conversely, any matrix g satisfying this condition corresponds to some function $h\colon k \to l$. Consider the family G of matrix representations of maps of the form $(\{x \mapsto \mathbf{d}_x\} \circ h)$. We give an algorithm decomposing f to a convex sum of g:

1. Let $r_0 = 1$ and $\tilde{f}_0 = f$. We have $\sum_j (\tilde{f}_0)_{i,j} = r_0$ for all $1 \le i \le l$.

2. For given $0 \le r_m \le 1$ and $\tilde{f}_m$ satisfying $\sum_j (\tilde{f}_m)_{i,j} = r_m$ for all $1 \le i \le l$,

we define $g_{m+1} \in G$, $\alpha_{m+1} \in [0,1]$, $\tilde{f}_{m+1}$ and $r_{m+1} \in [0,1]$ as follows:

$$\alpha_{m+1} = \min_s \max_t (\tilde{f}_m)_{s,t}, \qquad r_{m+1} = r_m - \alpha_{m+1},$$

$$(g_{m+1})_{i,j} = \begin{cases} 1 & j = \operatorname*{argmax}_s (\tilde{f}_m)_{i,s} \\ 0 & \text{(otherwise)} \end{cases}, \qquad \tilde{f}_{m+1} = \tilde{f}_m - \alpha_{m+1} \cdot g_{m+1}.$$

3. If $r_{s+1} = 0$ then we terminate. Otherwise, we repeat the previous step.

In each step, we obtain the following conditions:

- We have $g_{m+1} \in G$ because g_{m+1} can be written as $g_{m+1} = \{x \mapsto \mathbf{d}_x\} \circ (\lambda i.\ \operatorname*{argmax}_s (\tilde{f}_m)_{i,s})$.
- We have $0 < \alpha_{m+1}$ whenever $0 < r_m$ because

$$\alpha_{m+1} = 0 \iff \exists i.\ \max_j (\tilde{f}_m)_{i,j} = 0 \implies \exists i.\ r_m = \sum_j (\tilde{f}_m)_{i,j} = 0.$$

- We have $0 \le (\tilde{f}_{m+1})_{i,j} \le 1$ for any (i,j) from the following equation:

$$(\tilde{f}_{m+1})_{i,j} = \begin{cases} (\tilde{f}_m)_{i,j} - \min_t \max_s (\tilde{f}_m)_{t,s} & \text{if } j = \operatorname*{argmax}_s (\tilde{f}_m)_{i,s} \\ (\tilde{f}_m)_{i,j} & \text{otherwise} . \end{cases}$$

  When $i = \operatorname*{argmin}_s \max_t (\tilde{f}_m)_{s,t}$ and $j = \operatorname*{argmax}_s (\tilde{f}_m)_{i,s}$, we obtain $(\tilde{f}_{m+1})_{i,j} = 0$ while $0 < (\tilde{f}_{m+1})_{i,j}$. This implies that the number of 0 in $\tilde{f}_m$ increases in this operation.

- We also have $\sum_j (\tilde{f}_{m+1})_{i,j} = r_{m+1}$ for all $1 \le i \le k$ because

$$\sum_j (\tilde{f}_{m+1})_{i,j} = \sum_j (\tilde{f}_m)_{i,j} - \alpha_{m+1} \cdot \sum_j (\tilde{g}_{m+1})_{i,j} = r_m - \alpha_{m+1} \cdot 1 = r_{m+1}.$$

Therefore the construction of $g_l \in G$, $\alpha_l \in [0,1]$, $\tilde{f}_l$ and $r_l \in [0,1]$ terminates within $k \cdot l$ steps. When the construction terminates at the step N ($r_N = 0$ also holds), we have a convex decomposition of f by $f = \sum_{m=1}^N \alpha_m \cdot g_m$ where $\sum_{m=1}^N \alpha_m = 1$. This implies By taking $\gamma_1, \gamma_2, \dots, \gamma_N \colon k \to l$ such that g_m is a matrix representation of $(\{x \mapsto \mathbf{d}_x\} \circ \gamma_m)$, we obtain $\gamma(i) = \sum_{m=1}^N a_m \mathbf{d}_{\gamma_m(i)}$ for any $1 \le i \le k$ with $0 \le a_1, a_2, \dots, a_N \le 1$ and $\sum_{m=1}^N a_m = 1$. □

B Omitted Proofs

B.1 Compositions of probabilistic processes

For simplicity, we introduce the composition operator of probabilistic processes (inspired from [Giry, 1982]). For any $\gamma_1 \colon X \to \mathrm{Prob}(Z)$ and $\gamma \colon Z \to \mathrm{Prob}(Y)$, we define their composition $(\gamma \bullet \gamma_1) \colon X \to \mathrm{Prob}(Z)$ by $(\gamma \bullet \gamma_1)(x) \overset{\mathsf{def}}{=} \gamma(\gamma_1(x))$. It is easy to check that the composition $(\gamma \bullet \gamma_1)$ satisfies $(\gamma \bullet \gamma_1)(\mu) = \gamma(\gamma_1(\mu))$ for every $\mu \in \mathrm{Prob}(X)$.

- The composititon operator $\bullet$ is associative: $\gamma \bullet (\gamma_1 \bullet \gamma_2) = (\gamma \bullet \gamma_1) \bullet \gamma_2$ holds for all $\gamma_2 \colon W \to \mathrm{Prob}(X)$, $\gamma_1 \colon X \to \mathrm{Prob}(Z)$, and $\gamma \colon Z \to \mathrm{Prob}(Y)$.
- The function $\eta_X \colon X \to \mathrm{Prob}(X)$ defined by $\eta_X = \{x \mapsto \mathbf{d}_X\}$ is the unit of operator $\bullet$: we have $\gamma \bullet \eta_X = \gamma$ and $\eta_Y \bullet \gamma = \gamma$ for all $\gamma \colon X \to \mathrm{Prob}(Y)$

Thanks to the unit law and associativity of $\bullet$ as an abuse of notations, we define

- $(\gamma \bullet \gamma_1) \colon X \to \mathrm{Prob}(Z)$ for $\gamma_1 \colon X \to Z$ and $\gamma \colon Z \to \mathrm{Prob}(Y)$ by $\gamma \bullet (\eta_Z \circ \gamma_1)$.
- $(\gamma \bullet \gamma_1) \colon X \to \mathrm{Prob}(Z)$ for $\gamma_1 \colon X \to \mathrm{Prob}(Z)$ and $\gamma \colon Z \to Y$ by $(\eta_Y \circ \gamma) \bullet \gamma_1$.
- $(\gamma \bullet \gamma_1) \colon X \to \mathrm{Prob}(Z)$ for $\gamma_1 \colon X \to Z$ and $\gamma \colon Z \to Y$ by $(\eta_Y \circ \gamma) \bullet (\eta_Z \circ \gamma_1)$, which is equal to $\eta_Y \circ (\gamma \circ \gamma_1)$.

Notice that, $\gamma(\mu) \in \mathrm{Prob}(Y)$ defined under $\gamma \colon X \to Y$ and $\mu \in \mathrm{Prob}(X)$ is exactly $(\eta_Y \circ \gamma)(\mu)$.

B.2 Proof of the data-processing inequality of k-cuts

Lemma 2. *For any divergence Δ, every k-cut $\overline{\Delta}^k$ satisfies data-processing inequality.*

Proof. We consider the k-cut of Δ with respect to a set Y satisfying $|Y| = k$

$$\overline{\Delta}^k_X(\mu_1 || \mu_2) \stackrel{\mathsf{def}}{=} \sup_{\gamma \colon X \to \mathrm{Prob}(Y)} \Delta_Y(\gamma(\mu_1) || \gamma(\mu_2)).$$

For every pair $\mu_1, \mu_2 \in \mathrm{Prob}(X)$, and any function $\gamma_1 \colon X \to \mathrm{Prob}(Z)$, we obtain the data-processing inequality

$$\begin{aligned}
\overline{\Delta}^k_Z(\gamma_1(\mu_1) || \gamma_1(\mu_2)) &= \sup_{\gamma \colon Z \to \mathrm{Prob}(Y)} \Delta_Y(\gamma(\gamma_1(\mu_1)) || \gamma(\gamma_1(\mu_2))) \\
&= \sup_{\gamma \colon Z \to \mathrm{Prob}(Y)} \Delta_Y((\gamma \bullet \gamma_1)(\mu_1) || (\gamma \bullet \gamma_1)(\mu_2)) \\
&\leq \sup_{\gamma' \colon X \to \mathrm{Prob}(Y)} \Delta_Y(\gamma'(\mu_1) || \gamma'(\mu_2)) = \overline{\Delta}^k_X(\mu_1 || \mu_2).
\end{aligned}$$

The inequality is obtained by the inclusion

$$\{\, (\gamma \bullet \gamma_1) \colon X \to \mathrm{Prob} Y \mid \gamma \colon Z \to \mathrm{Prob}(Y) \,\} \subseteq \{\, \gamma' \colon X \to \mathrm{Prob}(Y) \,\}.$$

□

B.3 Proof of Lemma 10

Lemma 3 (Lemma 10)**.** *If a divergence Δ has the data-processing inequality, we have the inequality $\overline{\Delta}^k \leq \Delta$ and the equality $\overline{\Delta}^k_Y = \Delta_Y$ for any set Y with $|Y| = k$.*

Proof. We consider the k-cut of Δ with respect to a set W satisfying $|W| = k$

$$\overline{\Delta}^k_X(\mu_1 || \mu_2) \stackrel{\mathsf{def}}{=} \sup_{\gamma \colon X \to \mathrm{Prob}(W)} \Delta_W(\gamma(\mu_1) || \gamma(\mu_2)).$$

Thanks to the data-processing inequality of Δ, we have $\overline{\Delta}^k \leq \Delta$: for every pair $\mu_1, \mu_2 \in \mathrm{Prob}(X)$, we obtain

$$\overline{\Delta}^k_X(\mu_1||\mu_2) = \sup_{\gamma\colon X\to\mathrm{Prob}(W)} \Delta_W(\gamma(\mu_1)||\gamma(\mu_2)) \leq \Delta_X(\mu_1,\mu_2).$$

Now, we consider a set Y with $|Y| = k$. We already have $\overline{\Delta}^k_Y \leq \Delta_Y$. We want to prove $\Delta_Y \leq \overline{\Delta}^k_Y$ Since $|Y| = |W| = k$, there is a bijection $f\colon Y \to W$. We then obtain for every pair $\nu_1, \nu_2 \in \mathrm{Prob}(Y)$,

$$\Delta_Y(\nu_1||\nu_2) = \Delta_Y(f^{-1}(f(\nu_1))||f^{-1}(f(\nu_2))) \leq \Delta_W(f(\nu_1)||f(\nu_2)) \leq \overline{\Delta}^k_Y(\nu_1||\nu_2)$$

The first and second inequalities are obtained by the dataprocessing inequality and the definition of k-cut respectively. □

B.4 Proof of Lemma 13

Lemma 4 (Lemma 13). *If $\Delta = \{\Delta_X\}_{X\colon \mathrm{set}}$ is k-generated, for* any *set $|Y|$ with $|Y| = k$, we have*

$$\Delta_X(\mu_1||\mu_2) = \sup_{\gamma\colon X\to\mathrm{Prob}(Y)} \Delta_Y(\gamma(\mu_1)||\gamma(\mu_2)).$$

Proof. Suppose that Δ is equal to the k-cut of Δ with respect to a set W satisfying $|W| = k$.

$$\Delta_X(\mu_1||\mu_2) = \overline{\Delta}^k_X(\mu_1||\mu_2) = \sup_{\gamma\colon X\to\mathrm{Prob}(W)} \Delta_W(\gamma(\mu_1)||\gamma(\mu_2)).$$

Since $\overline{\Delta}^k$ always satisfies data-processing inequality, the divergence Δ itself do so. We fix an arbitrary set $|Y|$ with $|Y| = k$. Since $|Y| = |W| = k$, there is a bijection $f\colon Y \to W$. For every pair $\mu_1, \mu_2 \in \mathrm{Prob}(X)$, we obtain

$$\begin{aligned}
\Delta_X(\mu_1||\mu_2) &= \sup_{\gamma\colon X\to\mathrm{Prob}(W)} \Delta_W(\gamma(\mu_1)||\gamma(\mu_2)) \\
&= \sup_{\gamma\colon X\to\mathrm{Prob}(W)} \Delta_W(f(f^{-1}(\gamma(\mu_1)))||f(f^{-1}(\gamma(\mu_1)))) \\
&\leq \sup_{\gamma\colon X\to\mathrm{Prob}(W)} \Delta_Y(f^{-1}(\gamma(\mu_1))||f^{-1}(\gamma(\mu_1))) \\
&= \sup_{\gamma\colon X\to\mathrm{Prob}(W)} \Delta_Y((f^{-1}\bullet\gamma)(\mu_1)||(f^{-1}\bullet\gamma)(\mu_1)) \\
&\leq \sup_{\gamma\colon X\to\mathrm{Prob}(Y)} \Delta_Y(\gamma(\mu_1)||\gamma(\mu_2)) \leq \Delta_X(\mu_1||\mu_2).
\end{aligned}$$

Here, the first and last inequalities are obtained from the data-processing inequality of Δ. The second inequality is proved from the inclusion

$$\left\{ (f^{-1}\bullet\gamma)\colon X \to \mathrm{Prob}(Y) \;\middle|\; \gamma\colon X \to \mathrm{Prob}(W) \right\} \subseteq \{ \gamma\colon X \to \mathrm{Prob}(Y) \}.$$

□

B.5 Proof of Basic Properties of k-generatedness (Lemma 14)

Lemma 5 (Lemma 14 (1))**.** *If Δ is 1-generated, then Δ is constant, i.e. there exists $c \in [0, \infty]$ such that for every X and every $\mu_1, \mu_2 \in \mathrm{Prob}(X)$, we have $\Delta_X(\mu_1 || \mu_2) = c$.*

Proof. When Δ is 1-generated, there is a singleton set $\{a\}$ such that, for every pair $\mu_1, \mu_2 \in \mathrm{Prob}(X)$,

$$\Delta_X(\mu_1 || \mu_2) = \sup_{\gamma\colon X \to \mathrm{Prob}(\{a\})} \Delta_{\{a\}}(\gamma(\mu_1) || \gamma(\mu_2)).$$

Now, the set $\mathrm{Prob}(\{a\})$ is a singleton set $\{\mathbf{d}_a\}$, and therefore both $\gamma(\mu_1)$ and $\gamma(\mu_1)$ are equal to $\mathbf{d}_a$ for every $\gamma\colon X \to \mathrm{Prob}(\{a\})$ and every pair $\mu_1, \mu_2 \in \mathrm{Prob}(X)$. Hence, $\Delta_X(\mu_1 || \mu_2) = c$ where $c = \Delta_{\{a\}}(\mathbf{d}_a || \mathbf{d}_a)$. $\square$

Lemma 6 (Lemma 14 (2))**.** *If Δ is k-generated, then it is also $k+1$-generated.*

Proof. Suppose that Δ is equal to the k-cut of Δ with respect to a set W satisfying $|W| = k$.

$$\Delta_X(\mu_1 || \mu_2) = \overline{\Delta}^k_X(\mu_1 || \mu_2) = \sup_{\gamma\colon X \to \mathrm{Prob}(W)} \Delta_W(\gamma(\mu_1) || \gamma(\mu_2)).$$

Let V be an arbitrary set with $|V| = k+1$. We define the $k+1$-cut of $\overline{\Delta}^k$ with respect to the set V.

$$\overline{\overline{\Delta}^k}^{k+1}_X(\mu_1 || \mu_2) \overset{\mathsf{def}}{=} \sup_{\gamma\colon X \to \mathrm{Prob}(V)} \overline{\Delta}^k_V(\gamma(\mu_1) || \gamma(\mu_2)).$$

We then have

$$\begin{aligned}
\overline{\overline{\Delta}^k}^{k+1}_X(\mu_1 || \mu_2) &= \sup_{\gamma\colon X \to \mathrm{Prob}(V)} \overline{\Delta}^k_V(\gamma(\mu_1) || \gamma(\mu_2)) \\
&= \sup_{\gamma\colon X \to \mathrm{Prob}(V)} \sup_{\gamma_1\colon V \to \mathrm{Prob}(W)} \Delta_W(\gamma_1(\gamma(\mu_1)) || \gamma_1(\gamma(\mu_2))) \\
&= \sup_{\gamma\colon X \to \mathrm{Prob}(V)} \sup_{\gamma_1\colon V \to \mathrm{Prob}(W)} \Delta_W((\gamma_1 \bullet \gamma)(\mu_1) || (\gamma_1 \bullet \gamma)(\mu_2)) \\
&\overset{(*)}{=} \sup_{\gamma'\colon X \to \mathrm{Prob}(W)} \Delta_W(\gamma'(\mu_1) || \gamma'(\mu_2)) = \overline{\Delta}^k_X(\mu_1 || \mu_2)
\end{aligned}$$

The equality is obtained by the equality

$$\left\{ (\gamma_1 \bullet \gamma)\colon X \to \mathrm{Prob}(W) \;\middle|\; \begin{array}{l} \gamma\colon X \to \mathrm{Prob}(V), \\ \gamma_1\colon V \to \mathrm{Prob}(W) \end{array} \right\} = \{\, \gamma'\colon X \to \mathrm{Prob}(W) \,\}$$

The inclusion $\subseteq$ is obvious. We show the reverse inclusion $\supseteq$. Since $|V| \geq |W|$, there is a pair of function $f\colon W \to V$ and $g\colon V \to W$ such that $g \circ f = \mathrm{id}_W$. Then, every $\gamma'\colon X \to \mathrm{Prob}(W)$ can be decomposed into $\gamma' = \gamma_1 \bullet \gamma$ where $\gamma = (f \bullet \gamma)$ and $\gamma_1 = g$ (strictly, $\gamma = ((\eta_V \circ f) \bullet \gamma)$ and $\gamma_1 = \eta_W \circ g$). $\square$

Lemma 7 (Lemma 14 (3))**.** *If Δ has the data-processing inequality, then it is at least ∞-generated.*

Proof. We fix a pair $\mu_1, \mu_2 \in \mathrm{Prob}(X)$. The set $Y = \mathsf{supp}(\mu_1) \cup \mathsf{supp}(\mu_1)$ is at most countable. Hence there are two functions $f\colon X \to \mathbb{N}$ and $g\colon \mathbb{N} \to X$ such that $(g \circ f)(x) = x$ for every $x \in Y$. We then have $\mu_1 = (g \circ f)(\mu_1)$ and $\mu_2 = (g \circ f)(\mu_2)$. Thus,

$$\begin{aligned}
\Delta_X(\mu_1||\mu_2) &= \Delta_X((g \circ f)(\mu_1)||(g \circ f)(\mu_2)) \\
&\leq \Delta_{\mathbb{N}}(f(\mu_1)||f(\mu_2)) \\
&\leq \sup_{\gamma\colon X \to \mathrm{Prob}(\mathbb{N})} \Delta_{\mathbb{N}}(\gamma(\mu_1)||\gamma(\mu_2)) \\
&\leq \Delta_X(\mu_1||\mu_2).
\end{aligned}$$

The last part is an ∞-cut. The first and last inequality is obtained by the data-processing inequality. The second one is obvious ($f\colon X \to \mathbb{N}$ is regarded as $\{x \mapsto \mathbf{d}_{f(x)}\}\colon X \to \mathrm{Prob}(\mathbb{N})$). $\square$

Lemma 8 (Lemma 14 (4))**.** *Every k-cut of a divergence Δ is always k-generated.*

Proof. We can prove $\overline{\overline{\Delta}^k}^k = \overline{\Delta}^k$ in a almost the same way as Lemma 14 (2). $\square$

Continuity of divergence (Lemma 14(3) in general setting) We can extend the results on divergences in the discrete setting to general measurable setting using the continuity of divergences. We say that a divergence Δ is continuous if for any pair $\mu_1, \mu_2 \in \mathrm{Prob}(X)$,

$$\Delta_X(\mu_1||\mu_2) = \sup_{n \in \mathbb{N}} \sup_{\gamma\colon X \to \{0,1,2,\ldots,n\}} \Delta_{\{0,1,2,\ldots,n\}}(\gamma(\mu_1)||\gamma(\mu_2)).$$

If Δ is continuous and satisfies data-processing inequality we have ∞-generatedness (moreover we show the "countable"-generatedness) as follows:

$$\begin{aligned}
&\Delta_X(\mu_1||\mu_2) \\
&= \sup_{n \in \mathbb{N}} \sup_{\gamma\colon X \to \{0,1,2,\ldots,n-1\}} \Delta_{\{0,1,2,\ldots,n-1\}}(\gamma(\mu_1)||\gamma(\mu_2)) \\
&= \sup_{n \in \mathbb{N}} \sup_{\gamma\colon X \to \{0,1,2,\ldots,n-1\}} \Delta_{\{0,1,2,\ldots,n-1\}}((g_n \circ f_n)(\gamma(\mu_1))||(g_n \circ f_n)(\gamma(\mu_2))) \\
&= \sup_{n \in \mathbb{N}} \sup_{\gamma\colon X \to \{0,1,2,\ldots,n\}} \Delta_{\{0,1,2,\ldots,n-1\}}(g_n((f_n \bullet \gamma)(\mu_1))||g_n((f_n \bullet \gamma)(\mu_1))) \\
&\leq \sup_{n \in \mathbb{N}} \sup_{\gamma\colon X \to \{0,1,2,\ldots,n-1\}} \Delta_{\mathbb{N}}((f_n \bullet \gamma)(\mu_1)||(f_n \bullet \gamma)(\mu_1)) \\
&\leq \sup_{\gamma\colon X \to \mathbb{N}} \Delta_{\mathbb{N}}(\gamma(\mu_1)||\gamma(\mu_2)) \leq \overline{\Delta}^{\infty}_X(\mu_1, \mu_2) \\
&\leq \Delta_X(\mu_1, \mu_2).
\end{aligned}$$

Here $f_n\colon \{0,1,2,\ldots,n-1\} \to \mathbb{N}$ is the inclusion mapping, and $g_n\colon \mathbb{N} \to \{0,1,2,\ldots,n-1\}$ is defined by $g_n(k) = k$ if $(k < n)$ and $g_n(k) = n-1$ otherwise. We have $(g_n \circ f_n) = \mathrm{id}_{\{0,1,2,\ldots,n-1\}}$.

The first and last inequalities are obtained from data-processing inequality. The second inequality is obvious.

B.6 Proof of Lemma 15

Lemma 9 (Lemma 15)**.** *Consider a divergence Δ and a k-generated divergence Δ'. For any k-cut $\overline{\Delta}^k$ of Δ,*

$$\Delta' \leq \Delta \implies \Delta' \leq \overline{\Delta}^k.$$

Also, if Δ has the data-processing inequality, the k-cut is the greatest k-generated divergence below Δ:

$$\Delta' \leq \Delta \iff \Delta' \leq \overline{\Delta}^k \leq \Delta.$$

Proof. Since Δ' is k-generated, for any choice of Y with $|Y| = k$, we have

$$\Delta' \leq \Delta \implies \Delta'_Y \leq \Delta_Y \implies \overline{\Delta'}^k \leq \overline{\Delta}^k \iff \Delta' \leq \overline{\Delta}^k.$$

The second statement is proved as follows: From the first statement of this lemma and Lemma 3 (Lemma 10 in the paper), We have

$$\Delta' \leq \Delta \implies \Delta' \leq \overline{\Delta}^k \leq \Delta$$

The converse direction is obvious. $\square$

An extended version. We can extend this theorem to more suitable for conversion laws of differential privacy.

Lemma 10 (Lemma 15, extended)**.** *Consider a divergence Δ satisfying data-processing inequality and a k-generated divergence Δ'.*

$$\begin{aligned}&\forall X.\forall \mu_1, \mu_2 \in \mathrm{Prob}(X).(\Delta_X(\mu_1||\mu_2) \leq \delta \implies \Delta'_X(\mu_1||\mu_2) \leq \rho)\\ \iff &\forall X.\forall \mu_1, \mu_2 \in \mathrm{Prob}(X).(\overline{\Delta}^k_X(\mu_1||\mu_2) \leq \delta \implies \Delta'_X(\mu_1||\mu_2) \leq \rho)\end{aligned}$$

Proof. ($\Longleftarrow$) Obvious from Lemma 3 (Lemma 10 in the paper). ($\Longrightarrow$) From the assumption, we obtain

$$\begin{aligned}&\forall X.\forall \mu_1, \mu_2 \in \mathrm{Prob}(X).\forall \gamma\colon X \to \mathrm{Prob}(Y).\\ &\qquad \Delta_Y(\gamma(\mu_1)||\gamma(\mu_2)) \leq \delta \implies \Delta'_Y(\gamma(\mu_1)||\gamma(\mu_2)) \leq \rho.\end{aligned}$$

This implies

$$\forall X.\forall \mu_1, \mu_2 \in \mathrm{Prob}(X).(\overline{\Delta}^k_X(\mu_1||\mu_2 \leq \delta \implies \overline{\Delta'}^k_X(\mu_1||\mu_2) \leq \rho)$$

Thanks to the k-generatedness of Δ', we conclude the statement of this lemma. $\square$

B.7 Proof of 2-generatedness of ε-divergence

Theorem 11. *The ε-divergence Δ^ε is 2-generated for all ε.*

Proof. We recall that the ε-divergence Δ^ε is quasi-convex (moreover, jointly convex) and satisfies data-processing inequality. We choose a set $Y = \{\mathtt{Acc}, \mathtt{Rej}\}$, and take the 2-cut of Δ^ε by

$$\overline{\Delta^\varepsilon}^2_X(\mu_1||\mu_2) = \sup_{\gamma\colon X \to \mathrm{Prob}(\{\mathtt{Acc},\mathtt{Rej}\})} \Delta^\varepsilon_Y(\gamma(\mu_1)||\gamma(\mu_2))$$

We show this is equal to the original $\Delta^{\varepsilon}_{X}(\mu_1||\mu_2)$. Without loss of generality we may assume X is at most countable. If X is an arbitrary set, we can restrict it to countable set in a similar way as the proof of Lemma 7 (Lemma 14(3) in the paper).

By the weak Birkhoff-von Neumann Theorem (Theorem 1 in the appendix), each $\gamma\colon X \to \mathrm{Prob}(\{\mathtt{Acc},\mathtt{Rej}\})$ can be decomposed into a convex combination $\gamma(x) = \sum_{i\in I}\alpha_i \mathbf{d}_{\gamma_i(x)}$ of functions $\gamma_i\colon X \to \{\mathtt{Acc},\mathtt{Rej}\}$ $(i \in I)$ where I is a countable set and $\sum_{i\in I}\alpha_i = 1$. By combining this and quasi-convexity and data-processing inequality of Δ^{ε}, we obtain

$$\begin{aligned}\Delta^{\varepsilon}(\gamma(\mu_1)||\gamma(\mu_2)) &= \Delta^{\varepsilon}(\textstyle\sum_{i\in I}\alpha_i\gamma_i(\mu_1)||\sum_{i\in I}\alpha_i\gamma_i(\mu_1))\\ &= \sup_{i\in I}\Delta^{\varepsilon}(\gamma_i(\mu_1)||\gamma_i(\mu_1))\\ &\le \sup_{\gamma\colon X\to\{\mathtt{Acc},\mathtt{Rej}\}}\Delta^{\varepsilon}_{X}(\gamma(\mu_1)||\gamma(\mu_2)).\end{aligned}$$

This implies

$$\begin{aligned}\overline{\Delta^{\varepsilon}}^{2}_{X}(\mu_1||\mu_2) &= \sup_{\gamma\colon X\to\mathrm{Prob}(\{\mathtt{Acc},\mathtt{Rej}\})}\Delta^{\varepsilon}(\gamma(\mu_1)||\gamma(\mu_2))\\ &= \sup_{\gamma\colon X\to\{\mathtt{Acc},\mathtt{Rej}\}}\Delta^{\varepsilon}_{\{\mathtt{Acc},\mathtt{Rej}\}}(\gamma(\mu_1)||\gamma(\mu_2))\\ &= \sup_{\gamma\colon X\to\{\mathtt{Acc},\mathtt{Rej}\}}\sup_{A\subseteq\{\mathtt{Acc},\mathtt{Rej}\}}(\mathsf{Pr}[\gamma(\mu_1)\in A] - e^{\varepsilon}\mathsf{Pr}[\gamma(\mu_2)\in A])\\ &= \sup_{\gamma\colon X\to\{\mathtt{Acc},\mathtt{Rej}\}}\sup_{A\subseteq\{\mathtt{Acc},\mathtt{Rej}\}}(\mathsf{Pr}[\mu_1\in\gamma^{-1}(A)] - e^{\varepsilon}\mathsf{Pr}[\mu_2\in\gamma^{-1}(A)])\\ &\overset{(*)}{=} \sup_{S\subseteq X}(\mathsf{Pr}[\mu_1\in S] - e^{\varepsilon}\mathsf{Pr}[\mu_2\in S])\\ &= \Delta^{\varepsilon}_{X}(\mu_1||\mu_2)\end{aligned}$$

We have the 2-generatedness: $\overline{\Delta^{\varepsilon}}^{2} = \Delta^{\varepsilon}$. The equality $(*)$ is proved as follows: for given γ and A, we take $S = \gamma^{-1}$. Conversely, for any $S \subseteq X$ we take $A = \{\mathtt{Acc}\}$ and $\gamma = \chi_S$, which is the indicator function of S defined by $\chi_S(x) = 1$ if $x \in S$ and $\chi_S(x) = 0$ otherwise. $\square$

General version We can extend this result to general measurable setting by using the continuity of Δ^{ε} (see also [Liese and Vajda, 2006]), which is obtained by f-divergence characterization of Δ^{ε} [Barthe and Olmedo, 2013]. For general measurable sapce X and every pair $\mu_1,\mu_2 \in \mathrm{Prob}(X)$ we have

$$\begin{aligned}\Delta^{\varepsilon}_{X}(\mu_1||\mu_2) &= \sup_{\gamma\colon X\to\mathbb{N}}\Delta^{\varepsilon}_{\mathbb{N}}(\gamma(\mu_1)||\gamma(\mu_2))\\ &= \sup_{\gamma\colon X\to\mathbb{N}}\sup_{\gamma'\colon\mathbb{N}\to\mathrm{Prob}(\{\mathtt{Acc},\mathtt{Rej}\})}\Delta^{\varepsilon}_{\{\mathtt{Acc},\mathtt{Rej}\}}((\gamma'\bullet\gamma)(\mu_1)||(\gamma'\bullet\gamma)(\mu_2))\\ &= \sup_{\gamma\colon X\to\{\mathtt{Acc},\mathtt{Rej}\}}\Delta^{\varepsilon}_{\{\mathtt{Acc},\mathtt{Rej}\}}(\gamma(\mu_1)||\gamma(\mu_2))\end{aligned}$$

Functions are assumed to be measurable.

B.8 Counterexample: Rényi-divergence is not 2-generated

Theorem 12. *There are* $\mu_1, \mu_2 \in \texttt{Prob}(\{a,b,c\})$ *such that*

$$\overline{D^{\alpha}}^{2}_{\{a,b,c\}}(\mu_1||\mu_2) < D^{\alpha}_{\{a,b,c\}}(\mu_1||\mu_2)$$

Proof. Let $p = (1/2)^{\beta/(\alpha-1)}$ and $\alpha + 1 < \beta$ and define

$$\mu_1 = \frac{1}{3}\mathbf{d}_a + \frac{1}{3}\mathbf{d}_b + \frac{1}{3}\mathbf{d}_c,$$
$$\mu_2 = \frac{p^2}{p^2+p+1}\mathbf{d}_a + \frac{p}{p^2+p+1}\mathbf{d}_b + \frac{1}{p^2+p+1}\mathbf{d}_c$$

Since Rényi divergence is quasi-convex and satisfies data-processing inequality, it suffices to show the proper inequality $D^{\alpha}_{\{\texttt{Acc},\texttt{Rej}\}}(\gamma(\mu_1)||\gamma(\mu_2)) < D^{\alpha}_{\{a,b,c\}}(\mu_1||\mu_2)$ holds for any *deterministic decision rule* $\gamma\colon \{a,b,c\} \to \{\texttt{Acc},\texttt{Rej}\}$. There are 8 cases of $\gamma\colon \{a,b,c\} \to \{\texttt{Acc},\texttt{Rej}\}$, but thanks to the data-processing inequality and reflexivity of Rényi divergence, it suffices to consider 3 cases: $(\gamma(a),\gamma(b),\gamma(c)) = (\texttt{Acc},\texttt{Acc},\texttt{Rej}), (\texttt{Acc},\texttt{Rej},\texttt{Acc}), (\texttt{Rej},\texttt{Acc},\texttt{Acc})$. Hence,

$$\begin{aligned}
&\frac{\exp((\alpha-1)D^{\alpha}_{\{a,b,c\}}(\mu_1||\mu_2))}{\exp((\alpha-1)D^{\alpha}_{\{\texttt{Acc},\texttt{Rej}\}}(\gamma(\mu_1)||\gamma(\mu_2))}\\
&\geq \min\left(\frac{p^{2(1-\alpha)}+p^{1-\alpha}+1}{2^{\alpha}(p^2+p)^{1-\alpha}+1}, \frac{p^{2(1-\alpha)}+p^{1-\alpha}+1}{2^{\alpha}(p^2+1)^{1-\alpha}+p^{1-\alpha}}, \frac{p^{2(1-\alpha)}+p^{1-\alpha}+1}{2^{\alpha}(p+1)^{1-\alpha}+p^{2(1-\alpha)}}\right)\\
&\geq \min\left(\frac{2^{\beta}+2^{-\beta}+1}{2^{\alpha}(p+1)^{1-\alpha}+2^{-\beta}}, \frac{2^{\beta}+2^{-\beta}+1}{2^{\alpha-\beta}(p^2+1)^{1-\alpha}+1}, \frac{2^{\beta}+2^{-\beta}+1}{2^{\beta}+2^{\alpha-\beta}(p+1)^{1-\alpha}}\right)\\
&\geq \min\left(\frac{2^{\beta}+2^{-\beta}+1}{2^{\alpha+1}}, \frac{2^{\beta}+2^{-\beta}+1}{2^{\beta}+1}\right) > 1.
\end{aligned}$$

Hence,

$$\begin{aligned}
&D^{\alpha}_{\{\texttt{Acc},\texttt{Rej}\}}(\gamma(\mu_1)||\gamma(\mu_2)) + \frac{1}{\alpha-1}\log\min(\frac{2^{\beta}+2^{-\beta}+1}{2^{\alpha+1}}, \frac{2^{\beta}+2^{-\beta}+1}{2^{\beta}+1})\\
&\leq D^{\alpha}_{\{a,b,c\}}(\mu_1||\mu_2).
\end{aligned}$$

holds for any $\gamma\colon \{a,b,c\} \to \{\texttt{Acc},\texttt{Rej}\}$. By the data-processing inequality of Rényi divergence, this discussion does not depend on the choice of $\{\texttt{Acc},\texttt{Rej}\}$. By weak Birkhoff-von Neumann theorem, and the quasi-convexity Rényi divergence, we conclude

$$\overline{D^{\alpha}}^{2}_{X}(\mu_1||\mu_2) + \frac{1}{\alpha-1}\log\min(\frac{2^{\beta}+2^{-\beta}+1}{2^{\alpha+1}}, \frac{2^{\beta}+2^{-\beta}+1}{2^{\beta}+1}) \leq D^{\alpha}_{X}(\mu_1||\mu_2).$$

□

B.9 Proof of ∞-generatedness of Rényi-divergence

f-divergences is a class of divergences that are characterized by convex functions. For a given convex function $f\colon [0,\infty) \to \mathbb{R}$ satisfying $\lim_{t\to 0+} tf(0/t) = 0$ (this function is called weight function), we define an f-divergence Δ^f corresponding the function f,

$$\Delta^f_X(\mu_1||\mu_2) \stackrel{\mathsf{def}}{=} \sum_{x\in X} \mu_2(x) f\left(\frac{\mu_1(x)}{\mu_2(x)}\right).$$

The α-Rényi divergence D^α can also be characterized using f-divergence as follows:

$$D^\alpha(\mu_1||\mu_2) = \frac{1}{\alpha-1}\log\sum_{x\in X}\mu_2(x)\left(\frac{\mu_1(x)}{\mu_2(x)}\right)^\alpha = \frac{1}{\alpha-1}\log\Delta_X^{t\mapsto t^\alpha}(\mu_1||\mu_2).$$

Remark that every f-divergence is quasi-convex (moreover jointly convex) and continuous, and satisfies data-processing inequality (see also [Liese and Vajda, 2006, Theorems 14–16]).

Since the mapping $t \mapsto \frac{1}{\alpha-1}\log t$ is monotone, every α-Rényi divergence D^α is also quasi-convex and satisfies data-processing inequality. Thanks to the data-processing inequality, every α-Rényi divergence D^α is at least ∞-generated. We need to prove that for every finite k, every α-Rényi divergence D^α is not k-generated. To prove this, we use that the mapping $t \mapsto t^\alpha$ is strictly convex.

Lemma 13. *If a weight function is strictly convex, its f-divergence Δ^f is not k-generated for every finite k.*

Proof. Without loss of generality, we may assume $k > 1$.

Consider a pair $\mu_1, \mu_2 \in \mathrm{Prob}(\{0,1,2,\dots,k\})$ satisfying $\mathsf{supp}(\mu_1) = \mathsf{supp}(\mu_2) = \{0,1,2,\dots,k\}$ and $\mu_1(i)/\mu_2(i) \neq \mu_1(j)/\mu_2(j)$ where $1 \leq i,j \leq k+1$ and $i \neq j$. We can give such distributions. Then we obtain,

$$\begin{aligned}
&\overline{\Delta^f}^k_{\{0,1,2,\dots,k\}}(\mu_1||\mu_2)\\
&= \sup_{\gamma\colon \{0,1,2,\dots,k\}\to\mathrm{Prob}(\{0,1,2,\dots,k-1\})} \Delta^f_{\{0,1,2,\dots,k-1\}}(\gamma(\mu_1)||\gamma(\mu_2))\\
&\quad \{\text{Weak Birkhoff-von Neumann theorem and the joint convexity of } \Delta^f\}\\
&= \max_{\gamma\colon \{0,1,2,\dots,k\}\to\{0,1,2,\dots,k-1\}} \Delta^f(\gamma(\mu_1)||\gamma(\mu_2))\\
&= \max_{\gamma\colon \{0,1,2,\dots,k\}\to\{0,1,2,\dots,k-1\}} \sum_{j=0}^{k-1} f\left(\frac{\sum_{\gamma(i)=j}\mu_1(i)}{\sum_{\gamma(i)=j}\mu_2(i)}\right)\left(\textstyle\sum_{\gamma(i)=j}\mu_2(i)\right)\\
&\quad \{\text{Jensen's inequality with the strict convexity of the weight function } f\}\\
&< \sum_{i=0}^{k} f\left(\frac{\mu_1(i)}{\mu_2(i)}\right)\mu_2(i) = \Delta^f(\mu_1||\mu_2).
\end{aligned}$$

Since $k+1 > k$, by Dirichlet's pigeonhole principle, for any $\gamma\colon \{0,1,2,\dots,k\} \to \{0,1,2,\dots,k-1\}$, for some $j \in \{0,1,2,\dots,k\}$, there are at least two different $i_1, i_2 \in \{0,1,2,\dots,k-1\}$ such that $\gamma(i_1) = j$ and $\gamma(i_2) = j$. From the assumption on μ_1 and μ_2, we have $(\mu_1(i_1)/\mu_2(i_1)) \neq (\mu_1(i_2)/\mu_2(i_2))$ Since the function f is strictly convex, by the condition for equality of Jensen's inequality, we have the strict inequality

$$f\left(\frac{\mu_1(i_1)+\mu_1(i_2)}{\mu_2(i_1)+\mu_2(i_2)}\right)(\mu_2(i_1)+\mu_2(i_2)) < f\left(\frac{\mu_1(i_1)}{\mu_2(i_1)}\right)\mu_2(i_1) + f\left(\frac{\mu_1(i_2)}{\mu_2(i_2)}\right)\mu_2(i_2).$$

Therefore, for any $\gamma\colon \{0,1,2,\dots,k\} \to \{0,1,2,\dots,k-1\}$, we have

$$\sum_{j=1}^{k} \left(\frac{\sum_{\gamma(i)=j} \mu_1(i)}{\sum_{\gamma(i)=j} \mu_2(i)} \right)^{\alpha} (\textstyle\sum_{\gamma(i)=j} \mu_2(i)) < \displaystyle\sum_{i=1}^{k+1} \left(\frac{\mu_1(i)}{\mu_2(i)} \right)^{\alpha} \mu_2(i).$$

Since there only finite case of $\gamma\colon \{0,1,2,\dots,k\} \to \{0,1,2,\dots,k-1\}$, we conclude $\overline{\Delta^f}^{k}_{\{0,1,2,\dots,k\}}(\mu_1||\mu_2) < \Delta^f_{\{0,1,2,\dots,k\}}(\mu_1||\mu_2)$. Since every f-divergence satisfies data-processing inequality, this discussion does not depend on the choice of set Y with $|Y| = k$ in the construction of the k-cut $\overline{\Delta^f}^{k}$. Thus, Δ^f is not k-generated for any finite k. □

Since the mapping $t \mapsto \frac{1}{\alpha - 1} \log t$ is strict, we conclude,

Corollary 14. *For any alpha > 1, the α-Rényi divergence D^{α} is not k-generated for every finite k.*

B.10 Proof of Theorem 18

Theorem 15 (Theorem 18)**.** *Let $\mu_1, \mu_2 \in \mathrm{Prob}(X)$. $\overline{\Delta}^{2}_{X}(\mu_1||\mu_2) \le \rho$ holds if and only if for any $\gamma\colon X \to \mathrm{Prob}(\{\mathtt{Acc}, \mathtt{Rej}\})$,*

$$(\mathsf{Pr}[\gamma(\mu_1) = \mathtt{Rej}], \mathsf{Pr}[\gamma(\mu_2) = \mathtt{Acc}]) \in R^{\Delta}(\rho).$$

Proof. We fix a 2-cut $\overline{\Delta}^{2}$ of a divergence Δ. Suppose that it is defined with a set W satisfying $|W| = 2$.

$$\overline{\Delta}^{2}_{X}(\mu_1||\mu_2) = \sup_{\gamma\colon X \to \mathrm{Prob}(W)} \Delta_W(\gamma(\mu_1)||\gamma(\mu_2)).$$

We recall the definition of privacy region

$$R^{\Delta}(\rho) = \left\{ (x,y) \;\middle|\; \overline{\Delta}^{2}_{\{\mathtt{Acc},\mathtt{Rej}\}}((1-x)\mathbf{d}_{\mathtt{Acc}} + x\mathbf{d}_{\mathtt{Rej}}||y\mathbf{d}_{\mathtt{Acc}} + (1-y)\mathbf{d}_{\mathtt{Rej}}) \le \rho \right\}.$$

Since every probability distribution $\nu \in \mathrm{Prob}(\{\mathtt{Acc}, \mathtt{Rej}\})$ can be rewritten as $\nu = \mathsf{Pr}[\nu = \mathtt{Acc}]\mathbf{d}_{\mathtt{Acc}} + \mathsf{Pr}[\nu = \mathtt{Rej}]\mathbf{d}_{\mathtt{Rej}}$, we obtain

$$\begin{aligned} &\overline{\Delta}^{2}_{\{\mathtt{Acc},\mathtt{Rej}\}}(\gamma(\mu_1)||\gamma(\mu_2)) \le \rho \\ &\qquad \iff (\mathsf{Pr}[\gamma(\mu_1) = \mathtt{Rej}], \mathsf{Pr}[\gamma(\mu_2) = \mathtt{Acc}]) \in R^{\Delta}(\rho). \end{aligned}$$

Hence, it suffices to show

$$\begin{aligned} &(\overline{\Delta}^{2}_{X}(\mu_1||\mu_2) \le \rho) \\ &\iff \forall \gamma\colon X \to \mathrm{Prob}(\{\mathtt{Acc}, \mathtt{Rej}\}).(\overline{\Delta}^{2}_{\{\mathtt{Acc},\mathtt{Rej}\}}(\gamma(\mu_1)||\gamma(\mu_2)) \le \rho) \end{aligned}$$

($\Longrightarrow$) Obvious by the data-processing inequality of the 2-cut $\overline{\Delta}^{2}$.
($\Longleftarrow$) The assumption is equivalent to

$$\begin{aligned} &\forall \gamma\colon X \to \mathrm{Prob}(\{\mathtt{Acc}, \mathtt{Rej}\}).\forall \gamma'\colon \{\mathtt{Acc}, \mathtt{Rej}\} \to \mathrm{Prob}(W). \\ &\qquad (\Delta_W(\gamma'(\gamma(\mu_1))||\gamma'(\gamma(\mu_2))) \le \rho) \end{aligned}$$

Since $|W| = |\{\mathtt{Acc}, \mathtt{Rej}\}| = 2$, this is equivalent to

$$\gamma'' \colon X \to \mathrm{Prob}(W). \Delta_W(\gamma''(\mu_1)||\gamma''(\mu_2)) \leq \rho.$$

For any $\gamma \colon X \to \mathrm{Prob}(\{\mathtt{Acc}, \mathtt{Rej}\})$. and $\gamma' \colon \{\mathtt{Acc}, \mathtt{Rej}\} \to \mathrm{Prob}(W)$. we take $\gamma'' = \gamma' \bullet \gamma$. Conversely for any $\gamma'' \colon X \to \mathrm{Prob}(W)$ we take $\gamma = f \bullet \gamma''$ and $\gamma' = f^{-1}$ where $f \colon \{\mathtt{Acc}, \mathtt{Rej}\} \to W$ is a bijection. □

B.11 Detailed Proof of Theorem 20

Theorem 16 (Theorem 20)**.** *If a mechanism M is (α, ρ)-RDP then it is $(\rho + \log((\alpha-1)/\alpha) - (\log \delta + \log \alpha)/(\alpha - 1), \delta)$-DP for any $0 < \delta < 1$.*

Proof. The privacy region of Rényi divergence is given by

$$R^{D^\alpha}(\rho) = \left\{ (x, y) \;\middle|\; x^\alpha (1-y)^{1-\alpha} + (1-x)^\alpha y^{1-\alpha} \leq e^{\rho(\alpha-1)} \right\}.$$

Here we assume $0^{1-\alpha} = 0$.

By Lemma 10 (an extension of Lemma 15 in the paper), to find ε satisfying

$$\forall X. \forall \mu_1, \mu_2 \in \mathrm{Prob}(X).\ D^\alpha_X(\mu_1||\mu_2) \leq \rho \implies \Delta^\varepsilon_X(\mu_1||\mu_2) \leq \delta,$$

it is necessary and sufficient to find ε satisfying

$$\forall X. \forall \mu_1, \mu_2 \in \mathrm{Prob}(X).\ \overline{D^\alpha}^2_X(\mu_1||\mu_2) \leq \rho \implies \Delta^\varepsilon_X(\mu_1||\mu_2) \leq \delta.$$

By Thorem 14 (Theorem 18 in the paper), this is equivalent to find ε satisfying $R^{D^\alpha}(\rho) \subseteq R^{\Delta^\varepsilon}(\delta)$. Inspired from Mironov's proof of conversion law from RDP to DP [Mironov, 2017, Propisition 3]: we obtain,

$$\begin{aligned}
&x^\alpha (1-y)^{1-\alpha} + (1-x)^\alpha y^{1-\alpha} \leq e^{\rho(\alpha-1)} \\
&\implies (1-x)^\alpha y^{1-\alpha} \leq e^{\rho(\alpha-1)} \\
&\implies (1-x) \leq (e^\rho y)^{\frac{\alpha-1}{\alpha}} &&(\dagger)\\
&\implies (e^\rho y > \delta^{\frac{\alpha}{\alpha-1}} \implies (1-x) \leq e^{\rho - \log d/(\alpha-1)} y) \\
&\quad \wedge (e^\rho y \leq \delta^{\frac{\alpha}{\alpha-1}} \implies (1-x) \leq \delta) \\
&\implies (1-x) \leq e^{\rho - \log d/(\alpha-1)} y + \delta.
\end{aligned}$$

The last part of $(1-x) \leq e^{\rho - \log d/(\alpha-1)} y + \delta$ derives Mironov's result [Mironov, 2017, Propisition 3]. Now, starting from (†), we have a better bound for DP as follows: consider a curve C given by the equation

$$1 - x = (e^\rho y)^{\frac{\alpha-1}{\alpha}} \iff x = 1 - (e^\rho y)^{\frac{\alpha-1}{\alpha}}$$

. We have the derivative of x as follows:

$$\frac{dx}{dy} = -\frac{\alpha-1}{\alpha} e^{\frac{\alpha-1}{\alpha}\rho} y^{-\frac{1}{\alpha}}$$

We can take the tangent of the curve C by

$$x = \frac{dx}{dy}(t)(y-t) + (e^\rho(1-t))^{\frac{\alpha-1}{\alpha}}$$

We will find parameters that a tangent of C meets $(1-x) = e^{\varepsilon} y + \delta$. $x = -e^{\varepsilon} y - \delta + 1$ We first solve

$$-e^{\varepsilon} = \frac{dx}{dy}(t) = -\frac{\alpha-1}{\alpha} e^{\frac{\alpha-1}{\alpha}\rho} t^{-\frac{1}{\alpha}} \iff \varepsilon = \log(\frac{\alpha-1}{\alpha}) + \frac{\alpha-1}{\alpha}\rho - \frac{1}{\alpha}\log t.$$

Next we solve

$$1-\delta = -t\frac{dx}{dy}(t) + 1 - (e^{\rho}t)^{\frac{\alpha-1}{\alpha}} \iff 1-\delta = \frac{\alpha-1}{\alpha} e^{\frac{\alpha-1}{\alpha}\rho} t^{-\frac{1}{\alpha}} t + 1 - (e^{\rho}t)^{\frac{\alpha-1}{\alpha}}$$

We then have

$$\delta = (e^{\rho}t)^{\frac{\alpha-1}{\alpha}} - \frac{\alpha-1}{\alpha} e^{\frac{\alpha-1}{\alpha}\rho} t^{-\frac{1}{\alpha}} t = \frac{1}{\alpha}(e^{\rho}t)^{\frac{\alpha-1}{\alpha}} \iff t = (\delta\alpha e^{-\frac{\alpha-1}{\alpha}\rho})^{\frac{\alpha}{\alpha-1}}$$

Simple computations give the following:

$$\varepsilon = \log(\frac{\alpha-1}{\alpha}) + \rho - \frac{\log\delta + \log\alpha}{\alpha-1}.$$

By the symmetry of $R^{D^{\alpha}}(\rho)$ and $R^{\Delta^{\varepsilon}}(\delta)$, we have

$$R^{D^{\alpha}}(\rho) \subseteq R^{\Delta^{\varepsilon}}(\delta).$$

As we mentioned, it is equivalent to

$$\forall X. \forall \mu_1, \mu_2 \in \mathrm{Prob}(X).\ D^{\alpha}_X(\mu_1||\mu_2) \le \rho \implies \Delta^{\varepsilon}_X(\mu_1||\mu_2) \le \delta.$$

This completes the proof. □

As a conjecture, if we calculate tangents of the boundary of the privacy region $R^{D^{\alpha}}(\rho)$, we have *optimal* conversion law from (α,ρ)-RDP to DP. The boundary of $R^{D^{\alpha}}(\rho)$ is given by the equation

$$x^{\alpha}(1-y)^{1-\alpha} + (1-x)^{\alpha} y^{1-\alpha} = e^{\rho(\alpha-1)}.$$

B.12 Proof of Theorem 22

Theorem 17 (Theorem 22)**.** *Let $F\colon [0,1]^{2k} \to [0,\infty]$ be a quasi-convex function. Then the divergence Δ^F defined below is k-generated and quasi-convex.*

$$\Delta^F_X(\mu_1||\mu_2) \stackrel{\mathsf{def}}{=} \sup_{\substack{\{A_i\}_{i=1}^k \\ \textit{partition of } X}} F(\mu_1(A_1), \cdots, \mu_1(A_k), \mu_2(A_1), \cdots, \mu_2(A_k)).$$

Proof. The quasi-convexity is obvious from the quasi-convexity of $F\colon [0,1]^{2k} \to [0,\infty]$. We show the k-generatedness. We take the k-cut with respect to the k-element set $\{1, 2, \ldots, k\}$. We may assume X is countable. For any

$\mu_1, \mu_2 \in \mathrm{Prob}(X)$,

$$\begin{aligned}
&\overline{\Delta^F}^k_X(\mu_1||\mu_2)\\
&= \sup_{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})} \Delta^F_{\{1,2,\ldots,k\}}(\gamma(\mu_1)||\gamma(\mu_2))\\
&= \sup_{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})} \sup_{\substack{\{A_i\}_{i=1}^k\\ \text{partition of}\\ \{1,2,\ldots,k\}}} F\begin{pmatrix} (\gamma(\mu_1))(A_1), \cdots, (\gamma(\mu_1))(A_k),\\ (\gamma(\mu_2))(A_1), \cdots, (\gamma(\mu_2))(A_k) \end{pmatrix}\\
&= \sup_{\substack{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})\\ p\colon \{1,2,\ldots,k\}\to\{1,2,\ldots,k\}}} F\begin{pmatrix} (\gamma(\mu_1))(p^{-1}(1)), \cdots, (\gamma(\mu_1))(p^{-1}(k)),\\ (\gamma(\mu_2))(p^{-1}(1)), \cdots, (\gamma(\mu_2))(p^{-1}(k)) \end{pmatrix}\\
&= \sup_{\substack{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})\\ p\colon \{1,2,\ldots,k\}\to\{1,2,\ldots,k\}}} F\begin{pmatrix} (((p\bullet\gamma)(\mu_1))(1), \cdots, ((p\bullet\gamma)(\mu_1))(k),\\ ((p\bullet\gamma)(\mu_2))(1), \cdots, ((p\bullet\gamma)(\mu_2))(k) \end{pmatrix}\\
&= \sup_{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})} F\begin{pmatrix} ((\gamma(\mu_1))(1), \cdots, (\gamma(\mu_1))(k),\\ (\gamma(\mu_2))(1), \cdots, (\gamma(\mu_2))(k) \end{pmatrix}.
\end{aligned}$$

Here by weak Birkhoff-von Neumann thorem (countable version), every function $\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})$ is decomposed into a (countable) convex combination $\sum_{i\in I} a_i(\eta_{\{1,2,\ldots,k\}} \circ \gamma_i)$ of $\gamma_i\colon X \to \{1,2,\ldots,k\}$. Hence,

$$\begin{aligned}
&\overline{\Delta^F}^k_X(\mu_1||\mu_2)\\
&= \sup_{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})} F\begin{pmatrix} ((\gamma(\mu_1))(1), \cdots, (\gamma(\mu_1))(k),\\ (\gamma(\mu_2))(1), \cdots, (\gamma(\mu_2))(k) \end{pmatrix} \qquad (\dagger)\\
&= \sup_{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})} F\begin{pmatrix} ((\sum_{i\in I} a_i(\eta_{\{1,2,\ldots,k\}} \circ \gamma_i)(\mu_1))(1),\\ \cdots, (\sum_{i\in I} a_i(\eta_{\{1,2,\ldots,k\}} \circ \gamma_i)(\mu_1))(k),\\ (\sum_{i\in I} a_i(\eta_{\{1,2,\ldots,k\}} \circ \gamma_i)(\mu_2))(1),\\ \cdots, (\sum_{i\in I} a_i(\eta_{\{1,2,\ldots,k\}} \circ \gamma_i)(\mu_2))(k) \end{pmatrix}\\
&= \sup_{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})} F\begin{pmatrix} (\sum_{i\in I} a_i(\gamma_i(\mu_1))(1), \cdots, \sum_{i\in I} a_i(\gamma_i(\mu_1))(k),\\ \sum_{i\in I} a_i(\gamma_i(\mu_2))(1), \cdots, \sum_{i\in I} a_i(\gamma_i(\mu_2))(k) \end{pmatrix}\\
&\leq \sup_{\gamma\colon X \to \mathrm{Prob}(\{1,2,\ldots,k\})} \sup_{i\in I} F\begin{pmatrix} ((\gamma_i(\mu_1))(1), \cdots, (\gamma_i(\mu_1))(k),\\ (\gamma_i(\mu_2))(1), \cdots, (\gamma_i(\mu_2))(k) \end{pmatrix}\\
&\leq \sup_{\gamma\colon X \to \{1,2,\ldots,k\}} F\begin{pmatrix} ((\gamma(\mu_1))(1), \cdots, (\gamma(\mu_1))(k),\\ (\gamma(\mu_2))(1), \cdots, (\gamma(\mu_2))(k) \end{pmatrix}\\
&= \sup_{\gamma\colon X \to \{1,2,\ldots,k\}} F\begin{pmatrix} (\mu_1(\gamma^{-1}(1)), \cdots, \mu_1(\gamma^{-1}(k)),\\ \mu_2(\gamma^{-1}(1)), \cdots, \mu_2(\gamma^{-1}(k)) \end{pmatrix}\\
&\leq \sup_{\substack{\{A_i\}_{i=1}^k\\ \text{partition of}\\ \{1,2,\ldots,k\}}} F\begin{pmatrix} (\gamma(\mu_1))(A_1), \cdots, (\gamma(\mu_1))(A_k),\\ (\gamma(\mu_2))(A_1), \cdots, (\gamma(\mu_2))(A_k) \end{pmatrix}\\
&= \Delta^F_X(\mu_1||\mu_2).
\end{aligned}$$

We have $\overline{\Delta^F}^k_X(\mu_1||\mu_2) \leq \Delta^F_X(\mu_1||\mu_2)$. Conversely, by equality (†), we also have $\overline{\Delta^F}^k_X(\mu_1||\mu_2) \geq \Delta^F_X(\mu_1||\mu_2)$. This completes the proof. □

General version If the quasi-convex function $F\colon [0,1]^{2k} \to [0,\infty]$ is also continuous, we can extend Theorem 22 to general measurable setting.

Theorem 18 (k-generatedness in general setting)**.** *Assume that $F\colon [0,1]^{2k} \to [0,\infty]$ is quasi-convex and continuous. For any measurable space X, we have*

$$\Delta_X(\mu_1,\mu_2) = \sup_{\substack{\gamma\colon X\to \mathrm{Prob}(\{1,2,\dots k\}) \\ \textit{measurable function}}} \Delta_X(\gamma(\mu_1),\gamma(\mu_2)).$$

Proof. We easily calculate as follows (functions are assumed to be measurable):

$$\begin{aligned}
&\Delta_X(\mu_1,\mu_2)\\
&= \sup\left\{ F(\mu_1(A_1),\cdots,\mu_1(A_k),\mu_2(A_1),\cdots,\mu_2(A_k)) \;\middle|\; \{A_i\}_{i=1}^k\colon \text{ m'ble partition of } X \right\}\\
&= \sup\left\{ F(\mu_1(f^{-1}(1)),\cdots,\mu_1(f^{-1}(k)),\mu_2(f^{-1}(1)),\cdots,\mu_1(f^{-1}(k))) \;\middle|\; f\colon X\to\{1,2,\dots,k\} \right\}\\
&= \sup\{ F((f(\mu_1))(1),\cdots,(f(\mu_1))(k),(f(\mu_2))(1),\cdots,(f(\mu_2))(k)) \mid f\colon X\to\{1,2,\dots,k\} \}\\
&\le \sup\{ F((\gamma(\mu_1))(1),\cdots,(\gamma(\mu_1))(k),(\gamma(\mu_2))(1),\cdots,(\gamma(\mu_2))(k)) \mid \gamma\colon X\to \mathrm{Prob}(\{1,2,\dots,k\}) \}\\
&\le \sup\left\{ F\begin{pmatrix} (\gamma(\mu_1))(A_1),\cdots,(\gamma(\mu_1))(A_k), \\ (\gamma(\mu_2))(A_1),\cdots,(\gamma(\mu_2))(A_k)\end{pmatrix} \;\middle|\; \begin{matrix}\gamma\colon X\to\mathrm{Prob}(\{1,2,\dots,k\}), \\ \{A_i\}_{i=1}^k\colon \text{ m'ble partition of } X\end{matrix} \right\}\\
&= \sup_{\gamma\colon X\to\mathrm{Prob}(\{1,2,\dots,k\})} \Delta_{\{}1,2,\dots,k\}(\gamma(\mu_1),\gamma(\mu_2))
\end{aligned}$$

Note that we treat $\{1,2,\dots,k\}$ as a finite discrete space. Consider the family $\{J_n\}_{n=1}^\infty$ of finite sets (discrete spaces) defined as follows:

$$J_n = \left\{ (j_1,\dots,j_k) \;\middle|\; j_1,\dots,j_k \in \{0,1,\dots,2^n-1\}, C^n_{j_1\cdots j_k} \neq \emptyset \right\}.$$

We fix a measurable function $\gamma\colon X\to\mathrm{Prob}(k)$ and treat $\mathrm{Prob}(k)$ as a subset of $[0,1]^k$. For each $n\in\mathbb{N}$, we define a measurable partition $\{C^n_{j_1\cdots j_k}\}_{j_1,\dots,j_k\in\{0,1,\dots,2^n-1\}}$ of X by

$$\begin{aligned}
C^n_{j_1\cdots j_k} &= \gamma^{-1}(B^n_{j_1\cdots j_k})\\
&\qquad \text{where } B^n_{j_1\cdots j_N} = D_{j_1}\times\cdots\times D_{j_k} \qquad ((j_1\dots j_k)\in J_n),\\
&\qquad\quad D^n_0 = \{0\} \text{ and } D^n_{l+1} = (l/2^n,(l+1)/2^n] \qquad (l=0,1,2,\dots,2^n-1).
\end{aligned}$$

We next define $m_n^*\colon X\to J_n$ and $m_n\colon J_n\to X$ as follows: $m_n^*(x)$ is the unique element $(j_1,\dots,j_k)\in J_n$ satisfying $x\in C^n_{j_1,\dots,j_k}$, and we choose $m_n(j_1,\dots,j_k)$ is an element of $C^n_{j_1,\dots,j_k}$. Thanks to the measurability of each $C^n_{j_1\cdots j_k}$, the function m_n^* is measurable, and the measurability of m_n follows from the discreteness of J_n. From the construction of $\{C^n_{j_1\cdots j_k}\}_{j_1,\dots,j_k\in\{0,1,\dots,2^n-1\}}$, for any $n\in\mathbb{N}$, $x\in X$, and $i\in I$, we have,

$$|\gamma(x)(i) - (\gamma\circ m_n\circ m_n^*)(x)(i)| \le 1/2^n$$

This implies that the sequence $\{\gamma\circ m_n\circ m_n^*\}_{n=1}^\infty$ of measurable function converges uniformly to γ. Hence, for any $n\in\mathbb{N}$ and $D\subseteq k$, we have

$$\left| \int \gamma(x)(D)\; d\mu_1(x) - \int (\gamma\circ m_n\circ m_n^*)(x)(D)\; d\mu_1(x) \right| \le 1/2^n$$

Hence the sequence of probability measures $\{(\gamma\circ m_n\circ m_n^*)(\mu_1)\}_{n=1}^\infty$ converges to the probability measure $\gamma(\mu_1)$. Similarly, $\{(\gamma\circ m_n\circ m_n^*)(\mu_2)\}_{n=1}^\infty$ converges to $\gamma(\mu_2)$.

By the continuity of F, we obtain

$$
\begin{aligned}
&F((\gamma(\mu_1))(A_1),\cdots,(\gamma(\mu_1))(A_k),(\gamma(\mu_2))(A_1),\cdots,(\gamma(\mu_2))(A_k))\\
&=\lim_{n\to\infty} F\begin{pmatrix}((\gamma\circ m_n\circ m_n^*)(\mu_1))(A_1),\cdots,((\gamma\circ m_n\circ m_n^*)(\mu_1))(A_k),\\ ((\gamma\circ m_n\circ m_n^*)(\mu_2))(A_1),\cdots,((\gamma\circ m_n\circ m_n^*)(\mu_2))(A_k)\end{pmatrix}\\
&=\lim_{n\to\infty} F\begin{pmatrix}((\gamma\circ m_n)(m_n^*(\mu_1)))(A_1),\cdots,((\gamma\circ m_n)(m_n^*(\mu_1)))(A_k),\\ ((\gamma\circ m_n)(m_n^*(\mu_2)))(A_1),\cdots,((\gamma\circ m_n)(m_n^*(\mu_2)))(A_k)\end{pmatrix}\\
&\le \sup_{n\in\mathbb{N}} \Delta_{\{1,2,\ldots,k\}}(((\gamma\circ m_n)(m_n^*(\mu_1))),((\gamma\circ m_n)(m_n^*(\mu_2))))\\
&\qquad \{\text{Since } J_n \text{ is finite (countable and discrete), we can apply Theorem 22.}\}\\
&\le \sup_{n\in\mathbb{N}} \Delta_{J_n}(m_n^*(\mu_1),m_n^*(\mu_2))\\
&=\sup_{n\in\mathbb{N}}\sup\left\{ F(\begin{pmatrix} f(m_n^*(\mu_1)))(1),\cdots,(f(m_n^*(\mu_1)))(k),\\ (f(m_n^*(\mu_2)))(1),\cdots,((f(m_n^*(\mu_1)))(k)\end{pmatrix} \;\middle|\; f\colon J_n\to\{1,2,\ldots,k\}\right\}\\
&\le \sup\{F((g(\mu_1))(1),\cdots,(g(\mu_1))(k),(g(\mu_2))(1),\cdots,(g(\mu_2))(k)) \mid g\colon X\to\{1,2,\ldots,k\}\}\\
&=\Delta_X(\mu_1,\mu_2).
\end{aligned}
$$

This implies $\sup_{\gamma\colon X\to \mathrm{Prob}(k)} \Delta_k(\gamma(\mu_1),\gamma(\mu_2)) \le \Delta_X(\mu_1,\mu_2)$. □

C Additional Results

C.1 Total variation distance is 2-generated

We recall the definition of the total variation distance

$$\mathtt{TV}_X(\mu_1||\mu_2)=\sup_{S\subseteq X}|\mathsf{Pr}[\mu_1\in S]-\mathsf{Pr}[\mu_2\in S]|.$$

In a similar way as ε-divergence Δ^ε, we can prove 2-generatedness of the total variation distance $\mathtt{TV}$, but we can prove it easily by applying Theorems 16–17 (Theorems 20 and 22 in the paper).

Define $F\colon[0,1]^4\to[0,\infty]$ by $F(x,x',y,y')=|x-y|$. It is easy to check that the function is obviously quasi-convex, and that we have $\mathtt{TV}=\Delta^F$.

C.2 An optimal conversion law from Hellinger to DP

We recall the definition of the Hellinger distance

$$\mathrm{HD}_X(\mu_1||\mu_2)=1-\sum_{x\in X}\sqrt{\mu_1(x)\mu_2(x)}.$$

Since it is the f-divergence of weight function $w(t)=\sqrt{t}-1$ (strict convex), the Hellinger distance is exactly ∞-generated, quasi-convex and continuous.

Here is the essense of an optimal conversion law from the Hellinger distance to DP.

Lemma 19. *We have* $R^{\mathrm{HD}}(\rho) \subseteq R^{\Delta^{\varepsilon}}(\delta(\varepsilon, \rho))$ *where*

$$\delta(\varepsilon, \rho) = 1 - t - \frac{f(t)}{g(t)} \tag{1}$$

$$t = \frac{z^2 + 4 - z\sqrt{z^2+4}}{2(z^2+4)} \tag{2}$$

$$z = \frac{1/e^{\varepsilon} - 2(1-\rho) + 1}{(1-\rho)\sqrt{\rho(2-\rho)}}$$

$$f(x) = (1-\rho)^2(1-2x) + x - 2(1-\rho)\sqrt{d(2-d)x(1-x)}$$

$$g(x) = \frac{df}{dx}(x) = (1-\rho)^2(1-2x) + xf \;\; - 2(1-\rho)\sqrt{d(2-d)x(1-x)}$$

Proof. We may regard

$$R^{\mathrm{HD}}(\rho) = \left\{ (x, y) \in [0,1]^2 \;\middle|\; 1 - \sqrt{x(1-y)} - \sqrt{(1-x)y} \leq \rho \right\},$$
$$R^{\Delta^{\varepsilon}}(\delta) = \left\{ (x, y) \in [0,1]^2 \mid \max((1-x) - e^{\varepsilon}y, x - e^{\varepsilon}(1-y)) \leq \delta \right\}.$$

We first calculate the boundary of $R^{\mathrm{HD}}(\rho)$. Thus, we solve the following equation for y:

$$1 - \sqrt{x(1-y)} - \sqrt{(1-x)y} = \rho.$$

We first have

$$\begin{aligned}
&1 - \sqrt{x(1-y)} - \sqrt{(1-x)y} = \rho \\
&\iff (1-\rho)^2 - x(1-y) - y(1-x) = 2\sqrt{x(1-x)y(1-y)} \\
&\iff (1-\rho)^4 + x^2(1-y)^2 + y^2(1-x)^2 - 2x(1-y)(1-\rho)^2 - 2y(1-x)(1-\rho)^2
\end{aligned}$$

The degree of this equation is 2, so we can solve it. For given $x \in [0,1]$, we have

$$y = (1-\rho)^2(1-2x) + x \pm 2(1-\rho)\sqrt{x(1-x)\rho(2-\rho)}.$$

Thanks to the Symmetry of $R^{\mathrm{HD}}(\rho)$ and $R^{\Delta^{\varepsilon}}(\delta)$, we may consider the curve:

$$y = (1-\rho)^2(1-2x) + x - 2(1-\rho)\sqrt{x(1-x)\rho(2-\rho)} = f(x).$$

The tangent of the curve $y = f(x)$ that passes the point $(t, f(t))$ is given by the equation $x - \frac{y}{g(t)} = t - \frac{f(t)}{g(t)}$ where $g(x) = \frac{df}{dx}(x)$. We next find t and δ that the lower boundary

$$(1-x) - e^{\varepsilon}y = \delta \iff x + e^{\varepsilon}y = 1 - \delta$$

of $R^{\Delta^{\varepsilon}}(\delta(\varepsilon, \rho))$ is the same as the line $x - \frac{y}{g(t)} = t - \frac{f(t)}{g(t)}$. We solve the equation $e^{\varepsilon} = \frac{1}{g(t)}$ on t about the slope as (2). Finally, we obtain δ as (1). □

We conclude an optimal conversion law from the Hellinger distance to DP.

Theorem 20. *We always have* $\mathrm{HD}_X(d_1, d_2) \leq \rho \implies \Delta^{\varepsilon}_X(d_1, d_2) \leq \delta(\varepsilon, \rho)$ *where* $\delta(\varepsilon, \rho)$ *is given by (1).*

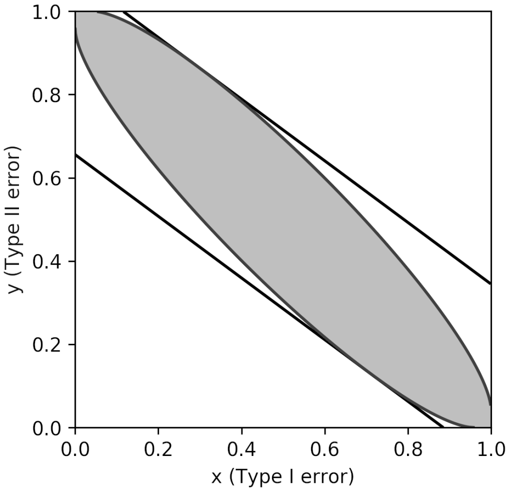

Figure 1: Comparison of the privacy region for DP and the one for 2-cut of Hellinger distance.

References

[Barthe and Olmedo, 2013] Barthe, G. and Olmedo, F. (2013). Beyond differential privacy: Composition theorems and relational logic for f-divergences between probabilistic programs. In *International Colloquium on Automata, Languages and Programming (ICALP), Riga, Latvia*, volume 7966 of *Lecture Notes in Computer Science*, pages 49–60. Springer-Verlag.

[Giry, 1982] Giry, M. (1982). A categorical approach to probability theory. In Banaschewski, B., editor, *Categorical Aspects of Topology and Analysis*, volume 915 of *Lecture Notes in Mathematics*, pages 68–85. Springer-Verlag.

[Liese and Vajda, 2006] Liese, F. and Vajda, I. (2006). On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412.

[Mironov, 2017] Mironov, I. (2017). Rényi differential privacy. In *IEEE Computer Security Foundations Symposium (CSF), Santa Barbara, California*, pages 263–275.