Universidad Nacional Autónoma de México

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y DE LA ESPECIALIZACIÓN EN ESTADÍSTICA MATEMÁTICA

UNA INTRODUCCIÓN A LA PRIVACIDAD DIFERENCIAL APROXIMADA Y DE RÉNYI

TESIS QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS

PRESENTA:

RICARDO FLORES LÓPEZ

DIRECTOR:

DR. MARIO ALBERTO DIAZ TORRES

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

CIUDAD DE MÉXICO 13 DE NOVIEMBRE DEL 2023.

Índice general

Capítulo 1

Introducción

En estos tiempos actuales la recopilación, análisis y protección de volúmenes masivos de datos juega un papel importante en diversas áreas de las ciencias de la computación y una rama que está brindando grandes resultados es el aprendizaje automático (*machine learning*). El aprendizaje automático es una rama de la inteligencia artificial que permite a un sistema aprender de los datos en lugar de aprender mediante la programación explícita. De manera informal, un modelo de aprendizaje automatizado es la salida de información que se genera cuando se entrena un algoritmo de aprendizaje automático con datos. Después del entrenamiento, al proporcionar un modelo con una entrada, se le dará una salida. A continuación, cuando proporcione el modelo predictivo con datos, recibirá un pronóstico basado en los datos que entrenaron al modelo.

Un problema fundamental que surge al aplicar el aprendizaje automático es que existe el riesgo de privacidad de datos. Las filtraciones de datos son cada vez más frecuentes y costosas de manejar. Además, los propios modelos para el aprendizaje automático pueden presentar vulnerabilidad, ya que se puede extraer datos confidenciales de ellos. De esta manera se busca brindar garantías en tres aspectos para un conjunto de datos: *Privacidad en la entrada* (la cual busca garantizar que otras partes, incluido el desarrollador del modelo, no podrá ver los datos de entrada de un usuario), *Privacidad en la salida* (la cual busca la seguridad de que el resultado de un modelo solo sea accesible para el cliente cuyos datos se infieren) y *Privacidad del modelo* (la cual busca la seguridad de que una parte hostil no podrá robar el modelo).

Dentro de los métodos que existen para garantizar que los datos no puedan ser robados está la *Privacidad Diferencial* la cual fue primero introducida en el año 2006 por Dwork et al., y ha sido abrazada por múltiples comunidades de investigación como una noción común de privacidad para algoritmos de bases de datos estadísticos. La privacidad diferencial es un tipo de privacidad que permite proporcionar información relevante sobre un conjunto de datos sin revelar ninguna información personal al respecto. En otras palabras, la presencia del registro de un individuo en el conjunto de datos no tiene un impacto en el resultado del análisis. Como resultado, el riesgo de privacidad es básicamente el mismo ya sea que una persona participe o no en el conjunto de datos.

Informalmente hablando privacidad diferencial se logra agregando ruido aleatorio al resultado, lo que se puede hacer a través de una variedad de procesos privados diferenciales, es decir, la privacidad diferencial limita un cambio en la distribución de salida de un algoritmo aleatorizante que puede ser inducido por una variación pequeña de su entrada.

La definición estándar de privacidad diferencial, la llamada ϵ-privacidad diferencial, pone un límite superior multiplicativo en el cambio en la densidad de la distribución como es el caso del mecanismo de Laplace. Una característica central que presenta es que es cerrada bajo composición; además, los mecanismos compuestos que son ϵ-diferencialmente privados simplemente se suman. Por otro lado una relajación a la ϵ-privacidad diferencial, llamada (ϵ, δ)-privacidad diferencial, es un método para expresar las garantías de privacidad de una variedad de algoritmos diferencialmente privados (DP), especialmente aquellos que introducen aleatoriedad bajo ruido de tipo Gaussiano. Una diferencia esencial entre estos dos conceptos se encuentra en el término aditivo δ el cual permite eliminar las colas largas de los mecanismos de distribución donde las garantías de ϵ-privacidad diferencial pueden no sustentarse. Otro uso a la (ϵ, δ)-privacidad diferencial se da en los teoremas de composición avanzada.

Motivados por los resultados y propiedades anteriormente expuestos, en esta tesis se analiza una alternativa para ϵ-privacidad diferencial y (ϵ, δ)-privacidad diferencial llamada *Privacidad Diferencial de Rényi*. Comparada con la (ϵ, δ)-privacidad diferencial, la privacidad diferencial de Rényi ofrece una manera operacionalmente conveniente y accesible para poder seguir la pérdida de privacidad acumulativa durante la ejecución de mecanismos

independientes. Además, la privacidad diferencial de Rényi se presenta como un concepto intuitivo para los resultados de composición avanzada de mecanismos.

Para lograr una exposición accesible a la privacidad diferencial de Rényi en el capítulo 2 introducimos las nociones básicas de la teoría de la información que nos servirán de referencia para poder abordar el concepto de *Divergencia de Rényi*, el cual es fundamental para el estudio de la privacidad diferencial de Rényi. En general estudiamos el concepto de divergencia de Kullback-Leibler y sus propiedades básicas, para posteriormente dar la definición de la divergencia de Rényi y realizar una comparativa de las propiedades que estos dos conceptos comparten. Finalizamos con el estudio de la divergencia de Rényi para ciertas distribuciones de probabilidad que se usan a lo largo de este trabajo.

En el capítulo 3 se introducen los conceptos de base de datos, distancia y adyacencia entre base de datos y de mecanismo aleatorizante. Seguidamente se definen los conceptos de ϵ-privacidad diferencial y (ϵ, δ)-privacidad diferencial que nos servirán para poder abordar la privacidad diferencial de Rényi, y para finalizar con las propiedades básicas que presenta: post-procesamiento, privacidad grupal, composición básica y composición avanzada.

En el capítulo 4 es donde se introduce el tema central de esta tesis que es privacidad diferencial de Rényi. Se inicia con la definición y posteriormente se estudian las propiedades que comparte con ϵ-privacidad diferencial y (ϵ, δ)-privacidad diferencial tales como composición adaptativa y privacidad grupal. Seguidamente se estudia el importante resultado del teorema de composición avanzada y se finaliza con una relación entre privacidad diferencial de Rényi y (ϵ, δ)-privacidad diferencial.

Finalmente, en el capítulo 5 se busca aplicar la privacidad diferencial de Rényi sobre tres mecanismos aleatorizantes básicos: mecanismo de respuesta aleatoria, mecanismo de Laplace y mecanismo Gaussiano, y realizar una comparativa con la privacidad diferencial aproximada.

Capítulo 2

Preliminares de Teoría de la Información

En este capítulo introducimos conceptos básicos de la teoría de la información que usaremos a lo largo de esta tesis. Comenzamos con la definición de la divergencia de Kullback-Leibler para dos distribuciones de probabilidad y recordamos sus propiedades básicas. Posteriormente introducimos la definición de la divergencia de Rényi y sus propiedades que usaremos en los capítulos 3 y 4. Finalmente concluimos con el cálculo de la divergencia de Rényi para las distribuciones de Bernoulli, Laplace y normal en algunos casos particulares.

2.1. Divergencia de Kullback-Leibler

La divergencia de Kullback-Leibler, también conocida como entropía relativa, es una medida de la diferencia entre dos distribuciones de probabilidad. Su definición se da a continuación.

Definición 2.1.1. *Sean P y Q distribuciones definidas sobre un conjunto contable $\mathcal{X}$. Se define la divergencia de Kullback-Leibler como*

$$D_{KL}(P\|Q) := \sum_{x\in\mathcal{X}} \log\left(\frac{P(x)}{Q(x)}\right) P(x), \tag{2.1}$$

donde convenimos que $0 \cdot \log\left(\frac{0}{0}\right) = 0$ *(lo cual se sigue por continuidad de la función logaritmo) y* $a \log(a/0) = \infty$ *si* $a > 0$.

Una expresión equivalente a (2.1) es

$$D_{KL}(P\|Q) = \mathbb{E}_{x\sim P}\left[\log \frac{P(x)}{Q(x)}\right]. \tag{2.2}$$

Como puede verse, la divergencia de Kullback-Leibler (KL) actúa como una medida de semejanza entre distribuciones, pero no es simétrica y no satisface la desigualdad triangular. Sin embargo, tiene algunas propiedades deseables las cuales se enuncian a continuación.

Proposición 2.1.1. *Sean* P *y* Q *distribuciones definidas sobre un conjunto contable* $\mathcal{X}$*, entonces*

1) $D_{KL}(P\|Q) \geq 0$ *con igualdad si y sólo si* $P(x) = Q(x)$ *para todo* $x \in \mathcal{X}$.

2) Para un par de distribuciones P_1, Q_1 *y* P_2, Q_2 *definidas sobre* $\mathcal{X}$ *se cumple que*

$$D_{KL}\left(\lambda P_1 + (1-\lambda)P_2 \| \lambda Q_1 + (1-\lambda)Q_2\right) \leq \lambda D_{KL}(P_1\|Q_1) + (1-\lambda) D_{KL}(P_2\|Q_2)$$

para cualquier $\lambda \in [0,1]$.

Demostración. Para la parte 1) consideramos la función $\varphi : (0,\infty) \to \mathbb{R}$ dada por $\varphi(x) = x\log(x)$ entonces partiendo de la ecuación (2.1) se obtiene que

$$D_{KL}\left(P\|Q\right) = \sum_{x\in\mathcal{X}} Q(x)\varphi\left(\frac{P(x)}{Q(x)}\right) = \mathbb{E}_{x\sim Q(x)}\left[\varphi\left(\frac{P(x)}{Q(x)}\right)\right]. \tag{2.3}$$

Aplicando la desigualdad de Jensen (ver apéndice) a la igualdad (2.3), pues φ es convexa, se obtiene

$$D_{KL}\left(P\|Q\right) \geq \varphi\left(\mathbb{E}_{x\sim Q(x)}\left(\frac{P(x)}{Q(x)}\right)\right), \tag{2.4}$$

pero $\mathbb{E}_{x\sim Q(x)}\left(\frac{P(x)}{Q(x)}\right) = 1$ obteniendo la desigualdad pedida.

Por otro lado de la igualdad (2.2)

$$\mathbb{E}_{x\sim P(x)}\left[\log\left(\frac{P(x)}{Q(x)}\right)\right] = 0 \Leftrightarrow \log\left(\frac{P(x)}{Q(x)}\right) = 0 \quad (c.s). \tag{2.5}$$

De esta forma de la última igualdad de (2.5) se sigue que $P(x) = Q(x)$ $(c.s)$. Con lo cual queda demostrada la parte 1).

Para la parte 2) partiendo de la definición 2.1.1

$$\begin{aligned} &D_{KL}\left(\lambda P_1 + (1-\lambda)P_2 \| \lambda Q_1 + (1-\lambda)Q_2\right) \\ &= \sum_{x\in\mathcal{X}} \log\left(\frac{\lambda P_1(x) + (1-\lambda)P_2(x)}{\lambda Q_1(x) + (1-\lambda)Q_2(x)}\right)\left(\lambda P_1(x) + (1-\lambda)P_2(x)\right). \end{aligned} \tag{2.6}$$

Aplicando la desigualdad de la suma logarítmica (ver apéndice) a un sumando de la suma del lado derecho de la igualdad (2.6) obtenemos que

$$\begin{aligned} &\left(\lambda P_1(x) + (1-\lambda)P_2(x)\right)\log\left(\frac{\lambda P_1(x) + (1-\lambda)P_2(x)}{\lambda Q_1(x) + (1-\lambda)Q_2(x)}\right) \\ &\leq \lambda P_1(x)\log\left(\frac{P_1(x)}{Q_1(x)}\right) + (1-\lambda)P_2(x)\log\left(\frac{P_2(x)}{Q_2(x)}\right). \end{aligned} \tag{2.7}$$

Finalmente sumando en ambos lados de la desigualdad (2.7) se obtiene lo pedido en 2) completando la demostración. ■

2.2. Divergencia de Rényi

Una generalización para la divergencia de Kullback-Leibler, introducida en la sección anterior, es la divergencia de Rényi sobre la cual estudiaremos privacidad diferencial. A continuación se establece la definición.

Definición 2.2.1. *Sean P y Q dos distribuciones de probabilidad sobre un conjunto contable $\mathcal{X}$, la divergencia de Rényi de orden $\alpha > 1$ se define como*

$$D_\alpha(P\|Q) := \frac{1}{\alpha - 1}\log\left[\mathbb{E}_{x\sim Q}\left(\left(\frac{P(x)}{Q(x)}\right)^\alpha\right)\right], \tag{2.8}$$

en donde el logaritmo es el natural.

Al igual que se realizó para la divergencia de Kullback-Leibler, la ecuación (2.8) puede escribirse como

$$D_\alpha\left(P\|Q\right) = \frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}}\left(\frac{P(x)}{Q(x)}\right)^\alpha Q(x). \tag{2.9}$$

De la definición se tiene que el orden α toma valores en el intervalo $(1,\infty)$. Para los casos $\alpha = 1$ y $\alpha = \infty$ la divergencia de Rényi es definida por continuidad. Concretamente,

$$D_1(P||Q) = \lim_{\alpha\to 1} D_\alpha(P||Q) \tag{2.10}$$

mientras que

$$D_\infty(P||Q) = \sup_{x\in\mathrm{supp}(Q)} \log\frac{P(x)}{Q(x)}. \tag{2.11}$$

2.3. Propiedades de la Divergencia de Rényi

Para una mejor comprensión de la privacidad diferencial de Rényi, la cual se estudiará en los capítulos 3 y 4, enunciamos y demostramos en esta sección las propiedades básicas de esta divergencia.

Lema 2.3.1. *Para $\alpha \geq 1$ y distribuciones P, Q arbitrarias se cumple que*

$$D_\alpha\left(P\|Q\right) = \frac{1}{\alpha-1}\log\left[\mathbb{E}_{x\sim P}\left(\left(\frac{Q(x)}{P(x)}\right)^{1-\alpha}\right)\right]. \tag{2.12}$$

Demostración. Realizando álgebra en la suma de la igualdad (2.9) se obtiene que

$$\begin{aligned} D_\alpha\left(P\|Q\right) &= \frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}}\left(\frac{Q(x)}{P(x)}\right)^{1-\alpha} P(x) \\ &= \frac{1}{\alpha-1}\log\left[\mathbb{E}_{x\sim P}\left(\left(\frac{Q(x)}{P(x)}\right)^{1-\alpha}\right)\right] \end{aligned} \tag{2.13}$$

lo cual completa la prueba. ■

La siguiente proposición relaciona la divergencia de Rényi y de Kullback-Leibler.

Proposición 2.3.1. *Para $\alpha = 1$ y P, Q distribuciones arbitrarias se cumple que*

$$D_1(P\|Q) = D_{KL}(P\|Q). \tag{2.14}$$

Demostración. Para $\alpha = 1$ consideremos las funciones $f, g : (1, \infty) \to \mathbb{R}$ dadas por

$$f(\alpha) = \log\left[\sum_{x\in\mathcal{X}} \left(\frac{P(x)}{Q(x)}\right)^{\alpha} Q(x)\right] \quad \text{y} \quad g(\alpha) = \alpha - 1$$

de las cuales usando (2.10) $D_1(P\|Q) = \lim\limits_{\alpha\to 1} \dfrac{f(\alpha)}{g(\alpha)} = \dfrac{0}{0}$.

Para evitar la indeterminación aplicamos la regla de L'Hopital. Por un lado

$$f'(\alpha) = \frac{\sum\limits_{x\in\mathcal{X}} \log\left(\dfrac{P(x)}{Q(x)}\right) P(x) \left(\dfrac{P(x)}{Q(x)}\right)^{\alpha-1}}{\sum\limits_{x\in\mathcal{X}} \left(\dfrac{P(x)}{Q(x)}\right)^{\alpha} Q(x)} \quad \text{y} \quad g'(\alpha) = 1$$

entonces

$$\lim_{\alpha\to 1} \frac{f(\alpha)}{g(\alpha)} = \lim_{\alpha\to 1} \frac{f'(\alpha)}{g'(\alpha)} = \sum_{x\in\mathcal{X}} \log\left(\frac{P(x)}{Q(x)}\right) P(x) = D_{KL}(P\|Q).$$

Obteniendo la igualdad pedida. ■

La propiedad de no negatividad también se aplica a la divergencia de Rényi.

Proposición 2.3.2. *Para $\alpha \geq 1$ y distribuciones P, Q arbitrarias*

$$D_\alpha(P\|Q) \geq 0. \tag{2.15}$$

Demostración. Supongamos que $\alpha > 1$. Definimos la función $\phi : (0, \infty) \to \mathbb{R}$ dada por

$\phi(x) = x^{1-\alpha}$. Entonces aplicando la ecuación (2.12) y la desigualdad de Jensen obtenemos

$$\begin{aligned} D_\alpha(P\|Q) &= \frac{1}{\alpha-1}\log\left\{\mathbb{E}_{x\sim P}\phi\left(\frac{Q(x)}{P(x)}\right)\right\} \\ &\geq \frac{1}{\alpha-1}\log\phi\left(\mathbb{E}_{x\sim P}\left(\frac{Q(x)}{P(x)}\right)\right). \end{aligned}$$

Donde la última desigualdad se sigue de la convexidad de ϕ y ya que $\mathbb{E}_{x\sim P}\left(\frac{Q(x)}{P(x)}\right) = 1$ se obtiene la desigualdad pedida.

Para $\alpha = 1$ tenemos de la proposición 2.3.1 que $D_1(P\|Q) = D_{KL}(P\|Q)$ y del inciso 1 de la proposición 2.1.1 obtenemos la desigualdad pedida. ■

La siguiente propiedad será de utilidad para las siguientes secciones.

Proposición 2.3.3 (Monotonía)**.** *Para $1 \leq \alpha < \beta$ y P, Q distribuciones arbitrarias se cumple que*

$$D_\alpha(P\|Q) \leq D_\beta(P\|Q). \tag{2.16}$$

Demostración. Asumamos que $\alpha > 1$. Consideremos la función $\varphi(x) = x^{\frac{\alpha-1}{\beta-1}}$ entonces manipulando el lado derecho de la igualdad (2.12) obtenemos

$$\begin{aligned} D_\alpha(P\|Q) &= \frac{1}{\alpha-1}\log\left[\mathbb{E}_{x\sim P}\left(\left(\frac{P(x)}{Q(x)}\right)^{(\beta-1)\cdot\frac{\alpha-1}{\beta-1}}\right)\right] \\ &= \frac{1}{\alpha-1}\log\left\{\mathbb{E}_{x\sim P}\left[\varphi\left(\left(\frac{P(x)}{Q(x)}\right)^{\beta-1}\right)\right]\right\}. \end{aligned}$$

Ya que la función φ es cóncava, entonces por desigualdad de Jensen se tiene que

$$\begin{aligned} D_\alpha(P\|Q) &\leq \frac{1}{\alpha-1}\log\left\{\varphi\left[\mathbb{E}_{x\sim P}\left(\left(\frac{P(x)}{Q(x)}\right)^{\beta-1}\right)\right]\right\} \\ &= \frac{1}{\alpha-1}\cdot\frac{\alpha-1}{\beta-1}\log\left[\mathbb{E}_{x\sim P}\left(\left(\frac{P(x)}{Q(x)}\right)^{\beta-1}\right)\right] \\ &= D_\beta(P\|Q) \end{aligned}$$

lo cual completa la prueba. ■

El siguiente resultado, como se verá en el capítulo 3, jugará un papel fundamental para el estudio del teorema de composición avanzada bajo la privacidad diferencial de Rényi.

Proposición 2.3.4 (Preservación de la probabilidad)**.** *Sea $\alpha > 1$, P y Q dos distribuciones definidas sobre $\mathcal{X}$ con idéntico soporte y A un evento arbitrario. Entonces*

$$P(A) \leq \left(\exp\left[D_\alpha\left(P\|Q\right)\right] \cdot Q(A)\right)^{\frac{\alpha-1}{\alpha}}. \tag{2.17}$$

Demostración. Tomemos $p = \alpha$, $q = \alpha/(\alpha - 1)$. Como el espacio $\mathcal{X}$ es contable entonces consideremos las funciones $f(x) = P(x)/\left(Q(x)\right)^{1/q}$ y $g(x) = \left(Q(x)\right)^{1/q}$ para las cuales se cumple que

$$P(A) = \sum_{x\in A} P(x) = \sum_{x\in A} f(x)g(x). \tag{2.18}$$

Aplicando la desigualdad de Hölder (ver apéndice) a la ultima suma de (2.18),

$$P(A) \leq \left(\sum_{x\in A} \left(P(x)\right)^\alpha \left(Q(x)\right)^{1-\alpha}\right)^{1/\alpha} \cdot \left(\sum_{x\in A} Q(x)\right)^{\frac{\alpha-1}{\alpha}}.$$

Simplificando se obtiene

$$\begin{aligned} P(A) &\leq \left(\sum_{x\in\mathcal{X}} \left(P(x)\right)^\alpha \left(Q(x)\right)^{1-\alpha}\right)^{1/\alpha} \cdot \left(Q(A)\right)^{\frac{\alpha-1}{\alpha}} \\ &= \left[\exp\left(D_\alpha\left(P\|Q\right)\right)\right]^{\frac{\alpha-1}{\alpha}} \left(Q(A)\right)^{\frac{\alpha-1}{\alpha}}. \end{aligned}$$

completando la prueba. ■

Hacemos la observación de que la divergencia de Rényi no es una métrica: Esta no es simétrica y no satisface la desigualdad triangular, pero una variante debilitada de la desigualdad triangular sí se cumple. Presentamos su versión general.

Proposición 2.3.5 (Desigualdad triangular débil.)**.** *Sean P, Q y R distribuciones definidas sobre $\mathcal{X}$. Entonces para $\alpha > 1$ y para cualesquiera $p, q > 1$ satisfaciendo $1/p + 1/q = 1$,*

se tiene

$$D_\alpha\left(P\|Q\right) \leq \frac{\alpha - 1/p}{\alpha - 1} D_{p\alpha}\left(P\|R\right) + D_{q(\alpha-1/p)}\left(R\|Q\right). \tag{2.19}$$

Demostración. Por un lado tenemos la igualdad

$$\exp\left[(\alpha-1)D_\alpha\left(P\|Q\right)\right] = \sum_{x\in\mathcal{X}} \frac{\left(P(x)\right)^\alpha}{\left(R(x)\right)^{\alpha-1/p}} \cdot \frac{\left(R(x)\right)^{\alpha-1/p}}{\left(Q(x)\right)^{\alpha-1}}. \tag{2.20}$$

Aplicando Hölder a la suma de (2.20)

$$\begin{aligned}\exp\left[(\alpha-1)D_\alpha\left(P\|Q\right)\right] &\leq \left\{\sum_{x\in\mathcal{X}} \frac{\left(P(x)\right)^{p\alpha}}{\left(R(x)\right)^{p\alpha-1}}\right\}^{1/p} \cdot \left\{\sum_{x\in\mathcal{X}} \frac{\left(R(x)\right)^{q\alpha-q/p}}{\left(Q(x)\right)^{q\alpha-q}}\right\}^{1/q} \\ &= \exp\left[(\alpha-1/p)\left(D_{p\alpha}\left(P\|R\right) + (\alpha-1)D_{q(\alpha-1/p)}\left(R\|Q\right)\right)\right].\end{aligned}$$

Tomando logaritmos y dividiendo ambos lados por $\alpha - 1$ obtenemos lo pedido. ■

Varios casos especiales importantes de la desigualdad triangular débil pueden ser obtenidos fijando los parámetros p y q como se muestra en el siguiente corolario.

Corolario 2.3.1. *Para P, Q y R con mismo soporte tenemos que:*

1) $D_\alpha\left(P\|Q\right) \leq \dfrac{\alpha - 1/2}{\alpha - 1} D_{2\alpha}\left(P\|R\right) + D_{2\alpha-1}\left(R\|Q\right)$.

2) $D_\alpha\left(P\|Q\right) \leq \dfrac{\alpha}{\alpha - 1} D_{\infty}\left(P\|R\right) + D_{\alpha}\left(R\|Q\right)$.

3) $D_\alpha\left(P\|Q\right) \leq D_\alpha\left(P\|R\right) + D_{\infty}\left(R\|Q\right)$.

Demostración. Todas las afirmaciones se siguen de la desigualdad triangular débil donde p y q se escogen, respectivamente, como

1) $p = q = 2$.

2) $p \to \infty$ y $q = p/(p-1) \to 1$.

3) $q \to \infty$ y $p = q/(q-1) \to 1$.

■

Para la siguiente propiedad de la divergencia de Rényi los siguientes resultados de análisis serán usados.

Lema 2.3.2. *Si* $x \in [0,1]$ *entonces*

$$e^x \leq 1 + x + x^2 \quad \text{y} \quad e^{-x} \leq 1 - x + x^2.$$

Demostración. Demostraremos la primera desigualdad ya que la segunda se demuestra de manera análoga. Sean las funciones $\varphi, \phi : [0,1] \to \mathbb{R}$ definidas como

$$\varphi(x) = x \quad \text{y} \quad \phi(x) = \log(1 + x + x^2)$$

las cuales cumplen que $\varphi(0) = \phi(0) = 0$. Observemos que $\varphi^{'}(x) \leq \phi^{'}(x)$ ya que $1+x+x^2 \leq 1 + 2x$, pues $x \in [0,1]$, por lo que $\varphi(x) \leq \phi(x)$ obteniendo la desigualdad pedida.

Para la segunda desigualdad basta tomar $\varphi(x) = -x$ y $\phi(x) = \log(1-x+x^2)$ y usamos el mismo argumento anteriormente dado. ■

Del lema 2.3.2 se desprende la siguiente

Proposición 2.3.6. *Sean* $x > y > 0$, $\lambda = \log(x/y)$, *y* $0 \leq \beta \leq 1/\lambda$ *entonces*

$$x^{\beta+1}y^{-\beta} + x^{-\beta}y^{\beta+1} \leq (1 + (\beta\lambda)^2)(x + y) + \beta\lambda(x - y). \tag{2.21}$$

Demostración. Realizando álgebra en lado derecho de (2.21) obtenemos

$$x^{\beta+1}y^{-\beta} + x^{-\beta}y^{\beta+1} = (x^{\beta}y^{-\beta})x + (x^{-\beta}y^{\beta})y. \tag{2.22}$$

Observemos que

$$x^{\beta}y^{-\beta} = e^{\log(x/y)^{\beta}} = e^{\beta\lambda} \quad \text{y} \quad x^{-\beta}y^{\beta} = e^{\log(y/x)^{\beta}} = e^{-\beta\lambda} \tag{2.23}$$

entonces sustituyendo las igualdades (2.23) en (2.22) obtenemos

$$\begin{aligned} x^{\beta+1}y^{-\beta} + x^{-\beta}y^{\beta+1} &= e^{\beta\cdot\lambda}x + e^{-\beta\cdot\lambda}y \\ &\leq (1+\beta\lambda+(\beta\lambda)^2)x + (1-\beta\lambda+(\beta\lambda)^2)y \\ &= (1+(\beta\lambda)^2)(x+y) + \beta\lambda(x-y) \end{aligned} \tag{2.24}$$

donde la desigualdad de (2.24) se da por lema 2.3.2 completando la prueba. ■

El siguiente lema nos será de ayuda para poder estudiar el teorema de composición avanzada bajo el concepto de la divergencia de Rényi.

Lema 2.3.3. *Si P y Q son distribuciones tales que $D_\infty(P\|Q) \leq \zeta$ y $D_\infty(Q\|P) \leq \zeta$, entonces para $\alpha \geq 1$*

$$D_\alpha(P\|Q) \leq 2\alpha\zeta^2. \tag{2.25}$$

Demostración. Comenzamos suponiendo que $\alpha > 1$. La prueba la realizamos por casos.

Caso 1 $\alpha \geq 1 + 1/\zeta$. Por la propiedad de monotonía $D_\alpha(P\|Q) \leq D_\infty(P\|Q)$ entonces

$$D_\alpha(P\|Q) \leq \frac{\alpha}{\alpha-1} D_\infty(P\|Q). \tag{2.26}$$

Ahora, por hipótesis $D_\infty(P\|Q) \leq \zeta$, entonces el lado derecho de (2.26) es menor o igual a $\frac{\alpha}{\alpha-1}\zeta$ y esta última expresión, por la desigualdad del caso 1 en el que estamos, es menor que $2\alpha\zeta^2$. Por lo tanto obtenemos (2.25).

Caso 2. $\alpha < 1 + 1/\zeta$. Por un lado tenemos la igualdad

$$\exp\left[(\alpha-1)D_\alpha(P\|Q)\right] = \sum_{x\in\mathcal{X}} (P(x))^\alpha (Q(x))^{1-\alpha} \tag{2.27}$$

y por propiedad de no negatividad para la divergencia de Rényi tenemos que

$$1 \leq \exp\left\{(\alpha-1)D_\alpha(Q\|P)\right\}, \tag{2.28}$$

entonces de (2.28) se obtiene la desigualdad

$$\exp\{(\alpha-1)D_\alpha(P\|Q)\} \leq \exp\{(\alpha-1)D_\alpha(Q\|P)\} + \exp\{(\alpha-1)D_\alpha(P\|Q)\} - 1. \quad (2.29)$$

Por lo que, de (2.29), se sigue que (2.27) cumple que

$$\exp[(\alpha-1)D_\alpha(P\|Q)] \leq \sum_{x\in\mathcal{X}} \left[(P(x))^\alpha (Q(x))^{1-\alpha} + (Q(x))^\alpha (P(x))^{1-\alpha}\right] - 1. \quad (2.30)$$

Aplicando a la suma de la desigualdad (2.30) la proposición 2.3.6, con $\beta = \alpha - 1$, $x = P(x)$ $y = Q(x)$ y $\lambda = \zeta$, obtenemos

$$\begin{aligned} &\sum_{x\in\mathcal{X}} \left[(P(x))^\alpha (Q(x))^{1-\alpha} + (Q(x))^\alpha (P(x))^{1-\alpha}\right] - 1 \\ &\leq \sum_{x\in\mathcal{X}} \left[(1+(\alpha-1)^2\zeta^2)(P(x)+Q(x)) + (\alpha-1)\zeta|P(x)-Q(x)|\right] - 1 \\ &= 1 + 2(\alpha-1)^2\zeta^2 + (\alpha-1)\zeta\|P(x)-Q(x)\|_1. \end{aligned}$$

Entonces hemos llegado a la desigualdad

$$\exp[(\alpha-1)D_\alpha(P\|Q)] \leq 1 + 2(\alpha-1)^2\zeta^2 + \zeta(\alpha-1)\|P(x)-Q(x)\|_1. \quad (2.31)$$

Tomando logaritmos en ambos lados de (2.31) y usando que $\log(1+x) < x$ para x positivo, encontramos que

$$D_\alpha(P\|Q) \leq 2(\alpha-1)\zeta^2 + \zeta\|P(x)-Q(x)\|_1.$$

Además, se puede verificar que $\|P-Q\|_1 \leq \min(2, e^\zeta - 1) \leq 2\zeta$, por lo que sustituyendo en la última desigualdad obtenemos lo pedido para $\alpha > 1$.

El caso $\alpha = 1$ se obtiene por continuidad. Específicamente, tenemos que

$$D_1(P\|Q) = \lim_{\alpha\to 1} D_\alpha(P\|Q) \leq \lim_{\alpha\to 1}(2\alpha\zeta^2) = 2\zeta^2. \quad (2.32)$$

Obteniendo la desigualdad pedida. ■

Finalizamos esta sección con un resultado que nos servirá para analizar el equivalente al post-procesamiento para la divergencia de Rényi. Para el concepto de kernel de una transformación aleatoria, el cual es usado en la siguiente proposición, se puede consultar el apéndice o bien referirse a [2].

Proposición 2.3.7. *Sean P y Q distribuciones definidas sobre un conjunto contable $\mathcal{X}$ y $\mathcal{K} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ un kernel. Entonces para $\alpha > 1$ se cumple que*

$$D_\alpha(\mathcal{K}P\|\mathcal{K}Q) \leq D_\alpha(P\|Q).$$

Demostración. Observemos primero que

$$\begin{aligned} \exp\left[(\alpha-1)D_\alpha(P\|Q)\right] &= 1 + \sum_{x\in\mathcal{X}} \left[\left(\frac{P(x)}{Q(x)}\right)^\alpha - 1\right] Q(x) \\ &= 1 + \sum_{x\in\mathcal{X}} f_\alpha\left(\frac{P(x)}{Q(x)}\right) Q(x) \\ &= 1 + D_{f_\alpha}(P\|Q). \end{aligned} \tag{2.33}$$

Ya que la función $f_\alpha : (0,\infty) \to \mathbb{R}$ dada por $f_\alpha(t) = t^\alpha - 1$ es una f-divergencia (ver apéndice). Entonces de (2.33) se sigue que

$$\begin{aligned} \exp\left[(\alpha-1)D_\alpha(\mathcal{K}P\|\mathcal{K}Q)\right] - 1 &= D_{f_\alpha}\left(\mathcal{K}P\|\mathcal{K}Q)\right) \\ &\leq D_{f_\alpha}(P\|Q) \end{aligned} \tag{2.34}$$

donde la desigualdad de (2.34) se sigue de la desigualdad para el procesamiento de datos. En consecuencia se obtiene la desigualdad

$$\exp\left[(\alpha-1)D_\alpha(\mathcal{K}P\|\mathcal{K}Q)\right] - 1 \leq \exp\left[(\alpha-1)D_\alpha(P\|Q)\right] - 1,$$

entonces eliminando el valor -1, la exponencial y dividiendo por $\alpha - 1$ obtenemos la desigualdad pedida. ■

2.4. Divergencia de Rényi de Distribuciones de Probabilidad Especiales

En esta sección se calcula la divergencia de Rényi para tres distribuciones de probabilidad para las cuales, como se verá en los capítulos 3 y 4, se analiza la privacidad diferencial bajo la divergencia de Rényi. Si bien estas expresiones pueden darse en forma general, en esta tesis, nos limitaremos sólo a ciertos casos especiales.

2.4.1. Distribución de Bernoulli

Recordemos la definición de una distribución de Bernoulli.

Definición 2.4.1. *Sea $0 < p < 1$. Una variable aleatoria discreta X se distribuye Bernoulli de parámetro p, denotado por $X \sim Ber(p)$, si su función de masa de probabilidad está dada por*

$$f_X(x) = p^x(1-p)^x, \quad x \in \{0, 1\}.$$

De la definición anterior tenemos que

Proposición 2.4.1. *Sea $\alpha \geq 1$. Si $P \sim Ber(p)$ y $Q \sim Ber(q)$ con $0 < p, q < 1$, entonces*

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \left[(1-p)^\alpha (1-q)^{1-\alpha} + p^\alpha q^{1-\alpha}\right]$$

para $\alpha > 1$, mientras que para $\alpha = 1$ se tiene

$$D_1(P\|Q) = \log\left(\frac{1-p}{1-q}\right)(1-p) + \log\left(\frac{p}{q}\right)p.$$

Demostración. Para $\alpha > 1$ partimos de la ecuación (2.9)

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \left[\left(\frac{P(0)}{Q(0)}\right)^\alpha Q(0) + \left(\frac{P(1)}{Q(1)}\right)^\alpha Q(1)\right]$$

sustituyendo los respectivos valores de $P(0) = 1 - p$, $Q(0) = 1 - q$ y $P(1) = p$, $Q(1) = q$ se obtiene las expresión pedida.

Para $\alpha = 1$ usamos la proposición 2.3.1

$$D_1(P\|Q) = \log\left(\frac{P(0)}{Q(0)}\right)P(0) + \log\left(\frac{P(1)}{Q(1)}\right)P(1)$$

y nuevamente sustituyendo los valores $P(0) = 1-p$, $Q(0) = 1-q$ y $P(1) = p$, $Q(1) = q$ obtenemos la igualdad. ■

2.4.2. Distribución Normal

Comenzamos recordando la definición de la distribución normal.

Definición 2.4.2. *Una variable aleatoria X absolutamente continua se distribuye normal con parámetros μ y σ, denotado por $X \sim N(\mu, \sigma^2)$, si su función de densidad esta dada por*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\mu - x)^2\right\}, \quad x \in (-\infty, \infty).$$

Los siguientes resultados nos servirán para estudiar la privacidad diferencial de Rényi para mecanismos aleatorizantes Gaussianos.

Proposición 2.4.2. *Para $\alpha > 1$ se tiene que*

$$D_\alpha\left(N(\mu_1, \sigma_1^2)\|N(\mu_2, \sigma_2^2)\right) = \frac{\alpha(\mu_1 - \mu_2)^2}{2\sigma_\alpha} + \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2(\alpha-1)}\log\left(\frac{\sigma_2^2}{\sigma_\alpha}\right),$$

con $\sigma_\alpha = (1-\alpha)\sigma_1^2 + \alpha\sigma_2^2 > 0$.

Demostración. Tomemos $P \sim N(\mu_1, \sigma_1^2)$ y $Q \sim N(\mu_2, \sigma_2^2)$ con densidades $p(x)$ y $q(x)$ respectivamente entonces

$$\begin{aligned} D_\alpha(P\|Q) &= \frac{1}{\alpha-1}\int_{-\infty}^{\infty}(p(x))^\alpha(q(x))^{1-\alpha}dx \\ &= \frac{1}{\alpha-1}\log\left[\frac{1}{(\sigma_1\sqrt{2\pi})^\alpha(\sigma_2\sqrt{2\pi})^{1-\alpha}}\right. \\ &\qquad \left.\times\int_{-\infty}^{\infty}\exp\left\{-\frac{\alpha}{2\sigma_1^2}(\mu_1 - x)^2 - \frac{(1-\alpha)}{2\sigma_2^2}(\mu_2 - x)^2\right\}dx\right]. \end{aligned} \tag{2.35}$$

Realizando álgebra en el argumento de la función exponencial de la ecuación (2.35) este toma la forma del polinomio $-(Ax^2 + Bx + C)$ donde

$$A = \left(\frac{\alpha}{2\sigma_1^2} + \frac{1-\alpha}{2\sigma_2^2}\right), \quad B = -\left(\frac{\alpha\mu_1}{\sigma_1^2} + \frac{(1-\alpha)\mu_2}{\sigma_2^2}\right) \quad \text{y} \quad C = \left(\frac{\alpha\mu_1^2}{2\sigma_1^2} + \frac{(1-\alpha)\mu_2^2}{2\sigma_2^2}\right).$$

Para hacer uso del teorema A.0.3 (ver apéndice) necesitamos que $A > 0$, pero

$$A = \frac{\sigma_\alpha}{4\sigma_1^2\sigma_2^2} \quad \text{con} \quad \sigma_\alpha = (1-\alpha)\sigma_1^2 + \alpha\sigma_2^2,$$

por lo que $A > 0$ si $\sigma_\alpha > 0$. Realizando la integral

$$\begin{aligned}\int_{-\infty}^{\infty} exp\left\{-\frac{\alpha}{2\sigma_1^2}(\mu_1 - x)^2 - \frac{(1-\alpha)}{2\sigma_2^2}(\mu_2 - x)^2\right\} dx &= \int_{-\infty}^{\infty} exp\left\{-(Ax^2 - Bx + C)\right\} dx \\ &= K \cdot exp\left(\frac{B^2 - 4AC}{4A}\right)\end{aligned}$$

donde

$$K = \sqrt{\frac{\pi}{A}} = \frac{\sigma_1\sigma_2\sqrt{2\pi}}{\sqrt{\sigma_\alpha}} \quad \text{y} \quad \frac{B^2 - 4AC}{4A} = \frac{\alpha(\alpha-1)(\mu_1 - \mu_2)^2}{2\sigma_\alpha}. \tag{2.36}$$

Por lo que sustituyendo las igualdades de (2.36) en (2.35) obtenemos

$$\begin{aligned}D_\alpha(P\|Q) &= \frac{1}{\alpha - 1}\log\frac{K}{(\sigma_1\sqrt{2\pi})^\alpha(\sigma_2\sqrt{2\pi})^{1-\alpha}} exp\left\{\frac{\alpha(\alpha-1)(\mu_1-\mu_2)^2}{2\sigma_\alpha}\right\} \\ &= \frac{1}{\alpha-1}\left[\frac{\alpha(\alpha-1)(\mu_1-\mu_2)^2}{2\sigma_\alpha} + \log\left(\frac{\sigma_2^\alpha\sigma_1^{1-\alpha}}{\sqrt{\sigma_\alpha}}\right)\right] \\ &= \frac{\alpha(\mu_1-\mu_2)^2}{2\sigma_\alpha} + \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2(\alpha-1)}\log\left(\frac{\sigma_2^2}{\sigma_\alpha}\right)\end{aligned}$$

obteniendo la igualdad pedida. ■

De la proposición (2.4.2) se obtienen los siguientes corolarios.

Corolario 2.4.1. *Para $\alpha > 1$ se cumple que*

$$D_\alpha\left(N(\mu_1, \sigma^2)\|N(\mu_2, \sigma^2)\right) = \frac{\alpha(\mu_1 - \mu_2)^2}{2\sigma^2}.$$

Demostración. En la proposición 2.4.2 se toma $\sigma_1 = \sigma_2 = \sigma$ y se obtiene el resultado. ■

Corolario 2.4.2. *Para $\alpha > 1$ se cumple que*

$$D_\alpha\left(N(\mu_1, \sigma^2)\|N(\mu_2, \sigma^2)\right) = D_\alpha\left(N(0, \sigma^2)\|N(\mu_2 - \mu_1, \sigma^2)\right).$$

Demostración. Aplicando el corolario 2.4.1 obtenemos

$$\begin{aligned} D_\alpha\left(N(0, \sigma^2)\|N(\mu_2 - \mu_1, \sigma^2)\right) &= \frac{\alpha(0 - (\mu_1 - \mu_2))^2}{2\sigma_\alpha} \\ &= \frac{\alpha(\mu_1 - \mu_2)^2}{2\sigma_\alpha} \\ &= D_\alpha\left(N(\mu_1, \sigma^2)\|N(\mu_2, \sigma^2)\right) \end{aligned}$$

obteniendo la igualdad que se pide. ■

2.4.3. Distribución de Laplace

Comenzamos con la definición de la distribución de Laplace.

Definición 2.4.3. *Una variable aleatoria X absolutamente continua se distribuye Laplace con parámetros μ y $\lambda > 0$, denotado por $X \sim \text{Lap}(\mu, \lambda)$, si su función de densidad esta dada por*

$$f_X(x) = \frac{1}{2\lambda} \exp\left\{-\frac{|x - \mu|}{\lambda}\right\}$$

con $x \in (-\infty, \infty)$.

Al igual que para la distribución normal los siguientes resultados, que se enuncian y demuestran, nos ayudarán para analizar la privacidad diferencial de Rényi para esta distribución.

Proposición 2.4.3. *Sea $\alpha > 1$. Si $P \sim \text{Lap}(\mu_1, \lambda)$ y $Q \sim \text{Lap}(\mu_2, \lambda)$, entonces*

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log\left[\frac{\alpha}{2\alpha - 1} \exp\left\{\frac{(\alpha - 1)|\mu_2 - \mu_1|}{\lambda}\right\} + \frac{\alpha - 1}{2\alpha - 1} \exp\left\{-\frac{\alpha|\mu_2 - \mu_1|}{\lambda}\right\}\right].$$

Demostración. Tomemos $P \sim Lap(\mu_1, \lambda)$ y $Q \sim Lap(\mu_2, \lambda)$ con densidades $p(x)$ y $q(x)$ respectivamente. Entonces

$$\begin{aligned} D_\alpha(P\|Q) &= \frac{1}{\alpha-1}\log\int_{-\infty}^{\infty}(p(x))^\alpha(q(x))^{1-\alpha}dx \\ &= \frac{1}{\alpha-1}\log\frac{1}{2\lambda}\int_{-\infty}^{\infty}\exp\left\{\frac{-\alpha|x-\mu_1|-(1-\alpha)|x-\mu_2|}{\lambda}\right\}dx. \end{aligned} \tag{2.37}$$

Para realizar la última integral de (2.37) consideramos dos casos.

Caso 1: $\mu_1 < \mu_2$. En este caso hacemos $\bar{\alpha} = 1-\alpha$ y dividimos la recta en los intervalos $[-\infty, \mu_1]$, $[\mu_1, \mu_2]$ y $[\mu_2, \infty]$ para tener

$$\begin{aligned} \int_{-\infty}^{\infty}\exp\left\{\frac{-\alpha|x-\mu_1|-(1-\alpha)|x-\mu_2|}{\lambda}\right\}dx &= \int_{-\infty}^{\mu_1}\exp\left\{\frac{-\alpha(\mu_1-x)-\bar{\alpha}(\mu_2-x)}{\lambda}\right\}dx \\ &+ \int_{\mu_1}^{\mu_2}\exp\left\{\frac{-\alpha(x-\mu_1)-\bar{\alpha}(\mu_2-x)}{\lambda}\right\}dx \\ &+ \int_{\mu_2}^{\infty}\exp\left\{\frac{-\alpha(x-\mu_1)-\bar{\alpha}(x-\mu_2)}{\lambda}\right\}dx. \end{aligned} \tag{2.38}$$

Realizando cada integral de (2.38) y simplificando se obtiene por resultado

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\left(\frac{\alpha}{2\alpha-1}\exp\left\{\frac{(\alpha-1)(\mu_2-\mu_1)}{\lambda}\right\} + \frac{\alpha-1}{2\alpha-1}\exp\left\{\frac{-\alpha(\mu_2-\mu_1)}{\lambda}\right\}\right). \tag{2.39}$$

Caso 2: $\mu_2 \leq \mu_1$. Al igual que en caso 1 hacemos $\bar{\alpha} = 1-\alpha$ y dividimos la recta en los intervalos $[-\infty, \mu_2]$, $[\mu_2, \mu_1]$ y $[\mu_1, \infty]$ obteniendo

$$\begin{aligned} \int_{-\infty}^{\infty}\exp\left\{\frac{-\alpha|x-\mu_1|-(1-\alpha)|x-\mu_2|}{\lambda}\right\}dx &= \int_{-\infty}^{\mu_2}\exp\left\{\frac{-\alpha(\mu_1-x)-\bar{\alpha}(\mu_2-x)}{\lambda}\right\}dx \\ &+ \int_{\mu_2}^{\mu_1}\exp\left\{\frac{-\alpha(\mu_1-x)-\bar{\alpha}(x-\mu_2)}{\lambda}\right\}dx \\ &+ \int_{\mu_1}^{\infty}\exp\left\{\frac{-\alpha(x-\mu_1)-\bar{\alpha}(x-\mu_2)}{\lambda}\right\}dx. \end{aligned} \tag{2.40}$$

Realizando cada integral de (2.40) se obtiene

$$D_\alpha\left(P\|Q\right) = \frac{1}{\alpha-1}\log\left(\frac{\alpha}{2\alpha-1}\exp\left\{\frac{(\alpha-1)(\mu_1-\mu_2)}{\lambda}\right\} + \frac{\alpha-1}{2\alpha-1}\exp\left\{\frac{-\alpha(\mu_1-\mu_2)}{\lambda}\right\}\right). \tag{2.41}$$

Por lo tanto de (2.41) y (2.39) se obtiene el resultado pedido. ■

En este capítulo se expusieron dos divergencias para distribuciones de probabilidad: la divergencia de Kullback-Leibler y la divergencia de Rényi, las cuales comparten propiedades tales como no negatividad y un tipo de "desigualdad triangular". También se obtuvo, como se demuestra en la proposición 1.3.1, que la divergencia de Kullback-Leibler es un caso límite de la divergencia de Rényi. Las propiedades que presenta la divergencia de Rényi, como se verá en los siguientes capítulos, nos ayudan a ver la privacidad diferencial de Rényi como una relajación de la privacidad diferencial aproximada que es muy adecuado para expresar garantías para la preservación de la privacidad bajo mecanismos heterogéneos.

Capítulo 3

Privacidad Diferencial Aproximada

En este capítulo introducimos las nociones y notaciones básicas de privacidad diferencial aproximada, que usaremos en el siguiente capítulo, para poder abordar la definición de privacidad diferencial de Rényi. Para esto, comenzamos definiendo el concepto de una base de datos cuyos elementos pertenecen a un conjunto contable $\mathcal{X}$, para después introducir los conceptos de distancia y adyacencia entre dos bases de datos. Posteriormente se continua con las definiciones clásicas de ϵ-privacidad diferencial y (ϵ, δ)-privacidad diferencial y sus resultados clásicos: post-procesamiento, privacidad grupal, composición básica y composición avanzada, los cuales serán contrastados con los resultados que se obtendrán con la privacidad diferencial de Rényi.

3.1. Mecanismos Aleatorizantes

Comenzamos con la definición de base de datos sobre un conjunto.

Definición 3.1.1. *Sea $\mathcal{X}$ un conjunto contable. Una base de datos D de $\mathcal{X}$ será entendida como una colección finita y ordenada de elementos de $\mathcal{X}$ de la forma*

$$D = (x_1, x_2, ..., x_n).$$

Será a veces conveniente representar a una base de datos $D = (x_1, ..., x_n)$ a través de

su histograma: $(d_x)_{x\in\mathcal{X}} \in \mathbb{N}^{|\mathcal{X}|}$ donde

$$d_x = \sum_{i=1}^{n} \mathbf{1}_{\{x_i = x\}}$$

el cual representa el número de elementos de $x \in \mathcal{X}$ que son del tipo $x_i \in D$.

Por ejemplo, consideremos el estudio de los grupos sanguíneos del ser humano. Hay cuatro tipos de sangre: A, B, AB y O. Asimismo, la sangre es Rh positivo o Rh negativo. Entonces el conjunto contable (también llamado conjunto objetivo) es $\mathcal{X} = \{A, B, AB, O\} \times \{+, -\}$, y sea la base de datos $D = \{x_1, x_2, ..., x_6, x_7, x_8\}$, la cual está representada en la tabla 3.1, mientras que el histograma asociado a la base D es $(d_x)_{x\in\mathcal{X}} = (2, 1, 1, 0, 1, 1, 2, 0)$. Por abuso de notación cuando mencionemos una base de datos los referiremos a su histograma, es decir, $D = (d_x)_{x\in\mathcal{X}}$. Bajo esta representación, tiene sentido hablar de la medida natural de la distancia entre dos bases de datos D y $D^{'}$ y para esto se necesita la siguiente definición.

Dato	Grupo sanguíneo
x_1	$(A, +)$
x_2	$(A, +)$
x_3	$(AB, -)$
x_4	$(B, +)$
x_5	$(A, -)$
x_6	$(AB, +)$
x_7	$(O, +)$
x_8	$(O, +)$

Cuadro 3.1: Base de datos D para del conjunto $\mathcal{X}$.

Definición 3.1.2. *La norma ℓ_1 de una base de datos $D = (d_x)_{x\in\mathcal{X}} \in \mathbb{N}^{|\mathcal{X}|}$ es denotada por $\|D\|_1$ y es definida como:*

$$\|D\|_1 := \sum_{x\in\mathcal{X}} d_x. \tag{3.1}$$

Además, se define la distancia entre dos bases de datos $D = (d_x)_{x\in\mathcal{X}}$ y $D^{'} = (d_x^{'})_{x\in\mathcal{X}}$ como

$$\|D - D^{'}\|_1 = \sum_{x\in\mathcal{X}} |d_x - d_x^{'}|.$$

Notemos que $\|D\|_1$ es el tamaño de la base de datos (el número de registros que contiene), mientras que $\|D - D^{'}\|_1$ es el número de registros diferentes entre D y $D^{'}$.

Definición 3.1.3. *Diremos que dos bases de datos D y $D^{'} \in \mathbb{N}^{|\mathcal{X}|}$ son adyacentes si $\|D - D^{'}\|_1 \leq 1$.*

Otro concepto que usaremos relacionado con bases de datos es el de sensitividad.

Definición 3.1.4. *Una función $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}$ se dice ser de sensitividad 1 si para cualesquiera bases de datos D y $D^{'}$ adyacentes:*

$$|f(D) - f(D^{'})| \leq 1,$$

mientras que la ℓ_2-sensitividad de f es:

$$\Delta_2 f := \text{máx}_{D,D^{'} \in \mathbb{N}^{|\mathcal{X}|}} \|f(D) - f(D^{'})\|_2$$

donde $\|D - D^{'}\|_1 = 1$.

Intuitivamente la privacidad diferencial aproximada proporcionará privacidad a la entrada de una base de datos, en particular, introducirá aleatoriedad la cual modelaremos a través de un mecanismo aleatorizante.

Definición 3.1.5. *Un mecanismo aleatorizante $\mathcal{M}$ es asociado con un mapeo con dominio $\mathbb{N}^{|\mathcal{X}|}$ y rango* $\text{Rang}(\mathcal{M})$.

Para simplificar la notación denotaremos por $\mathcal{R}$ al rango de un mecanismo aleatorizante. Además, hacemos la observación de que los objetos que pueden pertenecer a $\mathcal{R}$ son, para los propositos de este trabajo, ya sean valores numéricos o bien distribuciones de probabilidad. De esta manera, a partir de aquí en adelante, cuando escribamos $\mathcal{M}(D)$ con $D \in \mathbb{N}^{|\mathcal{X}|}$ nos referiremos a una distribución, con $\mathcal{M}(D) = x$ estaremos diciendo el valor que toma la distribución con x un número real y $S \subset \mathcal{R}$ denotará un subconjunto de números reales.

Los tres ejemplos importantes de mecanismos aleatorizantes con los cuales trabajaremos son: mecanismo de respuesta aleatoria, mecanismo de Laplace y el mecanismo de Gauss los cuales son definidos a continuación.

Definición 3.1.6. *El mecanismo aleatorizante* $\mathcal{M}_{RRM}^{f,\epsilon} : \mathbb{N}^{|\mathcal{X}|} \to \{0,1\}$ *para* $f : \mathbb{N}^{|\mathcal{X}|} \to \{0,1\}$ *es definido como*

$$\mathcal{M}_{RRM}^{f,\epsilon}(D) := \begin{cases} f(D) & \text{con probabilidad } e^{\epsilon}/(e^{\epsilon}+1) \\ 1-f(D) & \text{con probabilidad } 1/(e^{\epsilon}+1) \end{cases} \tag{3.2}$$

es llamado mecanismo de respuesta aleatoria(random response mechanism (RRM)).

Definición 3.1.7. *Se define el mecanismo de Laplace para* f *de sensitividad 1 como*

$$\mathcal{M}_{\text{Lap}}^{f,\epsilon}(D) = f(D) + Y \tag{3.3}$$

donde $Y \sim \text{Lap}(0, \Delta f/\epsilon)$.

Definición 3.1.8. *Se define el mecanismo Gaussiano para* f *de sensitividad* 2 *como*

$$\mathcal{M}_{\text{N}}^{f,\sigma}(D) = f(D) + Y \tag{3.4}$$

donde $Y \sim \text{N}(0, \sigma^2)$.

En base a estos conceptos, a continuación introducimos la noción de privacidad diferencial aproximada.

3.2. (ϵ, δ)-Privacidad Diferencial

Primero introducimos la definición estándar de ϵ-privacidad diferencial.

Definición 3.2.1. *Un mecanismo aleatorizante* $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}$ *satisface* ϵ*-privacidad diferencial, o más brevemente* ϵ*-DP, si para cualesquiera bases de datos adyacentes* D, $D' \in \mathbb{N}^{|\mathcal{X}|}$ *y* $S \subset \mathcal{R}$ *se tiene que*

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(D') \in S]. \tag{3.5}$$

La definición 3.2.1 concuerda con la noción de entradas adyacentes D y D', en la cual el dominio es especificado, y es típicamente elegido para capturar la contribución a

la entrada del mecanismo por un solo individuo. En otras palabras, si un mecanismo es diferencialmente privado, entonces es difícil decidir en base a la salida del mecanismo si un individuo perteneció a la base de datos utilizada.

Una relajación de ϵ-privacidad diferencial permite un término aditivo δ en la desigualdad (3.5) por lo que la interpretación común de (ϵ, δ)-privacidad diferencial es que esta es ϵ-privacidad diferencial "excepto con probabilidad δ".

Definición 3.2.2. *Un mecanismo aleatorizante $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}$ ofrece (ϵ, δ)-privacidad diferencial, o más brevemente (ϵ, δ)-DP, si para cualesquiera bases de datos D y $D' \in \mathbb{N}^{|X|}$ adyacentes y $S \subset \mathcal{R}$ se cumple que*

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(D') \in S] + \delta. \tag{3.6}$$

La definición de (ϵ, δ)- privacidad diferencial fue inicialmente propuesta para mecanismos aleatorizantes Gaussianos ya que análisis elementales muestran que este mecanismo no puede cumplir ϵ-DP para cualquier ϵ [1]. Otra razón para usar (ϵ, δ)-privacidad diferencial es la aplicación de los teoremas de composición avanzada que veremos en la capítulo 4.

3.3. Propiedades Básicas

Sabemos que la privacidad diferencial es inmune al post-procesamiento: un analista de datos, sin conocimiento adicional sobre la base de datos privada, no puede calcular una función de salida de un algoritmo aleatorizante y hacerlo menos diferencialmente privado. Formalmente, la composición de un mapeo con un mecanismo aleatorizante (ϵ, δ)-diferencialmente privado también es (ϵ, δ)-diferencialmente privado.

Proposición 3.3.1 (Post-procesamiento.)**.** *Sea $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}$ un mecanismo aleatorizante (ϵ, δ)-diferencialmente privado. Sea $f : \mathcal{R} \to \mathcal{R}'$ un mapeo arbitrario. Entonces $f \circ \mathcal{M} : \mathbb{N}^{|X|} \to \mathcal{R}'$ es (ϵ, δ)-diferencialmente privado.*

Demostración. Sean D y $D^{'}$ bases de datos adyacentes y $S \subset \mathcal{R}^{'}$ arbitrarios, entonces

$$\begin{aligned}\Pr\left[(f \circ \mathcal{M})(D) \in S\right] &= \Pr\left[\mathcal{M}(D) \in f^{-1}(S)\right] \\ &\leq e^{\epsilon} \cdot \Pr\left[\mathcal{M}(D^{'}) \in f^{-1}(S)\right] + \delta \\ &= e^{\epsilon} \cdot \Pr\left[(f \circ \mathcal{M})(D^{'}) \in S\right] + \delta.\end{aligned}$$

que es lo que queríamos demostrar. ■

Otra propiedad útil es que la privacidad diferencial proporciona protecciones para pequeños grupos de personas.

Proposición 3.3.2 (Privacidad grupal.)**.** *Si $\mathcal{M}$ es un mecanismo (ϵ, δ)-diferencialmente privado, entonces para cualesquiera parejas de bases de datos D y $D^{'} \in \mathbb{N}^{|\mathcal{X}|}$, $\mathcal{M}(D)$ y $\mathcal{M}(D^{'})$ son $(k\epsilon, ke^{k}\delta)$- indistinguibles para $\|D - D^{'}\|_1 \leq k$.*

Demostración. Tomemos D y $D^{'}$ bases de datos tales que $\|D - D^{'}\|_1 \leq k$. Sea la sucesión $D = D_0, D_1, ..., D_k = D^{'}$ formada de tal manera que D_{i-1} y D_i son adyacentes(es decir, D_i proviene añadiendo un elemento a D_{i-1}). Como $\mathcal{M}$ es (ϵ, δ)- diferencialmente privado entonces para $S \subset \mathcal{R}$ se tiene que

$$\begin{aligned}\Pr\left[\mathcal{M}(D_0) \in S\right] &\leq e^{\epsilon} \cdot \Pr\left[\mathcal{M}(D_1) \in S\right] + \delta \\ &\leq e^{\epsilon}\left(e^{\epsilon} \cdot \Pr\left[\mathcal{M}(D_2) \in S\right] + \delta\right) + \delta.\end{aligned}$$

Continuando con las sustituciones obtenemos

$$\begin{aligned}\Pr\left[\mathcal{M}(D_0) \in S\right] &\leq e^{k\epsilon} \Pr\left[\mathcal{M}(D_k) \in S\right] + (1 + e^{\epsilon} + e^{2\epsilon} + ... + e^{(k-1)\epsilon}) \cdot \delta \\ &\leq e^{k\epsilon} \Pr\left[\mathcal{M}(D_k) \in S\right] + ke^{k\epsilon} \cdot \delta,\end{aligned}$$

con lo cual se completa la prueba. ■

Por otro lado el lema de composición básica dice que la privacidad disminuye a lo más linealmente con el número de mecanismos ejecutados.

Lema 3.3.1. *Si $\mathcal{M}_1$ y $\mathcal{M}_2$ son mecanismos independientes tales que $\mathcal{M}_1$ es (ϵ_1, δ_1)-diferencialmente privado y $\mathcal{M}_2$ es (ϵ_2, δ_2)-diferencialmente privado, entonces el mecanismo $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$ es $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-diferencialmente privado.*

Demostración. Denotemos por $\mathcal{R}_1 = \text{Rang}(\mathcal{M}_1)$ y $\mathcal{R}_2 = \text{Rang}(\mathcal{M}_2)$. Sean D y $D^{'} \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes y $S \subset \mathcal{R}_1 \times \mathcal{R}_2$, entonces

$$\begin{aligned} \Pr\left[\mathcal{M}(D) \in S\right] &= \Pr\left[(\mathcal{M}_1(D), \mathcal{M}_2(D)) \in S\right] \\ &= \sum_{x \in \mathcal{R}_1} \sum_{y \in \mathcal{R}_2} \chi_S(x, y) \Pr\left[\mathcal{M}_1(D) = x\right] \Pr\left[\mathcal{M}_2(D) = y\right] \\ &= \sum_{x \in \mathcal{R}_1} \Pr\left[\mathcal{M}_1(D) = x\right] \sum_{y \in \mathcal{R}_2} \chi_S(x, y) \Pr\left[\mathcal{M}_2(D) = y\right]. \end{aligned} \tag{3.7}$$

Definamos para cada $y \in \mathcal{R}_2$ el conjunto $A_y = \{x \in \mathcal{R}_1 \ : \ (x, y) \in S\}$ entonces se tiene la igualdad

$$\sum_{y \in \mathcal{R}_2} \chi_S(x, y) \Pr\left[\mathcal{M}_2(D) = y\right] = \Pr\left[\mathcal{M}_2(D) \in A_y\right], \tag{3.8}$$

por lo que sustituyendo (3.8) en (3.7) obtenemos

$$\Pr\left[\mathcal{M}(D) \in S\right] = \sum_{x \in \mathcal{R}_1} \left(\Pr\left[\mathcal{M}_1(D) = x\right] \cdot \Pr\left[\mathcal{M}_2(D) \in A_y\right]\right). \tag{3.9}$$

Ya que $\mathcal{M}_2$ es (ϵ_2, δ_2)-diferencialmente privado entonces

$$\begin{aligned} \Pr\left[\mathcal{M}_2(D) \in A_y\right] &\leq e^{\epsilon_2} \Pr\left[\mathcal{M}_2(D^{'}) \in A_y +\right] + \delta_2 \\ &\leq \left(e^{\epsilon_2} \Pr\left[\mathcal{M}_2(D^{'}) \in A_y\right] + \delta_2\right) \wedge 1 \\ &\leq \left(e^{\epsilon_2} \Pr\left[\mathcal{M}_2(D^{'}) \in A_y\right] \wedge 1\right) + \delta_2 \end{aligned} \tag{3.10}$$

y la desigualdad de (3.9) junto con (3.10) llega a cumplir

$$\Pr\left[\mathcal{M}(D) \in S\right] \leq \sum_{x \in \mathcal{R}_1} \Pr\left[\mathcal{M}_1(D) = x\right] \left[e^{\epsilon_2} \Pr\left(\mathcal{M}_2(D^{'}) \in A_y\right) \wedge 1\right] + \delta_2. \tag{3.11}$$

Ahora, sumando y restando al lado derecho de la desigualdad en (3.11) la expresión

$$\sum_{x\in\mathcal{R}_1} e^{\epsilon_1} \Pr\left[\mathcal{M}_1(D^{'}) = x\right] \left(e^{\epsilon_2} \Pr\left[\mathcal{M}_2(D^{'}) \in A_y\right] \wedge 1\right)$$

y agrupando obtenemos

$$\Pr\left[\mathcal{M}(D) \in S\right] \leq e^{\epsilon_1+\epsilon_2} \sum_{x\in\mathcal{R}_1} \Pr\left[\mathcal{M}_1(D^{'}) = x\right] \left(\Pr\left[\mathcal{M}_2(D^{'}) \in A_y\right] \wedge 1\right) + A + \delta_2 \quad (3.12)$$

donde

$$A = \sum_{x\in\mathcal{R}_1} \left(\Pr\left[\mathcal{M}_1(D) = x\right] - e^{\epsilon_1} \Pr\left[\mathcal{M}_1(D^{'}) = x\right]\right) \left(e^{\epsilon_2} Pr\left(\mathcal{M}_2(D^{'}) \in A_y\right) \wedge 1\right). \quad (3.13)$$

Observemos que la suma del lado derecho de la desigualdad (3.12) es igual al valor $\Pr\left[\mathcal{M}(D^{'}) \in S\right]$ y como $\mathcal{M}_1$ es (ϵ_1, δ_1)-diferencialmente privado entonces

$$\Pr\left[\mathcal{M}_1(D) = x\right] - e^{\epsilon_1} \Pr\left[\mathcal{M}_1(D^{'}) = x\right] \leq \delta_1$$

por lo que $A \leq \delta_1$. Por lo tanto $\Pr\left[\mathcal{M}(D) \in S\right] \leq e^{\epsilon_1+\epsilon_2} \Pr\left[\mathcal{M}(D^{'}) \in S\right] + \delta_1 + \delta_2$. ■

Del lema 3.3.1 se sigue el siguiente resultado.

Teorema 3.3.1. *Sean $\mathcal{M}_i : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_i$ mecanismos independientes tales que cada uno es (ϵ_i, δ_i)-diferencialmente privado. Entonces el mecanismo $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \prod_{i=1}^{k} \mathcal{R}_i$ definido como $\mathcal{M} = (\mathcal{M}_1, ..., \mathcal{M}_k)$ es $\left(\sum_{i=1}^{k} \epsilon_i, \sum_{i=1}^{k} \delta_i\right)$-diferencialmente privado.*

Demostración. Por inducción sobre k.

i) Para $k = 2$ se cumple por lema 3.3.1.

ii) Supongamos para k y consideremos $\mathcal{M}_1, ..., \mathcal{M}_{k+1}$ mecanismos independientes tal que $\mathcal{M}_i$ es (ϵ_i, δ_i)-diferencialmente privado para $1 \leq i \leq k+1$. Sea $\mathcal{M} = (\mathcal{M}^{'}, \mathcal{M}_{k+1})$ con $\mathcal{M}^{'} = (\mathcal{M}_1, ..., \mathcal{M}_k)$ donde $\mathcal{M}^{'}$ y $\mathcal{M}_{k+1}$ son independientes.

iii) Por hipótesis inductiva $\mathcal{M}^{'}$ es $\left(\sum_{i=1}^{k} \epsilon_i, \sum_{i=}^{k} \delta_i\right)$-diferencialmente privado y como $\mathcal{M}_{k+1}$ es $(\epsilon_{k+1}, \delta_{k+1})$-diferencialmente privado entonces nuevamente por lema (3.3.1) ob-

tenemos que $\left(\mathcal{M}', \mathcal{M}_{k+1}\right)$ es $\left(\sum_{i=1}^{k} \epsilon_i + \epsilon_{k+1}, \sum_{i=}^{k} \delta_i + \delta_{k+1}\right)$-diferencialmente privado, por lo que $\mathcal{M}$ es $\left(\sum_{i=1}^{k+1} \epsilon_i, \sum_{i=1}^{k+1} \delta_i\right)$-diferencialmente privado.

Por lo tanto por el principio de inducción matemática se obtiene lo pedido. ■

El siguiente resultado es consecuencia inmediata del lema 3.3.1.

Proposición 3.3.3. *Si $\mathcal{M}_1$ y $\mathcal{M}_2$ son mecanismos independientes (ϵ, δ)-diferencialmente privados, entonces $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$ es $(2\epsilon, 2\delta)$-diferencialmente privado.*

Demostración. Del lema 3.3.1 basta hacer $\epsilon_1 = \epsilon_2$ y $\delta_1 = \delta_1$. ■

Finalizamos esta sección enunciando el teorema de composición avanzada, cuya demostración se omite pues escapa a los propósitos de esta tesis, sin embargo para ver una demostración puede consultarse en [1].

Teorema 3.3.2 (composición avanzada.)**.** *Sean $\epsilon \geq 0$ y $0 < \delta, \delta^{'}$ y $\mathcal{F}$ la familia de mecanismos (ϵ, δ)-diferencialmente privados. En el contexto de composición tenemos que $\mathcal{M} = (\mathcal{M}_1, ..., \mathcal{M}_k)$ satisface $(\epsilon^{'}, k\delta + \delta^{'})$-privacidad diferencial con:*

$$\epsilon^{'} = \sqrt{2k \ln\left(1/\delta'\right) \epsilon} + k\epsilon(e^{\epsilon} - 1).$$

Como se observa en este capítulo la privacidad diferencial aproximada, tanto ϵ-DP como (ϵ, δ)-DP, limita un cambio en la distribución de salida de un algoritmo aleatorizante que puede ser inducido por una variación pequeña de su entrada, y a diferencia de la definición clásica de privacidad diferencial, (ϵ, δ)-privacidad diferencial ofrece una pérdida acumulada asintóticamente más pequeña bajo composición y permite una mayor flexibilidad en la selección de mecanismos de preservación de la privacidad.

Capítulo 4

Privacidad Diferencial de Rényi

En este capítulo describimos una generalización de la noción de privacidad diferencial aproximada basado en el concepto de la divergencia de Rényi definida en el capítulo 1. En la sección 3.1 introducimos la definición de privacidad diferencial de Rényi y estudiamos las propiedades que hereda de la privacidad diferencial aproximada: composición secuencial adaptativa y privacidad grupal. En la sección 3.2 se estudia el teorema de composición avanzada para finalizar en la sección 3.3 donde se estudia una relación entre privacidad diferencial aproximada y privacidad diferencial de Rényi.

4.1. (α, ζ)-Privacidad Diferencial de Rényi

Definición 4.1.1. *Un mecanismo aleatorizante $\mathcal{M}$ con dominio $\mathbb{N}^{|\mathcal{X}|}$ se dice ser ζ-Rényi diferencialmente privado de orden α, o (α, ζ)-RDP, si para cualesquiera base de datos adyacentes D y D' se tiene que*

$$D_\alpha \left(\mathcal{M}(D) || \mathcal{M}(D') \right) \leq \zeta. \tag{4.1}$$

La relación entre la divergencia de Rényi cuando $\alpha = \infty$ y la privacidad diferencial aproximada es inmediata. Un mecanismo aleatorizante $\mathcal{M}$ es ϵ- diferencialmente privado si

y solo si su distribución sobre cualesquiera dos bases de datos D y D' adyacentes satisface

$$D_\infty(\mathcal{M}(D)||\mathcal{M}(D')) \leq \epsilon. \tag{4.2}$$

A continuación analizamos las propiedades de la privacidad diferencial aproximada del capítulo 2 que puede heredar la definición 4.1.1.

La primera propiedad a analizar es el *post-procesamiento*.

Proposición 4.1.1. *Sea $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}$ un mecanismo aleatorizante (α, ζ)-RDP. Si $g : \mathcal{R} \to \mathcal{R}'$ es un mapeo arbitrario entonces $g \circ \mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}'$ es (α, ζ)-RDP.*

Demostración. Sean D y $D' \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes y sea el kernel $\mathcal{K} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ dado por $\mathcal{K}(y|x) = \delta_{g(x)}(y)$. Por 2.3.7 se obtiene que

$$\begin{aligned} D_\alpha\left((g \circ \mathcal{M})(D) \| (g \circ \mathcal{M})(D')\right) &= D_\alpha(\mathcal{K}\mathcal{M}(D) \| \mathcal{K}\mathcal{M}(D')) \\ &\leq D_\alpha(\mathcal{M}(D) \| \mathcal{M}(D')). \end{aligned} \tag{4.3}$$

Donde la última desigualdad de (4.3) es menor o igual que ζ por hipótesis. Esto completa la prueba. ■

La segunda propiedad a analizar es la llamada *Preservación bajo composición secuencial adaptativa*. Recordemos que si $\mathcal{M}_1(\cdot)$ es ϵ_1-diferencialmente privado y $\mathcal{M}_2(\cdot)$ es ϵ_2- diferencialmente privado, entonces la realización simultanea de $\mathcal{M}_1(D)$ y $\mathcal{M}_2(D)$ es $\epsilon_1 + \epsilon_2$- diferencialmente privado. La garantía se extiende incluso cuando $\mathcal{M}_2$ se elige adaptativamente en función de una salida $\mathcal{M}_1$, si $\mathcal{M}_2$ es indexado por elementos de $\mathbb{N}^{|\mathcal{X}|}$ y $\mathcal{M}_2(D, \cdot)$ es ϵ_2-diferencialmente privado para cualquier $D \in \mathbb{N}^{|\mathcal{X}|}$, entonces (P_X, P_Y), donde $P_X = \mathcal{M}_1(D)$ y $P_Y = \mathcal{M}_2(D, \cdot)$, es $\epsilon_1 + \epsilon_2$-diferencialmente privado.

Probamos una afirmación similar para la composición de dos mecanismos RDP.

Proposición 4.1.2. *Sea $\mathcal{M}_1 : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_1$ un mecanismo aleatorizante (α, ζ_1)-RDP y $\mathcal{M}_2 : \mathcal{R}_1 \times \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_2$ un mecanismo aleatorizante (α, ζ_2)-RDP, entonces el mecanismo aleatorizante $\mathcal{M}_3 = (\mathcal{M}_1, \mathcal{M}_2)$ es $(\alpha, \zeta_1 + \zeta_2)$-RDP.*

Demostración. Tomemos D y $D^{'} \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes. Si llamamos

$$P_X = \mathcal{M}_1(D), \quad P_Y = \mathcal{M}_2(D) \quad \text{y} \quad P_Z = (\mathcal{M}_1(D), \mathcal{M}_2(D)),$$

mientras que

$$P_{X'} = \mathcal{M}_1(D^{'}), \quad P_{Y'} = \mathcal{M}_2(D^{'}) \quad \text{y} \quad P_{Z'} = (\mathcal{M}_1(D^{'}), \mathcal{M}_2(D^{'}))$$

entonces

$$\begin{aligned}
\exp[(\alpha-1)D_\alpha(P_Z\|P_{Z'})] &= \sum_{\mathcal{R}_1\times\mathcal{R}_2} \left(\frac{P_Z(x,y)}{P_{Z'}(x,y)}\right)^\alpha P_{Z'}(x,y) \\
&= \sum_{\mathcal{R}_1\times\mathcal{R}_2} (P_Z(x,y))^\alpha (P_{Z'}(x,y))^{1-\alpha} \\
&= \sum_{\mathcal{R}_1\times\mathcal{R}_2} (P_X(x))^\alpha \left(P_Y(x,y)\right)^\alpha \left(P_{X'}(x)\right)^{1-\alpha} \left(P_{Y'}(x,y)^{1-\alpha}\right). \quad (4.4)
\end{aligned}$$

Separando la última suma de (4.4) se tiene

$$\begin{aligned}
\exp[(\alpha-1)D_\alpha(P_Z\|P_{Z'})] &\leq \sum_{\mathcal{R}_1} (P_X(x))^\alpha \left(P_{X'}(x)\right)^{1-\alpha} \left\{\sum_{\mathcal{R}_2} (P_Y(x,y)^\alpha \left(P_{Y'}(x,y)\right)^{1-\alpha}\right\} \\
&= \exp\left[(\alpha-1)D_\alpha D_\alpha\left(P_X\|P_{X'}\right)\right] \cdot \exp\left[(\alpha-1)D_\alpha\left(P_Y\|P_{Y'}\right)\right] \\
&\leq \exp((\alpha-1)\zeta_1)\exp((\alpha-1)\zeta_2)
\end{aligned}$$

por lo que la última desigualdad nos da lo pedido. ■

La siguiente propiedad es la *privacidad de grupo*. Aunque la definición de privacidad diferencial aproximada restringe las salidas de un mecanismo sobre parejas adyacentes de entradas, su garantía se extiende, en forma progresivamente más débil, a entradas que están más separadas. Esta propiedad tiene dos importantes consecuencias. En primer lugar, las garantías de privacidad diferencial disminuyen enormemente si nuestras suposiciones acerca de la influencia de una persona en la entrada son incorrectas.

En segundo lugar, la propiedad de privacidad de grupo permite la entrada de pre-procesamiento en un mecanismo diferencialmente privado, posiblemente amplificando (de

manera controlada) el impacto de registro en la salida del cálculo. De esta manera definamos el concepto un mecanismo c-estable.

Definición 4.1.2. *Decimos que $g : \mathcal{D} \to \mathcal{D}'$ es c-estable si $g(A)$ y $g(B)$ son adyacentes en $\mathcal{D}'$ entonces existe una sucesión de longitud $c+1$ D_0, D_1,...,D_c tal que*

i) $D_0 = A$ y $D_c = B$.

ii) D_i y D_{i+1} son adyacentes en $\mathcal{D}$.

En base a la definición anterior, la privacidad diferencial de Rényi tiene la siguiente propiedad.

Proposición 4.1.3. *Si $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}$ es (α, ζ)-RDP, $g : \mathcal{D} \to \mathcal{D}'$ es 2^c-estable y $\alpha \geq 2^{c+1}$, entonces $\mathcal{M} \circ g$ es $(\alpha/2^c, 3^c\zeta)$- RDP.*

Demostración. La demostración se realiza por inducción sobre c. Definamos $h = \mathcal{M} \circ g$ y tomando $D, D' \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes tenemos que

i) $c = 1$. Por un lado $g(D)$ y $g(D')$ son adyacentes en $\mathcal{D}'$ y g es 2-estable entonces existe $A \in \mathcal{D}$, tal que $g(D)$ y A, A y $g(D')$ son adyacentes en $\mathcal{D}$.

Por inciso 1 de corolario 2.3.1 tenemos que

$$\begin{aligned} D_{\alpha/2}(h(D)\|h(D')) &\leq \frac{\alpha/2 - 1/2}{\alpha/2 - 1} D_\alpha(h(D)\|h(A)) + D_{\alpha-1}(h(A)\|h(D')) \\ &= \frac{\alpha - 1}{\alpha - 2} D_\alpha(h(D)\|h(A)) + D_{\alpha-1}(h(A)\|h(D')). \end{aligned} \tag{4.5}$$

Como $\alpha - 1 < \alpha$, entonces por proposición 2.3.3 se sigue que (4.5) cumple

$$\begin{aligned} D_{\alpha/2}(h(D)\|h(D')) &\leq \frac{\alpha - 1}{\alpha - 2} D_\alpha(h(D)\|h(A)) + D_\alpha(h(A)\|h(D')) \\ &\leq \frac{\alpha - 1}{\alpha - 2}\zeta + \zeta. \end{aligned} \tag{4.6}$$

Como por hipótesis $\alpha \geq 2^2$, entonces la suma del lado derecho de (4.6) es menor o igual a 3, y en consecuencia se sigue que $D_{\alpha/2}(h(D)\|h(D')) \leq 3\zeta$.

ii) Supongamos que se cumple para c y verifiquemos para $c+1$.

Por un lado aplicando 1 de corolario 2.3.1

$$D_{\alpha/2^{c+1}}(h(D)\|h(D^{'})) \leq \frac{\alpha/2^{c+1}-1/2}{\alpha/2^{c+1}-1} D_{\alpha/2^{c}}(h(D)\|h(A)) + D_{\alpha/2^{c}-1}(h(A)\|h(D^{'})). \quad (4.7)$$

Como $\alpha/2^c - 1 < \alpha/2^c$ entonces por proposición 2.3.3 la desigualdad (4.7) cumple

$$\begin{aligned} D_{\alpha/2^{c+1}}(h(D)\|h(D^{'})) &\leq \frac{\alpha/2^{c}-1}{\alpha/2^{c}-2} D_{\alpha/2^{c}}(h(D)\|h(A)) + D_{\alpha/2^{c}}(h(A)\|h(D^{'})) \\ &\leq \frac{\alpha/2^{c}-1}{\alpha/2^{c}-2} \cdot 3^{c}\zeta + 3^{c}\zeta. \end{aligned} \quad (4.8)$$

Además, por hipótesis se tiene la desigualdad $\alpha \geq 2^{c+1}$, entonces la última suma del lado derecho de (4.8) es menor o igual a 3. Por lo tanto $D_{\alpha/2^{c+1}}(h(D)\|h(D^{'})) \leq 3^{c+1}\zeta$. ■

4.2. Teorema de Composición Avanzada

En esta sección daremos una versión del teorema de composición avanzada de la privacidad diferencial aproximada, enunciada en el capítulo 3, con respecto a la privacidad diferencial de Rényi.

Proposición 4.2.1. *Sea $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}$ una composición adaptativa de n mecanismos ζ- diferencialmente privados. Sean D y $D^{'}$ dos bases de datos adyacentes. Entonces para cualquiera $S \subset \mathcal{R}$:*

$$\Pr\left[\mathcal{M}(D) \in S\right] \leq \exp\left(2\zeta\sqrt{n\log\left(\frac{1}{\Pr\left[\mathcal{M}(D^{'}) \in S\right]}\right)}\right) \cdot \Pr\left[\mathcal{M}(D^{'}) \in S\right]. \quad (4.9)$$

Demostración. Por la proposición 2.3.6 se tiene que $\mathcal{M}$ es $(\alpha, n\zeta)$-RDP y por lema 2.3.3 obtenemos que para todo $\alpha \geq 1$

$$D_{\alpha}\left(\mathcal{M}(D)\|\mathcal{M}(D^{'})\right) \leq 2\alpha n\zeta^2.$$

Denotemos por q a $\Pr\left[\mathcal{M}(D^{'}) \in S\right]$ y consideremos dos casos.

Caso 1: $\log(1/q) \geq \zeta^2 n$. En este caso se tiene que $\dfrac{\sqrt{\log(1/q)}}{\sqrt{n}\zeta} \geq 1$ y haciendo $\alpha =$

$\frac{\sqrt{\log(1/q)}}{\sqrt{n}\zeta}$, tenemos por preservación de la probabilidad que

$$\Pr[\mathcal{M}(D) \in S] \leq \left\{\exp\left[D_\alpha \mathcal{M}(D) \| \mathcal{M}(D')\right] \cdot q\right\}^{1-1/\alpha} \tag{4.10}$$

$$\leq \exp\left(2(\alpha-1)n\zeta^2\right) exp\left(-\frac{1}{\alpha}\log(q)\right) \cdot q$$

$$= \exp\left(2(\alpha-1)n\zeta^2 - \frac{1}{\alpha}\log(q)\right) \cdot q. \tag{4.11}$$

Se puede verificar sin dificultad que

$$\frac{1}{\alpha}\log(q) = \zeta\sqrt{n\log(1/q)} \tag{4.12}$$

$$2(\alpha-1)n\zeta^2 = 2\zeta\sqrt{n\log(1/q)} - 2n\zeta^2 \tag{4.13}$$

$$< 2\zeta\sqrt{n\log(1/q)}.$$

Entonces usando las igualdades (4.12) y (4.13) el argumento de la exponencial de la expresión (4.11)es menor que $2\zeta\sqrt{n\log(\frac{1}{q})}$ y de esta forma obteniendo la desigualdad pedida.

Caso 2: $\log(1/q) < \zeta^2 n$. Este caso se sigue fácilmente, pues el lado derecho de la desigualdad que se quiere probar es mayor que 1, entonces por un lado

$$1 < \frac{1}{q} = q \cdot \exp\left[\log(1/q)^2\right]$$

$$= q \cdot \exp\left[2 \cdot \log(1/q)\right].$$

Ahora, de $\log(1/q) < \zeta^2 n$ se implica que $\exp[2\log(1/q)] \leq \exp\left[2\zeta\sqrt{n\log(1/q)}\right]$ y de esta última desigualdad obtenemos lo pedido. ■

Un rasgo notable de la proposición 4.2.1 es que su garantía de privacidad se encuentra de tal forma que depende de los eventos de probabilidad.

El siguiente corolario nos da una variante más convencional de la composición avanzada.

Corolario 4.2.1. *Sea $\mathcal{M}$ una composición de n mecanismos ζ- diferencialmente privados. Sea $0 < \delta < 1$ tal que $\log(1/\delta) \geq \zeta^2 n$. Entonces $\mathcal{M}$ satisface (ζ', δ)-privacidad diferencial*

donde

$$\zeta^{'} = 4\zeta\sqrt{2n\log(1/\delta)}. \tag{4.14}$$

Demostración. Sean D y $D^{'}$ dos bases de datos adyacentes, y S un subconjunto del rango de $\mathcal{M}$.

Por proposición 4.2.1

$$\Pr\left[\mathcal{M}(D) \in S\right] \leq \exp\left(2\zeta\sqrt{n\log\left(\frac{1}{\Pr\left[\mathcal{M}(D') \in S\right]}\right)}\right) \cdot \Pr\left[\mathcal{M}(D^{'}) \in S\right].$$

Denotemos por q a $\Pr\left[\mathcal{M}(D^{'}) \in S\right]$ y consideremos dos casos.

Caso 1: $8\log(1/\delta) > \log(1/q)$. De la desigualdad que supone el caso 1 se puede verificar que

$$2\zeta\sqrt{n\log(1/q)} < 2\zeta\sqrt{8n\log(1/\delta)}, \tag{4.15}$$

por lo que de (4.15) se obtiene que

$$\begin{aligned}\Pr\left[\mathcal{M}(D) \in S\right] &\leq \exp\left(2\zeta\sqrt{n\log(1/q)}\right) \cdot q \\ &< \exp\left(2\zeta\sqrt{8n\log(\log(1/\delta))}\right) \cdot q \\ &= \exp\left(4\zeta\sqrt{2n\log(1/\delta)}\right) \cdot q\end{aligned}$$

por lo que $Pr\left[\mathcal{M}(D) \in S\right] \leq e^{\zeta^{'}} \cdot q + \delta$.

Caso 2. $8\log(1/\delta) \leq \log(1/q)$. De la desigualdad de este caso se infiere que

$$q^{1/8} \leq \delta, \tag{4.16}$$

entonces teniendo en cuenta (4.16) se obtiene que

$$\begin{aligned}\Pr\left[\mathcal{M}(D) \in S\right] &\leq \exp\left(2\zeta\sqrt{n\log\left(1/\delta\right)}\right)\cdot q \\ &= \exp\left(2\sqrt{n\zeta^2\log\left(1/\delta\right)}\right)\cdot q \\ &\leq \exp\left(\sqrt{\log\left(1/q\right)\cdot\log\left(1/\delta\right)}\right)\cdot q.\end{aligned}$$

Aplicando (4.16) se tiene que

$$\begin{aligned}\Pr\left[\mathcal{M}(D) \in S\right] &\leq \exp\left(2\sqrt{\left(1/8\right)\left(\log\left(1/q\right)\right)^2}\right)\cdot q \\ &= \exp\left(\left(1/\sqrt{2}\right)\cdot\log\left(1/q\right)\right)\cdot q \\ &= q^{1-\frac{1}{\sqrt{2}}} \\ &\leq \delta.\end{aligned}$$

Pero tenemos que $\delta \leq \text{máx}\left\{e^{\zeta'}, \delta\right\} \leq e^{\zeta'}\cdot q + \delta$ entonces se obtiene la desigualdad pedida. ■

La condición $\log\left(1/\delta\right) \geq \zeta^2 n$ corresponde a la así llamada *alta privacidad* régimen del teorema de composición, donde $\zeta' < (1+\sqrt{2})\log\left(1/\delta\right)$. Ya que δ típicamente se escoge pequeño, digamos, menor que 0.001, esto cubre el caso de $\zeta' < 11$. En otras palabras, si $\log\left(1/\delta\right) < \zeta^2 n$, estos y otros teoremas de composición son improbables que produzcan límites fuertes.

4.3. RDP y (ζ, δ)-DP

Como observamos en la desigualdad (4.2), la definición de ϵ-privacidad diferencial coincide con (∞, ζ)-RDP. Por monotonía de la divergencia de Rényi, (∞, ζ)-RDP implica (α, ζ)-RDP para todo α finito.

Proposición 4.3.1 (De (α, ζ)-RDP a (ζ, δ)-DP)**.** *Si $\mathcal{M}$ es un mecanismo (α, ζ)-RDP, entonces satisface $\left(\zeta + \frac{\log(1/\delta)}{\alpha-1}, \delta\right)$-DP para cualquier $0 < \delta < 1$.*

Demostración. Tomemos cualesquiera dos bases de datos D, $D^{'} \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes y $S \subset \mathcal{R}$. Mostremos que $\mathcal{M}$ es $(\zeta^{'}, \delta)$- diferencialmente privado, donde $\zeta^{'} = \zeta + \frac{1}{\alpha-1}\log(1/\delta)$.

Por la proposición 2.3.4 tenemos que

$$\Pr\left[\mathcal{M}(D) \in S\right] \leq \left\{e^{\zeta}\Pr\left[\mathcal{M}(D^{'}) \in S\right]\right\}^{1-1/\alpha}.$$

Denotemos por q a $\Pr\left[\mathcal{M}(D^{'}) \in S\right]$ y consideremos los siguientes casos.

Caso 1. $e^{\zeta}q > \delta^{\alpha/(\alpha-1)}$. Para este caso

$$\begin{aligned}\Pr\left[\mathcal{M}(D) \in S\right] &\leq \left\{e^{\zeta}q\right\}^{1-1/\alpha} \\ &= e^{\zeta}q \cdot \left(e^{\zeta}q\right)^{-1/\alpha} \\ &\leq e^{\zeta}q \cdot \delta^{-1/(\alpha-1)}.\end{aligned}$$

Usando la identidad $\delta^{-\frac{1}{\alpha-1}} = \exp\left(\dfrac{\log(1/\delta)}{\alpha-1}\right)$ obtenemos

$$\begin{aligned}\Pr\left[\mathcal{M}(D) \in S\right] &\leq \exp\left(\zeta + \frac{\log(1/\delta)}{\alpha-1}\right) \cdot q \\ &\leq \exp(\zeta^{'}) \cdot \Pr\left[\mathcal{M}(D^{'}) \in S\right] + \delta.\end{aligned}$$

Caso 2. $e^{\zeta}q \leq \delta^{\alpha/(1-\alpha)}$. En este caso es inmediato el resultado ya que

$$\begin{aligned}\Pr\left[\mathcal{M}(D) \in S\right] &\leq \left\{e^{\zeta}q\right\}^{1-1/\alpha} \\ &= e^{\zeta}q \cdot \left(e^{\zeta}q\right)^{-1/\alpha}.\end{aligned}$$

Usando la desigualdad que propone el caso 2 obtenemos

$$\begin{aligned}\Pr\left[\mathcal{M}(D) \in S\right] &\leq e^{\zeta}q \cdot \delta^{-1/(\alpha-1)} \\ &\leq \exp(\zeta^{'})\Pr\left[\mathcal{M}(D^{'}) \in S\right] + \delta\end{aligned}$$

lo cual completa la prueba. ■

En este capítulo se establecieron comparativas importantes entre privacidad diferencial aproximada y privacidad diferencial de Rényi tales como: RDP implica DP, mientras que a través del concepto de divergencia de Rényi, podemos enunciar un resultado alternativo para el teorema de composición avanzada bajo privacidad diferencial aproximada. También mostramos que los resultados de composición adaptativa y grupal son compatibles bajo privacidad diferencial de Rényi.

Capítulo 5

Mecanismos Básicos

En este capítulo analizamos la privacidad diferencial de Rényi de tres mecanismos básicos y de su auto composición: *mecanismo de respuesta aleatoria (RRM)*, *mecanismo de Laplace* y *mecanismo Gaussiano*.

5.1. Mecanismo de Respuesta Aleatoria

Comenzamos analizando el mecanismo de respuesta aleatoria. Recordemos que un mecanismo aleatorizante es de respuesta aleatoria si

$$\mathcal{M}_{RRM}^{f,\epsilon}(D) := \begin{cases} f(D) & \text{con probabilidad } e^{\epsilon}/(e^{\epsilon}+1) \\ 1-f(D) & \text{con probabilidad } 1/(e^{\epsilon}+1) \end{cases}$$

donde $f : \mathbb{N}^{|\mathcal{X}|} \to \{0,1\}$.

5.1.1. Privacidad Diferencial Aproximada

Proposición 5.1.1. *El mecanismo $\mathcal{M}_{RRM}$ es ϵ-diferencialmente privado.*

Demostración. Sean $D, D' \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes y x, y, los cuales pueden ser distintos, en-

tonces

$$\frac{\Pr[\mathcal{M}_{RRM}^{f,\epsilon}(D) = x]}{\Pr[\mathcal{M}_{RRM}^{f,\epsilon}(D') = y]} = \frac{e^{\epsilon}/(e^{\epsilon}+1)}{1/(e^{\epsilon}+1)} = e^{\epsilon}.$$

■

5.1.2. Privacidad Diferencial de Rényi

Las siguiente proposición puede ser verificada directamente de la definición de privacidad diferencial de Rényi.

Proposición 5.1.2. *El mecanismo de respuesta aleatoria (RRM) satisface*

$$(\alpha, p_\alpha)\text{-}RDP \tag{5.1}$$

si $\alpha > 1$ *con* $p_\alpha = \dfrac{1}{\alpha-1}\log\left(p^\alpha(1-p)^{1-\alpha} + (1-p)^\alpha p^{1-\alpha}\right)$, *y*

$$(\alpha, p_1)\text{-}RDP \tag{5.2}$$

si $\alpha = 1$ *con* $p_1 = (2p-1)\log\left(\dfrac{p}{1-p}\right)$.

Demostración. Sean D y $D^{'}$ dos bases de datos adyacentes. Analicemos los casos.

i) $\alpha > 1$. Este caso es inmediato utilizando el caso $\alpha > 1$ en proposición 2.4.1 y haciendo $q = 1-p$.

ii) $\alpha = 1$. Usando el caso $\alpha = 1$ en proposición 2.4.1 y haciendo $q = 1-p$ tenemos que

$$\begin{aligned} D_1(\mathcal{M}(D)\|\mathcal{M}(D^{'})) &= \log\left(\frac{p}{1-p}\right)p + \log\left(\frac{1-p}{p}\right)(1-p) \\ &= \log\left(\frac{p}{1-p}\right)p - \log\left(\frac{p}{1-p}\right)(1-p) \\ &\leq \log\left(\frac{p}{1-p}\right)(2p-1). \end{aligned}$$

Obteniendo lo pedido.

■

5.2. Mecanismo de Laplace

En esta sección analizamos el mecanismo de Laplace y su privacidad diferencial realizando una comparación con la privacidad diferencial bajo divergencia de Rényi.

5.2.1. Privacidad Diferencial Aproximada

Teorema 5.2.1. *El mecanismo de Laplace $\mathcal{M}_{\text{Lap}}^{f,\epsilon}$ preserva $(\epsilon, 0)$-privacidad diferencial.*

Demostración. Tomemos $D, D^{'} \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes y sean $P = \mathcal{M}_{\text{Lap}}^{f,\epsilon}(f(D))$ y $Q = \mathcal{M}_{\text{Lap}}^{f,\epsilon}(f(D^{'}))$ con $\lambda = \frac{\Delta f}{\epsilon}$. Tenemos que

$$\begin{aligned} P(z) &= \frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon|f(D) - z|}{\Delta f}\right) \\ Q(z) &= \frac{\epsilon}{2\Delta f} \exp\left(-\frac{\epsilon|f(D^{'}) - z|}{\Delta f}\right). \end{aligned} \tag{5.3}$$

Realizando el cociente de las expresiones de (5.3) se obtiene que

$$\begin{aligned} \frac{P(z)}{Q(z)} &= \exp\left(\frac{\epsilon}{\Delta f}\left(|f(D^{'}) - z| - |f(D) - z|\right)\right) \\ &\leq \exp\left(\frac{\epsilon}{\Delta f}|f(D) - f(D^{'})|\right) \\ &\leq \exp(\epsilon). \end{aligned}$$

■

5.2.2. Privacidad Diferencial de Rényi

Ya que el mecanismo de Laplace es aditivo, la divergencia de Rényi entre $\mathcal{M}_{\text{Lap}}^{f,\epsilon}(f(D))$ y $\mathcal{M}_{\text{Lap}}^{f,\epsilon}(f(D^{'}))$ depende solamente de α y de la distancia $|f(D) - f(D^{'})|$.

Proposición 5.2.1. *Si una función f tiene sensitividad 1, entonces el mecanismo de Laplace $\mathcal{M}_{\text{Lap}}^{f,\epsilon}$ satisface $(\alpha, \zeta_{\lambda,\alpha})$-RDP con*

$$\zeta_{\lambda,\alpha} = \frac{1}{\alpha - 1} \log \left\{ \frac{\alpha}{2\alpha - 1} \exp\left(\frac{\alpha - 1}{\lambda}\right) + \frac{\alpha - 1}{2\alpha - 1} \exp\left(-\frac{\alpha}{\lambda}\right) \right\}.$$

Demostración. Sean $D, D' \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes. De acuerdo a la proposición 2.4.3 tenemos que $\mu_1 = f(D)$ y $\mu_2 = f(D')$ por lo que $|\mu_1 - \mu_2| < 1$ obteniendo el resultado. ■

De la proposición anterior se sigue el siguiente corolario.

Corolario 5.2.1. *Si $P \sim \text{Lap}(0, \lambda)$ y $Q \sim \text{Lap}(1, \lambda)$, entonces $D_\infty(P\|Q) = \frac{1}{\lambda}$.*

Demostración. Sean $p(x)$ y $q(x)$ las distribuciones de P y Q respectivamente.

De la ecuación (2.11) sabemos que $D_\infty(P\|Q) = \underset{x \in \text{supp}(Q)}{\text{Sup}} \left(\log \left(\frac{P(x)}{Q(x)} \right) \right)$ por lo que

$$D_\infty(P\|Q) = \underset{x \in \text{supp}(Q)}{\text{Sup}} \left\{ -\frac{|x|}{\lambda} + \frac{|1 - x|}{\lambda} \right\}.$$

Como $-\frac{|x|}{\lambda} + \frac{|1-x|}{\lambda} \leq \frac{1}{\lambda}$ y tenemos que $\frac{1}{\lambda} \leq \left\{ -\frac{|x|}{\lambda} + \frac{|1 - x|}{\lambda} \right\}$, pues supp(Q) $= \mathbb{R}$, concluimos que $D_\infty(\text{Lap}(0, \lambda)\|\text{Lap}(1, \lambda)) = \frac{1}{\lambda}$. ■

Este es, por supuesto, consistente con el mecanismo de Laplace satisfaciendo $1/\lambda$-privacidad diferencial. El otro valor extremo sigue de la igualdad

$$\lim_{\alpha \to 1} D_\alpha(\text{Lap}(0, \lambda)\|\text{Lap}(1, \lambda)) = 1/\lambda + \exp(-1/\lambda) - 1,$$

donde para λ suficientemente grande se aproxima a $0.5/\lambda$.

5.3. Mecanismo Gaussiano

Una alternativa al ruido Laplaciano es añadir ruido Gaussiano. En este caso, en lugar de escalar el ruido a ℓ_1 de sensibilidad Δf, se escala a ℓ_2.

5.3.1. Privacidad Diferencial Aproximada

Como se ha mencionado anteriormente el mecanismo Gaussiano presenta privacidad diferencial para ciertos casos particulares. El siguiente teorema, cuya demostración se omite pues escapa a los propósitos de esta tesis, nos dice en que casos el mecanismo Gaussiano presentará privacidad diferencial aproximada. Para una demostración puede consultarse [1].

Teorema 5.3.1. *Sea $\epsilon \in (0,1)$ arbitrario. Para $c^2 > 2\ln(1.25/\delta)$, el mecanismo Gaussiano con parámetro $\sigma \geq c\Delta_2 f/\epsilon$ es (ϵ,δ)- diferencialmente privado.*

5.3.2. Privacidad Diferencial de Rényi

Bajo el concepto de divergencia de Rényi el mecanismo Gaussiano presenta la siguiente propiedad.

Proposición 5.3.1. *Si f tiene sensitividad 1, entonces el mecanismo Gaussiano $\mathcal{M}_{\mathrm{N}}^{f,\sigma}$ es $(\alpha, \alpha/(2\sigma^2))$-RDP.*

Demostración. Sean D y $D' \in \mathbb{N}^{|\mathcal{X}|}$ adyacentes. Como $\mathcal{M}_{\mathrm{N}}^{f,\sigma}(f(D))$ se distribuye normal con media $f(D)$ y varianza σ^2, mientras que $\mathcal{M}_{\mathrm{N}}^{f,\sigma}(f(D'))$ tiene distribución normal con media $f(D')$ y varianza σ^2, entonces por corolario 2.3.1 se tiene que $(\mu_1 - \mu_2)^2 < 1$ obteniendo la desigualdad pedida. ■

5.4. Gráficas y Tablas

En esta sección se realiza una comparativa entre la privacidad diferencial aproximada y la privacidad diferencial bajo la divergencia de Rényi de tres mecanismos aleatorizantes: *mecanismo de respuesta aleatoria, mecanismo de Laplace y mecanismo Gaussiano.* Como puede verse, la relación de la cota ζ en función del parámetro α es lineal para el mecanismo Gaussiano como se muestra en la figura 5.3 para distintos valores de σ la cual es más simple comparada con la relación $\sigma \geq \frac{c\Delta_2 f}{\epsilon}$ pedida en el teorema 5.3.1. Por otro lado para los mecanismos de Laplace y de respuesta aleatoria la relación de ζ en función de α es más

complicada por ejemplo, de acuerdo a la figura 5.2, para el mecanismo de Laplace la gráfica de ζ para distintos valores de λ en función de α sufre un crecimiento, mientras que, comparada con la relación $\lambda = \frac{\Delta f}{\epsilon}$ para el mecanismo de Laplace en teorema 5.2.1 sería de decrecimiento. En la tabla 5.1 se resume los resultados obtenidos en este capítulo.

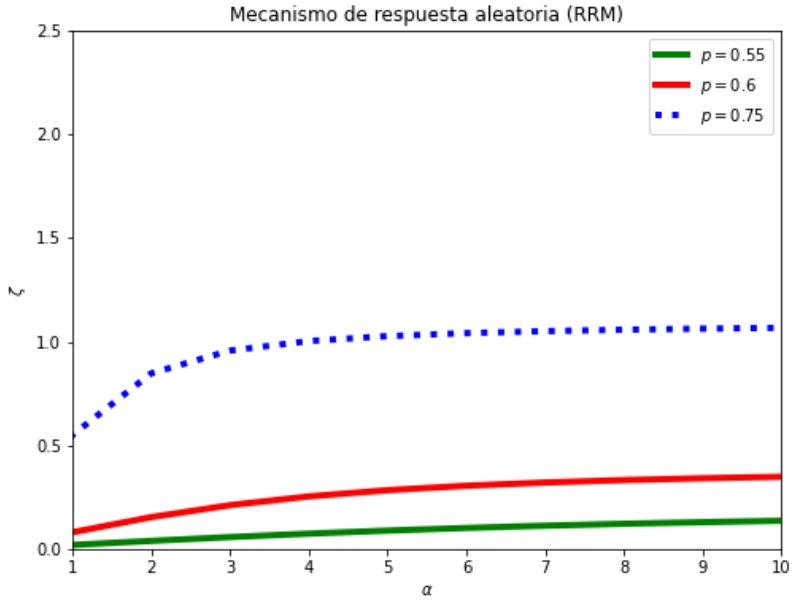

Figura 5.1: Gráfica de Presupuesto para el mecanismo RRM.

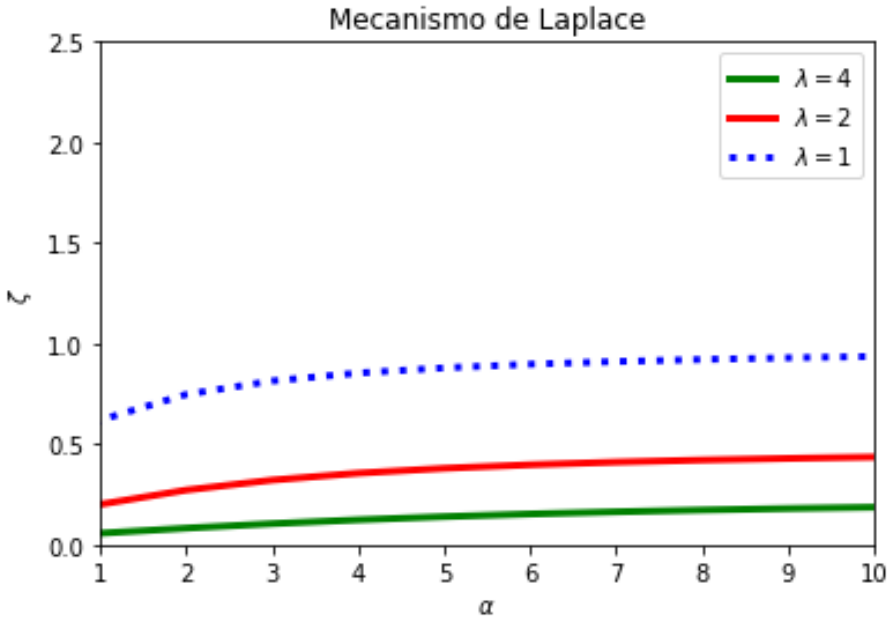

Figura 5.2: Gráfica de Presupuesto para el mecanismo de Laplace.

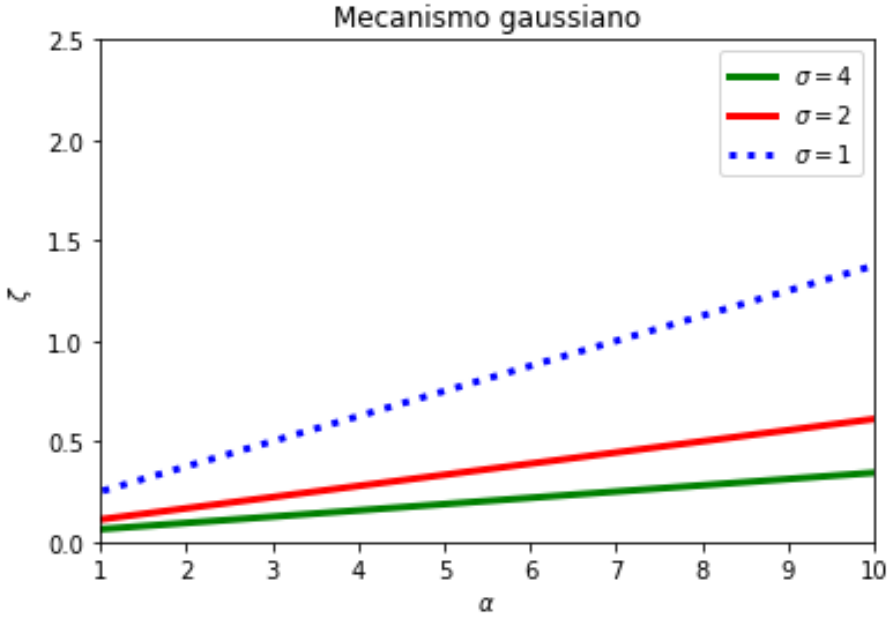

Figura 5.3: Gráfica de Presupuesto para el mecanismo de Gauss.

Propiedad	**Privacidad diferencial**	**Privacidad diferencial de Rényi**
Cambio en la probabilidad de S	$\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(D') \in S]$ (Def.3.2.1)	$\Pr[\mathcal{M}(D) \in S] \leq \left(e^{\epsilon} Pr[\mathcal{M}(D') \in S]\right)^{(\alpha-1)/\alpha}$ (Prop 2.3.4)
Post-procesamiento	$\mathcal{M}$ es ϵ-DP $\Rightarrow g \circ \mathcal{M}$ es ϵ-DP (Prop 3.3.1)	$\mathcal{M}$ es (α, ζ)-RDP $\Rightarrow g \circ \mathcal{M}$ es (α, ζ)-RDP (Prop 4.1.1)
Composición secuencial básica	$\mathcal{M}_1$ y $\mathcal{M}_2$ son ϵ-DP $\Rightarrow (\mathcal{M}_1, \mathcal{M}_2)$ es 2ϵ-DP (Lema 3.3.1)	$\mathcal{M}_1$ y $\mathcal{M}_2$ son (α, ζ)-RDP $\Rightarrow (\mathcal{M}_1, \mathcal{M}_2)$ es $(\alpha, 2\zeta)$-RDP (Prop 4.1.1)
Privacidad grupal (pre-procesamiento)	$\mathcal{M}$ es ϵ-DP, g es 2^c-estable $\Rightarrow \mathcal{M} \circ g$ es $2^c\epsilon$-DP (Prop 3.3.2)	$\mathcal{M}$ es (α, ζ)-RDP, g es 2^c-estable $\Rightarrow \mathcal{M} \circ g$ es $(\alpha/2^c, 3^c\zeta)$-RDP (Prop 4.1.3)
Mecanismo RRM	ϵ-DP (Prop 5.1.1)	(α, p_α)-RDP si $\alpha > 1$, (α, p_1)-RDP si $\alpha = 1$ (Prop 5.1.2)
Mecanismo de Laplace	$(\epsilon, 0)$-DP (Teo 5.2.1)	$(\alpha, \zeta_{\lambda,\alpha})$-RDP (Cor 5.2.1)
Mecanismo Gaussiano	(ϵ, δ)-DP (Teo 5.3.1)	$(\alpha, \alpha/2\sigma^2)$-RDP (Cor 5.3.1)

Cuadro 5.1: Comparación entre Privacidad diferencial y Privacidad diferencial de Rényi.

$\zeta_{\sigma,\alpha} = \dfrac{\alpha}{2\sigma^2}$.... Función con gráfica 5.3

$\zeta_{p,\alpha} = \dfrac{1}{\alpha - 1} \log\left(p^{\alpha}(1-p)^{1-\alpha} + (1-p)^{\alpha}p^{1-\alpha}\right)$....Función con gráfica 5.1

$\zeta_{\lambda,\alpha} = \dfrac{1}{\alpha - 1} \log\left\{\dfrac{\alpha}{2\alpha - 1} \exp\left(\dfrac{\alpha - 1}{\lambda}\right) + \dfrac{\alpha - 1}{2\alpha - 1} \exp\left(-\dfrac{\alpha}{\lambda}\right)\right\}$.... Función con gráfica 5.2

Capítulo 6

Conclusiones

En esta tesis se presentó la teoría necesaria para estudiar una alternativa a la privacidad diferencial aproximada llamada *Privacidad Diferencial de Rényi*. Particularmente, en este trabajo, se muestra que la privacidad diferencial de Rényi es una relajación natural tanto a ϵ-privacidad diferencial como a (ϵ, δ)-privacidad diferencial.

El concepto de divergencia de Rényi, el cual se expone en el capítulo 2, muestra propiedades operativas y teóricas las cuales vuelven a esta divergencia un buen candidato para estudiar privacidad diferencial bajo su concepto. De esta manera la privacidad diferencial de Rényi muestra mantener, bajo ciertas adaptaciones, las propiedades básicas que deben de presentar los mecanismos aleatorizantes: post-procesamiento, preservación bajo composición secuencial y privacidad grupal, las cuales son resumidas en el cuadro 5.1. Además, dos resultados a favor a la divergencia de Rényi son 1) la versión que ofrece al teorema de composición avanzada supone condiciones menos restrictivas para el parámetro ζ comparadas con las condiciones para el parámetro ϵ en el teorema 3.3.2 y 2) la conexión que ofrece la proposición 2.3.6 (i.e., privacidad diferencial de Rényi implica privacidad diferencial aproximada).

Una vez estudiadas las propiedades de la divergencia de Rényi parece natural aplicarla a los mecanismos aleatorizantes básicos: mecanismo de respuesta aleatoria, mecanismo de Laplace y mecanismo de respuesta aleatoria. En el capítulo 5 se ofrece una comparativa entre privacidad diferencial de Rényi y privacidad diferencial aproximada para estos mecanismos. En esta parte se resaltan las relaciones que que existen entre los parámetros

α y ζ. Principalmente, como muestra la gráfica 5.1, bajo Rényi existe una relación lineal para diversos valores de σ mientras que de acuerdo al teorema 5.3.1 se pide la relación $\sigma \geq c\Delta_2 f/\epsilon$.

También como puede verse se presentan ciertas preguntas. Del lema 1.3.3 un mecanismo (ϵ, δ)-RDP se muestra bastante limitado entonces, bajo estas restricciones, queda abierta la pregunta acerca de si las garantías de privacidad pueden mejorarse para ciertos valores de α. Otra cuestión se presenta con respecto a la preservación de la probabilidad (proposición 2.3.4) en la cual podemos preguntarnos si es posible hacer tender $P(A)$ hacia $Q(A)$ o si bien se puede mejorar la cota $P(A)^{(\alpha-1)/\alpha}$.

En general podemos decir que el concepto de privacidad diferencial de Rényi se muestra abierta a estudio y aplicaciones que podrán ayudar a abordar, y tal vez mejorar, las garantías de privacidad que ofrece la privacidad diferencial aproximada.

Bibliografía

[1] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211-407.

[2] Polyanskiy, Y., & Wu, Y. (2014). Lecture notes on information theory. Lecture Notes for ECE563 (UIUC) and, 6(2012-2016), 7.

[3] Cover, T. M. (1991). 1. A. Thomas, Elements of information theory.

[4] Mironov, I. (2017, August). Rényi differential privacy. In 2017 IEEE 30th computer security foundations symposium (CSF) (pp. 263-275). IEEE.

[5] Royden, H. L., & Fitzpatrick, P. (1988). Real analysis (Vol. 32). New York: Macmillan.

[6] Jacod, J., & Protter, P. (2004). Probability essentials. Springer Science & Business Media.

[7] Zwillinger, D., & Jeffrey, A. (Eds.). (2007). Table of integrals, series, and products. Elsevier.

Apéndice A

Resultados Auxiliares de Análisis

En este apéndice resumimos los resultados del análisis real y conceptos de teoría de la información los cuales son aplicados en la demostración y explicación de varios resultados que se presentan en este trabajo.

Definición A.0.1. *[2] Sea S conjunto convexo. Una función $f : S \to \mathbb{R}$ se dice ser convexa si para todo $x, y \in S$ y $\alpha \in [0, 1]$ se tiene:*

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

Teorema A.0.1 (Desigualdad de Jensen [6])**.** *Sea $\varphi : \mathbb{R} \to \mathbb{R}$ función convexa, y sean X y $\varphi(X)$ variables aleatorias integrables. Para cualquier sub-σ álgebra $\mathcal{G}$,*

$$\varphi\left(\mathbb{E}\left\{X|\mathcal{G}\right\}\right) \leq \mathbb{E}\left\{\varphi(X)|\mathcal{G}\right\}.$$

Lema A.0.1 (Desigualdad de Hölder [5])**.** *Si $p>1$ y $\frac{1}{p} + \frac{1}{q} = 1$, entonces para cualesquiera $x_1, ..., x_n; y_1, ..., y_n \in \mathbb{R}$ se tiene:*

$$\sum_{i=1}^{n} |x_i y_i| \leq \left\{\sum_{i=1}^{n} |x_i|^p\right\}^{1/p} \left\{\sum_{i=1}^{n} |y_i|^q\right\}^{1/q}.$$

Teorema A.0.2 (Desigualdad de la suma logarítmica [3])**.** *Para números reales no nega-*

tivos $a_1, ..., a_n$ *y* $b_1, b_2, ..., b_n$,

$$\sum_{i=1}^{n} a_i \log\left(a_i/b_i\right) \geq \left(\sum_{i=1}^{n} a_i\right) \log\left(\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}\right)$$

con igualdad si y solo si $\dfrac{a_i}{b_i} = constante.$

Teorema A.0.3. *[7] Si* $a > 0$ *entonces*

$$\int_{-\infty}^{\infty} \exp\left(-(ax^2 + bx + c)\right) dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2 - 4ac}{4a}\right).$$

Definición A.0.2 (f-divergencia [2])**.** *Sean* $f : (0, \infty) \to \mathbb{R}$ *una función convexa tal que para todo* $s, t \in (0, \infty)$ *y* $\alpha \in (0, 1)$ *tal que* $\alpha s + (1 - \alpha)t = 1$*, se cumple que* $\alpha f(s) + (1 - \alpha) f(t) > f(1)$ *y* P*,* Q *distribuciones definidas sobre un conjunto contable* $\mathcal{X}$*. Se define la* f*-divergencia entre* P *y* Q *como:*

$$D_f(P\|Q) := \mathbb{E}_{x\sim Q}\left[f\left(\frac{P(x)}{Q(x)}\right)\right].$$

Proposición A.0.1 (Data Processing Inequality [2])**.** *Sea* $P_{Y|X}$ *un kernel. Si* P_Y *es la distribución de* Y *cuando* X *es generado por* P_X *y* Q_Y *es la distribución de* Y *cuando* X *es generado por* Q_X*, entonces para cualquier* f*-divergencia* $D_f(\cdot\|\cdot)$*,*

$$D_f(P_Y\|Q_Y) \leq D_f(P_X\|Q_X).$$

Definición A.0.3 (Transformación aleatoria [2])**.** *Dada una probabilidad condicional definimos el kernel de probabilidad de transición como una función* $K(\cdot|\cdot)$ *de dos argumentos en cual con respecto al primer elemento toma conjuntos medibles de* $\mathcal{Y}$*, y con respecto al segundo argumento es un elemento de* $\mathcal{X}$ *la cual satisface*

1) Para cualquier $x \in \mathcal{X}$*:* $K(\cdot|x)$ *es una función medible sobre* $\mathcal{Y}$*.*

2) Para cualquier conjunto medible A*:* $K(A|\cdot)$ *es una función medible sobre* $\mathcal{X}$*.*