دانشگاه صنعتی شریف
دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد
ریاضی کاربردی

محرمانگی در یادگیری ماشین

نگارش

فیروزه ابریشمی

استاد راهنما

دکتر جواد ابراهیمی بروجنی

اسفند ۱۴۰۴


بسم الله الرحمن الرحیم

به نام خدا

دانشگاه صنعتی شریف

دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد

این پایان‌نامه به عنوان تحقق بخشی از شرایط دریافت درجه کارشناسی ارشد است.

عنوان: محرمانگی در یادگیری ماشین

نگارش: فیروزه ابریشمی

کمیته ممتحنین

استاد راهنما: دکتر جواد ابراهیمی امضاء:
بروجنی

استاد مدعو: دکتر امین السادات طالبی امضاء:

استاد مدعو: دکتر محمدحسین یاسائی امضاء:
میبدی

تاریخ:

اظهارنامه

(اصالت متن و محتوای پایان‌نامه کارشناسی ارشد)

عنوان پایان‌نامه: محرمانگی در یادگیری ماشین

استاد راهنما: دکتر جواد ابراهیمی بروجنی **استاد مشاور:** -

این‌جانب فیروزه ابریشمی اظهار می‌دارم:

۱. متن و نتایج علمی ارائه شده در این پایان‌نامه اصیل بوده و زیرنظر استادان نام‌برده شده در بالا تهیه شده است.

۲. متن پایان‌نامه به این صورت در هیچ جای دیگری منتشر نشده است.

۳. متن و نتایج مندرج در این پایان‌نامه، حاصل تحقیقات این‌جانب به عنوان دانشجوی کارشناسی ارشد دانشگاه صنعتی شریف است.

۴. کلیه مطالبی که از منابع دیگر در این پایان‌نامه مورد استفاده قرار گرفته، با ذکر مرجع مشخص شده است.

نگارنده: فیروزه ابریشمی

تاریخ:

امضاء:

نتایج تحقیقات مندرج در این پایان‌نامه و دستاوردهای مادی و معنوی ناشی از آن (شامل فرمول‌ها، توابع کتابخانه‌ای، نرم‌افزارها، سخت‌افزارها و مواردی که قابلیت ثبت اختراع دارد) متعلق به دانشگاه صنعتی شریف است. هیچ شخصیت حقیقی یا حقوقی بدون کسب اجازه از دانشگاه صنعتی شریف حق فروش و ادعای مالکیت مادی یا معنوی بر آن یا ثبت اختراع از آن را ندارد. همچنین، کلیه حقوق مربوط به چاپ، تکثیر، نسخه‌برداری، ترجمه، اقتباس و نظائر آن در محیط‌های مختلف اعم از الکترونیکی، مجازی یا فیزیکی برای دانشگاه صنعتی شریف محفوظ است. نقل مطلب با ذکر ماخذ بلامانع است.

استاد راهنما: دکتر جواد ابراهیمی بروجنی

تاریخ:

امضاء:

نگارنده: فیروزه ابریشمی

تاریخ:

امضاء:

سپاس

بِسْمِ اللَّهِ الرَّحْمنِ الرَّحیم

”مِنّت خدای را عزَّ وَ جَلّ که طاعتش موجبِ قُربَت است و به شُکر اندرش مَزیدِ نعمت. هر نفسی که فرو می‌رود مُمدِّ حیات است و چون بر می‌آید مُفَرِّحِ ذات.“

خداوند را سپاس‌گزارم که مرا در زیباترین و بهترین نقطه‌ی جهان، ایران عزیزم، آفرید و در آغوش مهربان‌ترین پدر و مادر دنیا قرار داد. و او را شاکرم که من را در مسیری قرار داد که بتوانم قطره‌ای از دریای نامتناهی علم و عظمتش را ببینم. امیدوارم بتوانم همواره در این راه و در خدمت خَلقش قدم بردارم و پایان این کار کوچک سرآغازی بر ادامه‌ی مسیر روشن علم و دانش باشد.

پیمودن این راه پرچالش بدون حمایت و راهنمایی‌های عزیزانم ممکن نبود؛

ابتدا از استاد عزیزم، جناب آقای دکتر جواد ابراهیمی، که با کمک‌ها و راهنمایی‌های بی‌دریغشان، مرا در به سرانجام رساندن این پایان‌نامه یاری دادند، تشکر و قدردانی می‌کنم. ایشان علاوه بر علم، به من درس اخلاق آموختند؛ منش و تأکید همیشگیِ ایشان بر گشاده‌رویی، گره‌گشایی و تسهیلِ مسیر دیگران، آموزه‌ای است که همواره در خاطر خواهم داشت.

از خانواده‌ی عزیزتر از جانم، خصوصاً مادر و پدرم که سال‌ها تلاش کردند و زحمت کشیدند تا فرزندانی نیک تربیت کنند، و از خواهر و برادران مهربانم که در تمام این سال‌ها، هم از نظر روحی و ذهنی و هم از نظر علمی همراهم بودند و من را از کمک‌هایشان محروم نکردند، سپاس‌گزارم.

هم‌چنین از دوستان خوبم که حمایت‌های عاطفی و مهرشان، سبب استقامت من در مسیر زندگی بوده است متشکرم.

تقدیم به

می‌خواهم این پایان‌نامه را به عزیزی تقدیم کنم
که پناهِ بی‌صدای دل‌های خسته است؛
همان مونسی که مهربانی‌اش چون نَفَسی آرام،
در هیاهوی زندگی، ما را از فراموشیِ آسمان نجات می‌دهد.

او که اگر یاد و نگاهش نبود،
تندبادِ حوادث، چراغِ راه را از ما می‌گرفت،
نااهلان و دشمنان بر ما چیره می‌شدند،
و گام‌هایمان در تاریکیِ روزگار گم می‌شد؛
او که اندوهِ ما را به دل می‌گیرد
و با رنجِ ما، بی‌آن‌که دیده شود، همراه می‌شود.

لحظاتی که صرف این کار کردم را به امامی تقدیم می‌کنم،
که اگرچه از نگاه‌ها پنهان است،
اما حضورش در تار و پودِ جهان جاری‌ست
و برکتِ نامش، امیدِ ایستادن و ادامه دادن است.

می‌دانم این نوشته در برابر عظمتِ او ناچیز است؛
پایان‌نامه که هیچ،
تمام زندگی‌ام را تقدیم به موعودی می‌کنم
که با آمدنش، صبحِ دانایی بر جهان خواهد دمید.

از خداوند، تنها یک آرزو دارم:
روزی که او بیاید،
نام ما نیز در شمارِ منتظرانِ راستینش نوشته شده باشد.

چکیده

چارچوب «محرمانگی تفاضلی» به عنوان استاندارد طلایی حفاظت از داده‌ها، با ارائه یک ضمانت ریاضی قوی تضمین می‌کند که خروجی الگوریتم‌ها حساسیت معناداری به حضور یا عدم حضور داده‌های یک فرد خاص نداشته باشند. با این وجود، مدل استاندارد و متمرکز این چارچوب نیازمند تجمیع داده‌های خام توسط یک نهاد واسط است. چارچوب «محرمانگی تفاضلی موضعی» به منظور رفع این وابستگی توسعه یافته است؛ در این مدل داده‌ها پیش از تجمیع و در سمت کاربر نویزدار می‌شوند. با این وجود، حل مسئله‌ی اعتماد منجر به چالش بنیادین دیگری می‌شود: افت شدید سودمندی آماری، به گونه‌ای که پیچیدگی تعداد نمونه از مرتبه‌ی $\mathcal{O}(n)$ در مدل متمرکز به تقریب $\mathcal{O}(n^{۲})$ در مدل موضعی افزایش می‌یابد.

پرسش اساسی این است که آیا این افت چشم‌گیر دقت یک محدودیت ذاتی است یا ناشی از ضعف الگوریتم‌های فعلی. این پایان‌نامه با هدف پاسخ به این پرسش، مروری تحلیلی بر ادبیات پیشرو این حوزه ارائه می‌دهد. در این راستا، دو رویکرد مکمل مطالعه و تلفیق شده‌اند: نخست، چارچوب نظریه مینی‌مکس آماری بر پایه کارهای دوچی، جردن و وین‌رایت، که با ابزارهایی نظیر لم اسود و نامساوی فانو کران‌های پایین خطا را اثبات می‌کند؛ دوم، چارچوب هندسه اطلاعات بر پایه کارهای آسوده و همکاران، که مکانیزم‌های موضعی را به مثابه «کانال‌های انقباض اطلاعات» تفسیر می‌کند.

در این رساله نشان داده می‌شود که شرط محرمانگی تفاضلی موضعی معادل است با انقباض f-واگرایی‌ها، و از این هم‌ارزی برای استخراج ضرایب انقباض دقیق برای واگرایی‌های کای‌ـدو، کولبک‌ـلایبلر و هاکی‌ـاستیک استفاده می‌شود. ترکیب این ضرایب با ابزارهای کلاسیک نظریه تصمیم آماری، کران‌های پایین مینی‌مکس را به شکلی منسجم و یک‌پارچه اثبات می‌کند و به صراحت نشان می‌دهد که افت سرعت همگرایی به $\mathcal{O}\left(\frac{۱}{\sqrt{n}}\right)$ یک مانع اطلاعاتی و هندسیِ اجتناب‌ناپذیر است. این چارچوب، معیاری دقیق برای ارزیابی بهینگی الگوریتم‌های آتی در حوزه‌ی حریم خصوصی فراهم می‌سازد.

کلمات کلیدی: محرمانگی تفاضلی موضعی، نظریه مینی‌مکس آماری، fـواگرایی‌ها، نامساوی پردازش داده قوی، موازنه دقت‌ـمحرمانگی

فهرست مطالب

۱ مقدمه **۲**

۱-۱ اهمیت داده‌ها و ضرورت حفظ حریم خصوصی ۲

۱-۱-۱ محرمانگی تفاضلی (DP) ۴

۱-۱-۲ محرمانگی تفاضلی موضعی (LDP) ۷

۱-۲ کارهای پیشین و مرور ادبیات ۸

۱-۲-۱ آغازگرها: از پیمایش‌های آماری تا تعریف مدرن محرمانگی ۸

۱-۲-۲ دسته‌بندی پروتکل‌های موضعی: تعاملی و غیرتعاملی ۱۲

۱-۲-۳ چالش سودمندی و موازنه دقت ـ محرمانگی ۱۲

۱-۲-۴ نگاهی آماری به LDP: چارچوب مینی‌مکس و حدود بنیادین ۱۳

۱-۳ بیان مسئله و اهداف پژوهش ۱۴

۱-۳-۱ رویکرد تحلیل: f ـ واگرایی‌ها به عنوان زبان مشترک ۱۵

۱-۳-۲ رویکرد تحلیل: چارچوب انقباض f ـ واگرایی‌ها ۱۶

۱-۳-۳ اهداف و ساختار پژوهش ۱۶

۱-۴ ساختار پایان‌نامه ۱۷

۲ پیش‌نیازها **۱۹**

۲-۱ محرمانگی تفاضلی متمرکز (CDP) ۱۹

۲-۱-۱ مدل اعتماد و تعریف رسمی ۱۹

۲-۱-۲ تفسیر پارامترهای محرمانگی ۲۲
۲-۱-۳ تعاریف معادل و صورت‌بندی‌های جایگزین ۲۳
۲-۱-۴ مکانیزم‌های پایه ۲۴
۲-۱-۵ مکانیزم‌های بنیادی محرمانگی تفاضلی ۲۶
۲-۱-۶ ترکیب‌پذیری ۲۹
۲-۱-۷ محرمانگی گروهی ۳۰
۲-۱-۸ محدودیت مدل متمرکز ۳۰
۲-۲ fـ واگرایی‌ها ۳۰
۲-۲-۱ تعریف رسمی در فضای اندازه‌پذیر ۳۱
۲-۲-۲ تعریف fـ واگرایی ۳۲
۲-۲-۳ نمونه‌های مهم و توابع مولد ۳۳
۲-۲-۴ خواص بنیادین و روابط بین fـ واگرایی‌ها ۳۶
۲-۳ مبانی آماری و کران‌های اطلاعاتی ۳۸
۲-۳-۱ نظریه تصمیم و ریسک مینی‌مکس ۳۸
۲-۳-۲ تقلیل به آزمون فرض (روش بسته‌بندی) ۳۹
۲-۳-۳ نامساوی‌های کران پایین ۴۲

۳ محرمانگی تفاضلی موضعی ۴۴

۳-۱ مقدمه ۴۴
۳-۲ تعاریف رسمی و مدل‌های محاسباتی ۴۵
۳-۲-۱ تعریف ریاضی LDP ۴۶
۳-۲-۲ محرمانگی تقریبی ۴۷
۳-۳ پروتکل‌های تعاملی و خواص ترکیب ۴۸
۳-۳-۱ پروتکل‌های غیرتعاملی ۴۸
۳-۳-۲ پروتکل‌های تعاملی (ترتیبی) ۴۹

۳-۳-۳ قضیه ترکیب ترتیبی ۴۹
۳-۴ مکانیزم‌های پایه در LDP ۵۰
۳-۴-۱ پاسخ تصادفی دودویی (RR) ۵۰
۳-۴-۲ پاسخ تصادفی تعمیم‌یافته (GRR) ۵۱
۳-۴-۳ مکانیزم‌های مبتنی بر کدگذاری یکانی (UE) ۵۲
۳-۴-۴ تحلیل مقایسه‌ای: چرا GRR در ابعاد بالا شکست می‌خورد؟ ۵۴
۳-۴-۵ مکانیزم لاپلاس موضعی ۵۵
۳-۴-۶ مکانیزم تخمین میانگین دوچی (۱ ـ بیتی) ۵۶
۳-۵ چالش سودمندی و هزینه عدم اعتماد ۵۸
۳-۵-۱ تعریف مسئله: تخمین میانگین دودویی ۵۸
۳-۵-۲ تحلیل در مدل متمرکز (CDP) ۵۸
۳-۵-۳ تحلیل در مدل موضعی (LDP) ۵۹
۳-۵-۴ نتیجه‌گیری: شکاف کارایی ۵۹

۴ تحلیل مینی‌مکس و هندسه اطلاعاتی در LDP ۶۱
۴-۱ مقدمه ۶۱
۴-۲ محرمانگی به عنوان انقباض ۶۲
۴-۲-۱ تفسیر رژیم‌های محرمانگی ۶۶
۴-۲-۲ تعمیم به n کاربر مستقل (خاصیت تنسوری شدن) ۶۶
۴-۳ نامساوی‌های پردازش داده قوی (SDPI) ۶۷
۴-۳-۱ کران انقباض برای واگرایی کای‌ـ‌دو (χ^2) ۶۸
۴-۳-۲ مزیت‌های تحلیلی واگرایی χ^2 نسبت به KL ۶۹
۴-۳-۳ تنسوری شدن واگرایی χ^2 برای n کاربر ۷۰
۴-۴ اثبات نرخ مینی‌مکس برای تخمین میانگین ۷۱
۴-۴-۱ تعریف مسئله ۷۱

۴-۴-۲ بحث و تفسیر ۷۳

۵ نتیجه‌گیری و پیشنهادها **۷۴**

۵-۱ جمع‌بندی و دستاوردهای اصلی ۷۴

۵-۱-۱ دستاوردهای اصلی پژوهش ۷۵

۵-۲ پیشنهادهایی برای تحقیقات آتی ۷۵

۵-۲-۱ تحلیل انقباض در مدل شافل ۷۶

۵-۲-۲ ارتباط با محرمانگی تفاضلی رنی ۷۶

۵-۲-۳ تخمین‌گرهای تطبیقی ۷۶

۵-۲-۴ تعمیم به داده‌های وابسته ۷۶

مراجع **۷۸**

واژه‌نامه **۸۱**

فهرست جداول

۲-۱ خانواده fـ واگرایی‌ها و توابع مولد آن‌ها ۳۸

فهرست تصاویر

۲-۱ مدل محرمانگی تفاضلی متمرکز (CDP) با یک متصدی مورد اعتماد. ۲۰

۲-۲ تجسم هندسی گوی یکه در فضای l_p دوبعدی برای مقادیر مختلف p. همان‌طور که در شکل پیداست، به ازای $p \geqslant ۱$ گوی‌ها تشکیل مجموعه‌هایی محدب می‌دهند، اما برای مقادیر $p < ۱$ (مانند $p = ۰/۵$) خاصیت تحدب از بین می‌رود. حالت $p = ۲$ نمایانگر فضای استاندارد اقلیدسی و حالت $p = ۱$ نمایانگر فاصله‌ی منهتن (مجموع قدرمطلق‌ها) است. ۲۱

۲-۳ نمایش هندسی روش بسته‌بندی برای تقلیل مسئله‌ی تخمین به آزمون فرض. فضای پارامتر Θ با مجموعه‌ای از گوی‌های مجزا $\{\theta^۱, \ldots, \theta^K\}$ پوشانده شده است که تشکیل یک ۲δ ـ بسته‌بندی می‌دهند. اگر تخمین‌گر $\widehat{\theta}$ (نقطه قرمز) دارای خطای کم‌تر از δ باشد، لزوماً درون یکی از این گوی‌ها قرار می‌گیرد. از آن‌جا که گوی‌ها مجزا هستند، حداکثر یک اندیس i وجود دارد که $\widehat{\theta}$ به آن نزدیک باشد؛ لذا مسئله تخمین به یافتن این اندیس (آزمون فرض چندگزینه‌ای) تقلیل می‌یابد. ۴۰

۳-۱ گذار از مدل متمرکز به موضعی؛ نویز به صورت موضعی روی دستگاه کاربر اضافه می‌شود. ۴۵

۳-۲ (الف) مدل گرافیکی نمایش‌دهنده‌ی روابط استقلال شرطی بین داده‌های خصوصی $\{X_i\}_{i=۱}^n$ و متغیرهای مشاهده‌شده $\{Z_i\}_{i=۱}^n$ در پروتکل‌های تعاملی؛ پیکان‌های افقی نشان‌دهنده‌ی وابستگی Z_i به تاریخچه‌ی پیشین هستند. (ب) مدل گرافیکی ساده‌تر در پروتکل‌های غیرتعاملی که در آن خروجی‌ها به شرط ورودی‌ها از یکدیگر مستقل هستند. ۴۸

فصل ۱

مقدمه

در این بخش به توضیح مسئله‌ی محرمانگی تفاضلی و اهمیت آن می‌پردازیم. سپس ادبیات موضوع را شرح داده و مسئله را بیان می‌کنیم.

۱-۱ اهمیت داده‌ها و ضرورت حفظ حریم خصوصی

در دهه‌های اخیر، جهان شاهد رشد انفجاری در تولید و جمع‌آوری داده‌ها بوده است. پیشرفت‌های چشم‌گیر در فناوری‌های ذخیره‌سازی، محاسبات ابری و اینترنت اشیاء، منجر به انباشت حجم عظیمی از داده‌ها شده است که اغلب تحت عنوان کلان‌داده[۱] شناخته می‌شوند. این داده‌ها سوخت اصلی موتورهای تصمیم‌گیری مدرن و سیستم‌های هوشمند هستند. امروزه، الگوریتم‌های یادگیری ماشین[۲] و تحلیل داده[۳] با بهره‌گیری از این مخازن عظیم اطلاعاتی، قادرند الگوهای پیچیده‌ای را شناسایی کنند که در حوزه‌هایی نظیر تشخیص پزشکی، بهینه‌سازی ترافیک شهری، توصیه‌گرهای تجاری و سیاست‌گذاری‌های کلان اقتصادی کاربرد حیاتی دارند.

با این حال، این استفاده‌ی گسترده از داده‌ها، نگرانی‌های جدی و فزاینده‌ای را در خصوص حریم خصوصی[۴] افراد به وجود آورده است. داده‌های خامی که برای آموزش مدل‌های هوشمند یا استخراج آماره‌ها استفاده می‌شوند، اغلب حاوی اطلاعات حساس[۵] و شخصی هستند. تاریخچه‌ی تراکنش‌های مالی، سوابق پزشکی، موقعیت‌های مکانی و حتی الگوهای جستجو در وب، همگی می‌توانند جزئیات دقیقی از زندگی

[1] Big Data
[2] Machine Learning
[3] Data Analytics
[4] Privacy
[5] Sensitive Information

خصوصی افراد را فاش کنند. بنابراین، یک چالش اساسی شکل می‌گیرد: چگونه می‌توان از سودمندی[۶] آماری داده‌ها بهره برد، بدون آنکه حریم خصوصی مشارکت‌کنندگان در داده‌ها نقض شود؟

در سال‌های ابتدایی عصر اطلاعات، تصور عمومی بر این بود که حذف شناسه‌های صریح[۷] (مانند نام، کد ملی و شماره تلفن) برای محافظت از هویت افراد کافی است. این فرایند که گمنام‌سازی[۸] نامیده می‌شود، با این فرض انجام می‌شد که داده‌های باقی‌مانده قابلیت ردیابی به فرد خاصی را ندارند. اما پژوهش‌های متعددی نشان داده‌اند که این روش‌های سنتی در برابر حملات بازشناسایی[۹] به شدت آسیب‌پذیر هستند. در این نوع حملات، مهاجم با استفاده از اطلاعات جانبی[۱۰] یا اتصال پایگاه‌داده‌های مختلف به یکدیگر، موفق به کشف هویت افراد در داده‌های به ظاهر گمنام می‌شود.

چندین رخداد مشهور در دو دهه‌ی گذشته، ناکارآمدی روش‌های سنتی گمنام‌سازی را اثبات کرده‌اند:

- **داده‌های پزشکی ماساچوست:** در یکی از اولین و مشهورترین موارد، لاتانیا سوئینی نشان داد که می‌توان با ترکیب داده‌های پزشکی گمنام‌سازی شده (که نام بیماران از آن حذف شده بود) با فهرست عمومی رأی‌دهندگان، هویت افراد را بازشناسایی کرد. او با استفاده از ترکیب تاریخ تولد، جنسیت و کد پستی (که به آن‌ها شبه‌شناسه[۱۱] می‌گویند)، موفق شد پرونده پزشکی فرماندار وقت ایالت ماساچوست را شناسایی کند [۱۸].
- **مجموعه داده‌ی نتفلیکس**[۱۲]: شرکت نتفلیکس مجموعه‌ای از امتیازهای کاربران به فیلم‌ها را منتشر کرد که در آن شناسه‌های کاربری با اعداد تصادفی جایگزین شده بودند. پژوهشگران نشان دادند که با استفاده از اطلاعات عمومی موجود در وب‌سایت IMDb و تطبیق الگوهای امتیازدهی، می‌توان هویت بسیاری از کاربران را با دقت بالا کشف کرد [۱۵].
- **داده‌های جستجوی AOL:** در سال ۲۰۰۶، شرکت AOL تاریخچه‌ی جستجوی هزاران کاربر خود را منتشر کرد. اگرچه نام کاربران حذف شده بود، اما تحلیل محتوای جستجوها منجر به شناسایی هویت افراد شد (از جمله پرونده مشهور تلما آرنولد) که نشان داد حتی خودِ داده‌ها نیز می‌توانند به عنوان شناسه عمل کنند [۴].

این شواهد تجربی و نظری نشان می‌دهند که تعاریف هیوریستیک و روش‌های موردی (مانند حذف ستون‌ها یا مخدوش‌سازی ساده) نمی‌توانند تضمین امنیتی پایداری ارائه دهند. مهاجمان همواره می‌توانند

[6] Utility
[7] Explicit Identifiers
[8] Anonymization
[9] Re-identification Attacks
[10] Auxiliary Information
[11] Quasi-identifier
[12] Netflix Prize Data

دانش پس‌زمینه‌ی پیش‌بینی‌نشده‌ای داشته باشند که مکانیزم‌های سنتی را دور بزند. در نتیجه، نیاز مبرمی به یک چارچوب ریاضی دقیق احساس شد که بتواند حریم خصوصی را به صورت کمی تعریف کرده و تضمین دهد که ریسک افشای اطلاعات، مستقل از توان محاسباتی یا دانش جانبی مهاجم، همواره محدود باقی می‌ماند. این نیاز، زمینه را برای ظهور مفهوم محرمانگی تفاضلی فراهم کرد که در بخش‌های آتی به تفصیل به آن خواهیم پرداخت.

۱-۱-۱ محرمانگی تفاضلی (DP)

در پاسخ به چالش‌های امنیتی و ناکارآمدی روش‌های سنتی گمنام‌سازی، دوُرک و همکاران در سال ۲۰۰۶ مفهوم محرمانگی تفاضلی[۱۳] را معرفی کردند [۸]. این چارچوب ریاضی دقیق، به جای تمرکز بر ویژگی‌های ظاهری داده‌ها (مانند حذف نام‌ها)، بر فرایند تولید خروجی تمرکز دارد و تضمین می‌کند که حضور یا عدم حضور یک فرد خاص در پایگاه‌داده، تأثیر ناچیزی بر خروجی نهایی الگوریتم داشته باشد.

پیش از آنکه به تعاریف صوری و ریاضی بپردازیم، ضروری است که درک عمیقی از چیستی محرمانگی و تمایز بنیادین آن با مفاهیم امنیتی کلاسیک پیدا کنیم. بسیاری از سوتفاهم‌ها در این حوزه ناشی از تمایز ندادن دو مفهوم امنیت داده (که قلمرو رمزنگاری[۱۴] است) و محرمانگی داده (که هدف ماست) می‌باشد. رمزنگاری اساساً سازوکاری برای کنترل دسترسی[۱۵] است و تضمین می‌کند که تنها افراد مجاز می‌توانند داده‌ها را ببینند؛ اما در برابر نشت اطلاعات از خروجی‌های مجاز سکوت می‌کند. تصور کنید یک پایگاه‌داده‌ی حساس پزشکی کاملاً رمزنگاری شده باشد و پژوهشگری مجاز، نتیجه‌ی یک تحلیل آماری ساده (مانند میانگین حقوق یا نرخ یک بیماری) را منتشر کند. رمزنگاری هیچ محافظتی در برابر استنتاج‌های ثانویه ارائه نمی‌دهد و مهاجم می‌تواند با ترکیب این خروجی مجاز با دانش پیشین[۱۶] خود، اطلاعات خصوصی افراد را بازسازی کند. بنابراین، رمزنگاری شرط لازم است، اما برای حفظ محرمانگی کافی نیست؛ چرا که خودِ نتیجه‌ی تحلیل، حامل اطلاعات است.

در پاسخ به این چالش، دوُرک مفهوم محرمانگی را با ایده‌ی امنیت معنایی[۱۷] پیوند می‌زند. در این دیدگاه، هدف محرمانگی جلوگیری از یادگیری حقایق کلی درباره‌ی جامعه نیست، بلکه هدف محافظت از حقایق خاص مربوط به یک فرد مشخص است. فلسفه‌ی مرکزی این است که نتیجه‌ی هر تحلیلی باید تقریباً یکسان باشد، چه یک فرد خاص در آن مطالعه مشارکت کند و چه نکند. این تعریف، محرمانگی را به مفهوم ریسک گره می‌زند؛ به این معنا که مشارکت در یک پایگاه‌داده نباید باعث شود ریسک افشای رازهای یک

[13] Differential Privacy
[14] Cryptography
[15] Access Control
[16] Auxiliary Knowledge
[17] Semantic Security

فرد به طور چشم‌گیری افزایش یابد.

برای مدل‌سازی این مفهوم، می‌توانیم از استعاره‌ی جهان‌های موازی استفاده کنیم. دو پایگاه‌داده‌ی همسایه[18] (مانند $\mathcal{D}$ و $\mathcal{D}'$) را به عنوان دو جهان موازی در نظر بگیرید که در یکی، داده‌های کاربر x وجود دارد و در دیگری، این داده‌ها حذف یا تغییر یافته‌اند. هدف نهایی این است که از دیدگاه یک ناظر بیرونی (مهاجم)، این دو جهان غیرقابل تفکیک[19] باشند. اگر مکانیزم محرمانگی بتواند کاری کند که مهاجم با مشاهده‌ی خروجی، نتواند تشخیص دهد که این خروجی از کدام جهان آمده است، آنگاه حریم خصوصی کاربر x حفظ شده است.

برای رسیدن به این هدف، ما از الگوریتم‌های تصادفی[20] بهره می‌بریم. این الگوریتم‌ها با تزریق نویز کنترل‌شده، توزیع خروجی‌ها را بین دو مجموعه‌داده‌ی همسایه چنان به هم نزدیک می‌کنند که تمایز قائل شدن میان آن‌ها از نظر آماری ناممکن می‌شود.

چرا روش‌های قطعی شکست می‌خورند؟

دستیابی به هدف فوق با روش‌های قطعی[21] ممکن نیست. برای درک بهتر، یک حمله‌ی تفاضلی[22] کلاسیک را در نظر بگیرید. فرض کنید f تابعی قطعی است که میانگین درآمد n فرد در پایگاه‌داده را برمی‌گرداند ($f(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} x_i$). مهاجم می‌تواند دو پرس‌وجو انجام دهد:

۱. میانگین درآمد n نفر حاضر در پایگاه‌داده ($M_1 = f(\mathcal{D})$).

۲. میانگین درآمد همان افراد، به جز فرد هدف k ($M_2 = f(\mathcal{D} \setminus \{x_k\})$).

از آن‌جا که خروجی بدون نویز است، مهاجم با یک محاسبه‌ی ساده‌ی جبری ($x_k = n \cdot M_1 - (n-1) \cdot M_2$)، مقدار دقیق درآمد فرد k را به دست می‌آورد. این مثال نشان می‌دهد که هر تغییر کوچکی در ورودی یک تابع قطعی، به تغییری مشخص و قابل‌ردیابی در خروجی منجر می‌شود که بلافاصله دو جهان موازی را از هم متمایز می‌کند.

در واقع مثال میانگین را می‌توان به هر تابع قطعی $f : \mathcal{X}^n \to \mathcal{Z}$ تعمیم داد. فرض کنید مهاجم به دنبال بازیابی داده‌ی فرد kـام (x_k) است. اگر سایر داده‌های موجود در پایگاه‌داده، یعنی $\mathcal{D}_{-k} = \{x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n\}$ برای مهاجم شناخته شده باشند (فرضی که در تحلیل‌های

[18] Neighboring Database
[19] Indistinguishable
[20] Randomized Algorithms
[21] Deterministic
[22] Differencing Attack

بدبینانه‌ی محرمانگی تفاضلی استاندارد است)، تابع خروجی را می‌توان تنها بر حسب متغیر مجهول x_k به صورت $g(x) = f(x, \mathcal{D}_{-k})$ بازنویسی کرد.

اگر تابع g روی دامنه‌ی $\mathcal{X}$ یک‌به‌یک[۲۳] (یا حتی در بازه‌ای مشخص وارون‌پذیر) باشد، محرمانگی به طور کامل از بین می‌رود؛ زیرا مهاجم با مشاهده‌ی خروجی z، می‌تواند ورودی را به صورت $x_k = g^{-1}(z)$ بازیابی کند. حتی اگر g کاملاً وارون‌پذیر نباشد، مشاهده‌ی z فضای جستجوی مقادیر ممکن برای x_k را به شدت کاهش می‌دهد:

$$x_k \in \{x \in \mathcal{X} \mid g(x) = z\}$$

از دیدگاه نظریه اطلاعات، مشکل مکانیزم‌های قطعی این است که توزیع احتمال خروجی آن‌ها به ازای یک ورودی مشخص، یک جرم احتمالی[۲۴] تک‌نقطه‌ای (تابع دلتای دیراک) است. اگر دو پایگاه‌داده‌ی همسایه‌ی $\mathcal{D}$ و $\mathcal{D}'$ چنان باشند که $f(\mathcal{D}) \neq f(\mathcal{D}')$، آن‌گاه تکیه‌گاه[۲۵] توزیع‌های خروجی کاملاً مجزا خواهد بود. در نتیجه، واگرایی کولبک–لایبلر[۲۶] بین آن‌ها بی‌نهایت می‌شود:

$$D_{KL}(\mathcal{M}(\mathcal{D}) || \mathcal{M}(\mathcal{D}')) = \infty$$

این رابطه اثبات می‌کند که هیچ سطح محدودی از محرمانگی ($\varepsilon < \infty$) با توابع قطعیِ غیرثابت قابل دستیابی نیست. بنابراین، همان‌طور که در ادبیات موضوع تأکید شده است [۹]، برای شکستن این وابستگی قطعی و ایجاد ابهام آماری، تصادفی‌سازی[۲۷] در فرآیند مکانیزم الزامی است.

کاربردهای محرمانگی تفاضلی

این چارچوب ریاضی امروزه به استاندارد طلایی در تحلیل داده‌های حساس تبدیل شده و کاربردهای آن فراتر از آمارهای ساده رفته است. برخی از مهم‌ترین کاربردهای آن عبارتند از:

- **تخمین میانگین و مجموع**[۲۸]: اساسی‌ترین کاربرد DP در محاسبه‌ی آماره‌های توصیفی است. سازمان‌های آماری (مانند اداره سرشماری آمریکا) از این روش برای انتشار میانگین درآمد، سن یا جمعیت مناطق استفاده می‌کنند، بدون آنکه داده‌های فردی شهروندان به خطر بیفتد.
- **انتشار هیستوگرام**[۲۹]: بسیاری از تحلیل‌ها نیازمند دانستن توزیع داده‌ها هستند. DP اجازه می‌دهد

[23] Injective
[24] Probability Mass
[25] Support
[26] Kullback-Leibler Divergence
[27] Randomization
[28] Mean and Sum Estimation
[29] Histogram Release

تا تعداد افراد در هر بازه (مثلاً گروه‌های سنی یا درآمدی) با دقت بالا منتشر شود، در حالی که نویز اضافه شده مانع از شناسایی افراد در گروه‌های کم‌جمعیت می‌شود.

- **یادگیری ماشین خصوصی**[۳۰]: در آموزش مدل‌های عمیق، خطر به‌خاطرسپاری[۳۱] داده‌های آموزشی وجود دارد. با استفاده از الگوریتم‌هایی نظیر DP-SGD، می‌توان مدل‌هایی آموزش داد که الگوهای کلی را یاد می‌گیرند اما قادر به بازتولید داده‌های آموزشی حساس (مانند تصاویر چهره یا متون خصوصی) نیستند.
- **سیستم‌های توصیه‌گر و داده‌های مکانی:** شرکت‌های فناوری از DP برای جمع‌آوری آمارهای رفتاری (مانند پربازدیدترین وب‌سایت‌ها یا مکان‌های پرتردد) استفاده می‌کنند تا بدون ردیابی لحظه‌ای کاربران، کیفیت خدمات خود را بهبود بخشند (مانند مکانیزم RAPPOR در گوگل کروم).

این توضیحات، زیربنای اصلی تعاریف ریاضی دقیقی است که در فصل بعد به آن‌ها خواهیم پرداخت.

۱-۱-۲ محرمانگی تفاضلی موضعی (LDP)

اگرچه محرمانگی تفاضلی متمرکز (CDP) استاندارد طلایی حفاظت از داده‌ها محسوب می‌شود، اما پاشنه‌ی آشیل آن در فرضیه‌ی وجود یک متصدی مورد اعتماد[۳۲] نهفته است که به تمام داده‌های خام دسترسی دارد. این مدل در دنیای واقعی با چالش‌های امنیتی و حقوقی جدی روبروست؛ چرا که تجربه نشان داده است اعتماد کامل به سرورهای مرکزی، حتی در صورت مدیریت توسط نهادهای بزرگ فناوری، همواره در معرض تهدید است. یکی از این خطرات، نفوذهای خارجی و سرقت انبوه داده‌هاست؛ به طوری که حتی پیشرفته‌ترین دیوارهای آتش[۳۳] نیز در برابر حملات پیچیده آسیب‌پذیرند. در چنین شرایطی، اگر داده‌ها به صورت خام ذخیره شده باشند، نشت اطلاعاتی مانند آنچه در واقعه‌ی Equifax رخ داد، مکانیزم‌های محرمانگی تفاضلی مرکزی را عملاً بی‌فایده می‌کند؛ زیرا مهاجم با دور زدن مکانیزم، مستقیماً به مخزن داده‌های حساس دست می‌یابد[۱۹].

علاوه بر تهدیدهای خارجی، خطر سوءاستفاده‌های داخلی توسط کارمندان یا مدیران سیستم با دسترسی‌های سطح بالا نیز وجود دارد که امنیت داده‌ها را نه به ریاضیات، بلکه به اخلاق انسانی گره می‌زند. از سوی دیگر، محدودیت‌های حقوقی و احضاریه‌های قضایی نیز متصدی را ملزم به افشای اطلاعات می‌کند. در تمام این سناریوها، مدل متمرکز با یک نقطه شکست مرکزی[۳۴] روبروست.

[30] Private Machine Learning
[31] Memorization
[32] Trusted Curator
[33] Firewalls
[34] Single Point of Failure

گذار به مدل موضعی، حذف نیاز به اعتماد

بنابراین، تنها راه تضمین قطعی حریم خصوصی، اتخاذ رویکردی است که در آن متصدی اساساً به داده‌های اصلی دسترسی نداشته باشد. این ضرورت، نقطه‌ی عزیمت ما از مدل متمرکز به سمت چارچوب محرمانگی تفاضلی موضعی (LDP) است که در آن فرآیند خصوصی‌سازی پیش از خروج داده از دستگاه کاربر انجام می‌شود.

در این معماری، مرز اعتماد از سرور مرکزی به دستگاه شخصی کاربر (موبایل یا لپ‌تاپ) منتقل می‌شود. پروتکل به گونه‌ای طراحی می‌شود که هیچ‌کس، نه نفوذگران، نه کارمندان کنجکاو و نه حتی دولت‌ها، هرگز داده‌ی واقعی کاربر را مشاهده نکنند. سرور تنها نسخه‌هایی مخدوش و نویزدار از داده‌ها را دریافت می‌کند که به تنهایی بی‌معنی هستند، اما در تجمیع با تعداد زیادی داده‌ی دیگر، الگوهای آماری دقیق را آشکار می‌سازند. این رویکرد، خطر نقض حریم خصوصی را بسیار کنترل می‌کند.

۱-۲ کارهای پیشین و مرور ادبیات

۱-۲-۱ آغازگرها: از پیمایش‌های آماری تا تعریف مدرن محرمانگی

اگرچه نگرانی پیرامون محرمانگی داده‌ها قدمتی به اندازه خودِ آمار دارد، اما فرمول‌بندی ریاضی دقیق آن دستاورد قرن بیست و یکم است. ادبیات کلاسیک این حوزه با تلاش برای کنترل افشای آماری[۳۵] آغاز شد، اما ناکارآمدی روش‌های مبتنی بر گمنام‌سازی در برابر دانش پس‌زمینه مهاجم، نیاز به یک تعریف معنایی قوی‌تر را ایجاب کرد.

نقطه عطف این تحول، معرفی مفهوم **محرمانگی تفاضلی**[۳۶] توسط دِوُرک و همکاران بود [۸]. این تعریف، برخلاف روش‌های پیشین که بر ویژگی‌های داده تمرکز داشتند، بر ویژگی‌های مکانیزم پردازش داده تمرکز دارد. در مدل استاندارد (متمرکز)، یک مکانیزم تصادفی $\mathcal{M}$ دارای شرایط ε-DP است اگر برای هر دو پایگاه‌داده همسایه $\mathcal{D}$ و $\mathcal{D}'$ (که تنها در داده یک فرد متفاوت‌اند) و برای هر زیرمجموعه از خروجی‌ها $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$، رابطه زیر برقرار باشد:

$$\Pr[\mathcal{M}(\mathcal{D}) \in \mathcal{S}] \leqslant e^{\varepsilon} \cdot \Pr[\mathcal{M}(\mathcal{D}') \in \mathcal{S}] + \delta \qquad (۱-۱)$$

که در آن ε پارامتری کلیدی به نام **بودجه محرمانگی**[۳۷] است و δ احتمال شکست ناچیز مکانیزم را نشان

[35] Statistical Disclosure Control
[36] Differential Privacy
[37] Privacy Budget

می‌دهد [۹]. برای درک شهودی این مفهوم، می‌توان ε را به عنوان یک «پیچ تنظیم» برای کنترل توازن میان امنیت و مطلوبیت داده‌ها در نظر گرفت. این پارامتر تعیین می‌کند که خروجی مکانیزم تا چه حد اجازه دارد بین دو جهان موازی (جهانی با حضور داده‌ی شما و جهانی بدون آن) تمایز قائل شود:

- **مقادیر کوچک ε (محرمانگی قوی):** زمانی که $\varepsilon \to 0$، توزیع‌های خروجی برای دو پایگاه‌داده همسایه تقریباً بر هم منطبق می‌شوند. در این حالت، مکانیزم مجبور است نویز بسیار زیادی به پاسخ اضافه کند تا تفاوت‌ها را بپوشاند. در نتیجه، مهاجم تقریباً هیچ توانی برای تشخیص حضور فرد ندارد، اما در مقابل، دقت آماری خروجی کاهش می‌یابد.
- **مقادیر بزرگ ε (محرمانگی ضعیف):** با افزایش ε، مکانیزم آزادی عمل بیش‌تری دارد تا خروجی‌های متمایزتری تولید کند (نویز کمتر). این امر دقت تحلیل را افزایش می‌دهد، اما هم‌زمان ریسک بازشناسایی فرد و نشت اطلاعات خصوصی نیز به صورت نمایی بالا می‌رود.

همان‌طور که در بخش قبل توضیح داده شد، پیاده‌سازی این تعریف نیازمند یک پیش‌فرض قوی است: وجود یک متصدی مورد اعتماد که تمام داده‌های خام را جمع‌آوری کرده و نویز را به صورت مرکزی اعمال کند. اما دیدیم که این مدل دارای نقطه ضعف‌هایی است. حذف این فرض و انتقال اعتماد از سرور به کاربر، منجر به شکل‌گیری مفهوم **محرمانگی تفاضلی موضعی**[۳۸] **(LDP)** شد. اگرچه اصطلاح LDP و صورت‌بندی مدرن آن در سال‌های اخیر توسط پژوهشگرانی نظیر کاسی‌یسواناتان و دیگران تدوین شد [۱۳]، اما ریشه‌های عملی آن به دهه‌ها قبل باز می‌گردد.

تعریف ۱-۱ (محرمانگی تفاضلی موضعی (α,δ)-LDP) *یک مکانیزم تصادفی $\mathcal{M} : \mathcal{X} \to \mathcal{Z}$ (تصادفی‌ساز موضعی) شرط «محرمانگی تفاضلی موضعی تقریبی» یا (α, δ)-LDP را برآورده می‌کند، اگر برای تمام جفت ورودی‌های ممکن $x, x' \in \mathcal{X}$ و هر زیرمجموعه‌ی خروجی $\mathcal{S} \subseteq \mathcal{Z}$، رابطه زیر برقرار باشد:*

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leqslant e^{\alpha} \cdot \Pr[\mathcal{M}(x') \in \mathcal{S}] + \delta \qquad (۲-۱)$$

در این تعریف:

- *α، بودجه محرمانگی است که میزان شباهت توزیع‌های خروجی را کنترل می‌کند.*
- *δ، احتمال ناچیز شکست مکانیزم در برقراری شرط محرمانگی است.*

اگر $\delta = 0$ باشد، تعریف به حالت استاندارد یا «محرمانگی تفاضلی موضعی خالص» (α-LDP) باز می‌گردد.

[38] Local Differential Privacy

مثال‌هایی از مکانیزم و کاربرد LDP

در واقع، ساده‌ترین و نخستین نمونه از یک مکانیزم LDP، روش «پاسخ تصادفی»[39] است که توسط وارنر در سال ۱۹۶۵ برای حذف سوگیری در نظرسنجی‌های حساس معرفی شد [۲۲]. وارنر این روش را نه برای حفاظت در برابر حملات سایبری، بلکه برای تشویق پاسخ‌دهندگان به صداقت در سوالات حساس (مانند مصرف مواد مخدر یا عقاید سیاسی خاص) طراحی کرد.

سازوکار کلاسیک این روش برای یک پرسش با پاسخ «بله/خیر» به صورت زیر است: فرض کنید از کاربر i خواسته می‌شود که ویژگی حساس $X_i \in \{۰, ۱\}$ را گزارش کند. کاربر به جای پاسخ مستقیم، طبق دستورالعمل زیر عمل می‌کند:

۱. یک سکه را پرتاب می‌کند. (می‌تواند سکه غیرمنصفانه[40] باشد)

۲. اگر سکه «شیر» آمد، پاسخ واقعی (X_i) را گزارش می‌کند.

۳. اگر سکه «خط» آمد، یک پاسخ تصادفی (با پرتاب سکه‌ی دوم) تولید و گزارش می‌کند.

در این سناریو، حتی اگر سرور پاسخ «بله» را دریافت کند، با قطعیت نمی‌داند که آیا کاربر واقعاً دارای ویژگی X بوده است (شیر آمده) یا صرفاً به دلیل تصادف (خط آمدن سکه‌ی اول و شیر آمدن سکه‌ی دوم) این پاسخ را ارسال کرده است. با این حال، از آن‌جایی که احتمالات سکه‌ها مشخص است، سرور می‌تواند با جمع‌آوری تعداد زیادی از پاسخ‌ها (n بسیار بزرگ)، اثر نویز را به صورت آماری حذف کرده و توزیع واقعی جامعه را با خطا تخمین بزند. به زبان ریاضی مدرن، اگر احتمال گزارش پاسخ واقعی p باشد، نسبت احتمال خروجی‌ها برای دو ورودی متفاوت x و x' به صورت زیر محدود می‌شود:

$$\frac{\Pr[\mathcal{M}(x) = z]}{\Pr[\mathcal{M}(x') = z]} \leqslant \frac{p}{۱ - p} \tag{۳-۱}$$

این رابطه دقیقاً منطبق بر تعریف α-LDP است و نشان می‌دهد که بودجه محرمانگی α چگونه مستقیماً از پارامترهای مکانیزم (p) مشتق می‌شود:

$$e^{\alpha} = \frac{p}{۱ - p} \quad \Rightarrow \quad \alpha = \ln\left(\frac{p}{۱ - p}\right) \tag{۴-۱}$$

این فرمول، تفسیر شهودی LDP را کامل می‌کند:

- اگر $p \approx ۰/۵$ (سکه کاملاً تصادفی)، آنگاه $\alpha \approx ۰$ می‌شود. یعنی خروجی هیچ اطلاعاتی از ورودی ندارد (محرمانگی کامل، اما بدون فایده آماری).

[39] Randomized Response (RR)
[40] Unfair

- اگر $p \to ۱$ (پاسخ تقریباً همیشه راست)، آنگاه $\alpha \to \infty$ می‌شود. یعنی داده‌ها دقیق هستند اما هیچ محرمانگی وجود ندارد.

بنابراین، کار وارنر را می‌توان سنگ‌بنای تاریخی این حوزه دانست که نشان داد چگونه می‌توان بدون اعتماد به گیرنده پیام، و با تنظیم دقیق پارامتر p (و در نتیجه α)، اطلاعات آماری مفیدی را مخابره کرد.

اما پاسخ تصادفی تنها راهکاری برای ایجاد محرمانگی در داده‌های دودویی است، و در کاربردهای مدرن با چالش دامنه‌ی بسیار بزرگ[۴۱] روبروست. شرکت‌های بزرگ فناوری نیاز دارند داده‌هایی نظیر «آدرس‌های اینترنتی بازدید شده» یا «کلمات جدید تایپ‌شده» را جمع‌آوری کنند که دامنه‌ی آن‌ها($\mathcal{X}$) می‌تواند شامل میلیون‌ها حالت باشد. اعمال مستقیم RR در این حالات منجر به نویز بسیار زیاد و کاهش شدید سودمندی می‌شود. در ادامه، راهکارهای اتخاذ شده توسط بزرگ‌ترین شرکت‌های فناوری را مرور می‌کنیم:

- **گوگل و پروتکل RAPPOR:** در سال ۲۰۱۴، گوگل برای جمع‌آوری آمار تنظیمات مرورگر کروم و شناسایی بدافزارها، پروتکل RAPPOR[۴۲] را معرفی کرد [۲۰]. چالش اصلی گوگل، جمع‌آوری رشته‌های متنی[۴۳] بود. راه‌حل آن‌ها ترکیب پاسخ تصادفی با فیلترهای بلوم[۴۴] بود. در این روش، داده‌ی ورودی ابتدا به یک بردار بیتی (با استفاده از توابع درهم‌ساز) نگاشت می‌شود و سپس پاسخ تصادفی روی تک‌تک بیت‌های این فیلتر اعمال می‌گردد. این معماری به گوگل اجازه داد تا بدون دانستن ورودی دقیق هر کاربر، الگوهای پرتکرار و ناهنجاری‌ها را در مقیاس میلیونی شناسایی کند.

- **اپل و جمع‌آوری داده‌های دایره‌لغات:** شرکت اپل از LDP برای بهبود کیبورد QuickType، شناسایی ایموجی‌های پرطرفدار و داده‌های سلامت در سیستم‌عامل‌های iOS و macOS استفاده می‌کند. مسئله‌ی اپل، مخابره‌ی کارآمد داده‌ها با حفظ حریم خصوصی بود. راه‌حل اپل استفاده از تکنیک‌های مبتنی بر طرح‌ریزی[۴۵] و تبدیل‌های ریاضی مانند تبدیل هادامارد[۴۶] است. این تبدیل‌ها به مکانیزم اجازه می‌دهند که اطلاعات را در ابعاد پایین‌تر فشرده کند تا هم بار ارتباطی کاهش یابد و هم واریانس تخمین‌گر در دامنه‌های بزرگ کنترل شود [۲۰].

- **مایکروسافت و داده‌های تله‌متری:** مایکروسافت برای جمع‌آوری داده‌های تله‌متری ویندوز (مانند مدت زمان استفاده از برنامه‌ها) با چالش تخمین هیستوگرام‌های پیوسته روبرو بود. آن‌ها از مکانیزم‌هایی نظیر نمونه‌برداری هیستوگرام و روش‌های تکرارکننده برای بازسازی توزیع داده‌ها

[41] High-Dimensional Domain
[42] Randomized Aggregatable Privacy-Preserving Ordinal Response
[43] String
[44] Bloom Filters
[45] Sketching
[46] Hadamard Transform

استفاده کردند. تمرکز اصلی در این‌جا، ایجاد تعادل بین دقت آماری در جمع‌آوری داده‌های سیستمی و عدم امکان بازشناسایی رفتار یک کاربر خاص در طول زمان است.

این نمونه‌ها نشان می‌دهند که محرمانگی تفاضلی موضعی (LDP) تنها یک مفهوم نظری نیست، بلکه یک ابزار حیاتی مهندسی است که با استفاده از تکنیک‌های پیشرفته‌ی آماری برای حل مسائل دنیای واقعی مقیاس‌دهی شده است.

۱-۲-۲ دسته‌بندی پروتکل‌های موضعی: تعاملی و غیرتعاملی

برای بالا بردن دقت کارهای انجام‌شده لازم است به دسته‌بندی پروتکل‌های LDP بر اساس «معماری ارتباطی» اشاره کنیم. پژوهش‌های انجام شده در این حوزه، مکانیزم‌ها را به دو دسته‌ی کلی تقسیم می‌کنند:

۱. **پروتکل‌های غیرتعاملی**[۴۷]: در این حالت، تمام کاربران به صورت هم‌زمان و مستقل عمل می‌کنند. هر کاربر i مکانیزم $\mathcal{M}$ را تنها بر اساس داده‌ی خودش X_i اجرا کرده و پیام Z_i را به سرور می‌فرستد. هیچ ارتباطی بین کاربران وجود ندارد و سرور نیز هیچ بازخوردی به کاربران نمی‌دهد. به دلیل سادگی پیاده‌سازی و مقیاس‌پذیری بالا، اکثر پروتکل‌های صنعتی (مانند RAPPOR گوگل یا سیستم‌های اپل) در این دسته قرار می‌گیرند.

۲. **پروتکل‌های تعاملی**[۴۸]: در این روش، کاربران به صورت متوالی با سرور ارتباط برقرار می‌کنند. کاربر i می‌تواند قبل از ارسال داده‌ی خود، خلاصه‌ای از داده‌های کاربران قبلی $(Z_1, \ldots, Z_{i-1})$ را از سرور دریافت کند و نویز خود را هوشمندانه‌تر تنظیم نماید. اگرچه به نظر می‌رسد این آزادی عمل باید دقت را افزایش دهد، اما دوچی و همکاران در نتایج حیرت‌انگیزی نشان دادند که برای دسته‌ی بزرگی از توابع محدب (مانند تخمین میانگین)، پروتکل‌های تعاملی هیچ مزیتی نسبت به روش‌های غیرتعاملی ندارند و نرخ مینی‌مکس را بهبود نمی‌بخشند [۶، ۱۱].

۱-۲-۳ چالش سودمندی و موازنه دقت‌ـ محرمانگی

اگرچه حذف متصدی مرکزی در مدل LDP، تضمین‌های امنیتی بسیار قوی‌تری را فراهم می‌کند، اما این امنیت رایگان به دست نمی‌آید. چالش بنیادین در این رویکرد، کاهش چشم‌گیر **سودمندی**[۴۹] داده‌ها یا همان دقت تحلیل‌های آماری است. این پدیده تحت عنوان **موازنه محرمانگی‌ـ دقت**[۵۰] شناخته می‌شود.

[47] Non-interactive / Simultaneous
[48] Interactive / Sequential
[49] Utility
[50] Privacy-Accuracy Trade-off

برای درک شهودی این چالش، مقایسه نحوه اعمال نویز در دو مدل ضروری است:

- **در مدل متمرکز (CDP):** نویز تنها «یک‌بار» و پس از تجمیع داده‌ها به نتیجه نهایی اضافه می‌شود. از آن‌جا که مجموع (یا میانگین) داده‌ها حساسیت کمی دارد، مقدار نویز معمولاً مستقل از تعداد کاربران (n) و بسیار کوچک است.
- **در مدل موضعی (LDP):** نویز باید به «تک‌تک» داده‌های ورودی اضافه شود (پیش از آنکه از دستگاه کاربر خارج شوند). وقتی تحلیل‌گر قصد دارد میانگین این داده‌ها را محاسبه کند، واریانس نویزهای n کاربر با هم جمع می‌شود.

این انباشت نویز باعث می‌شود که **نسبت سیگنال به نویز**[51] در مدل موضعی بسیار پایین‌تر از مدل متمرکز باشد. به بیان دیگر، برای دستیابی به همان سطح از دقت که در مدل متمرکز وجود دارد، در مدل LDP نیازمند تعداد بسیار بیش‌تری نمونه داده هستیم.

این مسئله در کاربردهای عملی بسیار حائز اهمیت است. برای مثال، اگر هدف تخمین فراوانی یک بیماری نادر باشد، نویز اضافه شده توسط مکانیزم‌های LDP ممکن است سیگنال اصلی را کاملاً بپوشاند. همین چالش بود که پژوهشگران را بر آن داشت تا به جای استفاده از روش‌های ساده (مثل وارنر)، به دنبال پاسخ این پرسش باشند که: «آیا می‌توان مکانیزم‌هایی طراحی کرد که با کم‌ترین میزان نویز، بیش‌ترین محرمانگی را فراهم کنند؟» و «حد نهایی این دقت کجاست؟» این پرسش‌ها زمینه را برای ورود تئوری‌های پیشرفته‌تر نظیر «تخمین مینی‌مکس» فراهم کرد.

۱-۲-۴ نگاهی آماری به LDP: چارچوب مینی‌مکس و حدود بنیادین

پاسخ به پرسش بالا، مسیر پژوهش‌های این حوزه را به سمت **نظریه مینی‌مکس آماری**[52] تغییر داد. نقطه عطف این تحول، سلسله مقالات جریان‌ساز دوچی، جردن و وین‌رایت[53] بود [۶، ۷]. آن‌ها با صورت‌بندی مسئله در قالب نظریه اطلاعات، نشان دادند که هزینه محرمانگی در مدل موضعی بسیار سنگین و غیرقابل اجتناب است.

در تحلیل مینی‌مکس، هدف یافتن «ریسک مینی‌مکس» ($\mathfrak{M}_n$) است؛ یعنی کم‌ترین «زیان مورد انتظاری» (MSE) که بهترین تخمین‌گر ممکن در بدترین توزیع داده مرتکب می‌شود. دوچی و همکاران با استفاده

[51] Signal-to-Noise Ratio (SNR)
[52] Statistical Minimax Theory
[53] Duchi, Jordan, and Wainwright

از ابزارهایی نظیر **نامساوی فانو**[۵۴] و **لم اسود**[۵۵] (که در فصل سوم به تفصیل بررسی خواهند شد)، ثابت کردند که رفتار مجانبی ریسک در مدل LDP تفاوتی بنیادین با مدل متمرکز دارد.

به طور مشخص، برای n کاربر و بودجه محرمانگی α، ریسک مینی‌مکس (واریانس خطا) از رابطه‌ی زیر پیروی می‌کند:

$$\mathfrak{M}_{\text{CDP}} \asymp \frac{۱}{n^۲\epsilon^۲} \quad ، \text{در حالی که} \quad \mathfrak{M}_{\text{LDP}} \asymp \frac{۱}{n\alpha^۲} \tag{۵-۱}$$

این نتیجه که نرخ هم‌گرایی بهینه در مدل موضعی را تعیین می‌کند، حاوی دو پیام مهم است:

۱. **کندی هم‌گرایی:** در حالی که خطای مدل متمرکز با سرعت $۱/n$ کاهش می‌یابد، خطای مدل موضعی با سرعت بسیار کندتر $۱/\sqrt{n}$ کم می‌شود.

۲. **اندازه نمونه مؤثر:** ضریب $\alpha^۲$ نشان می‌دهد که هر نمونه داده‌ی خصوصی‌سازی شده، عملاً حاوی اطلاعاتی معادل با $\alpha^۲$ نمونه داده‌ی خام است (برای $\alpha < ۱$). این یعنی برای جبران نویز α-LDP، حجم داده‌ها باید با ضریب $۱/\alpha^۲$ افزایش یابد.

پس از استقرار این چارچوب نظری، تمرکز جامعه علمی بر طراحی مکانیزم‌های بهینه[۵۶] قرار گرفت که بتوانند به این کران‌های نظری دست یابند. از جمله مهم‌ترین این تلاش‌ها می‌توان به معرفی مکانیزم‌های پله‌ای[۵۷] توسط کایروز و همکاران [۱۱] و توسعه پروتکل‌های پیشرفته‌ای نظیر کدگذاری یگانی[۵۸] و هشینگ محلی بهینه[۵۹] توسط وانگ و همکاران [۲۰] اشاره کرد. این روش‌ها تلاش می‌کنند با بهینه‌سازی ساختار نویز و استفاده از تکنیک‌های فشرده‌سازی اطلاعات، فاصله بین عملکرد عملی و حدود نظری مینی‌مکس را به حداقل برسانند.

۱-۳ بیان مسئله و اهداف پژوهش

محرمانگی تفاضلی موضعی (LDP) به عنوان یک حوزه‌ی میان‌رشته‌ای، محل تلاقی «علوم کامپیوتر» (با تمرکز بر طراحی الگوریتم و امنیت) و «آمار ریاضی» (با تمرکز بر نظریه تخمین و مینی‌مکس) است. این

[54] Fano's Inequality
[55] Assouad's Lemma
[56] Optimal Mechanisms
[57] Staircase Mechanisms
[58] Unary Encoding (UE)
[59] Optimal Local Hashing (OLH)

ماهیت دوگانه، اگرچه باعث غنای ادبیات موضوع شده، اما منجر به پراکندگی قابل‌توجهی در روش‌ها و ابزارهای تحلیلی گشته است.

همان‌طور که در مرور ادبیات اشاره شد، چارچوب مینی‌مکس که توسط دوچی و همکاران [۶] پایه‌گذاری شده، نشان می‌دهد که اعمال محدودیت محرمانگی منجر به کاهش نرخ هم‌گرایی در تخمین‌های آماری می‌شود. با این حال، اثبات این نتایج در مقالات مختلف با ابزارهای متفاوتی صورت گرفته است. برای مثال:

- در برخی مسائل تخمین چگالی، پژوهشگران عمدتاً از **واگرایی کولبک‌ـ‌لایبلر (KL)** و نامساوی فانو استفاده کرده‌اند.
- در مسائل آزمون فرض ساده، اغلب از **فاصله‌ی تغییرات کل (TV)** و لم لوکام بهره گرفته شده است.
- در تحلیل‌های اخیرتر، **فاصله‌ی کای‌ـ‌دو** (χ^2) به دلیل رفتار هموارتر در همسایگی صفر و ارتباط مستقیم با واریانس، مورد توجه قرار گرفته است.

در این پایان‌نامه تلاش شده است تا با گردآوری و بررسی جامع مطالعات پیشین، تحلیلی تطبیقی میان رویکردهای مختلف صورت گیرد و نتایج آن‌ها در چارچوبی منسجم و یک‌پارچه ارائه شود.

۱-۳-۱ رویکرد تحلیل: fـ‌واگرایی‌ها به عنوان زبان مشترک

برای غلبه بر چالش پراکندگی و ایجاد یکپارچگی نظری، این پایان‌نامه پیشنهاد می‌کند که تمام تحلیل‌ها بر مبنای خانواده‌ی عمومی **fـ‌واگرایی‌ها**[۶۰] بازنویسی و تفسیر شوند. fـ‌واگرایی‌ها (معرفی شده توسط سیزار[۶۱])، کلاسی جامع از معیارهای فاصله بین دو توزیع احتمال P و Q هستند:

$$D_f(P\|Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right] \qquad (۶-۱)$$

که در آن f یک تابع محدب با ویژگی $f(۱) = ۰$ است.

انتخاب این رویکرد به ما اجازه می‌دهد تا «محرمانگی» را نه صرفاً به عنوان یک ویژگی الگوریتمی، بلکه به عنوان یک **محدودیت هندسی** تفسیر کنیم. در این دیدگاه، هر مکانیزم LDP مانند یک «کانال انقباضی»[۶۲] عمل می‌کند که فاصله‌ی بین توزیع‌های ورودی را فشرده می‌سازد. هدف ما در این پژوهش،

[60] f-divergences
[61] Csiszár
[62] Contraction Channel

بررسی این است که چگونه ادبیات پیشرو، مفهوم «ضریب انقباض»[۶۳] را برای fـ واگرایی‌های مختلف محاسبه کرده و از آن برای استخراج کران‌های مینی‌مکس استفاده می‌کنند. این دیدگاه هندسی، پلی میان اثبات‌های پراکنده ایجاد می‌کند و نشان می‌دهد که انتخاب f مناسب، تابعی از هندسه‌ی مسئله است.

۱-۳-۲ رویکرد تحلیل: چارچوب انقباض fـ واگرایی‌ها

برای غلبه بر چالش پراکندگی مفاهیم و ارائه یک تحلیل منسجم، این پایان‌نامه بر مرور و تشریح رویکرد نوینی تمرکز دارد که در سال‌های اخیر توسط **آسوده، علی‌اکبری و کالمون** [۲] و **آسوده و ژانگ** [۳] توسعه یافته است. این پژوهشگران نشان داده‌اند که قیود محرمانگی تفاضلی موضعی (LDP) را می‌توان به‌صورت دقیق و ریاضی معادل با «خواص انقباضی»[۶۴] خانواده‌ی fـ واگرایی‌ها فرمول‌بندی کرد.

به‌طور مشخص، در این رویکرد اثبات می‌شود که شرط LDP معادل است با کران‌دار بودن ضریب انقباض برای «واگرایی هاکی‌ـ استیک» (E_γ-divergence). همچنین، تحلیل‌های دقیق‌تر نشان می‌دهند که چگونه می‌توان بهترین کران‌های انقباض را برای واگرایی‌های مهمی نظیر χ^2 و KL استخراج کرد. این دیدگاه هندسی، هر مکانیزم محرمانگی را به مثابه‌ی یک کانال نویزی می‌بیند که فاصله بین توزیع‌های احتمالی ورودی را کاهش می‌دهد. در این پایان‌نامه، ما تلاش می‌کنیم تا با مطالعه‌ی دقیق این مراجع، نشان دهیم چگونه این ضرایب انقباض (η_f) محاسبه می‌شوند و چگونه می‌توان از آن‌ها برای بازتولید و بهبود کران‌های مینی‌مکس در مسائل تخمین و آزمون فرض استفاده کرد.

۱-۳-۳ اهداف و ساختار پژوهش

هدف اصلی این پایان‌نامه، ارائه‌ی یک «مرور تحلیلی و آموزشی» از ادبیات مدرن نظریه مینی‌مکس تحت محدودیت محرمانگی است. ما با محوریت قرار دادن کارهای انجام‌شده توسط محققان نام‌برده، اهداف زیر را دنبال می‌کنیم:

۱. **تبیین هم‌ارزی LDP و انقباض E_γ:** مرور و بازنویسی اثبات‌های ریاضی که نشان می‌دهند محرمانگی تفاضلی موضعی دقیقاً معادل با انقباض در واگرایی هاکی‌ـ استیک است. این بخش بر اساس نتایج [۲] تنظیم شده و پایه‌ای برای تعمیم نتایج به سایر واگرایی‌ها فراهم می‌کند.

۲. **بررسی ضرایب انقباض برای واگرایی‌های χ^2 و KL:** مطالعه‌ی کران‌های دقیق[۶۵] برای ضریب

[63] Contraction Coefficient
[64] Contraction Properties
[65] Sharp Bounds

انقباض مکانیزم‌های LDP تحت واگرایی‌های کای‌ـدو و کولبک‌ـلایبلر. این بخش با تمرکز بر نتایج [۳]، نشان می‌دهد که چگونه می‌توان «نامساوی‌های پردازش داده قوی»[۶۶] را برای محیط‌های خصوصی استخراج کرد.

۳. **یک‌پارچه‌سازی روش‌های کران پایین:** تشریح اینکه چگونه ابزارهای انقباضی فوق، در ترکیب با متدهای کلاسیک آماری نظیر «لوکام»، «فانو» و «اسود»، منجر به استخراج حدود بهینه برای خطای تخمین مینی‌مکس می‌شوند. هدف ما نشان دادن این موضوع است که چگونه این چارچوب واحد، نتایج پراکنده در ادبیات را پوشش می‌دهد.

۴. **جمع‌بندی و مقایسه نتایج:** انجام مروری ساختاریافته بر کاربردهای این چارچوب در مسائلی نظیر «تخمین میانگین»، «تخمین چگالی» و «آزمون فرض»، و مقایسه نتایج حاصل از رویکرد انقباضی با سایر روش‌های موجود در ادبیات.

۱-۴ ساختار پایان‌نامه

ساختار ادامه‌ی این پایان‌نامه به شرح زیر سازمان‌دهی شده است:

در فصل دوم، مفاهیم بنیادی و ابزارهای ریاضی مورد نیاز برای تحلیل‌های نظری پیش رو معرفی می‌شوند. این فصل ابتدا به مرور تعاریف پایه در نظریه احتمال و اندازه می‌پردازد و سپس خانواده‌ی واگرایی‌های f (شامل واگرایی کولبک‌ـلایبلر و فاصله تغییرات کل) را به عنوان ابزاری برای سنجش شباهت توزیع‌های احتمال تعریف می‌کند. در ادامه، مبانی آزمون فرض آماری و نامساوی‌های اطلاعاتی مهم نظیر لم لی‌کم[۶۷] و لم فانو[۶۸] بررسی می‌شوند. بخش پایانی این فصل به معرفی مدل کلاسیک محرمانگی تفاضلی متمرکز (CDP) و ویژگی‌های آن اختصاص دارد.

فصل سوم به معرفی دقیق مدل محرمانگی تفاضلی موضعی (LDP) می‌پردازد. در این فصل، پس از ارائه تعریف و بررسی تفاوت‌های بنیادین آن با مدل متمرکز، مکانیزم‌های استاندارد این حوزه نظیر پاسخ تصادفی[۶۹] و خواص آن‌ها تشریح می‌شوند. هدف این فصل ایجاد درکی عمیق از محدودیت‌ها و الزامات طراحی مکانیزم در فضای موضعی است.

فصل چهارم به تحلیل کران‌های پایین خطر مینی‌مکس[۷۰] در مسائل تخمین آماری تحت قید محرمانگی

[66] Strong Data Processing Inequalities (SDPI)
[67] Le Cam's Lemma
[68] Fano's Lemma
[69] Randomized Response
[70] Minimax Risk

اختصاص دارد. در این فصل نشان داده می‌شود که چگونه می‌توان با ترکیب ابزارهای نظریه اطلاعات و محدودیت‌های محرمانگی، حدود بنیادین دقت تخمین‌گرها را تعیین کرد.

در فصل پنجم، چارچوب تحلیل خود را با استفاده از واگرایی‌های f تعمیم می‌دهیم. در این فصل ارتباط میان محرمانگی تفاضلی موضعی و مفهوم ضریب انقباض[۷۱] در واگرایی‌ها بررسی شده و نتایج جدیدی در خصوص بهینگی مکانیزم‌ها ارائه می‌گردد.

در نهایت، فصل ششم به جمع‌بندی دستاوردها، نتیجه‌گیری کلی و ارائه پیشنهادهایی برای پژوهش‌های آتی اختصاص دارد.

[71] Contraction Coefficient

فصل ۲

پیش‌نیازها

۲-۱ محرمانگی تفاضلی متمرکز (CDP)

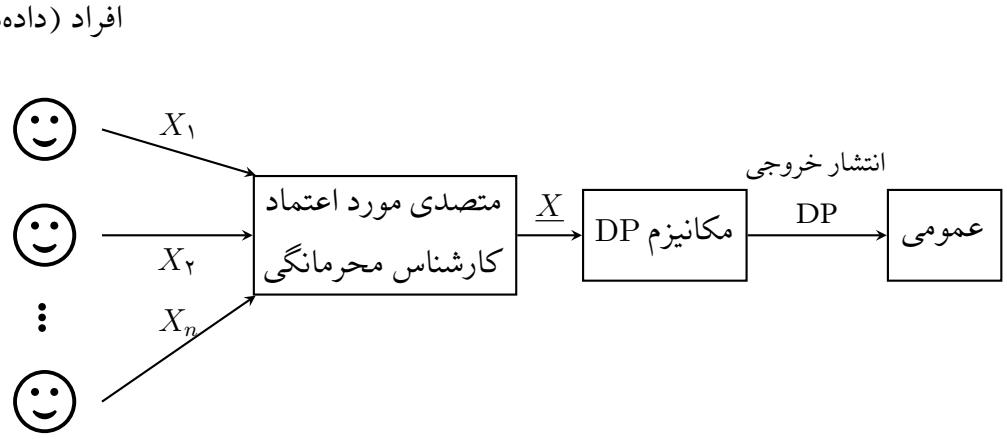
مفهوم محرمانگی تفاضلی[۱] یا به اختصار DP، اولین بار توسط دِوُرک و همکاران [۸] معرفی شد و به سرعت به استاندارد طلایی برای حفظ حریم خصوصی در تحلیل داده‌ها تبدیل گشت. این چارچوب، یک تعریف ریاضی قوی از حریم خصوصی ارائه می‌دهد که مبتنی بر پنهان‌سازی حضور یا عدم حضور یک فرد خاص در مجموعه‌داده است.

۲-۱-۱ مدل اعتماد و تعریف رسمی

در مدل متمرکز[۲]، فرض بر این است که یک متصدی مورد اعتماد[۳] وجود دارد. تمام افراد داده‌های خام و حساس خود را در اختیار این متصدی قرار می‌دهند (شکل ۲-۱ را ببینید). متصدی، مجموعه‌داده‌ی کامل $\mathcal{D}$ را در اختیار دارد. وظیفه‌ی متصدی این است که با اجرای یک مکانیزم تصادفی[۴] $\mathcal{M}$ بر روی مجموعه‌داده‌ی $\mathcal{D}$، نتایجی (مثلاً پاسخ به یک پرس‌وجو[۵]) را به صورت عمومی منتشر کند، به طوری که اطلاعات حساس افراد فاش نشود.

تعریف ۲-۱ (جهان داده‌ها و پایگاه‌داده) *مجموعه‌ی تمام مقادیر ممکن برای یک رکورد داده را «جهان*

[1] Differential Privacy
[2] Centralized
[3] Trusted Curator
[4] Randomized Mechanism
[5] Query



شکل ۲-۱: مدل محرمانگی تفاضلی متمرکز (CDP) با یک متصدی مورد اعتماد.

داده‌ها»[۶] می‌نامیم و آن را با $\mathcal{X}$ نمایش می‌دهیم (در ادامه‌ی پایان‌نامه نیز هرجا از $\mathcal{X}$ استفاده شد منظور جهان داده‌هاست). یک پایگاه‌داده‌ی $\mathcal{D}$ مجموعه‌ای از رکوردهاست که اعضای آن از $\mathcal{X}$ انتخاب شده‌اند. در ادبیات محرمانگی تفاضلی، پایگاه‌داده معمولاً به صورت یک بردار (هیستوگرام) $x \in \mathbb{N}^{|\mathcal{X}|}$ نمایش داده می‌شود که در آن هر مولفه x_i نشان‌دهنده‌ی تعداد تکرار عنصر iـ ام از $\mathcal{X}$ در پایگاه‌داده است.

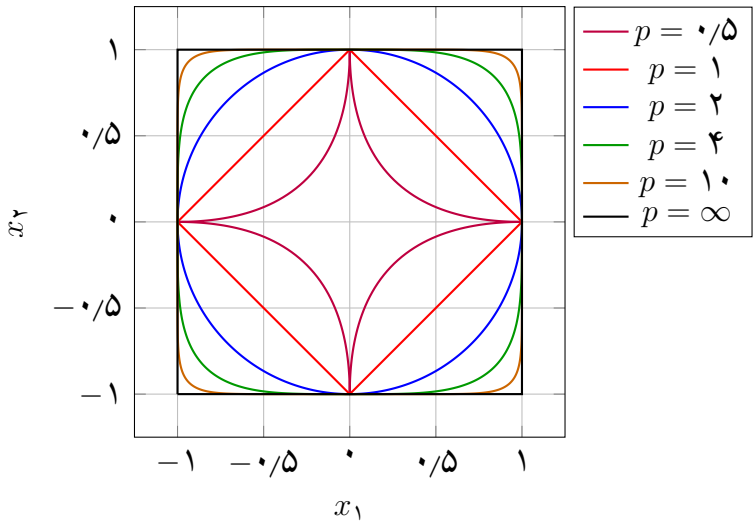
مثال ۲-۱ *فرض کنید می‌خواهیم وضعیت اشتغال افراد را بررسی کنیم. در اینجا جهان داده‌ها برابر است با $\mathcal{X} = \{$بیکار, شاغل$\}$. اگر در یک پایگاه‌داده ۳ نفر شاغل و ۱ نفر بیکار باشند، نمایش هیستوگرامی پایگاه‌داده $\mathcal{D}$ به صورت بردار زیر خواهد بود:*

$$\mathcal{D} = (۳, ۱)$$

تعریف ۲-۲ (الگوریتم تصادفی) *یک الگوریتم (مکانیزم) تصادفی $\mathcal{M}$ تابعی است که دامنه‌ی آن مجموعه‌ی تمام پایگاه‌داده‌های ممکن و برد آن مجموعه‌ی خروجی‌های ممکن $\mathcal{R}$ است. برخلاف الگوریتم‌های قطعی، خروجی $\mathcal{M}$ برای یک ورودی ثابت $\mathcal{D}$، یک متغیر تصادفی است. به عبارت دیگر، $\mathcal{M}$ یک توزیع احتمال روی $\mathcal{R}$ ایجاد می‌کند.*

مثال ۲-۲ *فرض کنید تابعی داریم که تعداد افراد بیمار را می‌شمارد. یک مکانیزم تصادفی ساده می‌تواند به این صورت باشد: «تعداد واقعی بیماران را بشمار و سپس نتیجه‌ی پرتاب یک سکه (۰ یا ۱) را به آن اضافه کن». در این حالت، خروجی یا تعداد واقعی بیماران است و یا یک عدد بیش‌تر از آن.*

[6] Data Universe



شکل ۲-۲: تجسم هندسی گوی یکه در فضای l_p دوبعدی برای مقادیر مختلف p. همان‌طور که در شکل پیداست، به ازای $p \geqslant ۱$ گوی‌ها تشکیل مجموعه‌هایی محدب می‌دهند، اما برای مقادیر $p < ۱$ (مانند $p = ۰/۵$) خاصیت تحدب از بین می‌رود. حالت $p = ۲$ نمایانگر فضای استاندارد اقلیدسی و حالت $p = ۱$ نمایانگر فاصله‌ی منهتن (مجموع قدرمطلق‌ها) است.

تعریف ۲-۳ (فاصله و نرم‌های ℓ_p) *برای سنجش میزان تفاوت دو پایگاه‌داده، از مفهوم نرم ℓ_p استفاده می‌شود. در حالت کلی برای $p \geqslant ۱$، فاصله‌ی ℓ_p بین دو پایگاه‌داده‌ی $\mathcal{D}_۱$ و $\mathcal{D}_۲$ (با بردارهای تکرار متناظر) به صورت زیر تعریف می‌گردد:*

$$\|\mathcal{D}_۱ - \mathcal{D}_۲\|_p = \left(\sum_{i=۱}^{|\mathcal{X}|} |x_{۱,i} - x_{۲,i}|^p \right)^{۱/p} \qquad (۱-۲)$$

*در ادبیات محرمانگی تفاضلی، **نرم** $\ell_۱$ (یا فاصله‌ی منهتن) به دلیل ارتباط مستقیم آن با تعداد رکوردها، معیار اصلی محسوب می‌شود. این فاصله دقیقاً تعداد رکوردهایی را می‌شمارد که باید تغییر کنند (اضافه یا حذف شوند) تا $\mathcal{D}_۱$ به $\mathcal{D}_۲$ تبدیل شود:*

$$\|\mathcal{D}_۱ - \mathcal{D}_۲\|_۱ = \sum_{i=۱}^{|\mathcal{X}|} |x_{۱,i} - x_{۲,i}|$$

مثال ۲-۳ *فرض کنید $\mathcal{D}_۱ = (۳, ۱)$ و $\mathcal{D}_۲ = (۳, ۰)$ باشند (یعنی در پایگاه‌داده دوم، یک نفر بیکار حذف شده است). فاصله‌ی $\ell_۱$ آن‌ها برابر است با:*

$$\|\mathcal{D}_۱ - \mathcal{D}_۲\|_۱ = |۳ - ۳| + |۱ - ۰| = ۱$$

تعریف ۲-۴ (پایگاه‌داده‌های همسایه) *دو پایگاه‌داده‌ی $\mathcal{D}_1$ و $\mathcal{D}_2$ را همسایه[7] می‌گوییم (و با $\mathcal{D}_1 \sim \mathcal{D}_2$ نشان می‌دهیم) اگر فاصله‌ی ℓ_1 آن‌ها حداکثر ۱ باشد:*

$$\|\mathcal{D}_1 - \mathcal{D}_2\|_1 \leqslant 1$$

این شرط تضمین می‌کند که دو پایگاه‌داده تنها در بود و نبودِ مشخصات یک فرد خاص با هم تفاوت دارند.

ایده‌ی اصلی محرمانگی تفاضلی این است که خروجی مکانیزم برای دو مجموعه‌داده‌ی همسایه باید از نظر آماری «شبیه» باشد، به طوری که مهاجم نتواند تشخیص دهد ورودی واقعی کدام بوده است.

تعریف ۲-۵ (εـ محرمانگی تفاضلی (ε-DP)) *یک مکانیزم تصادفی $\mathcal{M}$ با دامنه $\mathcal{X}^n$ و برد $\mathcal{R}$، ویژگی εـ محرمانگی تفاضلی را برآورده می‌سازد، اگر برای هر دو پایگاه‌داده‌ی همسایه‌ی $\mathcal{D}_1$ و $\mathcal{D}_2$ ($\mathcal{D}_1 \sim \mathcal{D}_2$) و برای هر زیرمجموعه از خروجی‌های ممکن $\mathcal{S} \subseteq \mathcal{R}$ (که در σـ جبر برد تعریف شده باشد)، داشته باشیم:*

$$\mathbb{P}\left[\mathcal{M}(\mathcal{D}_1) \in \mathcal{S}\right] \leqslant \exp(\varepsilon) \cdot \mathbb{P}\left[\mathcal{M}(\mathcal{D}_2) \in \mathcal{S}\right] \qquad (۲-۲)$$

در این نامساوی، پارامتر $\varepsilon \geqslant 0$ را «بودجه‌ی محرمانگی»[8] می‌نامیم.

۲-۱-۲ تفسیر پارامترهای محرمانگی

پارامتر ε نقش کنترل‌کننده‌ی توازن میان «محرمانگی» و «سودمندی» را ایفا می‌کند:

- مقادیر کوچک ε (مثلاً $\varepsilon \leqslant 1$) به معنای شباهت بسیار زیاد توزیع‌های خروجی است که منجر به محرمانگی قوی‌تر اما خطای بیش‌تر (نویز بیش‌تر) می‌شود.
- مقادیر بزرگ ε اجازه‌ی تمایز بیش‌تر بین توزیع‌ها را می‌دهد که به معنای دقت بالاتر اما ریسک افشای بیش‌تر است.
- اگر $\varepsilon = 0$ باشد، خروجی مکانیزم باید کاملاً مستقل از ورودی باشد (امنیت کامل اما بدون هیچ‌گونه فایده‌ی آماری).

تعریف ۲-۶ (DP ـ تقریبی یا (ε, δ)-DP) *در بسیاری از موارد (مانند مکانیزم گوسی)، ارضای شرط ε-DP خالص ممکن نیست. در این شرایط، از تعریف انعطاف‌پذیرتری به نام (ε, δ)-DP استفاده می‌شود که*

[7] Neighboring / Adjacent
[8] Privacy Budget

اجازه‌ی یک احتمال شکست کوچک δ را می‌دهد:

$$\mathbb{P}\left[\mathcal{M}(\mathcal{D}_1) \in \mathcal{S}\right] \leqslant \exp(\varepsilon) \cdot \mathbb{P}\left[\mathcal{M}(\mathcal{D}_2) \in \mathcal{S}\right] + \delta \qquad (۳-۲)$$

پارامتر $\delta \in [0, 1]$ را معمولاً «احتمال شکست»[9] یا احتمال نشت اطلاعات می‌نامند. تفسیر شهودی این است که مکانیزم با احتمال حداقل $\delta - 1$، تضمین ε-DP را رعایت می‌کند. در کاربردهای عملی، مقدار δ باید بسیار ناچیز (کم‌تر از معکوس چندجمله‌ای اندازه‌ی داده‌ها، مثلاً $\delta < 1/n$) در نظر گرفته شود.

۲-۱-۳ تعاریف معادل و صورت‌بندی‌های جایگزین

برای تسهیل تحلیل‌های ریاضی و درک عمیق‌تر، می‌توان تعریف اصلی ε-DP را به صورت‌های هم‌ارز دیگری بیان کرد. در ادامه دو دیدگاه مهم «نقطه‌ای» و «واگرایی» را بررسی می‌کنیم.

۱. **دیدگاه نقطه‌ای و چگالی احتمال (لم هم‌ارزی):**
اگرچه تعریف اصلی بر روی زیرمجموعه‌ها بنا شده است، اما لم زیر نشان می‌دهد که کنترل نسبت احتمالات در تک‌تک نقاط برای برقراری شرط کافی است.

لم ۲-۱ (هم‌ارزی نقطه‌ای) *یک مکانیزم $\mathcal{M}$ شرط ε-DP را برآورده می‌کند اگر و تنها اگر برای تمام همسایه‌های $\mathcal{D}_1 \sim \mathcal{D}_2$ شرایط زیر برقرار باشد:*

- *در فضای گسسته: برای هر خروجی $z \in \mathcal{R}$:*

$$\frac{\mathbb{P}\left[\mathcal{M}(\mathcal{D}_1) = z\right]}{\mathbb{P}\left[\mathcal{M}(\mathcal{D}_2) = z\right]} \leqslant e^{\varepsilon} \qquad (۴-۲)$$

- *در فضای پیوسته: با فرض وجود توابع چگالی $p(\cdot)$ و $q(\cdot)$، برای تمام $z \in \mathcal{R}$:*

$$p(z) \leqslant e^{\varepsilon} \cdot q(z) \qquad (۵-۲)$$

اثبات. اثبات بر پایه‌ی خاصیت جمع‌پذیری (در حالت گسسته) و خاصیت یکنوایی انتگرال (در حالت پیوسته روی میدان‌های σـجبر بورل) استوار است. فرض کنید شرط نقطه‌ای برقرار باشد؛ برای هر زیرمجموعه‌ی بورلی $\mathcal{S} \subseteq \mathcal{R}$:

$$\mathbb{P}\left[\mathcal{M}(\mathcal{D}_1) \in \mathcal{S}\right] = \int_{\mathcal{S}} p(z) d\mu(z) \leqslant \int_{\mathcal{S}} e^{\varepsilon} q(z) d\mu(z) = e^{\varepsilon} \mathbb{P}\left[\mathcal{M}(\mathcal{D}_2) \in \mathcal{S}\right]$$

(در حالت گسسته، انتگرال با عمل‌گر جمع جایگزین می‌شود). □

[9] Failure Probability

نکته: این هم‌ارزی تنها برای $\delta = 0$ صادق است. برای (ε, δ)-DP، بررسی نقطه‌به‌نقطه کافی نیست و شرط باید روی زیرمجموعه‌ها چک شود.

۲. **تعریف مبتنی بر واگرایی ماکزیمم:**
از دیدگاه نظریه اطلاعات، محرمانگی تفاضلی خالص محدودیتی بر روی «واگرایی ماکزیمم»[10] بین توزیع‌های خروجی است. واگرایی ماکزیمم به صورت $D_\infty(P||Q) = \sup_S \ln \frac{P(S)}{Q(S)}$ تعریف می‌شود. بنابراین تعریف ۲-۵ معادل است با:

$$\sup_{\mathcal{D}_1 \sim \mathcal{D}_2} D_\infty \left(\mathcal{M}(\mathcal{D}_1) \parallel \mathcal{M}(\mathcal{D}_2) \right) \leqslant \varepsilon \qquad (۶-۲)$$

در بخش‌های بعدی با تعریف دقیق واگرایی آشنا خواهیم شد.

۲-۱-۴ مکانیزم‌های پایه

برای دست‌یابی به محرمانگی تفاضلی، باید به پاسخ دقیق پرس‌وجو «نویز»[11] اضافه کنیم. میزان نویز به حساسیت[12] پرس‌وجو بستگی دارد.

تعریف ۲-۷ (حساسیت سراسری ℓ_p) *برای هر تابع پرس‌وجوی $f : \mathcal{X}^n \to \mathcal{R}^k$ که خروجی برداری دارد، «حساسیت سراسری ℓ_p»[13] که با $\Delta_p f$ نمایش داده می‌شود، برابر است با بیشینه‌ی مقدار تغییرات خروجی تابع، به ازای تغییر تنها یک رکورد در ورودی. با استفاده از تعریف نرم ℓ_p (تعریف ۲-۱) داریم:*

$$\Delta_p f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \| f(\mathcal{D}_1) - f(\mathcal{D}_2) \|_p \qquad (۷-۲)$$

که در آن ماکزیمم‌گیری روی تمام زوج پایگاه‌داده‌های همسایه $(\mathcal{D}_1 \sim \mathcal{D}_2)$ انجام می‌شود.

در میان انواع حساسیت‌ها، دو مورد زیر به دلیل کاربردشان در مکانیزم‌های پایه، از اهمیت ویژه‌ای برخوردارند:

تعریف ۲-۸ (حساسیت ℓ_1، ℓ_2 و ℓ_∞) *سه نوع حساسیت زیر بیش‌ترین کاربرد را در طراحی مکانیزم‌های محرمانگی دارند:*

[10] Max Divergence
[11] Noise
[12] Sensitivity
[13] ℓ_p-Global Sensitivity

۱. **حساسیت ℓ_1 ($\Delta_1 f$):** *این حساسیت برابر با ماکزیمم فاصله منهتن بین خروجی‌هاست و پارامتر اصلی در مکانیزم لاپلاس می‌باشد:*

$$\Delta_1 f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1 \qquad (۸-۲)$$

۲. **حساسیت ℓ_2 ($\Delta_2 f$):** *این حساسیت برابر با ماکزیمم فاصله اقلیدسی است و در مکانیزم گوسی کاربرد اساسی دارد. معمولاً استفاده از این حساسیت در ابعاد بالا منجر به خطای کم‌تری نسبت به ℓ_1 می‌شود:*

$$\Delta_2 f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_2 \qquad (۹-۲)$$

۳. **حساسیت ℓ_∞ ($\Delta_\infty f$):** *این حساسیت برابر با بیشینه‌ی تغییر در «تک‌تک مولفه‌های» خروجی است (نرم ماکزیمم). این معیار نشان می‌دهد که مقدار یک درایه خاص از خروجی چقدر می‌تواند تغییر کند:*

$$\Delta_\infty f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_\infty \qquad (۱۰-۲)$$

مثال ۲-۴ *فرض کنید f یک تابع هیستوگرام شمارشی باشد (تعداد افراد در دسته‌های مجزا). اگر مشخصات یک فرد تغییر کند، او از یک دسته خارج (تغییر -1) و به دسته‌ی دیگر وارد (تغییر $+1$) می‌شود. سایر دسته‌ها ثابت می‌مانند (0). بردار تغییرات برابر است با $(0, \ldots, +1, \ldots, -1, \ldots, 0)$. حال حساسیت‌ها را محاسبه می‌کنیم:*

- **حساسیت ℓ_1:** *مجموع قدرمطلق تغییرات:* $|1| + |-1| = 2$.
- **حساسیت ℓ_2:** *جذر مجموع مربعات:* $\sqrt{1^2 + (-1)^2} = \sqrt{2}$.
- **حساسیت ℓ_∞:** *ماکزیمم قدرمطلق تغییرات:* $\max(|1|, |-1|, 0) = 1$.

این مثال به وضوح رابطه $\Delta_\infty f \leqslant \Delta_2 f \leqslant \Delta_1 f$ را نشان می‌دهد.

انتخاب نوع حساسیت در طراحی مکانیزم، بستگی مستقیم به نوع نویز افزوده شده و ابعاد داده‌ها دارد. به طور خلاصه، حساسیت ℓ_1 برای کالیبره کردن مکانیزم لاپلاس و حساسیت ℓ_2 برای مکانیزم گوسی استفاده می‌شود.

۲-۱-۵ مکانیزم‌های بنیادی محرمانگی تفاضلی

در این بخش، سه مکانیزم اصلی را که بلوک‌های سازنده‌ی بسیاری از الگوریتم‌های پیچیده‌تر هستند، معرفی می‌کنیم.

مکانیزم لاپلاس

ساده‌ترین و پرکاربردترین روش برای توابع عددی، افزودن نویز از توزیع لاپلاس است. توزیع لاپلاس با پارامتر مقیاس b و میانگین μ دارای تابع چگالی احتمال $h(z) = \frac{۱}{۲b} \exp\left(-\frac{|z-\mu|}{b}\right)$ است.

قضیه‌ی ۲-۲ (محرمانگی مکانیزم لاپلاس) *فرض کنید $f : \mathcal{X}^n \to \mathcal{R}^k$ یک تابع پرس‌وجو با حساسیت سراسری $\Delta_۱ f$ باشد. مکانیزم لاپلاس که خروجی آن به صورت زیر تعریف می‌شود:*

$$\mathcal{M}_{\text{Lap}}(\mathcal{D}) = f(\mathcal{D}) + (Y_۱, \ldots, Y_k) \qquad (۱۱-۲)$$

که در آن $Y_i \overset{i.i.d}{\sim} \text{Lap}\left(\frac{\Delta_۱ f}{\varepsilon}\right)$، شرط ε-DP را برآورده می‌کند.

اثبات. فرض کنید $\mathcal{D}_۱ \sim \mathcal{D}_۲$ دو پایگاه‌داده‌ی همسایه باشند و خروجی تابع f یک بردار kـ بعدی باشد. نویز لاپلاس به هر مؤلفه به صورت مستقل اضافه می‌شود، بنابراین تابع چگالی احتمال توأم برابر با حاصل‌ضرب چگالی‌های مؤلفه‌هاست. با فرض $b = \frac{\Delta_۱ f}{\varepsilon}$، نسبت چگالی احتمال را برای یک بردار خروجی دلخواه $z = (z_۱, \ldots, z_k)$ بررسی می‌کنیم:

$$\begin{aligned}
\frac{p(z|\mathcal{D}_۱)}{p(z|\mathcal{D}_۲)} &= \frac{\prod_{i=۱}^{k} \frac{۱}{۲b} \exp\left(-\frac{|z_i - f(\mathcal{D}_۱)_i|}{b}\right)}{\prod_{i=۱}^{k} \frac{۱}{۲b} \exp\left(-\frac{|z_i - f(\mathcal{D}_۲)_i|}{b}\right)} \\
&= \prod_{i=۱}^{k} \exp\left(\frac{|z_i - f(\mathcal{D}_۲)_i| - |z_i - f(\mathcal{D}_۱)_i|}{b}\right) \\
&= \exp\left(\frac{۱}{b} \sum_{i=۱}^{k} \left(|z_i - f(\mathcal{D}_۲)_i| - |z_i - f(\mathcal{D}_۱)_i|\right)\right)
\end{aligned}$$

طبق نامساوی مثلثی ($|a| - |b| \leqslant |a - b|$) برای هر مؤلفه i داریم:

$$|z_i - f(\mathcal{D}_۲)_i| - |z_i - f(\mathcal{D}_۱)_i| \leqslant |f(\mathcal{D}_۱)_i - f(\mathcal{D}_۲)_i|$$

با اعمال این نامساوی در مجموع توان نمایی:

$$\sum_{i=۱}^{k} \left(|z_i - f(\mathcal{D}_۲)_i| - |z_i - f(\mathcal{D}_۱)_i|\right) \leqslant \sum_{i=۱}^{k} |f(\mathcal{D}_۱)_i - f(\mathcal{D}_۲)_i|$$

عبارت سمت راست دقیقاً برابر با نرم ℓ_1 تفاضل خروجی‌هاست:

$$\sum_{i=1}^{k} |f(\mathcal{D}_1)_i - f(\mathcal{D}_2)_i| = \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1$$

طبق تعریف حساسیت سراسری، می‌دانیم $\|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1 \leqslant \Delta_1 f$. بنابراین:

$$\frac{p(z|\mathcal{D}_1)}{p(z|\mathcal{D}_2)} \leqslant \exp\left(\frac{\Delta_1 f}{b}\right) = \exp\left(\frac{\Delta_1 f}{\Delta_1 f/\varepsilon}\right) = e^{\varepsilon}$$

و حکم ثابت می‌شود. □

مثال ۲-۵ (پرس‌وجوهای شمارشی) *پرس‌وجوهای شمارشی[14]، پرس‌وجوهایی به فرم «چه تعداد از اعضای پایگاه‌داده ویژگی P را دارند؟» هستند. این نوع توابع بلوک‌های سازنده‌ی بسیاری از تحلیل‌های آماری و داده‌کاوی (مانند هیستوگرام‌ها) هستند* [۹].

حالت تک پرس‌وجو: *حساسیت یک پرس‌وجوی شمارشی دقیقاً ۱ است ($\Delta_1 f = 1$)؛ زیرا افزودن یا حذف یک فرد، نتیجه‌ی شمارش را حداکثر ۱ واحد تغییر می‌دهد. بنابراین طبق قضیه ۲-۲، با افزودن نویز با مقیاس $1/\varepsilon$ (یعنی $\mathrm{Lap}(1/\varepsilon)$) به پاسخ واقعی، محرمانگی $(\varepsilon, 0)$-DP تضمین می‌شود.*

حالت برداری (چند پرس‌وجو): *فرض کنید لیستی از k پرس‌وجوی شمارشی $f = (f_1, \ldots, f_k)$ داریم (یک پرس‌وجوی برداری). بدون داشتن اطلاعات اضافی درباره‌ی ارتباط پرس‌وجوها، در بدترین حالت یک فرد مشخص می‌تواند در تمام k شمارش تأثیر بگذارد. بنابراین حساسیت ℓ_1 کل بردار برابر با مجموع تغییرات، یعنی k خواهد بود ($\Delta_1 f = k$). در این حالت برای دست‌یابی به ε-DP، باید به هر پاسخ نویزی با مقیاس k/ε اضافه کنیم.*

مکانیزم گوسی

زمانی که حساسیت ℓ_2 تابع بسیار کم‌تر از حساسیت ℓ_1 باشد (مثلاً در پرس‌وجوهای برداری)، مکانیزم گوسی ترجیح داده می‌شود. این مکانیزم به جای نویز لاپلاس، نویز گوسی (نرمال) به خروجی اضافه می‌کند.

قضیه‌ی ۲-۳ (محرمانگی مکانیزم گوسی) *فرض کنید $f : \mathcal{X}^n \to \mathcal{R}^k$ تابعی با حساسیت ℓ_2 برابر با $\Delta_2 f$ باشد. مکانیزم گوسی با افزودن نویز $Y \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_k)$ به خروجی تعریف می‌شود:*

$$\mathcal{M}_{\mathrm{Gauss}}(\mathcal{D}) = f(\mathcal{D}) + \mathcal{N}(0, \sigma^2 \mathbf{I}_k) \qquad (۱۲-۲)$$

[14] Counting Queries

اگر $\varepsilon \in (۰, ۱)$ باشد، با انتخاب انحراف معیار σ به صورت زیر، این مکانیزم شرط (ε, δ)-DP را برآورده می‌کند:

$$\sigma \geqslant \sqrt{۲ \ln (۱/۲۵/\delta)} \cdot \frac{\Delta_۲ f}{\varepsilon} \qquad (۱۳-۲)$$

اثبات. [طرح کلی] اثبات دقیق این قضیه نیازمند تحلیل «متغیر تصادفی زیان محرمانگی»[15] است که جزئیات کامل آن در [9, Appendix A] موجود است. ایده اصلی این است که برخلاف توزیع لاپلاس، در توزیع گوسی نسبت $\frac{p(z)}{q(z)}$ کران‌دار نیست، اما احتمال رخداد مقادیری که این نسبت را بزرگ می‌کنند (نواحی دمی توزیع)، با δ محدود می‌شود. □

مثال ۲-۶ (انتشار آماره‌های چندگانه و اثر ابعاد) *فرض کنید یک بیمارستان می‌خواهد میانگین d ویژگی حیاتی مختلف را منتشر کند. داده‌های هر بیمار را می‌توان به صورت یک بردار $v \in [۰, ۱]^d$ در نظر گرفت. اگر یک بیمار پرونده‌اش را تغییر دهد، در بدترین حالت تمام d ویژگی او می‌توانند از ۰ به ۱ تغییر کنند.*

تحلیل حساسیت:

- ***حساسیت $\ell_۱$:*** *مجموع قدرمطلق تغییرات برابر است با $\sum_{i=۱}^{d} |۱ - ۰| = d$.*
- ***حساسیت $\ell_۲$:*** *جذر مجموع مربعات تغییرات برابر است با $\sqrt{\sum_{i=۱}^{d} (۱ - ۰)^۲} = \sqrt{d}$.*

مقایسه نویز: *اگر تعداد ویژگی‌ها زیاد باشد (مثلاً $d = ۱۰۰$):*

- *مکانیزم لاپلاس باید نویزی متناسب با $d = ۱۰۰$ اضافه کند.*
- *مکانیزم گوسی نویزی متناسب با $\sqrt{۱۰۰} = ۱۰$ اضافه می‌کند.*

این کاهش چشم‌گیر نویز (با ضریب $\sqrt{d}$) دلیل اصلی استفاده از مکانیزم گوسی در الگوریتم‌هایی نظیر یادگیری عمیق با محرمانگی تفاضلی (DP-SGD) است.

مکانیزم نمایی

مکانیزم‌های قبلی برای خروجی‌های عددی بودند. اگر خروجی یک «عضو» از یک مجموعه باشد، از مکانیزم نمایی استفاده می‌شود. این مکانیزم بر اساس یک **تابع امتیاز**[16] $q(\mathcal{D}, r)$ کار می‌کند که میزان «خوبی» خروجی r را می‌سنجد. حساسیت این تابع به صورت $\Delta q = \max_r \max_{\mathcal{D} \sim \mathcal{D}'} |q(\mathcal{D}, r) - q(\mathcal{D}', r)|$ تعریف می‌شود.

[15] Privacy Loss Random Variable
[16] Score Function

قضیه‌ی ۲-۴ (محرمانگی مکانیزم نمایی) *مکانیزم نمایی، یک خروجی r از مجموعه ممکن $\mathcal{R}$ را با احتمالی متناسب با امتیاز آن انتخاب می‌کند:*

$$\mathbb{P}\left[\mathcal{M}_{\text{Exp}}(\mathcal{D}) = r\right] = \frac{\exp\left(\frac{\varepsilon \cdot q(\mathcal{D}, r)}{۲\Delta q}\right)}{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon \cdot q(\mathcal{D}, r')}{۲\Delta q}\right)} \tag{۲-۱۴}$$

این مکانیزم شرط ε-DP را برآورده می‌کند.

اثبات. فرض کنید $\mathcal{D}_۱ \sim \mathcal{D}_۲$. نسبت احتمالات برای یک خروجی ثابت r عبارت است از:

$$\frac{\mathbb{P}\left[\mathcal{M}(\mathcal{D}_۱) = r\right]}{\mathbb{P}\left[\mathcal{M}(\mathcal{D}_۲) = r\right]} = \frac{\exp\left(\frac{\varepsilon q(\mathcal{D}_۱, r)}{۲\Delta q}\right)}{\exp\left(\frac{\varepsilon q(\mathcal{D}_۲, r)}{۲\Delta q}\right)} \cdot \frac{\sum_{r'} \exp\left(\frac{\varepsilon q(\mathcal{D}_۲, r')}{۲\Delta q}\right)}{\sum_{r'} \exp\left(\frac{\varepsilon q(\mathcal{D}_۱, r')}{۲\Delta q}\right)}$$

ترم اول (صورت کسرها) با استفاده از خاصیت حساسیت q حداکثر $e^{\varepsilon/۲}$ است. ترم دوم (نسبت مخرج‌ها) نیز با استدلالی مشابه حداکثر $e^{\varepsilon/۲}$ خواهد بود. حاصل‌ضرب این دو مقدار حداکثر e^{ε} می‌شود. □

۲-۱-۶ ترکیب‌پذیری

در کاربردهای واقعی، معمولاً چندین پرس‌وجو روی یک پایگاه‌داده اجرا می‌شود. قضایای ترکیب‌پذیری[۱۷] نشان می‌دهند که چگونه بودجه‌ی محرمانگی انباشته می‌شود.

قضیه‌ی ۲-۵ (ترکیب‌پذیری ساده) *اگر k مکانیزم $\mathcal{M}_۱, \ldots, \mathcal{M}_k$ به ترتیب دارای بودجه‌های $\varepsilon_۱, \ldots, \varepsilon_k$ باشند، اجرای متوالی آن‌ها روی یک پایگاه‌داده‌ی واحد، تضمین $(\sum \varepsilon_i)$-DP را فراهم می‌کند* [۹].

این کران خطی در بسیاری موارد بدبینانه است. «قضیه ترکیب پیشرفته» نشان می‌دهد که با پذیرش اندکی احتمال شکست (δ' اضافه)، انباشت بودجه بسیار کندتر (با نرخ $\sqrt{k}$) رشد می‌کند [۹].

قضیه‌ی ۲-۶ (ترکیب‌پذیری پیشرفته) *برای هر $\delta' > ۰$، اجرای k مکانیزم که هرکدام ε-DP هستند، دارای تضمین $(\varepsilon', k\delta + \delta')$-DP است که در آن:*

$$\varepsilon' \approx \varepsilon\sqrt{۲k\ln(۱/\delta')} + k\varepsilon(e^{\varepsilon} - ۱) \tag{۲-۱۵}$$

برای مقادیر کوچک ε، جمله دوم ناچیز است و بودجه کل تقریباً با $\varepsilon\sqrt{k}$ رشد می‌کند.

[17] Composition

۲-۱-۷ محرمانگی گروهی

محرمانگی تفاضلی نه تنها از یک فرد، بلکه به صورت خودکار از گروه‌های کوچک نیز محافظت می‌کند [۹].

قضیه‌ی ۲-۷ (محرمانگی گروهی[۱۸]) *اگر دو پایگاه‌داده $\mathcal{D}_1$ و $\mathcal{D}_2$ در k رکورد متفاوت باشند (فاصله‌ی همسایگی k)، آنگاه هر مکانیزم ε-DP برای آن‌ها تضمین $(k\varepsilon)$-DP را ارائه می‌دهد:*

$$\mathbb{P}\left[\mathcal{M}(\mathcal{D}_1) \in S\right] \leqslant e^{k\varepsilon}\mathbb{P}\left[\mathcal{M}(\mathcal{D}_2) \in S\right] \qquad (۱۶-۲)$$

این خاصیت نشان می‌دهد که با بزرگ شدن گروه (k)، تضمین محرمانگی به صورت نمایی تضعیف می‌شود ($e^{k\varepsilon}$).

۲-۱-۸ محدودیت مدل متمرکز

با وجود تمام مزایا، مدل CDP یک نقطه‌ی ضعف اساسی دارد: نیاز به یک متصدی کاملاً مورد اعتماد. در بسیاری از سناریوهای دنیای واقعی (مانند جمع‌آوری داده از گوشی‌های هوشمند)، کاربران به سرور مرکزی اعتماد ندارند. این عدم اعتماد، ما را به سمت مدل جایگزین، یعنی «محرمانگی تفاضلی موضعی» سوق می‌دهد. در فصل بعد با محرمانگی تفاضلی موضعی و تعاریف و قضایای اساسی آن آشنا خواهیم شد.

۲-۲ fـ واگرایی‌ها

در بخش‌های پیشین، مکانیزم‌های محرمانگی تفاضلی را ابزاری برای ایجاد «شباهت آماری» بین خروجی‌های دو پایگاه‌داده‌ی همسایه معرفی کردیم. برای کمّی‌سازی دقیق این شباهت و اثبات کران‌های پایین در فصل‌های آینده، نیازمند معیاری هستیم که فاصله میان توزیع‌های احتمالی را در یک چارچوب عمومی بسنجد. این ابزار، خانواده‌ی fـ واگرایی‌ها[۱۹] است که نخستین بار توسط سیزار [۵] و علی و سیلوِی [۱] معرفی شد.

[19] f-divergences

۲-۲-۱ تعریف رسمی در فضای اندازه‌پذیر

برای ارائه تعریفی دقیق و مستقل از نوع متغیرهای تصادفی (پیوسته یا گسسته)، از زبان نظریه اندازه[۲۰] استفاده می‌کنیم. در این چارچوب، ابتدا باید ساختار فضایی که وقایع در آن رخ می‌دهند را مشخص کنیم.

تعریف ۲-۹ (**σـ‌جبر و فضای اندازه‌پذیر**[۲۱]) *فرض کنید Ω یک مجموعه دلخواه (فضای نمونه) باشد. یک خانواده $\mathcal{F}$ از زیرمجموعه‌های Ω را یک σـ‌جبر (یا میدان[۲۲]) می‌نامیم، هرگاه سه شرط زیر برقرار باشد:*

۱. *فضای کامل در آن باشد: $\Omega \in \mathcal{F}$.*

۲. *نسبت به متمم‌گیری بسته باشد: اگر $A \in \mathcal{F}$، آن‌گاه $A^c \in \mathcal{F}$.*

۳. *نسبت به اجتماع شمارا بسته باشد: اگر $A_1, A_2, \ldots$ دنباله‌ای از مجموعه‌ها در $\mathcal{F}$ باشند، آن‌گاه $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

در این صورت، زوج مرتب $(\Omega, \mathcal{F})$ را یک «فضای اندازه‌پذیر»[۲۳] و اعضای $\mathcal{F}$ را «مجموعه‌های اندازه‌پذیر» یا «پیشامد» می‌نامند.

تعریف ۲-۱۰ (اندازه[۲۴]**)** *تابع مجموعه‌ای $\mu : \mathcal{F} \to [0, \infty]$ را یک «اندازه» روی فضای $(\Omega, \mathcal{F})$ می‌نامیم، هرگاه $\mu(\emptyset) = 0$ باشد و خاصیت «جمع‌پذیری شمارا»[۲۵] را برآورده کند؛ یعنی برای هر دنباله از مجموعه‌های دو-به-دو جدا از هم $\{A_i\}_{i=1}^{\infty}$ در $\mathcal{F}$:*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{۲-۱۷}$$

اگر شرط نرمال‌سازی $\mu(\Omega) = 1$ برقرار باشد، μ را یک «اندازه احتمال» می‌نامیم.

حال فرض کنید $(\Omega, \mathcal{F})$ یک فضای اندازه‌پذیر باشد که روی آن دو اندازه احتمال P و Q و یک اندازه‌ی مرجع μ (مانند اندازه لبگ یا شمارشی) تعریف شده است.

تعریف ۲-۱۱ (پیوستگی مطلق[۲۶]**)** *می‌گوییم اندازه احتمال P نسبت به اندازه μ مطلقاً پیوسته است و می‌نویسیم $P \ll \mu$، اگر برای هر مجموعه اندازه‌پذیر $A \in \mathcal{F}$:*

$$\mu(A) = 0 \implies P(A) = 0. \tag{۲-۱۸}$$

[20] Measure Theory
[22] Field
[23] Measurable Space
[25] Countable Additivity

اهمیت این تعریف در قضیه رادون‌ـ نیکودیم نهفته است که وجود توابع چگالی را تضمین می‌کند.

قضیه‌ی ۲-۸ (رادون‌ـ نیکودیم[۲۷]) *اگر $P \ll \mu$ باشد (μ و یک اندازه σـ متناهی باشد)، آن‌گاه یک تابع منحصربه‌فرد (تقریباً همه‌جا) غیرمنفی و اندازه‌پذیر $p : \Omega \to [0, \infty)$ وجود دارد که برای هر $A \in \mathcal{F}$:*

$$P(A) = \int_A p \, d\mu. \tag{۲-۱۹}$$

این تابع p را «مشتق رادون‌ـ نیکودیم» P نسبت به μ می‌نامند و با $\frac{dP}{d\mu}$ نمایش می‌دهند.

۲-۲-۲ تعریف fـ واگرایی

با در دست داشتن چگالی‌های $p = \frac{dP}{d\mu}$ و $q = \frac{dQ}{d\mu}$، اکنون می‌توانیم تعریف عمومی fـ واگرایی را ارائه دهیم.

تعریف ۲-۱۲ (fـ واگرایی[۲۸]) *فرض کنید $f : (0, \infty) \to \mathcal{R}$ یک تابع محدب باشد که در $x = 1$ برابر صفر است ($f(1) = 0$). «fـ واگرایی» بین دو توزیع احتمال P و Q به صورت زیر تعریف می‌شود:*

$$D_f(P\|Q) \triangleq \int_\Omega f\left(\frac{p(x)}{q(x)}\right) q(x) \, d\mu(x). \tag{۲-۲۰}$$

برای کامل بودن تعریف در نقاط مرزی (وقتی $q(x) = 0$ یا $p(x) = 0$)، قراردادهای استاندارد زیر را اعمال می‌کنیم:

- $0 \cdot f\left(\frac{0}{0}\right) = 0$
- $0 \cdot f\left(\frac{a}{0}\right) = a \lim_{u\to\infty} \frac{f(u)}{u} = \lim_{t\to 0^+} t \cdot f(\frac{a}{t})$

این تعریف مستقل از انتخاب اندازه مرجع μ است. نکته مهم در مورد همگرایی انتگرال زمانی است که شرط پیوستگی مطلق ($P \ll Q$) برقرار نباشد (یعنی نقاطی وجود داشته باشند که Q صفر است اما P جرم دارد). در این نقاط، مقدار انتگرال تنها در صورتی متناهی می‌ماند که تابع f در بی‌نهایت دارای «رشد خطی» باشد (یعنی $\lim_{t\to\infty} \frac{f(t)}{t} < \infty$). به عنوان مثال، «فاصله تغییرات کل» به دلیل رشد خطی همواره متناهی است، اما «واگرایی KL» در صورت عدم برقراری شرط پیوستگی مطلق، بی‌نهایت خواهد شد.

تفسیر اجزاء:

- **تابع مولد f:** تابع f تعیین‌کننده نوع هندسه و خواص واگرایی است. تحدب f شرطی حیاتی برای خوش‌رفتاری ریاضی این اندازه است.

- **غیرمنفی بودن:** با استفاده از نامساوی ینسن[۲۹] برای تابع محدب f، می‌توان نشان داد که واگرایی همواره نامنفی است. از آن‌جا که $\int p(x)d\mu = ۱$ و $\mathbb{E}_Q[\frac{dP}{dQ}]$، داریم:

$$D_f(P\|Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right] \geqslant f\left(\mathbb{E}_Q\left[\frac{dP}{dQ}\right]\right) = f(۱) = ۰. \qquad (۲۱-۲)$$

بنابراین همواره $D_f(P\|Q) \geqslant ۰$. هم‌چنین اگر f در نقطه $t = ۱$ اکیداً محدب[۳۰] باشد، تساوی $D_f(P\|Q) = ۰$ برقرار است اگر و تنها اگر $P = Q$ (تقریباً همه‌جا).

در حالت‌های خاص که فضای نمونه Ω گسسته یا پیوسته (اقلیدسی) باشد و چگالی‌های p و q نسبت به یک اندازه پایه (مانند شمارشی یا لبگ) وجود داشته باشند، مشتق رادون‌ـنیکودیم به نسبت معمولی چگالی‌ها $\frac{p(x)}{q(x)}$ تبدیل می‌شود و تعریف انتگرالی بالا به فرم‌های آشنای زیر تقلیل می‌یابد:

$$D_f(P\|Q) = \int_{\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) dx \quad \text{یا} \quad \sum_{x\in\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) \qquad (۲۲-۲)$$

تفاوت واگرایی با متر ریاضی: اگرچه در ادبیات موضوع، گاهی با اغماض از واژه‌ی «فاصله»[۳۱] برای واگرایی‌ها استفاده می‌شود، اما باید توجه داشت که واگرایی‌ها لزوماً خواص یک متریک حقیقی را ندارند. یک تابع $d(P,Q)$ تنها در صورتی یک متر ریاضی است که علاوه بر غیرمنفی بودن و اصل همانی، دو شرط تقارن[۳۲] ($d(P,Q) = d(Q,P)$) و نامساوی مثلث[۳۳] ($d(P,R) \leqslant d(P,Q) + d(Q,R)$) را ارضا کند.

بسیاری از واگرایی‌های معرفی شده در این بخش، به‌ویژه واگرایی کولبک‌ـلایبلر (بخش ۲-۱۴)، این دو شرط را نقض می‌کنند. عدم تقارن در واگرایی‌ها دارای تفسیر آماری مهمی است؛ به عنوان مثال $\mathrm{KL}(P\|Q)$ بیانگر میزان اطلاعات از دست رفته زمانی است که از توزیع Q برای مدل‌سازی توزیع واقعی P استفاده می‌کنیم. این مفهوم ماهیتی «جهت‌دار» دارد و با جابجایی جایگاه فرضیه و واقعیت تغییر می‌کند. با این حال، fـ واگرایی‌ها هم‌چنان خواص توپولوژیکی و هندسی قدرتمندی دارند که در فصل‌های آتی برای استخراج کران‌های پایین مینی‌مکس از آن‌ها بهره خواهیم برد.

۲-۲-۳ نمونه‌های مهم و توابع مولد

با انتخاب‌های متفاوت برای تابع مولد محدب $f(t)$، می‌توان طیف وسیعی از اندازه‌های فاصله را تولید کرد. در ادامه، مهم‌ترین نمونه‌ها را معرفی می‌کنیم. در تعاریف زیر، فرض می‌کنیم P و Q دو اندازه احتمال باشند که دارای چگالی (یا جرم) احتمال $p(x)$ و $q(x)$ نسبت به یک اندازه پایه هستند.

[29] Jensen's Inequality
[30] Strictly Convex
[31] Distance
[32] Symmetry
[33] Triangle Inequality

تعریف ۲-۱۳ (تغییرات کل) *فاصله تغییرات کل[۳۴]، شهودی‌ترین متریک برای سنجش تمایزپذیری دو توزیع است و بیان‌گر بیش‌ترین تفاوت احتمالی است که دو توزیع می‌توانند روی یک پیشامد داشته باشند. تابع مولد آن* $f(t) = \frac{۱}{۲}|t-۱|$ *است.*

$$\mathrm{TV}(P,Q) = \frac{۱}{۲}\int_{\mathcal{X}} |p(x)-q(x)|\,dx \qquad (۲۳-۲)$$

$$= \sup_{A\subseteq\mathcal{X}} |P(A)-Q(A)| \qquad (۲۴-۲)$$

این واگرایی ممکن است با نمادهایی مانند d_{TV} *یا* $\|.\|_{\mathrm{TV}}$ *نیز نمایش داده شود.*

تعریف ۲-۱۴ (کولبک-لایبلر) *واگرایی کولبک-لایبلر[۳۵]، معروف‌ترین واگرایی در نظریه اطلاعات که آنتروپی نسبی[۳۶] نیز نامیده می‌شود. این معیار نامتقارن است و نقش اساسی در فشرده‌سازی داده‌ها و استنتاج بیزی دارد. تابع مولد آن* $f(t) = t\ln t$ *است.*

$$\mathrm{KL}(P\|Q) = \int_{\mathcal{X}} p(x)\ln\left(\frac{p(x)}{q(x)}\right)dx \qquad (۲۵-۲)$$

تعریف ۲-۱۵ (اطلاعات متقابل) *اگر* X *و* V *دو متغیر تصادفی باشند، اطلاعات متقابل[۳۷] بین آن‌ها به صورت امید ریاضی واگرایی* KL *بین توزیع شرطی و توزیع حاشیه‌ای تعریف می‌شود:*

$$I(X;V) = \mathrm{KL}(P_{X,V}\|P_X\otimes P_V) = \mathbb{E}_V\left[\mathrm{KL}(P_{X|V}\|P_X)\right] \qquad (۲۶-۲)$$

این معیار نقش کلیدی در نامساوی فانو و تحلیل کانال‌های اطلاعاتی ایفا می‌کند [۷].

تعریف ۲-۱۶ (کای-دو) *واگرایی کای-دو[۳۸]، اغلب برای تقریب‌زنی سایر فواصل (مانند* KL*) در همسایگی‌های کوچک استفاده می‌شود و به دلیل فرم مربعی، محاسبه آن معمولاً ساده‌تر است. تابع مولد آن* $f(t) = (t-۱)^۲$ *است.*

$$\chi^۲(P\|Q) = \int_{\mathcal{X}} \frac{(p(x)-q(x))^۲}{q(x)}dx \qquad (۲۷-۲)$$

تعریف ۲-۱۷ (هلینگر-دو) *فاصله هلینگر-دو[۳۹]، تابع مولد آن* $f(t) = (\sqrt{t}-۱)^۲$ *است. این فاصله به دلیل خواص ریاضی خوش‌رفتار (مانند متریک و کران‌دار بودن بین ۰ و ۲)، در نظریه برآورد[۴۰] و استخراج کران‌های مینی‌مکس (مانند روش* $Le\ Cam$*) کاربرد فراوان دارد.*

$$H^۲(P,Q) = \int_{\mathcal{X}} \left(\sqrt{p(x)}-\sqrt{q(x)}\right)^۲ dx \qquad (۲۸-۲)$$

[34] Total Variation (TV)
[35] Kullback-Leibler (KL)
[36] Relative Entropy
[37] Mutual Information
[38] Chi-Squared (χ^2)
[39] Squared Hellinger
[40] Estimation Theory

تعریف ۲-۱۸ (واگرایی E_γ یا واگرایی هاکی-استیک) *واگرایی E_γ یا واگرایی هاکی-استیک[41]، ابزاری کلیدی در تحلیل‌های مدرن حریم خصوصی و آزمون‌های فرضیه است. تابع مولد آن برای پارامتر $\gamma \geqslant 1$ به صورت $f_\gamma(t) = [t-\gamma]_+ = \max\{0, t-\gamma\}$ است.*

تعریف و صورت‌های معادل: *تعریف اصلی بر اساس انتگرالِ جرمِ اضافی نسبت درست‌نمایی است، اما صورت‌های معادل زیر بینش عملیاتی‌تری ارائه می‌دهند:*

$$E_\gamma(P\|Q) = \int_{\mathcal{X}} \max\{0, p(x) - \gamma q(x)\}\, dx \tag{۲-۲۹آ}$$

$$= \sup_{A \subseteq \mathcal{X}} \left(P(A) - \gamma Q(A)\right) \tag{۲-۲۹ب}$$

$$= P(Z > \gamma) - \gamma Q(Z > \gamma) \qquad \left(\text{که } Z = \frac{p(x)}{q(x)}\right) \tag{۲-۲۹ج}$$

رابطه (۲-۲۹ب) نشان می‌دهد که این واگرایی بیان‌گر بیشینه‌ی تفاضل وزن‌دار احتمالات است که مستقیماً با موازنه خطای نوع اول و دوم در آزمون فرضیه مرتبط است.

نکات تحلیلی و تاریخی:

نام‌گذاری توصیفی این واگرایی به «هاکی-استیک» که نخستین بار توسط ساسون و وِردو[۱۷] پیشنهاد شد، برخاسته از شکل هندسی نمودار تابع مولد $f(t)$ است. نمادگذاری E_γ و تدوین نقش بنیادی آن، پیش‌تر توسط پولیانسکی و همکاران [۱۶] صورت گرفته بود.

تابع $\gamma \mapsto E_\gamma(P\|Q)$ یک تابع نزولی و محدب است. هر چه γ بزرگ‌تر شود، "جریمه" اختصاص داده شده به Q بیش‌تر شده و واگرایی کم‌تر می‌شود.

کاربرد در حریم خصوصی: *شرط محرمانگی تفاضلی تقریبی ((ε, δ)-DP) دقیقاً معادل است با این‌که برای هر دو دیتابیس همسایه، واگرایی هاکی-استیک خروجی‌ها از مقدار δ تجاوز نکند:*

$$E_{e^\varepsilon}(P\|Q) \leqslant \delta \tag{۲-۳۰}$$

تعریف ۲-۱۹ (رنی) *واگرایی رنی[42]، با پارامتر $\alpha \in (1, \infty)$، به صورت زیر تعریف می‌شود:*

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \ln\left(\int_{\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} dx\right) \tag{۲-۳۱}$$

این واگرایی پلی میان واگرایی KL (در حد $\alpha \to 1$) و واگرایی ماکزیمم (در حد $\alpha \to \infty$) است و در تحلیل ترکیب‌پذیری مکانیزم‌ها کاربرد دارد [۱۴].

[41] Hockey-Stick Divergence
[42] Rényi Divergence

تعریف ۲-۲۰ (ماکزیمم (D_∞)) *این واگرایی متناظر با بدترین نسبت درست‌نمایی نقطه‌ای است و به عنوان حدِ واگرایی رنی به دست می‌آید:*

$$D_\infty(P\|Q) = \lim_{\alpha\to\infty} D_\alpha(P\|Q) = \sup_{x\in\mathcal{X}} \ln\left(\frac{p(x)}{q(x)}\right) \qquad (۳۲-۲)$$

کاربرد در حریم خصوصی: *شرط ε-DP (خالص) دقیقاً معادل است با کران‌دار بودن این واگرایی توسط بودجه حریم خصوصی: $D_\infty(P\|Q) \leqslant \varepsilon$.*

۲-۲-۴ خواص بنیادین و روابط بین fـ واگرایی‌ها

خانواده‌ی fـ واگرایی‌ها تنها مجموعه‌ای از فرمول‌های انتگرالی نیستند، بلکه دارای خواص ساختاری عمیقی هستند که آن‌ها را برای تحلیل سیستم‌های اطلاعاتی و حریم خصوصی ایده‌آل می‌سازد. در این بخش، سه ویژگی حیاتی این معیارها را بررسی می‌کنیم.

نامساوی پردازش داده (DPI)

مهم‌ترین ویژگی fـ واگرایی‌ها در نظریه اطلاعات، خاصیت یک‌نواختی[۴۳] آن‌ها تحت پردازش است. این ویژگی بیان می‌کند که هیچ عملیات پردازشی روی داده‌ها (اعم از قطعی یا تصادفی) نمی‌تواند تمایزپذیری بین دو توزیع را افزایش دهد.

قضیه‌ی ۲-۹ (نامساوی پردازش داده[۴۴]) *فرض کنید P و Q دو توزیع احتمال روی فضای $\mathcal{X}$ باشند و $\kappa : \mathcal{X} \to \mathcal{Y}$ یک هسته‌ی احتمالاتی (کانال مارکوف)[۴۵] باشد که داده‌ها را از فضای $\mathcal{X}$ به $\mathcal{Y}$ نگاشت می‌کند. اگر $P\kappa$ و $Q\kappa$ توزیع‌های خروجی پس از اعمال کرنل باشند، آنگاه برای هر fـ واگرایی داریم:*

$$D_f(P\kappa\|Q\kappa) \leqslant D_f(P\|Q) \qquad (۳۳-۲)$$

تفسیر در حریم خصوصی: این قضیه تضمین می‌کند که اگر یک مهاجم نتواند دو دیتابیس را بر اساس خروجی مکانیزم از هم تشخیص دهد، با انجام هیچ‌گونه پس‌پردازشی[۴۶] روی آن خروجی نیز قادر به بهبود توان تشخیص خود نخواهد بود. به عبارت دیگر، اطلاعات (و حریم خصوصی) با پردازش بیش‌تر، «خلق» یا «تخریب» نمی‌شود.

[43] Monotonicity
[45] Markov Kernel / Probability Kernel
[46] Post-processing

تحدب مشترک

تابع fـ واگرایی نسبت به جفت توزیع‌های ورودی خود، محدب است.

قضیه‌ی ۲-۱۰ (تحدب مشترک[۴۷]) *نگاشت $(P,Q) \mapsto D_f(P\|Q)$ یک تابع محدب مشترک است. یعنی برای هر $\lambda \in [۰,۱]$ و توزیع‌های $P_۱, P_۲, Q_۱, Q_۲$:*

$$D_f\left(\lambda P_۱ + (۱-\lambda)P_۲ \| \lambda Q_۱ + (۱-\lambda)Q_۲\right) \leqslant \lambda D_f(P_۱\|Q_۱) + (۱-\lambda) D_f(P_۲\|Q_۲) \quad (۲\text{-}۳۴)$$

این ویژگی در تحلیل مکانیزم‌هایی که ترکیبی از چند مکانیزم ساده‌تر هستند، بسیار کاربرد دارد.

روابط بین واگرایی‌ها

اگرچه انتخاب‌های مختلف f معیارهای متفاوتی تولید می‌کنند، اما این معیارها مستقل نیستند و می‌توان آن‌ها را با یکدیگر مرتبط ساخت.

- **نامساوی پینسکر[۴۸]:** این نامساوی مشهور، ارتباط هندسه (فاصله تغییرات کل) و اطلاعات (واگرایی KL) را برقرار می‌کند و نشان می‌دهد که همگرایی در آنتروپی نسبی، همگرایی در نرم $L_۱$ را تضمین می‌کند:

$$\|P-Q\|_{TV} \leqslant \sqrt{\frac{۱}{۲}\,\mathrm{KL}(P\|Q)} \quad (۲\text{-}۳۵)$$

- **واگرایی E_γ (چوب هاکی) و ارتباط با TV:** واگرایی E_γ که به صورت $E_\gamma(P\|Q) = \int (dP - \gamma dQ)^+$ تعریف می‌شود، تعمیمی از فاصله تغییرات کل است. به طور مشخص، در نقطه $\gamma = ۱$ این دو معیار بر هم منطبق می‌شوند:

$$E_۱(P\|Q) = \|P-Q\|_{TV} = \frac{۱}{۲}\|P-Q\|_۱ \quad (۲\text{-}۳۶)$$

این ویژگی نشان می‌دهد که E_γ طیفی از فواصل را می‌سازد که از هندسه محض ($\gamma = ۱$) شروع شده و به معیارهای اطلاعاتی می‌رسد.

- **نمایش انتگرالی fـ واگرایی‌ها[۴۹]:** یکی از عمیق‌ترین نتایج نظری در مقاله ساسون و وردو [۱۷] (قضیه ۱۱)، بیان می‌کند که E_γ به عنوان یک «مؤلفه‌ی سازنده» یا «پایه» برای تمامی fـ واگرایی‌ها

[48] Pinsker's Inequality
[49] Integral Representation of f-Divergences

عمل می‌کند. هر fـ واگرایی محدب D_f (برای تابع f دوبار مشتق‌پذیر با $f(1) = 0$) را می‌توان به صورت ترکیب انتگرالی خطی از واگرایی‌های E_γ بازنویسی کرد:

$$D_f(P\|Q) = \int_1^\infty \left(f''(\gamma)E_\gamma(P\|Q) + \gamma^{-3}f''(\gamma^{-1})E_\gamma(Q\|P)\right) d\gamma \qquad (۳۷-۲)$$

اهمیت ریاضی این رابطه در آن است که اثبات قضایا و نامساوی‌ها را بسیار ساده می‌کند؛ اگر بتوانیم یک ویژگی (مانند کران‌دار بودن یا تحدب) را برای E_γ اثبات کنیم، به دلیل مثبت بودن f'' (ناشی از تحدب f) و خطی بودن انتگرال، آن ویژگی به صورت خودکار برای تقریباً تمام خانواده‌ی f- واگرایی‌ها (شامل KL، χ^2 و هلینگر) تعمیم می‌یابد.

جدول ۲-۱: خانواده fـ واگرایی‌ها و توابع مولد آن‌ها

نام واگرایی	تابع مولد $f(t)$	نماد ریاضی
کالبک‌ـ لیبلر (KL)	$t\log t$	$\mathrm{KL}(P\|Q)$
کای‌ـ دو (χ^2)	$(t-1)^2$	$D_{\chi^2}(P\|Q)$
تغییرات کل (TV)	$\frac{1}{2}\|t-1\|$	$\mathrm{TV}(P\|Q)$
هاکی‌ـ استیک	$\max\{0, t-e^\varepsilon\}$	$E_{e^\varepsilon}(P\|Q)$

۲-۳ مبانی آماری و کران‌های اطلاعاتی

در بخش ۲-۲، ابزارهای سنجش فاصله بین توزیع‌ها (خانواده fـ واگرایی‌ها) را معرفی کردیم. در این بخش، قصد داریم با ایجاد پلی میان این مفاهیم انتزاعی و مسأله‌ی عملیاتی «تخمین پارامتر»، زیربنای ریاضی لازم برای تحلیل عملکرد تخمین‌گرها را بنا کنیم. تعاریف و لم‌های ارائه شده در ادامه، ابزارهای اصلی ما برای اثبات کران‌های پایین مینی‌مکس در فصل‌های آتی خواهند بود.

۲-۳-۱ نظریه تصمیم و ریسک مینی‌مکس

در چارچوب «نظریه تصمیم آماری»[50]، مسئله‌ی تخمین را می‌توان به عنوان یک بازی بین «طبیعت» (که پارامتر θ را انتخاب می‌کند) و «آمارگر» (که تخمین‌گر $\hat{\theta}$ را انتخاب می‌کند) مدل‌سازی کرد.

فرض کنید فضای نمونه $\mathcal{X}^n$ و کلاس مدل‌های آماری $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ داده شده‌اند. داده‌های مشاهده‌شده $X^n = (X_1, \ldots, X_n)$ متغیرهای تصادفی مستقلی هستند که از توزیع P_θ تولید شده‌اند. برای

[50] Statistical Decision Theory

سنجش کیفیت یک تخمین‌گر $\hat{\theta}$، به یک معیار فاصله‌ی متریک (یا شبه‌متریک) $\rho : \Theta \times \Theta \to \mathbb{R}_{\geqslant}$. روی فضای پارامتر نیاز داریم.

تعریف ۲-۲۱ (تابع زیان[۵۱]) *میزان جریمه‌ی ناشی از تخمین پارامتر θ با مقدار $\hat{\theta}$، توسط تابع زیان اندازه‌گیری می‌شود که معمولاً تابعی صعودی از فاصله متریک است:*

$$L\left(\hat{\theta}(X^n), \theta\right) \triangleq \Phi\left(\rho\left(\hat{\theta}(X^n), \theta\right)\right), \tag{۳۸-۲}$$

که در آن $\Phi : [۰, \infty) \to [۰, \infty)$ تابعی غیرنزولی با شرط $\Phi(۰) = ۰$ است. انتخاب‌های رایج شامل $\Phi(t) = t^۲$ (زیان مربعات) و $\Phi(t) = |t|$ (زیان قدرمطلق) است.

از آن‌جا که مشاهدات X^n تصادفی هستند، مقدار زیان نیز یک متغیر تصادفی است. برای حذف تصادف، از امیدریاضی زیان استفاده می‌کنیم.

تعریف ۲-۲۲ (ریسک نقطه‌ای[۵۲]) *ریسک یک تخمین‌گر ثابت $\hat{\theta}$ در نقطه پارامتری مشخص θ، برابر است با میانگین زیان وارده تحت توزیع P_θ:*

$$R(\hat{\theta}, \theta) \triangleq \mathbb{E}_{X^n \sim P_\theta^n}\left[L\left(\hat{\theta}(X^n), \theta\right)\right]. \tag{۳۹-۲}$$

از آن‌جا که پارامتر واقعی θ ناشناخته است، نمی‌توان تخمین‌گری یافت که ریسک نقطه‌ای آن برای تمام مقادیر θ کمینه باشد. برای حل این مشکل، رویکرد «مینی‌مکس»[۵۳] بدترین حالت ممکن را در نظر می‌گیرد.

تعریف ۲-۲۳ (ریسک مینی‌مکس) *ریسک مینی‌مکس برای کلاس $\mathcal{P}$ و متریک ρ، برابر است با کم‌ترین مقدارِ بیشینه‌ریسک ممکن بر روی تمامی تخمین‌گرهای اندازه‌پذیر. به بیان دقیق ریاضی:*

$$\begin{aligned}\mathfrak{M}_n(\Theta, \Phi \circ \rho) &\triangleq \inf_{\hat{\theta}} \sup_{\theta \in \Theta} R(\hat{\theta}, \theta) \\ &= \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{X^n \sim P_\theta^n}\left[\Phi\left(\rho\left(\hat{\theta}(X^n), \theta\right)\right)\right].\end{aligned} \tag{۴۰-۲}$$

هدف اصلی نظریه مینی‌مکس، یافتن کران‌های پایین و بالا برای این کمیت بر حسب اندازه نمونه n است.

۲-۳-۲ تقلیل به آزمون فرض (روش بسته‌بندی)

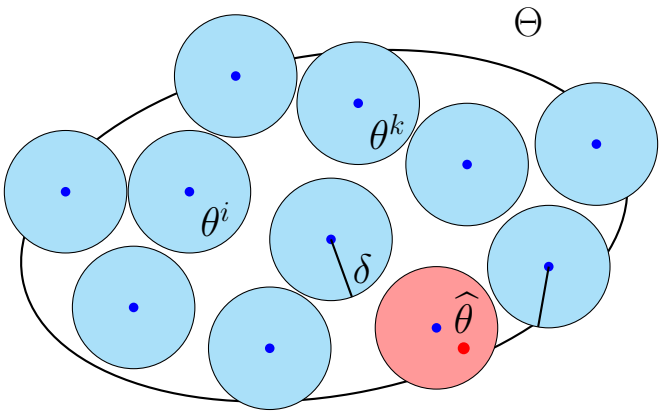
محاسبه مستقیم مقدار (۲-۴۰) دشوار است. استراتژی استاندارد ما برای غلبه بر این دشواری، «تبدیل فضای پیوسته پارامتر به یک مجموعه گسسته» است. به عبارت دیگر، مسئله‌ی «تخمین» را به مسئله‌ی «آزمون فرض چندگزینه‌ای» تقلیل می‌دهیم.

[53] Minimax Approach

این فرآیند بر پایه این شهود استوار است که هر تخمین‌گر دقیق، باید بتواند بین پارامترهایی که فاصله‌ی تقریباً زیادی از هم دارند، تمایز قائل شود. برای رسمی‌سازی این ایده، از مفهوم «بسته‌بندی» استفاده می‌کنیم.

تعریف ۲-۲۴ (مجموعه بسته‌بندی) *مجموعه‌ی متناهی $\mathcal{V} = \{\theta_1, \dots, \theta_M\} \subset \Theta$ یک 2δ ـ بسته‌بندی*[54] *برای فضای (Θ, ρ) نامیده می‌شود اگر اعضای آن حداقل به اندازه 2δ از یک‌دیگر فاصله داشته باشند:*

$$\min_{i \neq j} \rho(\theta_i, \theta_j) \geqslant 2\delta \tag{۴۱-۲}$$



شکل ۲-۳: نمایش هندسی روش بسته‌بندی برای تقلیل مسئله‌ی تخمین به آزمون فرض. فضای پارامتر Θ با مجموعه‌ای از گوی‌های مجزا $\{\theta^1, \dots, \theta^K\}$ پوشانده شده است که تشکیل یک 2δ ـ بسته‌بندی می‌دهند. اگر تخمین‌گر $\widehat{\theta}$ (نقطه قرمز) دارای خطای کم‌تر از δ باشد، لزوماً درون یکی از این گوی‌ها قرار می‌گیرد. از آن‌جا که گوی‌ها مجزا هستند، حداکثر یک اندیس i وجود دارد که $\widehat{\theta}$ به آن نزدیک باشد؛ لذا مسئله تخمین به یافتن این اندیس (آزمون فرض چندگزینه‌ای) تقلیل می‌یابد.

می‌توانیم قضیه‌ی زیر را که به صورت عمومی برای کران پایین ریسک مینی‌مکس ثابت کنیم:

قضیه‌ی ۲-۱۱ (کران پایین عمومی) *فرض کنید $\{\theta_1, \dots, \theta_M\}$ یک 2δ ـ بسته‌بندی برای Θ باشد. آن‌گاه ریسک مینی‌مکس به خطای آزمون فرض محدود می‌شود:*

$$\mathfrak{M}_n(\Theta) \geqslant \Phi(\delta) \cdot \inf_{\hat{\theta}} \sup_{\theta \in \mathcal{V}} \Pr_{P_\theta^n}\left(\rho\left(\hat{\theta}, \theta\right) \geqslant \delta\right) \tag{۴۲-۲}$$

$$\geqslant \Phi(\delta) \cdot \inf_{\psi} \bar{P}_{err}(\psi) \tag{۴۳-۲}$$

که در آن $\bar{P}_{err}(\psi) = \frac{1}{M}\sum_{v=1}^{M} P_{\theta_v}^n(\psi(X^n) \neq v)$ میانگین احتمال خطا روی فرضیه‌های گسسته است و ψ آزمون‌گری است که سعی در بازیافت اندیس v دارد.

[54] Packing

اثبات. برای اثبات این قضیه، از تعریف ریسک مینی‌مکس و خواص مجموعه‌های بسته‌بندی استفاده می‌کنیم. روند اثبات طی چند مرحله‌ی منطقی به شرح زیر انجام می‌شود:

۱. محدود کردن فضای پارامتر: طبق تعریف ریسک مینی‌مکس (رابطه ۲-۴۰)، ماکزیمم ریسک روی تمام فضای Θ محاسبه می‌شود. از آن‌جا که مجموعه‌ی بسته‌بندی $\mathcal{V} = \{\theta_1, \dots, \theta_M\}$ زیرمجموعه‌ای از Θ است ($\mathcal{V} \subset \Theta$)، سوپریمم روی Θ همواره بزرگ‌تر یا مساوی سوپریمم روی $\mathcal{V}$ خواهد بود. هم‌چنین، ماکزیمم خطا همواره از میانگین خطا بزرگ‌تر است. بنابراین داریم:

$$\begin{aligned}\mathfrak{M}_n(\Theta) &= \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{P_\theta}\left[\Phi(\rho(\hat{\theta}, \theta))\right] \\ &\geqslant \inf_{\hat{\theta}} \sup_{\theta \in \mathcal{V}} \mathbb{E}_{P_\theta}\left[\Phi(\rho(\hat{\theta}, \theta))\right] \\ &\geqslant \inf_{\hat{\theta}} \frac{1}{M} \sum_{v=1}^{M} \mathbb{E}_{P_{\theta_v}}\left[\Phi(\rho(\hat{\theta}, \theta_v))\right] \end{aligned} \tag{۲-۴۴}$$

۲. استفاده از نامساوی مارکوف: از آن‌جا که Φ تابعی صعودی و غیرمنفی است، می‌توانیم از نامساوی مارکوف استفاده کنیم. برای هر θ ثابت داریم:

$$\begin{aligned}\mathbb{E}\left[\Phi(\rho(\hat{\theta}, \theta))\right] &\geqslant \mathbb{E}\left[\Phi(\rho(\hat{\theta}, \theta)) \cdot \mathbb{I}(\rho(\hat{\theta}, \theta) \geqslant \delta)\right] \\ &\geqslant \Phi(\delta) \cdot \Pr\left(\rho(\hat{\theta}, \theta) \geqslant \delta\right) \end{aligned} \tag{۲-۴۵}$$

۳. ساخت آزمون‌گر از روی تخمین‌گر: فرض کنید متغیر تصادفی V به صورت یکنواخت از بین اندیس‌های $\{1, \dots, M\}$ انتخاب شود. برای هر تخمین‌گر دل‌خواه $\hat{\theta}$، می‌توانیم یک آزمون‌گر (تست) $\psi : \mathcal{X}^n \to \{1, \dots, M\}$ بر اساس قاعده‌ی «کم‌ترین فاصله» بسازیم:

$$\psi(X^n) = \arg \min_{k \in \{1, \dots, M\}} \rho(\hat{\theta}(X^n), \theta_k) \tag{۲-۴۶}$$

۴. تحلیل خطا با نامساوی مثلث: نکته‌ی کلیدی این است که نشان دهیم اگر تخمین‌گر $\hat{\theta}$ به پارامتر واقعی نزدیک باشد، آزمون‌گر لزوماً اندیس درست را باز می‌گرداند. فرض کنید اندیس واقعی v باشد. اگر آزمون‌گر دچار خطا شود (یعنی $\psi(X^n) \neq v$)، به این معنی است که یک اندیس $k \neq v$ وجود دارد که تخمین‌گر به θ_k نزدیک‌تر (یا مساوی) است تا به θ_v:

$$\rho(\hat{\theta}, \theta_k) \leqslant \rho(\hat{\theta}, \theta_v) \tag{۲-۴۷}$$

از طرفی، چون $\mathcal{V}$ یک 2δـ بسته‌بندی است، طبق نامساوی مثلث داریم:

$$\begin{aligned} 2\delta &\leqslant \rho(\theta_v, \theta_k) \quad \text{(تعریف بسته‌بندی)} \\ &\leqslant \rho(\theta_v, \hat{\theta}) + \rho(\hat{\theta}, \theta_k) \quad \text{(نامساوی مثلث)} \\ &\leqslant \rho(\theta_v, \hat{\theta}) + \rho(\hat{\theta}, \theta_v) \quad \text{(شرط خطای آزمون)} \\ &= 2\rho(\hat{\theta}, \theta_v) \end{aligned} \tag{۴۸-۲}$$

نتیجه می‌گیریم که وقوع خطای آزمون ($\psi \neq v$) مستلزم وقوع خطای تخمین بزرگ ($\rho \geqslant \delta$) است:

$$\{\psi(X^n) \neq v\} \subseteq \{\rho(\hat{\theta}, \theta_v) \geqslant \delta\} \tag{۴۹-۲}$$

و در نتیجه برای احتمالات داریم:

$$\Pr_{P_{\theta_v}}(\psi(X^n) \neq v) \leqslant \Pr_{P_{\theta_v}}(\rho(\hat{\theta}, \theta_v) \geqslant \delta) \tag{۵۰-۲}$$

۵. جمع‌بندی: با جایگذاری ۲-۵۰ در نامساوی مارکوف (۲-۴۵) و میانگین‌گیری روی تمام vها (مرحله ۱)، داریم:

$$\begin{aligned} \mathfrak{M}_n(\Theta) &\geqslant \frac{1}{M} \sum_{v=1}^{M} \Phi(\delta) \cdot \Pr(\rho(\hat{\theta}, \theta_v) \geqslant \delta) \\ &\geqslant \Phi(\delta) \cdot \frac{1}{M} \sum_{v=1}^{M} \Pr(\psi(X^n) \neq v) \\ &= \Phi(\delta) \cdot \bar{P}_{err}(\psi) \end{aligned} \tag{۵۱-۲}$$

چون این رابطه برای هر تخمین‌گر $\hat{\theta}$ (و آزمون‌گر متناظر ψ) برقرار است، پس برای اینفیمم آن‌ها نیز صادق است. □

۲-۳-۳ نامساوی‌های کران پایین

اکنون که مسئله را به آزمون فرض روی M نقطه تقلیل دادیم، برای اثبات کران‌های پایین نهایی نیاز به ابزارهایی داریم که $\bar{P}_{err}$ را از پایین محدود کنند. سه روش اصلی که بر پایه fـ واگرایی‌ها بنا شده‌اند عبارتند از:

قضیه‌ی ۲-۱۲ (نامساوی لو کم[۵۵]) *این روش معمولاً برای آزمون بین دو توزیع P_1 و P_2 ($M = 2$) استفاده می‌شود و برای «کران‌های محلی» حول یک نقطه مناسب است.*

$$\inf_{\psi} \bar{P}_{err}(\psi) \geqslant \frac{1}{2}\left(1 - \|P_1^n - P_2^n\|_{TV}\right) \tag{۵۲-۲}$$

تفسیر: *این نامساوی بیان می‌کند که اگر فاصله TV بین دو توزیع کم باشد، همپوشانی آن‌ها زیاد است و هیچ آزمون‌گری نمی‌تواند با خطای ناچیز آن‌ها را تفکیک کند.*

قضیه‌ی ۲-۱۳ (نامساوی فانو[۵۶]) *زمانی که پارامتر مورد نظر متعلق به مجموعه‌ای بزرگ‌تر $\mathcal{V}$ باشد $(M = |V| > ۲)$، نامساوی فانو کران پایین قوی‌تری ارائه می‌دهد:*

$$\inf_{\psi} \bar{P}_{err}(\psi) \geqslant ۱ - \frac{I\left(X^n; V\right) + \log ۲}{\log M} \tag{۵۳-۲}$$

که در آن $I(X^n;V)$ اطلاعات متقابل بین داده‌ها و اندیس پارامتر است.

تفسیر: *نامساوی فانو مسئله خطا را به اطلاعات متقابل $I(X^n;V)$ گره می‌زند. اگر داده‌ها حاوی اطلاعات کافی درباره اندیس واقعی V نباشند (ظرفیت کانال نسبت به تعداد فرضیه‌ها $\log M$ کم باشد)، خطا اجتناب‌ناپذیر است.*

لم ۲-۱۴ (لم اسود[۵۷]) *این لم ابزاری قدرتمند برای فضاهای پارامتر با ابعاد بالا (مانند $\{-۱, ۱\}^d$) است. قدرت لم اسود در شکستن یک مسئله dـ بعدی دشوار به d مسئله ۱ ـ بعدی مستقل است.*

$$\mathfrak{M}_n(\Theta) \geqslant \frac{\delta}{۲} \sum_{j=۱}^{d} \left[۱ - \left\| M_{+j}^n - M_{-j}^n \right\|_{TV} \right] \tag{۵۴-۲}$$

که در آن M_{+j}^n و M_{-j}^n توزیع‌های مخلوط حاشیه‌ای هستند (میانگین توزیع‌هایی که بیت jـ ام آن‌ها به ترتیب $+۱$ و -۱ است). اگر در هر بُعد تمایز قائل شدن سخت باشد، مجموع خطاها با تعداد ابعاد d جمع شده و کران دقیقی[۵۸] می‌سازد.

[58] Tight Bound

فصل ۳

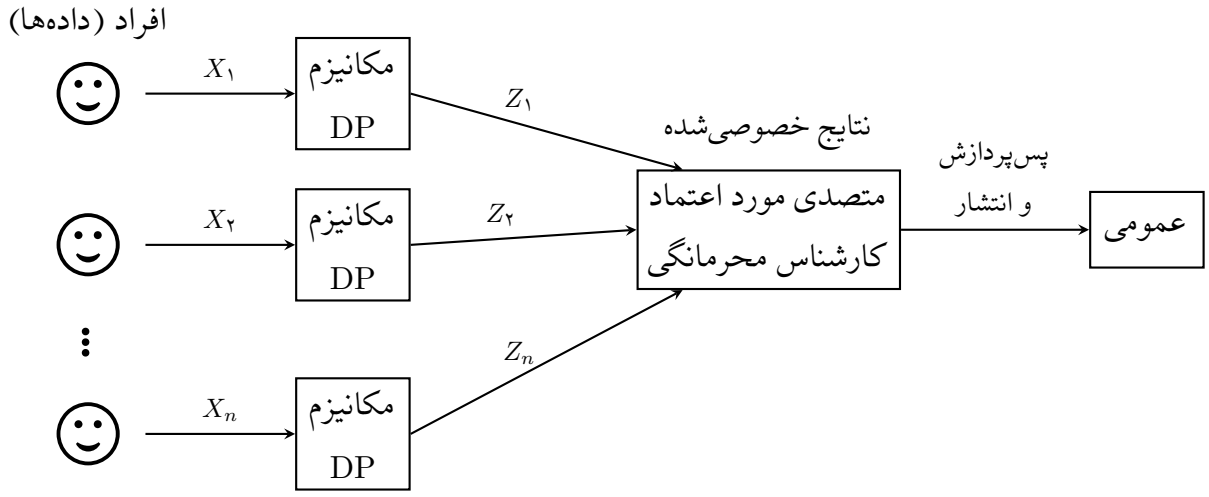
محرمانگی تفاضلی موضعی

۳-۱ مقدمه

در فصل پیشین (۲)، مفاهیم بنیادی نظریه اندازه و مدل محرمانگی تفاضلی متمرکز (CDP) را بررسی کردیم. همان‌طور که در بخش ۲-۱ مشاهده شد، مدل متمرکز (CDP) بر فرض وجود یک متصدی مورد اعتماد استوار است که به داده‌های خام تمامی کاربران دسترسی دارد ($M : \mathcal{X}^n \to \mathcal{R}$). اگرچه این مدل دقت آماری بالایی را فراهم می‌کند، اما ذخیره‌سازی متمرکز داده‌ها یک «نقطه شکست مرکزی» ایجاد می‌کند؛ به این معنا که نفوذ به سرور یا خیانت متصدی، حریم خصوصی تمامی کاربران را به خطر می‌اندازد. برای مثال در بسیاری از کاربردهای مدرن، مانند جمع‌آوری داده‌های تله‌متری مرورگرها یا اپلیکیشن‌های موبایل، اعتماد به سرور مرکزی خطرات امنیتی و چالش‌های حقوقی را به همراه دارد.

در پاسخ به این چالش، مدل «محرمانگی تفاضلی موضعی»[1] یا به اختصار LDP پارادایم اعتماد را تغییر می‌دهد. در این مدل، هیچ موجودیتی (حتی سرور) به داده‌ی خام X_i دسترسی ندارد؛ بلکه هر کاربر به صورت مستقل مکانیزم تصادفی $\mathcal{M}_i$ را روی داده‌ی خود اجرا کرده و تنها خروجی نویزدار Z_i را منتشر می‌کند (شکل ۳-۱).

[1] Local Differential Privacy



شکل ۳-۱: گذار از مدل متمرکز به موضعی؛ نویز به صورت موضعی روی دستگاه کاربر اضافه می‌شود.

۳-۲ تعاریف رسمی و مدل‌های محاسباتی

در مدل محرمانگی موضعی، فرض می‌کنیم n کاربر در سیستم حضور دارند. داده‌های خصوصی این کاربران را می‌توان به صورت دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع[۲] $X_1, \ldots, X_n$ مدل‌سازی کرد که هر یک از توزیع احتمال ناشناخته‌ی P روی دامنه‌ی $\mathcal{X}$ نمونه‌برداری شده‌اند:

$$X_i \overset{i.i.d.}{\sim} P, \quad X_i \in \mathcal{X}, \quad i = 1, \ldots, n. \tag{۳-۱}$$

هدف نهایی تحلیل‌گر در این بستر، معمولاً تخمین ویژگی‌های آماری این توزیع P (مانند میانگین یا چگالی احتمال) بر اساس خروجی‌های نویزدار است.

تفاوت بنیادین این مدل با مدل متمرکز (CDP) در قلمروی تعریف «همسایگی» نهفته است. در حالی که در مدل متمرکز، شرط محرمانگی روی «پایگاه‌داده‌های همسایه» (که تنها در یک رکورد تفاوت دارند) اعمال می‌شود، در مدل موضعی این شرط باید برای هر جفت ورودی ممکن $x, x' \in \mathcal{X}$ برقرار باشد. به عبارت دیگر، مکانیزم باید چنان عمل کند که تمایز قائل شدن بین هر دو مقدار ورودی برای مهاجم (که تنها خروجی مکانیزم را مشاهده می‌کند) دشوار گردد.

تعریف ۳-۱ (تصادفی‌ساز موضعی[۳]) *یک مکانیزم تصادفی $\mathcal{M} : \mathcal{X} \to \mathcal{Z}$ را یک تصادفی‌ساز موضعی می‌نامیم که توسط یک هسته‌ی انتقال احتمال[۴] یا توزیع شرطی Q مشخص می‌شود. به بیان دقیق ریاضی، $Q = \{Q(\cdot|x) : x \in \mathcal{X}\}$ خانواده‌ای از توزیع‌های احتمال روی فضای خروجی $\mathcal{Z}$ است که برای هر ورودی*

[2] Independent and Identically Distributed (i.i.d.)
[4] Probability Transition Kernel

ثابت $x \in \mathcal{X}$، رفتار احتمالی مکانیزم را تعیین می‌کند:

$$Q(z|x) \triangleq \Pr\left[\mathcal{M}(x) = z\right], \quad \forall z \in \mathcal{Z}. \tag{۲-۳}$$

در حالتی که فضاهای $\mathcal{X}$ و $\mathcal{Z}$ متناهی باشند (با اندازه‌های $|\mathcal{X}| = k$ و $|\mathcal{Z}| = m$)، Q را می‌توان به صورت یک ماتریس تصادفی سطری[۵] با ابعاد $k \times m$ نمایش داد که در آن درایه (x, z) برابر با احتمال خروجی دادن z به شرط ورودی x است و مجموع عناصر هر سطر برابر با یک می‌باشد.

مثال ۳-۱ (کانال دودویی متقارن به عنوان تصادفی‌ساز) *ساده‌ترین نمونه از یک تصادفی‌ساز موضعی، مدلی است که در آن فضای ورودی و خروجی هر دو دودویی هستند ($\mathcal{X} = \mathcal{Z} = \{۰, ۱\}$). فرض کنید مکانیزم $\mathcal{M}$، بیت ورودی را با احتمال p حفظ کرده و با احتمال $p - ۱$ تغییر می‌دهد (فلیپ می‌کند). در این صورت ماتریس انتقال Q به شکل زیر خواهد بود:*

$$Q = \begin{pmatrix} Q(۰|۰) & Q(۱|۰) \\ Q(۰|۱) & Q(۱|۱) \end{pmatrix} = \begin{pmatrix} p & ۱-p \\ ۱-p & p \end{pmatrix}. \tag{۳-۳}$$

این ماتریس دقیقاً بیانگر عملکرد مکانیزم مشهور «پاسخ تصادفی»[۶] است که در ادامه به تعریف رسمی آن می‌پردازیم.

۳-۲-۱ تعریف ریاضی LDP

هسته‌ی اصلی این مدل، تضمین این نکته است که توزیع‌های خروجی برای هر دو ورودی متمایز، از نظر آماری بسیار به هم نزدیک باشند.

تعریف ۳-۲ (α-LDP) *فرض کنید $\mathcal{M} : \mathcal{X} \to \mathcal{Z}$ یک مکانیزم تصادفی باشد که به ازای هر ورودی $x \in \mathcal{X}$، خروجی z را بر اساس توزیع احتمال شرطی $Q(\cdot|x)$ تولید می‌کند. مکانیزم $\mathcal{M}$ دارای ویژگی α-LDP است، اگر برای تمام جفت ورودی‌های ممکن $x, x' \in \mathcal{X}$ و برای هر زیرمجموعه‌ی اندازه‌پذیر $\mathcal{S} \subseteq \mathcal{Z}$ (در σـ جبرِ برد)، نامساوی زیر برقرار باشد:*

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leqslant e^{\alpha} \cdot \Pr[\mathcal{M}(x') \in \mathcal{S}] \tag{۴-۳}$$

یا به بیان دقیق‌تر ریاضی با استفاده از سوپریمم روی تمام پیشامدها:

$$\sup_{x, x' \in \mathcal{X}} \sup_{\mathcal{S} \in \sigma(\mathcal{Z})} \frac{Q(\mathcal{S}|x)}{Q(\mathcal{S}|x')} \leqslant e^{\alpha} \tag{۵-۳}$$

[5] Row-stochastic Matrix
[6] Randomized Response

اگر توزیع‌های خروجی $Q(\cdot|x)$ نسبت به یک اندازه پایه μ مطلقاً پیوسته باشند (یعنی تابع چگالی $q(\cdot|x)$ وجود داشته باشد)، شرط بالا معادل است با کران‌دار بودن نسبت چگالی‌ها در تمام نقاط [۶]:

$$\sup_{z\in\mathcal{Z}}\sup_{x,x'\in\mathcal{X}}\frac{q(z|x)}{q(z|x')}\leqslant e^{\alpha} \tag{۶-۳}$$

این تعریف تضمین می‌کند که برای هر خروجی مشاهده‌شده‌ی z، نسبت درست‌نمایی[۷] اینکه ورودی x بوده است یا x'، با مقدار e^{α} محدود می‌شود و در نتیجه تمایز قائل شدن بین این دو ورودی برای مهاجم دشوار می‌گردد.

نکته: در متون آماری این حوزه (مانند [۶])، معمولاً از پارامتر α برای بودجه‌ی محرمانگی موضعی استفاده می‌شود تا تمایز آن با پارامتر ε در مدل متمرکز مشخص گردد. ما نیز در این فصل و فصول بعدی از این نمادگذاری پیروی می‌کنیم.

این تعریف را می‌توان با استفاده از مفهوم «واگرایی ماکزیمم» (D_{∞}) که پیش‌تر معرفی شد، به صورت فشرده‌تری بیان کرد. شرط (۳-۵) دقیقاً معادل است با:

$$\sup_{x,x'\in\mathcal{X}} D_{\infty}\left(Q(\cdot|x)\parallel Q(\cdot|x')\right)\leqslant\alpha \tag{۷-۳}$$

این رابطه نشان می‌دهد که α-LDP محدودیتی سخت‌گیرانه بر روی «نسبت درست‌نمایی»[۸] توزیع‌های خروجی اعمال می‌کند و تضمین می‌دهد که مشاهده‌ی خروجی z، اطلاعات اندکی درباره‌ی ورودی x افشا می‌کند.

۳-۲-۲ محرمانگی تقریبی

مشابه مدل متمرکز، در برخی کاربردها نیاز است که تعریف α-LDP را تضعیف کنیم تا اجازه‌ی یک احتمال شکست ناچیز δ داده شود. این حالت معمولاً زمانی رخ می‌دهد که دامنه یا برد مکانیزم نامتناهی باشد (مانند مکانیزم گوسی).

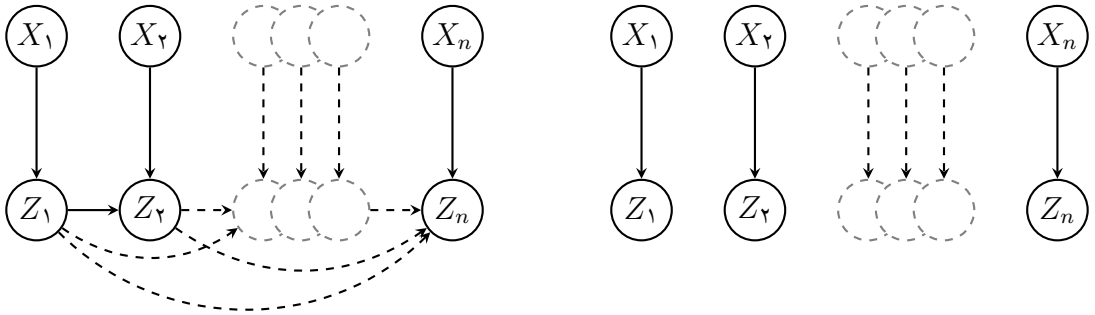
تعریف ۳-۳ ((α,δ)-LDP) *یک مکانیزم $\mathcal{M}$ دارای محرمانگی تفاضلی موضعی تقریبی است اگر برای تمام ورودی‌های $x,x'\in\mathcal{X}$ و تمام زیرمجموعه‌های خروجی $\mathcal{S}\subseteq\mathcal{Z}$ داشته باشیم:*

$$\mathbb{P}\left[\mathcal{M}(x)\in\mathcal{S}\right]\leqslant e^{\alpha}\cdot\mathbb{P}\left[\mathcal{M}(x')\in\mathcal{S}\right]+\delta \tag{۸-۳}$$

بدیهی است که اگر $\delta=0$ باشد، این تعریف به حالت α-LDP خالص باز می‌گردد [۲۰].

[7] Likelihood Ratio
[8] Likelihood Ratio

۳-۳ پروتکل‌های تعاملی و خواص ترکیب



شکل ۳-۲: (الف) مدل گرافیکی نمایش‌دهنده‌ی روابط استقلال شرطی بین داده‌های خصوصی $\{X_i\}_{i=1}^n$ و متغیرهای مشاهده‌شده $\{Z_i\}_{i=1}^n$ در پروتکل‌های تعاملی؛ پیکان‌های افقی نشان‌دهنده‌ی وابستگی Z_i به تاریخچه‌ی پیشین هستند. (ب) مدل گرافیکی ساده‌تر در پروتکل‌های غیرتعاملی که در آن خروجی‌ها به شرط ورودی‌ها از یکدیگر مستقل هستند.

برای تحلیل دقیق حدود مینی‌مکس و درک محدودیت‌های بنیادین LDP، نیازمند مدل‌سازی دقیق نحوه‌ی تعامل کاربران با سرور (یا جمع‌آورنده داده) هستیم. دوچی و همکاران [۶] پروتکل‌های موضعی را بر اساس ساختار وابستگی آماری خروجی‌ها به دو دسته‌ی کلی تقسیم می‌کنند (شکل ۳-۲ را ببینید): غیرتعاملی و تعاملی.

۱-۳-۳ پروتکل‌های غیرتعاملی

در پروتکل‌های غیرتعاملی[9]، تمام کاربران $i = 1, \dots, n$ مکانیزم‌های خود را به صورت هم‌زمان و مستقل از یک‌دیگر اجرا می‌کنند. اگر Z_i خروجی کاربر iـام باشد، توزیع آن تنها به داده‌ی خصوصی X_i وابسته است و هیچ وابستگی‌ای به خروجی سایر کاربران ندارد. به بیان ریاضی، توزیع مشترک خروجی‌ها به صورت حاصل‌ضرب توزیع‌های حاشیه‌ای فاکتور می‌شود:

$$\mathbb{P}(Z_1, \dots, Z_n | X_1, \dots, X_n) = \prod_{i=1}^{n} Q_i(Z_i | X_i) \qquad (۹-۳)$$

بسیاری از پیاده‌سازی‌های صنعتی فعلی، از جمله RAPPOR گوگل [۱۰]، در این دسته قرار می‌گیرند.

[9] Non-interactive

۳-۳-۲ پروتکل‌های تعاملی (ترتیبی)

در پروتکل‌های تعاملی[10]، کاربران به نوبت داده‌های خود را ارسال می‌کنند و مکانیزم کاربر iـام می‌تواند به خروجی‌های مشاهده‌شده از کاربران پیشین $(Z_1, \dots, Z_{i-1})$ وابسته باشد. این مدل، آزادی عمل بیشتری را برای طراحی الگوریتم‌های تطبیقی فراهم می‌کند.

از دیدگاه آنالیز ریاضی، این فرآیند با استفاده از «هسته‌های احتمالاتی»[11] مدل‌سازی می‌شود. فرض کنید $Z_{1:i-1} = (Z_1, \dots, Z_{i-1})$ بردار خروجی‌های پیشین باشد که σـ میدانِ $\mathcal{F}_{i-1}$ را تولید می‌کند. مکانیزم کاربر iـام، یک هسته احتمالاتی Q_i است که خروجی $Z_i \in \mathcal{Z}$ را مشروط بر داده‌ی خصوصی X_i و تاریخچه‌ی عمومی $Z_{1:i-1}$ تولید می‌کند:

$$Z_i \sim Q_i(dz_i | x_i, z_{1:i-1}) \qquad (۳-۱۰)$$

شرط اساسی محرمانگی در اینجا این است که با شرطی‌سازی روی X_i و $Z_{1:i-1}$، متغیر Z_i باید از سایر داده‌های خصوصی $X_{j \neq i}$ مستقل باشد (شرط مارکوفی). این ساختار به پروتکل اجازه می‌دهد تا پارامترهای پرس‌وجو را به صورت پویا بر اساس اطلاعات کسب‌شده از کاربران قبلی تنظیم کند.

۳-۳-۳ قضیه ترکیب ترتیبی

یکی از ویژگی‌های بنیادین LDP، پایداری آن در برابر ترکیب است. اگر یک پروتکل شامل چندین مرحله‌ی تعاملی باشد، بودجه‌های محرمانگی با یک‌دیگر جمع می‌شوند. قضیه‌ی زیر، کران بالای محرمانگی را برای یک پروتکل ترتیبی بیان می‌کند [۶].

قضیه‌ی ۳-۱ (ترکیب ترتیبی[12]) *فرض کنید در یک پروتکل تعاملی، برای هر کاربر $i \in \{1, \dots, n\}$ و به ازای هر تاریخچه‌ی ممکن $z_{1:i-1} \in \mathcal{Z}^{i-1}$، مکانیزم $Q_i(\cdot|\cdot, z_{1:i-1})$ دارای خاصیت α_i-LDP نسبت به ورودی x_i باشد. آنگاه توزیع مشترک کل خروجی‌ها $(Z_1, \dots, Z_n)$، دارای محرمانگی تفاضلی موضعی با بودجه‌ی مجموع است:*

$$\alpha_{total} = \sum_{i=1}^{n} \alpha_i \qquad (۳-۱۱)$$

اثبات. اثبات بر پایه خاصیت زنجیره‌ای واگرایی ماکزیمم (D_∞) یا تجزیه‌ی نسبت‌های درست‌نمایی استوار است. اگر P و P' دو توزیع احتمال روی دنباله‌ی خروجی‌ها Z^n باشند که ناشی از دو دنباله ورودی x^n و

[10] Interactive / Sequential
[11] Probability Kernels

x'^n هستند، نسبت احتمال توأم به حاصل‌ضرب نسبت‌های شرطی تجزیه می‌شود:

$$\frac{P(z^n)}{P'(z^n)} = \prod_{i=1}^{n} \frac{Q_i(z_i|x_i, z_{1:i-1})}{Q_i(z_i|x'_i, z_{1:i-1})}$$

از آن‌جا که هر گام طبق فرض با e^{α_i} کران‌دار است، کل حاصل‌ضرب با $e^{\sum \alpha_i}$ کران‌دار خواهد بود. □

۳-۴ مکانیزم‌های پایه در LDP

در این بخش، مکانیزم‌های بنیادین LDP را با رویکردی آماری تحلیل می‌کنیم. هدف اصلی در طراحی این مکانیزم‌ها، یافتن نگاشتی تصادفی است که علاوه بر ارضای شرط محرمانگی، «خطای تخمین» (که معمولاً با واریانس سنجیده می‌شود) را کمینه کند. فرض بنیادی در تمام این مکانیزم‌ها این است که برای بازیابی اطلاعات آماری (مانند هیستوگرام)، از یک «تخمین‌گر ناریب»[۱۳] معکوس استفاده می‌شود.

۳-۴-۱ پاسخ تصادفی دودویی (RR)

پایه‌ای‌ترین مکانیزم α-LDP، پاسخ تصادفی[۱۴] (RR) برای دامنه‌ی دودویی $\mathcal{X} = \{0, 1\}$ است. این مکانیزم را می‌توان به صورت یک کانال متقارن دودویی مدل‌سازی کرد که ورودی x را با احتمال p حفظ کرده و با احتمال $1 - p$ قرینه می‌کند:

$$\mathbb{P}[y = z|x] = \begin{cases} p & \text{if } z = x \\ 1 - p & \text{if } z \neq x \end{cases} \tag{۱۲-۳}$$

اثبات α-LDP: برای اینکه این مکانیزم شرط α-LDPرا برآورده کند، طبق تعریف ۳-۲ باید نسبت درست‌نمایی برای هر دو ورودی متمایز x, x' و هر خروجی ممکن y، با e^{α} کران‌دار شود. بدترین حالت زمانی رخ می‌دهد که صورت کسر بیشترین احتمال (p) و مخرج کسر کمترین احتمال ($1 - p$) باشد:

$$\sup_{y \in \{0,1\}} \frac{\mathbb{P}[y|x]}{\mathbb{P}[y|x']} = \frac{p}{1-p} \leqslant e^{\alpha} \tag{۱۳-۳}$$

با حل این نامساوی برای p (و با فرض $p > 1/2$ برای بی‌معنی نشدن نتیجه)، مقدار بهینه احتمال حفظ پاسخ برای بودجه‌ی α به دست می‌آید:

$$p = \frac{e^{\alpha}}{e^{\alpha} + 1} \tag{۱۴-۳}$$

[13] Unbiased Estimator
[14] Randomized Response

تحلیل واریانس: برای تحلیل دقیق خطا، ابتدا تخمین‌گر نااریب را استخراج می‌کنیم. اگر $y \in \{۰, ۱\}$ خروجی مکانیزم باشد، هدف یافتن تابعی $\hat{f}(y)$ است که $\mathbb{E}_{\hat{f}(y)} = x$ باشد.

لم ۳-۲ (واریانس پاسخ تصادفی) *برای مکانیزم RR با پارامتر p، تخمین‌گر نااریب ورودی x به صورت زیر است:*

$$\hat{x} = \frac{y - (۱ - p)}{۲p - ۱} \tag{۳-۱۵}$$

و واریانس این تخمین‌گر بر حسب بودجه محرمانگی α برابر است با:

$$\mathrm{Var}[\hat{x}] = \frac{e^{\alpha}}{(e^{\alpha} - ۱)^{۲}} \tag{۳-۱۶}$$

اثبات. ابتدا نااریبی را بررسی می‌کنیم. امید ریاضی y برابر است با:

$$\mathbb{E}_y = p \cdot x + (۱ - p)(۱ - x) = (۲p - ۱)x + (۱ - p)$$

با جایگذاری در معادله تخمین‌گر:

$$\mathbb{E}_{\hat{x}} = \frac{\mathbb{E}_y - (۱ - p)}{۲p - ۱} = \frac{(۲p - ۱)x}{۲p - ۱} = x$$

برای محاسبه واریانس، چون y یک متغیر برنولی است، واریانس آن $p(۱ - p)$ می‌شود. با اعمال خواص واریانس ($\mathrm{Var}[aY + b] = a^{۲}\mathrm{Var}[Y]$) داریم:

$$\mathrm{Var}[\hat{x}] = \frac{\mathrm{Var}[y]}{(۲p - ۱)^{۲}} = \frac{p(۱ - p)}{(۲p - ۱)^{۲}}$$

حال با جایگذاری $p = \frac{e^{\alpha}}{e^{\alpha}+۱}$ در رابطه بالا:

$$\mathrm{Var}[\hat{x}] = \frac{\frac{e^{\alpha}}{(e^{\alpha}+۱)^{۲}}}{\left(\frac{۲e^{\alpha}}{e^{\alpha}+۱} - ۱\right)^{۲}} = \frac{\frac{e^{\alpha}}{(e^{\alpha}+۱)^{۲}}}{\left(\frac{e^{\alpha}-۱}{e^{\alpha}+۱}\right)^{۲}} = \frac{e^{\alpha}}{(e^{\alpha} - ۱)^{۲}}$$

□

۳-۴-۲ پاسخ تصادفی تعمیم‌یافته (GRR)

پاسخ تصادفی تعمیم‌یافته[۱۵]، برای دامنه‌های گسسته با $k > ۲$ عنصر ($\mathcal{X} = \{۱, \ldots, k\}$)، به عنوان تعمیم مستقیم RR معرفی می‌شود. این مکانیزم را می‌توان با یک «ماتریس گذار»[۱۶] تصادفی $\mathbf{Q} \in [۰, ۱]^{k \times k}$ توصیف کرد.

[15] Generalized Randomized Response
[16] Transition Matrix

تعریف ۳-۴ (ماتریس احتمال GRR) *در مکانیزم GRR ، ماتریس احتمال شرطی $\mathbf{Q}$ که درایه (i,j) آن برابر با $\mathbb{P}[z=j|x=i]$ است، به صورت زیر تعریف می‌شود:*

$$\mathbf{Q} = \begin{pmatrix} p & q & \dots & q \\ q & p & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p \end{pmatrix}_{k\times k} \tag{۳-۱۷}$$

که در آن p احتمال گزارش صادقانه و q احتمال گزارش هر یک از $k-1$ گزینه‌ی دیگر است.

به صورت مشابه RR می‌توان نشان داد که مقادیر بهینه برای ارضای شرط α-LDP عبارتند از:

$$p = \frac{e^\alpha}{e^\alpha + k - 1}, \quad q = \frac{1}{e^\alpha + k - 1} \tag{۳-۱۸}$$

برای تخمین فراوانی یک آیتم خاص $v \in \mathcal{X}$، از تخمین‌گر ناریب $\hat{c}_v = \frac{\mathbb{I}(z=v)-q}{p-q}$ استفاده می‌شود. با تحلیلی مشابه لم ۳-۲، واریانس این تخمین‌گر برابر خواهد بود با:

$$\mathrm{Var}_{GRR} = \frac{p(1-p)}{(p-q)^2} = \frac{(e^\alpha)(k-1)+(k-1)^2}{(e^\alpha-1)^2} \approx \mathcal{O}(k) \tag{۳-۱۹}$$

این رابطه نشان می‌دهد که واریانس GRR وابستگی خطی به اندازه دامنه k دارد که نقطه ضعف این روش در ابعاد بالاست.

۳-۴-۳ مکانیزم‌های مبتنی بر کدگذاری یکانی (UE)

برای غلبه بر مشکل کاهش دقت GRR در دامنه‌های بزرگ، خانواده‌ای از مکانیزم‌ها تحت عنوان «کدگذاری یکانی»[17] توسعه یافته‌اند. این رویکرد اساس پروتکل مشهور RAPPOR گوگل را تشکیل می‌دهد [۱۰، ۲۰].

تعریف ۳-۵ (کدگذاری یکانی) *در این روش، فرآیند خصوصی‌سازی طی دو مرحله انجام می‌شود:*

۱. ***کدگذاری قطعی:*** *ورودی $x \in \{1,\dots,k\}$ به یک بردار بیتی $v \in \{0,1\}^k$ تبدیل می‌شود که تنها در موقعیت x برابر با ۱ و در سایر مکان‌ها ۰ است (کدگذاری وان‌ـ هات[18]).*

[17] Unary Encoding (UE)
[18] One-Hot Encoding

۲ . ***اختلال***[19] ***مستقل:*** *هر بیت این بردار به صورت مستقل با استفاده از یک مکانیزم باینری معکوس می‌شود. اگر v_i بیت iـ ام بردار کدگذاری شده باشد، خروجی z_i به صورت زیر تولید می‌شود:*

$$\mathbb{P}[z_i = ۱] = \begin{cases} p & \text{if } v_i = ۱ \\ q & \text{if } v_i = ۰ \end{cases} \qquad (۲۰-۳)$$

مثال ۳-۲ *فرض کنید دامنه شامل ۴ آیتم باشد ($\mathcal{X} = \{۱, ۲, ۳, ۴\}$) و ورودی کاربر ۲ $x = ۲$ باشد.*

۱ . *بردار کدگذاری شده:* $v = [۰, ۱, ۰, ۰]$

۲ . *اعمال نویز: هر بیت مستقل پرتاب می‌شود. ممکن است خروجی نهایی* $z = [۰, ۱, ۱, ۰]$ *شود (بیت سوم از ۰ به ۱ تغییر کرده است).*

اثبات α-LDP: برای بررسی شرط α-LDP، نسبت احتمال خروجی برداری $z = (z_۱, \ldots, z_k)$ را برای دو ورودی متمایز x و x' محاسبه می‌کنیم. بردارهای متناظر v و v' تنها در دو موقعیت تفاوت دارند: موقعیت x (که $v_x = ۱, v'_x = ۰$) و موقعیت x' (که $v_{x'} = ۰, v'_{x'} = ۱$). در سایر موقعیت‌ها ($j \neq x, x'$) بیت‌ها یکسان و برابر صفر هستند و در نسبت احتمالات ساده می‌شوند. از آن‌جا که بیت‌ها مستقل هستند:

$$\frac{\mathbb{P}[z|x]}{\mathbb{P}[z|x']} = \frac{\mathbb{P}[z_x|v_x = ۱]}{\mathbb{P}[z_x|v'_x = ۰]} \cdot \frac{\mathbb{P}[z_{x'}|v_{x'} = ۰]}{\mathbb{P}[z_{x'}|v'_{x'} = ۱]} \qquad (۲۱-۳)$$

بیشینه‌ی این کسر زمانی رخ می‌دهد که صورت کسر ماکزیمم و مخرج مینیمم شود؛ یعنی زمانی که $z_x = ۱$ (حفظ بیت ۱) و $z_{x'} = ۰$ (حفظ بیت ۰) باشد. در این حالت:

$$\frac{\mathbb{P}[z|x]}{\mathbb{P}[z|x']} \leqslant \frac{p}{q} \cdot \frac{۱-q}{۱-p} = \frac{p(۱-q)}{q(۱-p)} \qquad (۲۲-۳)$$

بنابراین شرط α-LDP معادل است با:

$$\alpha = \ln\left(\frac{p(۱-q)}{q(۱-p)}\right) \qquad (۲۳-۳)$$

لم ۳-۳ (تحلیل واریانس UE) *در پروتکل‌های UE ، واریانس تخمین فراوانی برای هر آیتم، تنها به پارامترهای p و q وابسته است و از رابطه زیر پیروی می‌کند:*

$$\mathrm{Var}_{UE} = \frac{q(۱-q)}{(p-q)^۲} \qquad (۲۴-۳)$$

(نکته: این واریانس برای حالتی است که ورودی واقعی کاربر آن آیتم نباشد، که در دامنه‌های بزرگ حالت غالب است).

[19] Perturbation

دو استراتژی اصلی برای تنظیم p و q بر اساس رابطه (۳-۲۳) وجود دارد:

۱. کدگذاری یکانی متقارن (SUE): در این روش $p+q=۱$ در نظر گرفته می‌شود. با جایگذاری در شرط α-LDP، مقادیر بهینه عبارتند از $p=\frac{e^{\alpha/۲}}{e^{\alpha/۲}+۱}$ و $q=\frac{۱}{e^{\alpha/۲}+۱}$. واریانس در این حالت برابر است با:

$$\mathrm{Var}_{SUE}=\frac{e^{\alpha/۲}}{(e^{\alpha/۲}-۱)^{۲}} \tag{۳-۲۵}$$

۲. کدگذاری یکانی بهینه (OUE): وانگ و همکاران [۲۰] نشان دادند که برای کمینه‌سازی واریانس در دامنه‌های بزرگ، باید اطلاعات بیت‌های ۱ (سیگنال) حفظ شود ($p=۱/۲$) و نویز روی بیت‌های ۰ (که اکثر بردار را تشکیل می‌دهند) کنترل شود ($q=\frac{۱}{e^{\alpha}+۱}$). واریانس حاصل برابر است با:

$$\mathrm{Var}_{OUE}=\frac{۴e^{\alpha}}{(e^{\alpha}-۱)^{۲}} \tag{۳-۲۶}$$

۳-۴-۴ تحلیل مقایسه‌ای: چرا GRR در ابعاد بالا شکست می‌خورد؟

یکی از مهم‌ترین نتایج نظری در ادبیات LDP، مقایسه رفتار مجانبی GRR و OUE نسبت به اندازه دامنه k است.

مثال ۳-۳ (ناکارآمدی GRR در دامنه‌های بزرگ) *فرض کنید می‌خواهیم کلمات پرکاربرد را از یک لغت‌نامه با $k=۱۰۰,۰۰۰$ کلمه استخراج کنیم.*

- *در مکانیزم GRR، طبق رابطه (۳-۱۹)، واریانس تقریباً با k رشد می‌کند:*

$$\mathrm{Var}_{GRR}\approx\frac{k}{(e^{\alpha}-۱)^{۲}}$$

به عبارتی، نویز اضافه شده متناسب با کل اندازه دیکشنری است که سیگنال کلمات نادر را کاملاً محو می‌کند.

- *در مکانیزم OUE، واریانس مستقل از k است:*

$$\mathrm{Var}_{OUE}=\frac{۴e^{\alpha}}{(e^{\alpha}-۱)^{۲}}$$

این استقلال از k باعث می‌شود که خانواده UE گزینه‌ی برتر برای دامنه‌های بزرگ باشند. با این حال، OUE هزینه مخابراتی بالایی دارد (ارسال بردار با طول k). برای رفع این مشکل، نسخه بهبودیافته‌ای به نام «درهم‌سازی موضعی بهینه[۲۰]» توسط وانگ و همکاران [۲۱] معرفی شده است. این روش با استفاده از توابع درهم‌ساز، ورودی را فشرده کرده و بدون افزایش واریانس، هزینه مخابراتی را کاهش می‌دهد.

[20] Optimized Local Hashing (OLH)

۳-۴-۵ مکانیزم لاپلاس موضعی

برای داده‌های عددی پیوسته، رویکرد استاندارد تعمیم مکانیزم لاپلاس از مدل متمرکز به مدل موضعی است. فرض کنید دامنه ورودی یک بازه‌ی کران‌دار $\mathcal{X} \subset \mathbb{R}$ باشد. بدون کاستن از کلیت[۲۱]، فرض می‌کنیم داده‌ها با یک تبدیل خطی به بازه‌ی $[-۱, ۱]$ نگاشت شده‌اند. در مدل موضعی، شرط محرمانگی باید برای هر جفت ورودی $x, x' \in \mathcal{X}$ برقرار باشد. بنابراین «حساسیت سراسری» (Δ) برابر با قطر دامنه است. در این صورت Δ، بیشترین مقدار ممکن را خواهد داشت:

$$\Delta = \sup_{x,x' \in [-۱,۱]} |x - x'| = |۱ - (-۱)| = ۲ \tag{۳-۲۷}$$

تعریف ۳-۶ (مکانیزم لاپلاس موضعی) *مکانیزم $\mathcal{M}_{Lap}$ ورودی نرمال شده $x \in [-۱, ۱]$ را دریافت کرده و خروجی z را طبق رابطه زیر تولید می‌کند:*

$$z = x + \eta, \quad \eta \sim \mathrm{Lap}\left(\frac{\Delta}{\alpha}\right) = \mathrm{Lap}\left(\frac{۲}{\alpha}\right) \tag{۳-۲۸}$$

تابع چگالی احتمال (PDF) خروجی برای ورودی x به صورت زیر است:

$$f(z|x) = \frac{\alpha}{۴} \exp\left(-\frac{\alpha|z - x|}{۲}\right) \tag{۳-۲۹}$$

اثبات α-LDP: برای هر دو ورودی x, x' و هر خروجی z، نسبت چگالی‌ها عبارت است از:

$$\frac{f(z|x)}{f(z|x')} = \frac{\exp\left(-\frac{\alpha}{۲}|z - x|\right)}{\exp\left(-\frac{\alpha}{۲}|z - x'|\right)} \tag{۳-۳۰}$$

$$= \exp\left(\frac{\alpha}{۲}(|z - x'| - |z - x|)\right) \tag{۳-۳۱}$$

طبق نامساوی مثلثی ($|a| - |b| \leqslant |a - b|$):

$$|z - x'| - |z - x| \leqslant |(z - x') - (z - x)| = |x - x'| \leqslant ۲$$

بنابراین نسبت احتمال با $\exp\left(\frac{\alpha}{۲} \cdot ۲\right) = e^{\alpha}$ کران‌دار می‌شود.

مثال ۳-۴ *فرض کنید می‌خواهیم دمای بدن یک بیمار را گزارش کنیم. اگر دامنه تغییرات دما $[۳۵, ۴۲]$ درجه باشد، طول بازه $۴۲ - ۳۵ = ۷$ است. اگر بدون نرمال‌سازی از لاپلاس استفاده کنیم، باید نویزی متناسب با $\frac{۷}{\alpha}$ اضافه کنیم. اما در روش استاندارد، ابتدا دما را به $[-۱, ۱]$ نگاشت می‌کنیم (که حساسیت ۲ شود)، نویز با مقیاس $\frac{۲}{\alpha}$ اضافه می‌کنیم و در سمت سرور مجدداً نتیجه را به مقیاس اصلی برمی‌گردانیم. حال اگر بخواهیم دمای بدن یک بیمار را که به بازه $[-۱, ۱]$ نرمال شده است، با بودجه $\alpha = ۱$ منتشر کنیم. اگر مقدار واقعی $x = ۰/۵$ باشد:*

[21] Without Loss of Generality

- *مقیاس نویز برابر است با* $b = \frac{۲}{۱} = ۲$.
- *یک نمونه تصادفی ممکن است* $z = ۰/۵ + (-۱/۲) = -۰/۷$ *باشد.*
- *واریانس خطا برابر است با* $۲b^۲ = ۸$ *. این واریانس برای یک مقدار در بازه* $[-۱, ۱]$ *بسیار زیاد است و نشان می‌دهد که* LDP *برای داده‌های عددی تک‌بعدی خطای زیادی تحمیل می‌کند مگر اینکه* n *(تعداد کاربران) بسیار زیاد باشد.*

۳-۴-۶ مکانیزم تخمین میانگین دوچی (۱ـ بیتی)

همان‌طور که در مثال قبل دیدیم، مکانیزم لاپلاس به دلیل دامنه‌ی نامتناهی خروجی، واریانس قابل توجهی را تحمیل می‌کند. برای حل این مشکل در داده‌های کران‌دار (مثلاً $[-۱, ۱]$)، دوچی و همکاران [۷] یک مکانیزم بهینه ارائه کردند که خروجی را به شدت کوانتیده (باینری) می‌کند.

این مکانیزم که به «مکانیزم ۱ـ بیتی» نیز شهرت دارد، ورودی پیوسته $x \in [-۱, ۱]$ را به یک متغیر تصادفی $z \in \{-B, +B\}$ تبدیل می‌کند، به طوری که نااریب باقی بماند ($\mathbb{E}_z = x$).

الگوریتم ۱ مکانیزم ۱ـ بیتی دوچی

۱: **ورودی:** داده کاربر $x \in [-۱, ۱]$ و بودجه محرمانگی α.

۲: **گام ۱ (کوانتزاسیون):** متغیر تصادفی برنولی u را تولید کنید:

$$u = \begin{cases} ۱ & \text{با احتمال } \frac{۱+x}{۲} \\ -۱ & \text{با احتمال } \frac{۱-x}{۲} \end{cases} \tag{۳۲-۳}$$

۳: **گام ۲ (اختلال):** مقدار u را با مکانیزم RR مخدوش کنید تا z' به دست آید (با احتمال $\frac{e^\alpha}{e^\alpha+۱}$ خود u و با احتمال مکمل، قرینه‌ی آن).

۴: **گام ۳ (تجمیع):** خروجی نهایی z را مقیاس‌دهی کنید:

$$z = z' \cdot \frac{e^\alpha + ۱}{e^\alpha - ۱} \tag{۳۳-۳}$$

۵: **خروجی:** مقدار z.

تحلیل واریانس: اگرچه این مکانیزم اطلاعات را به شدت فشرده (۱ بیت) می‌کند، اما دوچی ثابت کرد که این روش از نظر نرخ مینی‌مکس برای تخمین میانگین بهینه است. واریانس آن برابر است با:

$$\mathrm{Var}_{Duchi} \leqslant \left(\frac{e^\alpha + ۱}{e^\alpha - ۱}\right)^۲ \approx \mathcal{O}\left(\frac{۱}{\alpha^۲}\right) \tag{۳۴-۳}$$

مزیت اصلی این روش نسبت به لاپلاس، کران‌دار بودنِ قطعیِ خروجی است که در تحلیل‌های بدترین حالت[۲۲] اهمیت زیادی دارد.

چالش ابعاد بالا (نفرین ابعاد)

چرا در ابعاد بالا ($d > ۱$) از مکانیزم لاپلاس استفاده نمی‌شود؟ فرض کنید ورودی کاربر یک بردار $x \in \mathbb{R}^d$ باشد که در گوی واحد اقلیدسی قرار دارد ($\|x\|_۲ \leqslant ۱$).

مشکل بنیادین این است که مکانیزم لاپلاس متکی بر حساسیت در فضای $\ell_۱$ است، در حالی که هندسه فضای برداری اقلیدسی منطبق بر $\ell_۲$ می‌باشد. برای پوشش دادن گوی واحد $\ell_۲$ با نویز لاپلاس، باید حساسیت $\ell_۱$ را در بدترین حالت در نظر بگیریم. می‌دانیم برای هر بردار با نرم اقلیدسی ۱، نرم $\ell_۱$ می‌تواند تا $\sqrt{d}$ رشد کند. بنابراین قطر دامنه در متر $\ell_۱$ برابر است با:

$$\Delta_{\ell_۱} = \sup_{x,x' \in B_{\ell_۲}} \|x - x'\|_۱ \leqslant \sqrt{d} \cdot \sup \|x - x'\|_۲ = ۲\sqrt{d} \qquad (۳۵\text{-}۳)$$

برای تأمین α-LDP، باید به هر مؤلفه نویزی مستقل با مقیاس $\frac{۲\sqrt{d}}{\alpha}$ اضافه کنیم. در نتیجه:

- واریانس خطا در هر بُعد: $۲ \times \left(\frac{۲\sqrt{d}}{\alpha}\right)^۲ = \frac{۸d}{\alpha^۲}$
- خطای میانگین مربعات کل[۲۳] برای d بُعد: $d \times \frac{۸d}{\alpha^۲} = \frac{۸d^۲}{\alpha^۲}$

این نرخ رشد $\mathcal{O}(d^۲)$ برای خطا، بسیار ناکارآمد است. دوچی و همکاران [۶] ثابت کرده‌اند که حد پایین نظری مینی‌مکس برای این مسأله $\mathcal{O}(d)$ است. به همین دلیل، در ابعاد بالا از مکانیزم‌های پیشرفته‌تری مانند «توزیع برنولی چندبعدی» یا «نمونه‌برداری هایپرکیوب» استفاده می‌شود که با هندسه فضا سازگارترند.

مقایسه با مدل متمرکز: در مثال ۲-۶ فصل قبل دیدیم که در مدل متمرکز، مکانیزم لاپلاس با افزایش ابعاد (d) دچار افت کارایی می‌شود و نویز با ضریب d رشد می‌کند (که با استفاده از مکانیزم گوسی به $\sqrt{d}$ کاهش می‌یابد). اما در مدل موضعی، این پدیده بسیار شدیدتر است. در اینجا نه تنها نویز با d رشد می‌کند، بلکه به دلیل عدم تمرکز و جمع شدن خطاهای تک‌تک کاربران، خطای نهایی (MSE) با $d^۲$ افزایش می‌یابد. به همین دلیل، راهکارهای ساده‌ی مدل متمرکز (مانند افزودن نویز به هر بعد) در مدل موضعی تقریباً بلااستفاده هستند.

[22] Worst-case
[23] Mean Squared Error (MSE)

۳-۵ چالش سودمندی و هزینه عدم اعتماد

همان‌طور که دیدیم، مدل LDP گلوگاه اعتماد به سرور مرکزی را حذف می‌کند. اما این افزایش امنیت بدون هزینه نیست. در این بخش، با یک تحلیل دقیق ریاضی نشان می‌دهیم که حذف متصدی مورد اعتماد منجر به کاهش شدید دقت آماری (سودمندی) می‌شود. برای این منظور، ساده‌ترین مسئله آماری یعنی «تخمین میانگین جامعه» را در دو مدل متمرکز و موضعی مقایسه می‌کنیم.

۳-۵-۱ تعریف مسئله: تخمین میانگین دودویی

فرض کنید n کاربر وجود دارند و هر کاربر i دارای یک بیت خصوصی $X_i \in \{۰, ۱\}$ است. هدف تخمین‌گر، محاسبه‌ی میانگین واقعی جامعه است:

$$p = \frac{۱}{n} \sum_{i=۱}^{n} X_i \tag{۳۶-۳}$$

معیار ارزیابی ما، خطای میانگین مربعات تخمین‌گر $\hat{p}$ خواهد بود:

$$\mathrm{MSE}(\hat{p}) = \mathbb{E}\left[(\hat{p} - p)^۲\right] = \mathrm{Var}[\hat{p}] + (\mathrm{Bias}[\hat{p}])^۲ \tag{۳۷-۳}$$

ما در هر دو مدل از تخمین‌گرهای نااریب ($\mathrm{Bias} = ۰$) استفاده می‌کنیم، بنابراین خطا صرفاً ناشی از واریانس نویز تزریق‌شده است.

۳-۵-۲ تحلیل در مدل متمرکز (CDP)

در مدل متمرکز با بودجه محرمانگی ε-DP، متصدی به تمام X_iها دسترسی دارد. او ابتدا مقدار دقیق مجموع $\sum X_i$ را محاسبه می‌کند. چون تغییر یک بیت حداکثر مجموع را ۱ واحد تغییر می‌دهد، حساسیت سراسری برابر با $\Delta = ۱$ است. طبق مکانیزم لاپلاس، نویزی با مقیاس $۱/\varepsilon$ به مجموع اضافه شده و سپس بر n تقسیم می‌شود تا میانگین به دست آید:

$$\hat{p}_{CDP} = \frac{۱}{n}\left(\sum_{i=۱}^{n} X_i + \eta\right), \quad \eta \sim \mathrm{Lap}(۱/\varepsilon) \tag{۳۸-۳}$$

خطای این تخمین‌گر برابر است با:

$$\mathrm{MSE}_{CDP} = \mathrm{Var}\left[\frac{\eta}{n}\right] = \frac{۱}{n^۲}\mathrm{Var}[\eta] = \frac{۱}{n^۲} \cdot \frac{۲}{\varepsilon^۲} = \mathcal{O}\left(\frac{۱}{n^۲\varepsilon^۲}\right) \tag{۳۹-۳}$$

این رابطه نشان می‌دهد که در مدل متمرکز، در مسئله‌ی تخمین میانگین، خطا با سرعت $۱/n$ به سمت صفر میل می‌کند (یا انحراف معیار با سرعت $۱/n$).

۳-۵-۳ تحلیل در مدل موضعی (LDP)

در مدل موضعی با محدودیت LDP، هیچ‌کس به X_iهای خام دسترسی ندارد. هر کاربر به صورت مستقل مکانیزم پاسخ تصادفی (RR) را روی داده خود اجرا می‌کند و $\hat{X}_i$ را گزارش می‌دهد. طبق نتایج بخش ۳-۴-۱، واریانس تخمین‌گر هر کاربر برای αهای کوچک ($\alpha < ۱$) تقریباً برابر است با:

$$\mathrm{Var}[\hat{X}_i] \approx \frac{۱}{\alpha^۲} \tag{۳-۴۰}$$

متصدی برای تخمین میانگین کل، میانگین گزارش‌های دریافتی را محاسبه می‌کند: $\hat{p}_{LDP} = \frac{۱}{n}\sum_{i=۱}^{n} \hat{X}_i$. از آن‌جا که نویز کاربران مستقل از یکدیگر است، واریانس مجموع برابر با مجموع واریانس‌هاست:

$$\mathrm{MSE}_{LDP} = \mathrm{Var}\left[\frac{۱}{n}\sum_{i=۱}^{n}\hat{X}_i\right] = \frac{۱}{n^۲}\sum_{i=۱}^{n}\mathrm{Var}[\hat{X}_i] = \frac{۱}{n^۲}\cdot n \cdot \mathcal{O}\left(\frac{۱}{\alpha^۲}\right) = \mathcal{O}\left(\frac{۱}{n\alpha^۲}\right) \tag{۳-۴۱}$$

در اینجا خطا با سرعت $۱/\sqrt{n}$ به سمت صفر میل می‌کند.

۳-۵-۴ نتیجه‌گیری: شکاف کارایی

با مقایسه روابط (۳-۳۹) و (۳-۴۱)، تفاوت بنیادین دو مدل آشکار می‌شود (با فرض ثابت بودن بودجه‌های محرمانگی $\varepsilon \approx \alpha$):

- **مدل متمرکز:** نرخ همگرایی خطا $\mathcal{O}(۱/n)$ است.
- **مدل موضعی:** نرخ همگرایی خطا $\mathcal{O}(۱/\sqrt{n})$ است.

این تفاوت در نرخ همگرایی پیامد بسیار مهمی در حجم نمونه[۲۴] مورد نیاز دارد. برای رسیدن به یک دقت ثابت مشخص (مثلاً خطای τ)، تعداد کاربران مورد نیاز در هر مدل عبارت است از:

$$n_{CDP} \propto \frac{۱}{\tau}, \qquad n_{LDP} \propto \frac{۱}{\tau^۲} \tag{۳-۴۲}$$

بنابراین رابطه بین حجم داده مورد نیاز در دو مدل به صورت زیر است:

$$n_{LDP} \approx \mathcal{O}(n_{CDP}^۲) \tag{۳-۴۳}$$

این رابطه که در ادبیات موضوع به «هزینه عدم اعتماد» شهرت دارد، نشان می‌دهد که مدل موضعی برای رسیدن به دقتی مشابه مدل متمرکز، نیازمند داده‌های بسیار بیشتری است. برای مثال، اگر در مدل متمرکز با

[24] Sample Complexity

۱,۰۰۰ کاربر به دقت مطلوبی برسیم، در مدل موضعی برای همان دقت به ۱,۰۰۰,۰۰۰ کاربر نیاز خواهیم داشت [۶، ۱۲].

همین شکاف عظیم است که انگیزه اصلی فصل‌های آینده‌ی این پایان‌نامه را شکل می‌دهد: «چگونه می‌توان با استفاده از تحلیل‌های دقیق‌تر (مانند واگرایی‌های f) و الگوریتم‌های بهینه، ثابت‌های پنهان در این حدود را بهبود بخشید؟»

فصل ۴

تحلیل مینی‌مکس و هندسه اطلاعاتی در LDP

۴-۱ مقدمه

در فصل قبل (۳)، چارچوب محرمانگی تفاضلی موضعی (LDP) را به عنوان جایگزینی برای مدل متمرکز معرفی کردیم. مشاهده شد که مکانیزم‌های استاندارد مانند «پاسخ تصادفی تعمیم‌یافته» (GRR) و «کدگذاری‌های یکانی» (UE)، اگرچه امنیت داده‌ها را در سطح کاربر تضمین می‌کنند، اما بهای سنگینی را از منظر دقت آماری تحمیل می‌نمایند. به طور خاص، همان‌طور که در تحلیل چالش سودمندی دیدیم، در مسئله‌ی تخمین میانگین، واریانس خطا از مرتبه $\mathcal{O}\left(\frac{۱}{n^۲\varepsilon^۲}\right)$ در مدل متمرکز به مرتبه $\mathcal{O}\left(\frac{۱}{n\alpha^۲}\right)$ در مدل موضعی افزایش می‌یابد، که در آن n نشان‌دهنده تعداد کاربران و ε و α به ترتیب بودجه‌های محرمانگی متمرکز و موضعی هستند.

این مشاهده یک پرسش بنیادین را مطرح می‌سازد: آیا این کاهش دقت، ناشی از ضعف در طراحی مکانیزم‌های موجود است، یا یک محدودیت ذاتی و غیرقابل اجتناب در طبیعت محرمانگی موضعی به شمار می‌رود؟ به بیان دیگر، آیا می‌توان مکانیزم هوشمندانه‌تری طراحی کرد که ضمن ارضای محدودیت LDP، به نرخ خطای مدل متمرکز نزدیک شود؟

هدف اصلی این فصل، پاسخ به این پرسش با بهره‌گیری از ابزارهای قدرتمند نظریه مینی‌مکس است. نشان خواهیم داد که هزینه پرداختی در مدل موضعی (که معمولاً به صورت ضریبی از $\sqrt{n}$ در تقلیل نرخ همگرایی ظاهر می‌شود)، یک مانع اطلاعاتی بنیادین است و هیچ الگوریتمی قادر به عبور از آن نیست.

برای اثبات این ادعا، از زیرساخت‌های ریاضی بنا شده در فصل‌های پیشین، از جمله fـ واگرایی‌ها، لم لوکم، نامساوی فانو و لم اسود استفاده خواهیم کرد. با این وجود، به‌کارگیری مستقیم این ابزارها در محیط

LDP امکان‌پذیر نیست. نوآوری اصلی این فصل، بررسی چارچوب «محرمانگی به عنوان انقباض»[1] است که به طور گسترده توسط دوچی، جردن و وین‌رایت [۶، ۷] توسعه یافته است.

ایده مرکزی این چارچوب آن است که مکانیزم‌های α-LDP همانند یک فیلتر اطلاعاتی عمل می‌کنند که فاصله آماری بین توزیع‌های ورودی را منقبض کرده و در نتیجه، تشخیص آن‌ها را دشوارتر می‌سازند. بر این اساس، در روند این فصل نخست خاصیت انقباضی مکانیزم‌های موضعی را بر حسب واگرایی KL و فاصله TV فرمول‌بندی می‌کنیم. سپس، این ویژگی را با نامساوی‌های پردازش داده[2] ترکیب می‌نماییم تا کران‌های پایین مینی‌مکس را برای مسائل کلاسیک آماری استخراج کنیم. در نهایت، با استناد به این کران‌ها ثابت می‌کنیم که مکانیزم‌های معرفی شده در فصل قبل، در حقیقت بهینه مینی‌مکس بوده و به بهترین نرخ ممکن دست می‌یابند.

۲-۴ محرمانگی به عنوان انقباض

همان‌طور که در فصل پیش‌نیازها (۲-۲) بحث شد، fـ واگرایی‌ها ابزاری قدرتمند برای سنجش تمایز بین توزیع‌های احتمالی هستند. ایده مرکزی در تحلیل مینی‌مکس تحت LDP این است که مکانیزم‌های محرمانگی به عنوان «عملگرهای انقباضی»[3] عمل می‌کنند. به عبارت دیگر، شرط α-LDP باعث می‌شود که توزیع‌های خروجی مکانیزم برای هر دو ورودی دل‌خواه، به یک‌دیگر بسیار نزدیک شوند و در نتیجه واگرایی بین آن‌ها محدود شود.

در این بخش، ما لم اساسی دوچی و همکاران [۷] را بیان و اثبات می‌کنیم. این لم نشان می‌دهد که واگرایی KL بین توزیع‌های خروجی، توسط فاصله TV توزیع‌های ورودی و ضریبی از بودجه محرمانگی α کران‌دار می‌شود.

قضیه‌ی ۴-۱ (کران انقباض قوی برای LDP) *فرض کنید $\mathcal{M} : \mathcal{X} \to \mathcal{Z}$ یک مکانیزم α-LDP باشد. برای هر جفت توزیع احتمالی P_1 و P_2 روی فضای ورودی $\mathcal{X}$، اگر M_1 و M_2 توزیع‌های حاشیه‌ای القا شده روی خروجی $\mathcal{Z}$ باشند (یعنی $M_i(S) = \int_{\mathcal{X}} \mathcal{M}(S|x)dP_i(x)$)، آنگاه نامساوی زیر برقرار است:*

$$D_{KL}(M_1 \| M_2) + D_{KL}(M_2 \| M_1) \leqslant (e^{\alpha} - 1)^2 (\|P_1 - P_2\|_{TV})^2 \qquad (۱-۴)$$

اثبات. برای اثبات این قضیه، از خاصیت تحدب مشترک fـ واگرایی‌ها و تعریف α-LDP استفاده می‌کنیم. اثبات در دو گام انجام می‌شود: ابتدا مسئله را به ورودی‌های نقطه‌ای تقلیل می‌دهیم و سپس کران را برای

[1] Privacy as Contraction
[2] Data Processing Inequalities
[3] Contraction Operators

آن محاسبه می‌کنیم.

گام اول: تقلیل به ورودی‌های نقطه‌ای.گام اول: تقلیل به ورودی‌های نقطه‌ای. هدف ما در این گام این است که نشان دهیم برای یافتن کران بالای واگرایی، بررسی توزیع‌های پیچیده و کلی P_1 و P_2 ضرورتی ندارد و می‌توان مسئله را به توزیع‌های متمرکز روی تک‌نقطه‌ها تقلیل داد.

همان‌طور که در بخش پیش‌نیازها (۲-۲-۱) بررسی شد، نگاشت واگرایی $(P,Q) \mapsto D_{KL}(P\|Q)$ نسبت به هر دو ورودی خود از خاصیت تحدب مشترک[۴] برخوردار است. از آن‌جا که توزیع‌های حاشیه‌ای M_i در واقع ترکیب‌های خطی (مخلوط‌هایی) از توزیع‌های شرطی $\mathcal{M}(\cdot|x)$ با وزن‌های P_i هستند، کل عبارت واگرایی متقارن در سمت چپ نامساوی (۴-۱) نسبت به جفت توزیع‌های ورودی (P_1, P_2) تابعی محدب خواهد بود.

بر اساس اصول بنیادین آنالیز محدب، بیشینه یک تابع محدب بر روی یک مجموعه (در این‌جا سیمپلکس تمام توزیع‌های احتمالی)، همواره در نقاط فرین[۵] آن مجموعه رخ می‌دهد. نقاط فرین در فضای توزیع‌های احتمالی، همان توزیع‌های دیراک[۶] متمرکز بر یک نقطه منفرد (یعنی δ_x برای $x \in \mathcal{X}$) هستند. هم‌چنین می‌دانیم که فاصله تغییرات کل $\|\cdot\|_{TV}$ نیز محدب است و بیشینه آن زمانی حاصل می‌شود که توزیع‌های P_1 و P_2 کم‌ترین هم‌پوشانی را با یک‌دیگر داشته باشند (یعنی تکیه‌گاه آن‌ها کاملاً مجزا یا متعامد باشند).

بنابراین، با بهره‌گیری از تحدب مشترک واگرایی و خطی بودن عملگر مکانیزم $\mathcal{M}$ روی مخلوط‌های احتمالی، می‌توان به لحاظ ریاضی ثابت کرد که کران بالا همواره با ارزیابی توزیع‌های نقطه‌ای به دست می‌آید [۶]. فرض کنید $x, x' \in \mathcal{X}$ دو نقطه دل‌خواه از فضای ورودی باشند و $Q_x = \mathcal{M}(\cdot|x)$ و $Q_{x'} = \mathcal{M}(\cdot|x')$ توزیع‌های خروجی متناظر با آن‌ها فرض شوند. برای این ورودی‌های نقطه‌ای متمایز (یعنی $P_1 = \delta_x$ و $P_2 = \delta_{x'}$)، فاصله تغییرات کل دقیقاً برابر با بیشینه مقدار خود است: $\|\delta_x - \delta_{x'}\|_{TV} = 1$.

در نتیجه، اگر بتوانیم نشان دهیم که برای هر جفت خروجی Q_x و $Q_{x'}$ نامساوی زیر برقرار است:

$$D_{KL}(Q_x\|Q_{x'}) + D_{KL}(Q_{x'}\|Q_x) \leqslant (e^{\alpha} - 1)^2 \qquad (۲-۴)$$

آنگاه با توجه به این‌که برای ورودی‌های نقطه‌ای $\|\delta_x - \delta_{x'}\|_{TV} = 1$ است، حکم قضیه برای حالت کلی توزیع‌ها به طور مستقیم از طریق اعمال نامساوی ینسن[۷] و تحدب نتیجه می‌شود؛ به این صورت که مربع فاصله تغییرات کل توزیع‌های اولیه، یعنی $(\|P_1 - P_2\|_{TV})^2$، به عنوان ضریب جریمه (یا عامل نرمال‌ساز) در کران نهایی ظاهر می‌گردد.

[4] Joint Convexity
[5] Extreme Points
[6] Dirac Distributions
[7] Jensen's Inequality

گام دوم: محاسبه کران برای ورودی‌های ثابت. طبق تعریف α-LDP، برای هر خروجی $z \in \mathcal{Z}$ و هر جفت ورودی x, x' داریم:

$$e^{-\alpha} \leqslant \frac{q(z|x)}{q(z|x')} \leqslant e^{\alpha} \qquad (۳-۴)$$

که در آن $q(\cdot|x)$ تابع چگالی (یا جرم) احتمال مکانیزم است. برای درک چگونگی به دست آمدن کران نهایی، کافی است به تعریف انتگرالی واگرایی کولبک‌ـلایبلر رجوع کنیم. واگرایی متقارن (یا واگرایی جفریز[۸]) حاصل جمع دو واگرایی $D_{KL}(Q_x \| Q_{x'})$ و $D_{KL}(Q_{x'} \| Q_x)$ است. با استفاده از خاصیت جبری لگاریتم یعنی $\ln(A/B) = -\ln(B/A)$، می‌توانیم کسر داخل لگاریتم در انتگرال دوم را معکوس کرده و یک علامت منفی ایجاد کنیم. با فاکتورگیری از عبارت لگاریتمی مشترک، دو انتگرال به سادگی تجمیع می‌شوند:

$$\begin{aligned} D_{KL}(Q_x \| Q_{x'}) + D_{KL}(Q_{x'} \| Q_x) &= \int_{\mathcal{Z}} q(z|x) \ln\left(\frac{q(z|x)}{q(z|x')}\right) \mu(dz) - \int_{\mathcal{Z}} q(z|x') \ln\left(\frac{q(z|x)}{q(z|x')}\right) \mu(dz) \\ &= \int_{\mathcal{Z}} (q(z|x) - q(z|x')) \ln\left(\frac{q(z|x)}{q(z|x')}\right) \mu(dz) \qquad (۴-۴) \end{aligned}$$

اهمیت بنیادین این بسط فشرده در آن است که واگرایی را به حاصل‌ضرب دو بخش تفکیک می‌کند: «اختلاف چگالی‌ها» (که با فاصله L_1 مرتبط است) و «لگاریتم نسبت چگالی‌ها» (که مستقیماً توسط محدودیت α-LDPدر بازه $[-\alpha, \alpha]$ کران‌دار است).

برای محاسبه‌ی دقیق این کران، از یک نامساوی کمکی استفاده می‌کنیم که ارتباط میان واگرایی متقارن و فاصله‌ی L_1 را تحت محدودیت‌های نسبت چگالی مشخص می‌کند. این نتیجه که معادل با «لم ۱» در مقاله‌ی دوچی و همکاران [۶] است، به صورت زیر بیان و اثبات می‌شود:

لم ۴-۲ (کران واگرایی جفریز تحت نسبت چگالی محدود [۶]) *فرض کنید P و Q دو توزیع احتمالی روی فضای $\mathcal{Z}$ با توابع چگالی p و q باشند، به طوری که برای تمام $z \in \mathcal{Z}$ شرط $e^{-\alpha} \leqslant \frac{p(z)}{q(z)} \leqslant e^{\alpha}$ برقرار باشد. در این صورت، واگرایی متقارن میان آن‌ها به صورت زیر کران‌دار می‌شود:*

$$D_{KL}(P\|Q) + D_{KL}(Q\|P) \leqslant \alpha \|P - Q\|_1 \leqslant (e^{\alpha} - 1) \|P - Q\|_1 \qquad (۵-۴)$$

اثبات. همان‌طور که می‌دانیم، با استفاده از تقارن، مجموع دو واگرایی را می‌توان به صورت یک انتگرال واحد نوشت:

$$D_{KL}(P\|Q) + D_{KL}(Q\|P) = \int_{\mathcal{Z}} (p(z) - q(z)) \ln\left(\frac{p(z)}{q(z)}\right) \mu(dz) \qquad (۶-۴)$$

[8] Jeffreys Divergence

برای هر z مشخص، عبارت داخل انتگرال را بررسی می‌کنیم. از آنجا که این عبارت نسبت به جابجایی p و q کاملاً متقارن است، بدون کاستن از کلیت مسئله فرض می‌کنیم $p(z) \geqslant q(z)$. در این حالت، اختلاف آن‌ها برابر با قدر مطلقشان خواهد بود، یعنی $p(z) - q(z) = |p(z) - q(z)|$. بر اساس فرضِ محدود بودن نسبت چگالی‌ها، می‌دانیم که $1 \leqslant \frac{p(z)}{q(z)} \leqslant e^{\alpha}$. با اعمال تابع صعودی لگاریتم طبیعی بر این نامساوی، نتیجه می‌شود که $\ln\left(\frac{p(z)}{q(z)}\right) \leqslant \alpha$. اکنون با ضرب کردن این نتیجه در مقدار نامنفی $|p(z) - q(z)|$ به دست می‌آوریم:

$$(p(z) - q(z)) \ln\left(\frac{p(z)}{q(z)}\right) \leqslant \alpha\, |p(z) - q(z)| \tag{۴-۷}$$

علاوه بر این، با استفاده از بسط تیلور برای تابع نمایی می‌دانیم که نامساوی پایه $\alpha \leqslant e^{\alpha} - 1$ برای تمام $\alpha \geqslant 0$ همواره برقرار است. بنابراین می‌توانیم کران فوق را بسط دهیم:

$$(p(z) - q(z)) \ln\left(\frac{p(z)}{q(z)}\right) \leqslant (e^{\alpha} - 1)\, |p(z) - q(z)| \tag{۴-۸}$$

با انتگرال‌گیری از هر دو طرف این نامساوی بر روی کل فضای $\mathcal{Z}$، اثبات لم کامل می‌گردد. □

اکنون به ادامه‌ی اثبات قضیه‌ی اصلی بازمی‌گردیم. طبق تعریف ذاتی مکانیزم α-LDP، توزیع‌های خروجی Q_x و $Q_{x'}$ شرط محدودیتِ نسبت چگالی را در بازه‌ی $[e^{-\alpha}, e^{\alpha}]$ به طور کامل ارضا می‌کنند. بنابراین با استناد مستقیم به لم ۴-۲ داریم:

$$D_{KL}\left(Q_x \| Q_{x'}\right) + D_{KL}\left(Q_{x'} \| Q_x\right) \leqslant (e^{\alpha} - 1)\, \|Q_x - Q_{x'}\|_1 \tag{۴-۹}$$

همچنین می‌دانیم که تحت شرط α-LDP، فاصله L_1 خروجی‌ها نیز محدود است (زیرا جرم احتمال نمی‌تواند سریع‌تر از ضریب e^{α} جابجا شود):

$$\|Q_x - Q_{x'}\|_1 \leqslant e^{\alpha} - 1 \tag{۴-۱۰}$$

با ترکیب این دو نتیجه برای ورودی‌های نقطه‌ای داریم:

$$\begin{aligned} D_{KL}\left(Q_x \| Q_{x'}\right) + D_{KL}\left(Q_{x'} \| Q_x\right) &\leqslant (e^{\alpha} - 1)\, \|Q_x - Q_{x'}\|_1 \\ &\leqslant (e^{\alpha} - 1)^2 \end{aligned} \tag{۴-۱۱}$$

این اثبات برای حالت ورودی‌های نقطه‌ای کامل می‌شود. تعمیم نهایی به توزیع‌های کلی P_1 و P_2 از طریق اعمال نامساوی پردازش داده (DPI) و تحدب مشترک حاصل می‌شود که در آن، ضریب $(\|P_1 - P_2\|_{TV})^2$ به عنوان جریمه‌ی فاصله اولیه توزیع‌ها ظاهر می‌گردد. □

۴-۲-۱ تفسیر رژیم‌های محرمانگی

نامساوی (۴-۱) پیامدهای بسیار مهمی برای نرخ‌های همگرایی آماری دارد. رفتار ضریب $(e^\alpha - ۱)^۲$ در دو رژیم حدی قابل توجه است:

- **رژیم محرمانگی بالا (رژیم $\alpha \leqslant ۱$):** در این حالت، می‌توان از بسط تیلور استفاده کرد که نتیجه می‌دهد $e^\alpha - ۱ \approx \alpha$. بنابراین کران انقباض به صورت زیر تقریب زده می‌شود:

$$D_{KL}\left(M_۱ \| M_۲\right) \lesssim \alpha^۲ \left\| P_۱ - P_۲ \right\|_{TV}^۲ \qquad (۴-۱۲)$$

  این نتیجه به وضوح نشان می‌دهد که میزان اطلاعات (و به تبع آن اطلاعات فیشر) با توان دوم α کاهش می‌یابند. این پدیده، دلیل اصلی و ریاضیاتیِ نرخ همگرایی کندتر در مدل موضعی (ظهور عامل $۱/\alpha^۲$) نسبت به مدل متمرکز است.

- **رژیم محرمانگی پایین (رژیم $\alpha \to \infty$):** در این حالت، e^α به سرعت رشد کرده و کران به سمت بی‌نهایت میل می‌کند. این رفتار کاملاً منطقی است؛ زیرا وقتی α بسیار بزرگ باشد، مکانیزم α-LDP تقریباً هیچ محدودیتی اعمال نمی‌کند و اجازه می‌دهد توزیع‌های خروجی کاملاً متمایز از یک‌دیگر باقی بمانند (عدم وقوع انقباض اطلاعاتی).

این خاصیت انقباضی، ابزار اصلی ما در بخش‌های بعدی این فصل برای اثبات کران‌های پایین مینی‌مکس خواهد بود؛ جایی که نشان می‌دهیم به دلیل این انقباض، تمایز قائل شدن میان پارامترهای مدل حتی با در اختیار داشتن حجم عظیمی از داده‌ها، هم‌چنان دشوار باقی می‌ماند.

۴-۲-۲ تعمیم به n کاربر مستقل (خاصیت تنسوری شدن)

پیش از ورود به استخراج کران‌های مینی‌مکس، باید بررسی کنیم که این انقباضِ موضعی چگونه بر روی توزیع توأم یک مجموعه‌داده‌ی کامل اثر می‌گذارد. فرض کنید n کاربر به طور کاملاً مستقل داده‌های خود را از طریق مکانیزم‌های α-LDP (یعنی $\mathcal{M}_۱, \ldots, \mathcal{M}_n$) پردازش و منتشر می‌کنند.

اگر M_v^n نشان‌دهنده‌ی توزیع توأم مشاهدات خروجی $Z^n = (Z_۱, \ldots, Z_n)$ به شرط توزیع ورودی P_v باشد، به دلیل استقلال شرطی خروجی کاربران، واگرایی کولبک‌ـلایبلرِ توأم دقیقاً برابر با مجموع واگرایی‌های حاشیه‌ای هر یک از آن‌ها خواهد بود. این ویژگی در نظریه اطلاعات به عنوان «خاصیت

تنسوری شدن»[9] شناخته می‌شود:

$$D_{KL}\left(M_1^n \| M_2^n\right) = \sum_{i=1}^{n} D_{KL}\left(M_{1,i} \| M_{2,i}\right) \qquad (۱۳-۴)$$

اکنون با اعمال قضیه انقباض قوی (قضیه ۴-۱) به صورت مجزا برای مکانیزم هر کاربر، کران انقباض برای کل مجموعه‌داده به شکل زیر در می‌آید:

$$D_{KL}\left(M_1^n \| M_2^n\right) \leqslant n\left(e^{\alpha}-1\right)^2 \|P_1 - P_2\|_{TV}^2 \qquad (۱۴-۴)$$

این نامساوی بسیار کلیدی است؛ زیرا نشان می‌دهد که حتی با جمع‌آوری داده از n کاربر، اطلاعات موجود در شبکه تنها با ضریبِ تضعیف‌شده‌ی $n\alpha^2$ رشد می‌کند. این رابطه همان پل ارتباطی مهمی است که در بخش‌های آتی، مستقیماً در نامساوی‌های لوکم و فانو جای‌گذاری خواهد شد.

۴-۳ نامساوی‌های پردازش داده قوی (SDPI)

در بخش‌های پیشین مشاهده کردیم که اعمال مکانیزم‌های α-LDP به طور ذاتی باعث کاهش و انقباض فاصله میان توزیع‌های آماری می‌شود. این مفهوم بنیادین را می‌توان به صورت صوری‌تر و در چارچوب نظریه اطلاعات، تحت عنوان «نامساوی پردازش داده قوی»[10] فرمول‌بندی کرد. در حالی که نامساوی پردازش داده‌ی استاندارد (DPI) صرفاً بیان می‌کند که اعمال یک کانال تصادفی واگرایی را افزایش نمی‌دهد (یعنی $D_f\left(\mathcal{M}(P)\|\mathcal{M}(Q)\right) \leqslant D_f\left(P\|Q\right)$)، نسخه‌ی قویِ آن تضمین می‌کند که در حضور محدودیت‌هایی نظیر محرمانگی، این واگرایی با ضریبی مشخص، اکیداً کاهش می‌یابد.

تعریف ۴-۱ (ضریب انقباض دوبروشین[11]) *برای یک مکانیزم یا کانال مارکوف $\mathcal{M}$ و یک واگرایی دل‌خواه f ، ضریب انقباض $\eta_f(\mathcal{M})$ به صورت زیر تعریف می‌شود:*

$$\eta_f(\mathcal{M}) = \sup_{P \neq Q} \frac{D_f\left(\mathcal{M}(P)\|\mathcal{M}(Q)\right)}{D_f\left(P\|Q\right)} \qquad (۱۵-۴)$$

که اگر برای یک کانال $\eta_f(\mathcal{M}) < 1$ باشد، اصطلاحاً می‌گوییم آن مکانیزم دارای خاصیت $SDPI$ است.

برای خانواده‌ی مکانیزم‌های α-LDP، می‌توان به لحاظ ریاضی ثابت کرد که این ضریب همواره اکیداً کم‌تر از یک است، که این امر نشان‌دهنده‌ی اتلاف حتمی و غیرقابل‌بازگشت اطلاعات در فرآیند خصوصی‌سازی است. با این وجود، در تحلیل‌های مینی‌مکس ما غالباً به کران‌هایی نیاز داریم که واگرایی توزیع‌های خروجی را مستقیماً به فاصله‌ی تغییرات کل (TV) توزیع‌های ورودی مرتبط سازند؛ چرا که در بسیاری از مسائل آماری، فاصله‌ی TV مترِ طبیعی و استاندارد بر روی فضای پارامترها محسوب می‌شود.

[9] Tensorization Property
[10] Strong Data Processing Inequality (SDPI)

۴-۳-۱ کران انقباض برای واگرایی کای‌ـدو (χ^2)

اگرچه واگرایی کولبک‌ـلایبلر (KL) ابزارِ استاندارد و رایجی در نظریه اطلاعات است، اما کار با واگرایی χ^2 در مسائل مربوط به LDP غالباً به مراتب ساده‌تر و کارآمدتر است. قضیه‌ی زیر که برگرفته از تحلیل‌های دقیق دوچی و همکاران [۶] است، کران انقباض را به طور خاص برای واگرایی χ^2 بیان می‌کند.

قضیه‌ی ۴-۳ (انقباض χ^2 تحت محدودیت LDP) *فرض کنید $\mathcal{M}$ یک مکانیزم α-LDP باشد. برای هر دو توزیع دل‌خواه P_1 و P_2 روی فضای ورودی $\mathcal{X}$، اگر $M_1 = \mathcal{M}(P_1)$ و $M_2 = \mathcal{M}(P_2)$ توزیع‌های القاشده در فضای خروجی باشند، نامساوی زیر همواره برقرار است:*

$$D_{\chi^2}(M_1 \| M_2) \leqslant (e^\alpha + 1)(e^\alpha - 1)^2 (\|P_1 - P_2\|_{TV})^2 \qquad (۱۶-۴)$$

اثبات. برای استخراج دقیق این کران، از تکنیک تجزیه توزیع‌ها (مبتنی بر فاصله تغییرات کل) و اعمال مستقیم محدودیت‌های α-LDP بهره می‌گیریم.

فرض کنید $L = \|P_1 - P_2\|_{TV}$ فاصله تغییرات کل میان دو توزیع ورودی باشد. بر اساس اصول نظریه اندازه، می‌توانیم هر دو توزیع P_1 و P_2 را به صورت یک مخلوط احتمالی[۱۲] از یک بخش مشترک و یک بخش کاملاً متمایز بازنویسی کنیم. به عبارت دیگر، توزیع مشترک P_0 و توزیع‌های مجزای P'_1 و P'_2 وجود دارند به طوری که:

$$P_1 = (1 - L)P_0 + LP'_1, \qquad P_2 = (1 - L)P_0 + LP'_2 \qquad (۱۷-۴)$$

با توجه به خطی بودن عملگر مکانیزم $\mathcal{M}$ روی توزیع‌های احتمالی، توزیع‌های خروجی (M_1 و M_2) نیز از همین ساختار مخلوط پیروی می‌کنند:

$$M_1 = (1 - L)M_0 + LM'_1, \qquad M_2 = (1 - L)M_0 + LM'_2 \qquad (۱۸-۴)$$

که در آن M_0, M'_1 و M'_2 توزیع‌های خروجی حاصل از اعمال مکانیزم روی P_0, P'_1 و P'_2 هستند.

اکنون با کم کردن این دو رابطه از یکدیگر، بخش مشترک حذف شده و تفاضل خروجی‌ها دقیقاً متناسب با L به دست می‌آید:

$$M_1(z) - M_2(z) = L\,(M'_1(z) - M'_2(z)) \qquad (۱۹-۴)$$

با جایگذاری این تفاضل در تعریف انتگرالی واگرایی کای‌ـدو، عامل L^2 (مربع فاصله تغییرات کل) به طور طبیعی از انتگرال خارج می‌شود:

$$D_{\chi^2}(M_1 \| M_2) = \int_{\mathcal{Z}} \frac{(M_1(z) - M_2(z))^2}{M_2(z)} \nu(dz) = L^2 \int_{\mathcal{Z}} \frac{(M'_1(z) - M'_2(z))^2}{M_2(z)} \nu(dz) \qquad (۲۰-۴)$$

[12] Probability Mixture

در گام بعدی، باید عبارت داخل انتگرال را به کمک شرط α-LDP کران‌دار کنیم. از آن‌جا که خروجی‌های $M'_۱$ و $M'_۲$ از یک مکانیزم αـموضعی امن عبور کرده‌اند، برای هر $z \in \mathcal{Z}$ نسبت چگالی آن‌ها محدود است: $M'_۱(z) \leqslant e^{\alpha} M'_۲(z)$. این نامساوی ایجاب می‌کند که اختلاف آن‌ها به صورت زیر کران‌دار شود:

$$|M'_۱(z) - M'_۲(z)| \leqslant (e^{\alpha} - ۱)\, M'_۲(z) \qquad (۲۱\text{-}۴)$$

بنابراین، صورت کسر در رابطه‌ی (۴-۲۰) حداکثر برابر با $(e^{\alpha} - ۱)^۲ M'_۲(z)^۲$ خواهد بود.

از سوی دیگر، برای مخرج کسر نیز باید ارتباطی میان $M_۲(z)$ و $M'_۲(z)$ بیابیم. از آن‌جا که هر دوی این توزیع‌ها خروجی مکانیزم $\mathcal{M}$ (به ازای ورودی‌های متفاوت) هستند، مجدداً شرط α-LDP تضمین می‌کند که $M'_۲(z) \leqslant e^{\alpha} M_۲(z)$. با جایگذاری این کران در مخرج، کسر داخل انتگرال به شکل زیر ساده می‌شود:

(۲۲-۴)

$$\frac{(M'_۱(z) - M'_۲(z))^۲}{M_۲(z)} \leqslant \frac{(e^{\alpha} - ۱)^۲ M'_۲(z)^۲}{M_۲(z)} \leqslant (e^{\alpha} - ۱)^۲ e^{\alpha} \left(\frac{M'_۲(z) M_۲(z)}{M_۲(z)} \right) = e^{\alpha} (e^{\alpha} - ۱)^۲ M'_۲(z)$$

در نهایت، با جایگذاری این کران در انتگرال (۴-۲۰) و استفاده از این واقعیت که انتگرال یک توزیع احتمال برابر با یک است ($\int M'_۲(z)\nu(dz) = ۱$)، به دست می‌آوریم:

$$D_{\chi^۲}(M_۱ \| M_۲) \leqslant L^۲ \cdot e^{\alpha} (e^{\alpha} - ۱)^۲ \int_{\mathcal{Z}} M'_۲(z)\nu(dz) = e^{\alpha} (e^{\alpha} - ۱)^۲ (\|P_۱ - P_۲\|_{TV})^۲ \qquad (۲۳\text{-}۴)$$

از آن‌جا که به صورت بدیهی $e^{\alpha} \leqslant e^{\alpha} + ۱$ است، کران استخراج‌شده به طور کامل در نامساوی بیان‌شده در قضیه صدق می‌کند و اثبات کامل می‌گردد. □

۴-۳-۲ مزیت‌های تحلیلی واگرایی $\chi^۲$ نسبت به KL

در ادامه‌ی این فصل و به ویژه هنگام استخراج کران‌های پایین مینی‌مکس برای مسائلی نظیر تخمین میانگین در ابعاد بالا، ما رویکرد خود را از KL به سمت واگرایی $\chi^۲$ تغییر خواهیم داد. این انتخابِ استراتژیک ریشه در سه ویژگی بنیادین دارد:

نخست، «رفتار متقارن حول صفر»؛ واگرایی KL ذاتاً نامتقارن است و بسط تیلورِ آن شامل جملات پیچیده‌ی لگاریتمی می‌شود. در مقابل، $D_{\chi^۲}$ به صورت محلی رفتاری کاملاً شبیه به یک فرم مربعی (فاصله‌ی اقلیدسی وزن‌دار) از خود نشان می‌دهد. از آن‌جا که مکانیزم‌های α-LDPتوزیع‌ها را به شدت منقبض کرده و به یکدیگر نزدیک می‌سازند (قرارگیری در رژیم محلی)، تقریب مرتبه دومِ $\chi^۲$ برای تحلیل این مجاورت بسیار دقیق‌تر و از نظر جبری خوش‌رفتارتر است.

دلیل دوم، «سادگی در محاسبه‌ی مخلوط‌های احتمالی» است. در تکنیک‌های پیشرفته‌ای مانند لم اسود یا روش فانو، ما نیازمند محاسبه‌ی واگرایی میان یک فرض منفرد و «مخلوطی از توزیع‌ها» (مانند $\bar{P} = \mathbb{E}[P_\theta]$) هستیم. به دلیل ساختار انتگرالیِ ساده‌تر در واگرایی کای‌ـدو که به فرم $D_{\chi^2}(P\|Q) = \int (P^2/Q) - 1$ نوشته می‌شود، محاسبات جبری برای مخلوط‌ها به مراتب سرراست‌تر از KL است که عملگر لگاریتم را درون انتگرال خود حبس کرده است.

در نهایت، «ارتباط ذاتی با خطای میانگین مربعات (MSE)»؛ در مسائل تخمین پارامتر، هدفِ غایی معمولاً کمینه‌سازی ریسک مربعات است. واگرایی χ^2 به طور طبیعی و مستقیم با واریانس تخمین‌گرها، کران کرامر‌ـ رائو[13] و ماتریس اطلاعات فیشر[14] پیوند دارد که این امر مسیر استخراج کران‌های مینی‌مکس را هموارتر می‌سازد.

۴-۳-۳ تنسوری شدن واگرایی χ^2 برای n کاربر

پیش از ورود به قضایای اصلی، باید به یک تفاوت اساسی میان KL و χ^2 در مواجهه با داده‌های مستقل توجه کنیم. در واگرایی KL، اطلاعاتِ n کاربر مستقل به صورت خطی با هم جمع می‌شوند. اما واگرایی χ^2 خاصیت جمع‌پذیری ندارد و به صورت ضرب‌شونده[15] عمل می‌کند. اگر P^n و Q^n نشان‌دهنده‌ی توزیع توأم مشاهدات مستقل n کاربر باشند، رابطه‌ی زیر برقرار است:

$$1 + D_{\chi^2}(P^n\|Q^n) = \left(1 + D_{\chi^2}(P\|Q)\right)^n \qquad (۲۴-۴)$$

این ویژگیِ نمایی در نگاه اول ممکن است نگران‌کننده به نظر برسد، اما در رژیم محرمانگی بالا (جایی که $D_{\chi^2}(P\|Q)$ به دلیل وجود ضریب α^2 بسیار کوچک و نزدیک به صفر است)، با استفاده از بسط دوجمله‌ای تقریب $(1+x)^n \approx 1+nx$ برقرار می‌شود. بنابراین، واگرایی توأمِ مجموعه‌داده به شکل زیر مهار می‌گردد:

$$D_{\chi^2}(P^n\|Q^n) \approx nD_{\chi^2}(P\|Q) \leqslant n(e^\alpha+1)(e^\alpha-1)^2(\|P_1-P_2\|_{TV})^2 \qquad (۲۵-۴)$$

همین رابطه‌ی تنسوری است که به ما اجازه می‌دهد در لم اسود، تأثیر تجمیعی n کاربر را به شکلی تحلیلی وارد محاسبات کرده و کران‌های دقیق بر حسب n و α به دست آوریم.

[13] Cramér-Rao Bound
[14] Fisher Information
[15] Multiplicative Tensorization

۴-۴ اثبات نرخ مینی‌مکس برای تخمین میانگین

در این بخش نهایی، ما تمام ابزارهای توسعه یافته در این فصل (لم اسود و نامساوی‌های انقباض) را ترکیب می‌کنیم تا یک کران پایین بنیادین برای مسئله کلاسیک «تخمین میانگین» در چارچوب α-LDP اثبات کنیم. این نتیجه نشان می‌دهد که چرا روش‌هایی مانند پاسخ تصادفی (RR) که در فصل ۳ معرفی شدند، از نظر نرخ همگرایی بهینه هستند.

۴-۴-۱ تعریف مسئله

فرض کنید داده‌های ورودی $X_1, \dots, X_n \in \{-1, 1\}^d$ بردارهای تصادفی مستقل با میانگین $\theta = \mathbb{E}_{X_i}$ باشند، به طوری که $\theta \in [-1, 1]^d$. هدف ما تخمین پارامتر θ است. ما خطای تخمین‌گر $\hat{\theta}$ را با استفاده از تابع زیان «مربع خطای اقلیدسی» (L_2^2) می‌سنجیم. هدف یافتن نرخ مینی‌مکس زیر است:

$$\mathfrak{M}_n(\theta, \mathcal{M}) = \inf_{\mathcal{M}, \hat{\theta}} \sup_{\theta \in [-1,1]^d} \mathbb{E}\left[\|\hat{\theta}(Z_1, \dots, Z_n) - \theta\|_2^2\right] \tag{۲۶-۴}$$

که در آن اینفیمم روی تمام مکانیزم‌های α-LDP و تمام تخمین‌گرهای ممکن گرفته می‌شود.

قضیه‌ی ۴-۴ (کران پایین مینی‌مکس برای تخمین میانگین) *برای هر مکانیزم α-LDP با $\alpha \in (0, 1]$، خطای مینی‌مکس در تخمین میانگین یک توزیع روی مکعب $\{-1, 1\}^d$ حداقل از مرتبه زیر است:*

$$\inf_{\hat{\theta}} \sup_{P} \mathbb{E}\left[\|\hat{\theta} - \theta(P)\|_2^2\right] \geqslant c \cdot \frac{d}{n\alpha^2} \tag{۲۷-۴}$$

که $c > 0$ یک ثابت عددی مطلق است.

اثبات. برای اثبات این قضیه، از روش «لم اسود»[16] (فصل ۲) استفاده می‌کنیم. استراتژی کلی این است که یک زیرمجموعه گسسته از فضای پارامتر (یک ابرمکعب) بسازیم و نشان دهیم که تشخیص رأس‌های این مکعب تحت محدودیت α-LDP دشوار است.

۱. ساختن فضای پارامتر گسسته: مجموعه رئوس ابرمکعب دودویی $\mathcal{V} = \{-1, 1\}^d$ را در نظر بگیرید. برای هر بردار $v \in \mathcal{V}$، یک توزیع احتمال P_v روی داده‌های ورودی $\{-1, 1\}^d$ تعریف می‌کنیم به طوری که مولفه‌های آن مستقل باشند. برای هر مؤلفه $j \in \{1, \dots, d\}$، میانگین را به صورت زیر تنظیم می‌کنیم:

$$\mathbb{E}_{P_v}[(X)_j] = v_j \Delta \tag{۲۸-۴}$$

[16] Assouad's Lemma

که $\Delta \in (۰, ۱]$ پارامتری است که بعداً مقدار دقیق آن را تعیین می‌کنیم. به عبارت دیگر، بردار میانگین متناظر با v برابر است با $\theta_v = \Delta \cdot v$.

۲. اعمال لم اسود: طبق لم اسود، برای هر تخمین‌گر $\hat{\theta}$، بیشینه ریسک روی این مجموعه متناهی با مجموع خطاهای آزمون فرضیه باینری در هر مؤلفه کران‌دار می‌شود:

$$\max_{v \in \mathcal{V}} \mathbb{E}\|\hat{\theta} - \theta_v\|_۲^۲ \geqslant \frac{d}{۲} \cdot \Delta^۲ \cdot \min_{v,v': H(v,v')=۱} (۱ - \|M_v^n - M_{v'}^n\|_{TV}) \tag{۲۹-۴}$$

در این‌جا $H(v, v')$ فاصله همینگ است و M_v^n توزیع مشترک n خروجی مشاهده شده مکانیزم $\mathcal{M}$ تحت توزیع ورودی P_v است.

۳. استفاده از خاصیت انقباض LDP: اکنون باید فاصله تغییرات کل $\|M_v^n - M_{v'}^n\|_{TV}$ را کران‌دار کنیم. دو همسایه v و v' را در نظر بگیرید که تنها در یک مؤلفه (مثلاً مؤلفه j) تفاوت دارند. طبق نامساوی پینسکر و خاصیت تانسوری واگرایی KL برای نمونه‌های مستقل $(Z_۱, \ldots, Z_n)$:

$$\|M_v^n - M_{v'}^n\|_{TV}^۲ \leqslant \frac{۱}{۲} D_{KL}(M_v^n \| M_{v'}^n) = \frac{n}{۲} D_{KL}(M_v \| M_{v'}) \tag{۳۰-۴}$$

در این‌جا M_v توزیع خروجی یک‌بار اجرای مکانیزم برای یک داده ورودی است. حال از «قضیه انقباض قوی» (قضیه ۴-۱) استفاده می‌کنیم. می‌دانیم که واگرایی خروجی توسط واگرایی ورودی و بودجه محرمانگی کنترل می‌شود:

$$D_{KL}(M_v \| M_{v'}) \leqslant (e^\alpha - ۱)^۲ \|P_v - P_{v'}\|_{TV}^۲ \tag{۳۱-۴}$$

چون P_v و $P_{v'}$ توزیع‌های برنولی ضربی هستند که تنها در مؤلفه j تفاوت دارند (با میانگین‌های $+\Delta$ و $-\Delta$)، فاصله تغییرات کل آن‌ها دقیقاً برابر است با تفاوت میانگین‌ها تقسیم بر دامنه (در این‌جا ساده‌سازی شده):

$$\|P_v - P_{v'}\|_{TV} = \frac{(۱ + \Delta) - (۱ - \Delta)}{۲} = \Delta \tag{۳۲-۴}$$

با ترکیب این روابط و فرض $\alpha \leqslant ۱$ (که نتیجه می‌دهد $(e^\alpha - ۱)^۲ \approx \alpha^۲$):

$$\|M_v^n - M_{v'}^n\|_{TV}^۲ \leqslant \frac{n}{۲} \alpha^۲ \Delta^۲ \tag{۳۳-۴}$$

بنابراین:

$$\|M_v^n - M_{v'}^n\|_{TV} \leqslant \alpha \Delta \sqrt{\frac{n}{۲}} \tag{۳۴-۴}$$

۴. تنظیم پارامتر Δ: برای این‌که کران پایین در رابطه (۴-۲۹) غیر صفر و بزرگ باشد، باید عبارت داخل پرانتز مثبت باشد. ما Δ را چنان انتخاب می‌کنیم که فاصله TV برابر یک مقدار ثابت کوچک (مثلاً ۱/۲) شود:

$$\alpha\Delta\sqrt{\frac{n}{۲}} = \frac{۱}{۲} \implies \Delta = \frac{۱}{\alpha\sqrt{۲n}} \tag{۴-۳۵}$$

با جایگذاری این مقدار Δ در رابطه لم اسود:

$$\begin{aligned}\sup_{\theta} \mathbb{E}\|\hat{\theta} - \theta\|_۲^۲ &\geqslant \frac{d}{۲} \cdot \Delta^۲ \cdot \left(۱ - \frac{۱}{۲}\right) \\ &= \frac{d}{۴}\left(\frac{۱}{۲n\alpha^۲}\right) \\ &= \frac{d}{۸n\alpha^۲}\end{aligned} \tag{۴-۳۶}$$

این اثبات نشان می‌دهد که کران پایین از مرتبه $\Omega(\frac{d}{n\alpha^۲})$ است. □

۴-۴-۲ بحث و تفسیر

نتیجه به دست آمده در قضیه ۴-۴ دارای پیام‌های مهمی برای طراحی سیستم‌های خصوصی است:

۱. **هزینه محرمانگی:** در مقایسه با تخمین میانگین در حالت غیرخصوصی (یا متمرکز) که نرخ خطا $\frac{d}{n}$ است، در حالت α-LDP خطا با ضریب $\frac{۱}{\alpha^۲}$ افزایش می‌یابد. این یعنی برای جبران نویز اضافه شده توسط محرمانگی، حجم داده‌ها (n) باید متناسب با مربع بودجه محرمانگی افزایش یابد.

۲. **وابستگی به ابعاد:** خطا به صورت خطی با بعد داده (d) رشد می‌کند. این رفتار مشابه حالت کلاسیک است و نشان می‌دهد که α-LDP وابستگی به ابعاد را تغییر نمی‌دهد، بلکه تنها ضریب ثابت را بدتر می‌کند.

۳. **بهینگی مکانیزم‌ها:** در ۳-۴-۶ دیدیم که مکانیزم تخمین میانگین دوچی دقیقاً واریانسی از مرتبه $\frac{d \log d}{n\alpha^۲}$ (برای روش‌های تنک) یا $\frac{d}{n\alpha^۲}$ تولید می‌کند. تطابق کران پایین ثابت شده در این‌جا با کران بالای آن مکانیزم‌ها، ثابت می‌کند که الگوریتم‌های موجود نه تنها خوب، بلکه «بهینه مینی‌مکس» هستند و بهبود قابل توجهی در نرخ همگرایی آن‌ها ممکن نیست.

به این ترتیب، ما نشان دادیم که محدودیت‌های ذاتی α-LDP مانع از دست‌یابی به دقت‌های بالاتر می‌شود و این محدودیت ناشی از انقباض اطلاعاتی است که در ذات تعریف محرمانگی نهفته است.

فصل ۵

نتیجه‌گیری و پیشنهادها

۵-۱ جمع‌بندی و دستاوردهای اصلی

در این پایان‌نامه مروری تحلیلی بر حدود بنیادین دقت آماری در حضور محدودیت‌های محرمانگی ارائه شد. پرسش محوری این بود که هزینه اطلاعاتی که LDP بر داده‌ها تحمیل می‌کند، ذاتی است یا ناشی از ضعف الگوریتم‌های موجود.

در فصل دوم، زیربنای ریاضی لازم بنا نهاده شد. تعریف LDP نه صرفاً به عنوان یک الگوریتم، بلکه به عنوان محدودیتی روی کانال‌های ارتباطی معرفی شد و خانواده fـ واگرایی‌ها به عنوان ابزار اصلی سنجش فاصله میان توزیع‌ها مرور گردید. در فصل سوم، مکانیزم‌های استاندارد LDP تشریح شد و نشان داده شد که هزینه عدم اعتماد به سرور مرکزی، افزایش پیچیدگی نمونه از مرتبه $\mathcal{O}(n)$ به $\mathcal{O}(n^۲)$ است.

در فصل چهارم، رویکرد پیشگامانه دوچی، جردن و وین‌رایت [۷] به تفصیل بررسی شد. در این رویکرد تحلیل‌ها بر مبنای انقباض واگرایی KL بنا شده و با ابزارهایی نظیر لم اسود و نامساوی فانو، کران‌های پایین مینی‌مکس اثبات می‌شوند. در کنار مزایای این رویکرد، محدودیت آن نیز تحلیل شد: واگرایی KL تقارن و کران‌های ذاتی α-LDP را به طور کامل بازتاب نمی‌دهد و در رژیم‌های محرمانگی متوسط یا بالا، به ضرایب نادقیق در کران‌های مینی‌مکس منجر می‌شود.

در فصل پنجم، چارچوب مکمل «انقباض E_γـ واگرایی» بر پایه کارهای آسوده و همکاران [؟، ؟] مرور و تحلیل شد. نشان داده شد که واگرایی E_γ با پارامتر $\gamma = e^\alpha$ زبان طبیعی DP است: شرط α-LDP دقیقاً معادل با صفر شدن این واگرایی است.

۵-۱-۱ دستاوردهای اصلی پژوهش

مهم‌ترین دستاوردهای این پایان‌نامه به شرح زیر است:

- **گذار از تقریب به دقت کامل:** در رویکرد سنتی (فصل ۴)، استخراج کران‌های پایین مستلزم بسط‌های تیلور واگرایی KL حول صفر بود که تنها برای αهای کوچک معتبر است. رویکرد فصل ۵ نشان داد که با استفاده از E_γ می‌توان مستقیماً و بدون تقریب، کران‌هایی استخراج کرد که برای تمام مقادیر $\alpha \in (۰, \infty)$ معتبرند.
- **اصلاح ضرایب انقباض:** ضریب انقباض واقعی برای مکانیزم‌های α-LDP از مرتبه $(e^\alpha - ۱)^۲$ است، در حالی که تحلیل‌های مبتنی بر KL ضریب اضافه $(e^\alpha + ۱)$ را نیز حمل می‌کردند. این اصلاح به ویژه در رژیم‌های محرمانگی متوسط تفاوتی معنادار در تخمین حجم داده مورد نیاز ایجاد می‌کند.
- **دیدگاه هندسی:** مسئله طراحی مکانیزم‌های بهینه α-LDP را می‌توان به عنوان یافتن توزیع‌هایی تفسیر کرد که E_{e^α}ـ واگرایی آن‌ها صفر باشد. این دیدگاه درک عمیق‌تری نسبت به تعریف جبری «نسبت احتمالات» فراهم می‌کند و راه را برای طراحی مکانیزم‌های جدید هموار می‌سازد.

در مجموع، این پایان‌نامه نشان داد که برای تحلیل دقیق سیستم‌های خصوصی، باید ابزار واگرایی را با ماهیت ذاتی محدودیت محرمانگی هم‌راستا کرد و هزینه اطلاعاتی LDP یک مانع هندسی اجتناب‌ناپذیر است.

۵-۲ پیشنهادهایی برای تحقیقات آتی

پژوهش حاضر گامی در جهت یک‌پارچه‌سازی نظریه محرمانگی تفاضلی موضعی و بررسی کران‌های مینی‌مکس برداشت. در ادامه مسیرهای پژوهشی‌ای که می‌توانند امتداد طبیعی این پایان‌نامه باشند معرفی می‌شوند.

۵-۲-۱ تحلیل انقباض در مدل شافل

مدل شافل[1] به عنوان حد واسطی میان مدل‌های LDP و CDP ظهور کرده و ادبیات موجود نشان می‌دهد که عمل شافل کردن باعث تقویت محرمانگی[2] می‌شود. مسئله‌ای که همچنان باز مانده این است که آیا می‌توان با استفاده از چارچوب E_γ ـ واگرایی، کران‌های دقیق‌تری برای این پدیده نسبت به تحلیل‌های مبتنی بر KL استخراج کرد یا خیر.

پرسش مرتبط دیگری که پیش می‌آید این است که وقتی پارامتر محرمانگی ε_i برای هر کاربر متفاوت باشد، کران‌های انقباض چگونه تغییر می‌کنند و آیا می‌توان کران‌های مینی‌مکس را بر حسب توزیع پیشین روی ε_i بیان کرد.

۵-۲-۲ ارتباط با محرمانگی تفاضلی رنی

محرمانگی تفاضلی رنی[3] که امروزه استاندارد طلایی تحلیل ترکیب مکانیزم‌ها محسوب می‌شود، بر مبنای خانواده واگرایی‌های رنی بنا شده است. بررسی رابطه صریح میان ضرایب انقباض به‌دست‌آمده در این پژوهش و انقباض واگرایی‌های رنی، می‌تواند پلی میان این دو چارچوب ایجاد کند و بهره‌گیری از ابزارهای قدرتمند RDP را در تحلیل‌های موضعی ممکن سازد.

۵-۲-۳ تخمین‌گرهای تطبیقی

تمام تحلیل‌های این پایان‌نامه بر فرض ثابت بودن بودجه محرمانگی α استوار بودند. طراحی مکانیزم‌هایی که بودجه خود را به صورت تطبیقی[4] و بر اساس سختی داده تنظیم کنند مسیر پژوهشی ارزشمندی است. چالش اصلی، اثبات این نکته است که فرآیند تنظیم پارامتر خود باعث نقض شرط α-LDP نشود.

۵-۲-۴ تعمیم به داده‌های وابسته

تمامی کران‌های مینی‌مکس این پایان‌نامه بر فرض استقلال داده‌ها (i.i.d.) استوار بودند. تعمیم لم اسود[5] و نامساوی‌های انقباض به فرآیندهای تصادفی وابسته مانند زنجیره‌های مارکوف نیازمند پژوهش‌های بیشتری

[1] Shuffle Model
[2] Privacy Amplification
[3] Rényi Differential Privacy (RDP)
[4] Adaptive Mechanisms
[5] Assouad's Lemma

است؛ زیرا در حضور وابستگی خاصیت تانسوری واگرایی‌ها برقرار نیست و باید ابزارهای جدیدی برای سنجش نرخ انباشت اطلاعات توسعه یابد.

Bibliography

[1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):131–142, 1966.

[2] Shahab Asoodeh, Maryam Aliakbarpour, and Flavio P. Calmon. Local differential privacy is equivalent to contraction of an f-divergence. In *2021 IEEE International Symposium on Information Theory (ISIT)*, page 545–550. IEEE Press, 2021.

[3] Shahab Asoodeh and Huanyu Zhang. Contraction of locally differentially private mechanisms. *IEEE Journal on Selected Areas in Information Theory*, 5:385–395, 2024.

[4] Michael Barbaro and Tom Zeller. A face is exposed for aol searcher no. 4417749. *New York Times*, 01 2006.

[5] Imre Csiszár. Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 8(1-2):85–108, 1963.

[6] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438, 2013.

[7] John C Duchi, Michael I Jordan, and Martin J Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

[8] Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.

[9] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*, 9(3–4):211–407, 2014.

[10] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, page 1054–1067, New York, NY, USA, 2014. Association for Computing Machinery.

[11] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.*, 17(1):492–542, January 2016.

[12] Gautam Kamath. CS860: Algorithms for private data analysis – lecture 17: Local differential privacy. Lecture Notes, University of Waterloo, 2020. http://www.gautamkamath.com/CS860notes/lec17.pdf (Accessed: 2026-02-14).

[13] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 531–540, 2008.

[14] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE, 2017.

[15] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125, 2008.

[16] Yury Polyanskiy, H. Vincent Poor, and Sergio Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theor.*, 56(5):2307–2359, May 2010.

[17] Igal Sason and Sergio Verdu. f-divergence inequalities. *IEEE Trans. Inf. Theor.*, 62(11):5973–6006, November 2016.

[18] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[19] U.S. House of Representatives Committee on Oversight and Government Reform. The equifax data breach. Majority staff report, U.S. House of Representatives, December 2018.

[20] Teng Wang, Xuefeng Zhang, Jingyu Feng, and Xinyu Yang. A comprehensive survey on local differential privacy toward data statistics and analysis. *Sensors*, 20:1–48, 2020.

[21] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In *Proceedings of the 26th*

USENIX Conference on Security Symposium, SEC'17, page 729–745, USA, 2017. USENIX Association.

[22] Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American statistical association*, 60(309):63–69, 1965.

واژه‌نامه

الف

احتمال ..... Probability
اسود ..... Assouad
اطلاعات ..... Information
اطلاعات متقابل ..... Mutual Information
اکیداً ..... Strictly
الگوریتم ..... Algorithm
امنیت ..... Security
اندازه ..... Measure
اندازه‌پذیر ..... Measurable
انقباض ..... Contraction
آنتروپی نسبی ..... Relative Entropy

ب

بازشناسایی ..... Re-identification
بسته‌بندی ..... Packing
به‌خاطرسپاری ..... Memorization
بودجه محرمانگی ..... Privacy Budget
بهینه ..... Optimal

پ

پاسخ تصادفی ..... Randomized Response
پاسخ تصادفی تعمیم‌یافته Generalized Randomized Response
پایگاه‌داده ..... Database
پرس‌وجو ..... Query
پرس‌وجوهای شمارشی ..... Counting Queries
پس‌پردازش ..... Post-processing
پیشین ..... Prior
پینسکر ..... Pinsker
پیوستگی مطلق ..... Absolute Continuity

ت

تابع ..... Function
تبدیل هادامارد ..... Hadamard Transform
تحدب مشترک ..... Joint Convexity
تحلیل ..... Analytics
تخمین ..... Estimation
تخمین‌گر نااریب ..... Unbiased Estimator
ترکیب ..... Composition
ترکیب ترتیبی ..... Sequential Composition
ترتیبی ..... Sequential
تصادفی‌ساز موضعی ..... Local Randomizer
تصادفی‌سازی ..... Randomization
تعاملی ..... Interactive
تقارن ..... Symmetry
تقویت محرمانگی ..... Privacy Amplification
تکیه‌گاه ..... Support

تنسوری شدن ..... Tensorization

توزیع ..... Distribution

ج

جانبی ..... Auxiliary

جبر ..... Algebra

جرم ..... Mass

جفریز ..... Jeffreys

جمع‌پذیری ..... Additivity

جهان ..... Universe

چ

چوب هاکی ..... Hockey-Stick

ح

حجم نمونه ..... Sample Complexity

حریم خصوصی ..... Privacy

حساس ..... Sensitive

حساسیت ..... Sensitivity

حساسیت سراسری ..... Global Sensitivity

حمله ..... Attack

حمله‌ی تفاضلی ..... Differencing Attack

خ

خاصیت ..... Property

خطر مینی‌مکس ..... Minimax Risk

خطای میانگین مربعات ..... Mean Squared Error (MSE)

د

داده ..... Data

دامنه ..... Domain

دانش ..... Knowledge

دقیق ..... Tight / Sharp

دسترسی ..... Access

دوبروشین ..... Dobrushin

دیراک ..... Dirac

دیوارهای آتش ..... Firewalls

ر

رادون‌ـ نیکودیم ..... Radon-Nikodym

رشته ..... String

رمزنگاری ..... Cryptography

رنی ..... Rényi

رویکرد ..... Approach

ریسک ..... Risk

ز

زیان ..... Loss

س

سراسری ..... Global

سودمندی ..... Utility

ش

شبه‌شناسه ..... Quasi-identifier

شکست ..... Failure

شل ..... Loose

شمارا ..... Countable

شناسه ..... Identifier

ص

صریح ..... Explicit

ض

ضرب‌شونده ..... Multiplicative

ضریب ..... Coefficient

ط

طرح‌ریزی ..... Sketching

ع

عملگر ..... Operator

غ

غیرتعاملی ..... Non-interactive

غیرقابل تفکیک ..... Indistinguishable

غیرمنصفانه ..... Unfair

ف

فاصله ..... Distance

فاصله تغییرات کل ..... Total Variation (TV)

فانو ..... Fano

فضا ..... Space

فیلترهای بلوم ..... Bloom Filters

ق

قطعی ..... Deterministic

ک

کانال ..... Channel

کانال انقباضی ..... Contraction Channel

کدگذاری وان‌ـهات ..... One-Hot Encoding

کدگذاری یکانی ..... Unary Encoding (UE)

کرامر‌ـرائو ..... Cramér-Rao

کران ..... Bound

کلان‌داده ..... Big Data

کنترل ..... Control

کنترل افشای آماری Statistical Disclosure Control

کولبک‌ـلایبلر ..... Kullback-Leibler (KL)

گ

گمنام‌سازی ..... Anonymization

ل

لم ..... Lemma

لوکم ..... Le Cam

م

ماتریس اطلاعات فیشر ..... Fisher Information

ماتریس تصادفی سطری ... Row-stochastic Matrix

متصدی مورد اعتماد ..... Trusted Curator

متغیر تصادفی ..... Random Variable

متمرکز ..... Centralized

مجموع ..... Sum

مجموعه داده‌ی نتفلیکس ..... Netflix Prize Data

محدب ..... Convex

محرمانگی به عنوان انقباض Privacy as Contraction

محرمانگی تفاضلی ..... Differential Privacy

محرمانگی تفاضلی موضعی ..... Local Differential

Privacy

محرمانگی گروهی ..... Group Privacy

مخلوط ..... Mixture

مستقل و هم‌توزیع .. Independent and Identically Distributed (i.i.d.)

مکانیزم تصادفی ..... Randomized Mechanism

مکانیزم‌های پله‌ای ..... Staircase Mechanisms

معنایی ..... Semantic

موازنه ..... Trade-off

میانگین ..... Mean

میدان ..... Field

ن

ناااریب ..... Unbiased

نامساوی ..... Inequality

نامساوی پردازش داده Data Processing Inequality (DPI)

نامساوی پردازش داده قوی Strong Data Processing Inequalities (SDPI)

نامساوی مثلث ..... Triangle Inequality

نسبت درست‌نمایی ..... Likelihood Ratio

نسبت سیگنال به نویز Signal-to-Noise Ratio (SNR)

نقاط فرین ..... Extreme Points

نقطه شکست مرکزی ..... Single Point of Failure

نقطه‌ای ..... Pointwise

نظریه برآورد ..... Estimation Theory

نظریه تصمیم آماری .. Statistical Decision Theory

نظریه مینی‌مکس آماری ..... Statistical Minimax Theory

نمایش انتگرالی ..... Integral Representation

نویز ..... Noise

و

واگرایی ..... Divergence

ه

هشینگ محلی بهینه ..... Optimal Local Hashing (OLH)

هسته ..... Kernel

هسته‌های احتمالاتی ..... Probability Kernels

هسته‌ی انتقال احتمال .... Probability Transition Kernel

هسته‌ی احتمالاتی (کانال مارکوف) / Markov Kernel / Probability Kernel

هلینگر ..... Hellinger

هم‌زمان ..... Simultaneous

همسایه ..... Neighboring / Adjacent

ی

یادگیری ماشین ..... Machine Learning

یادگیری ماشین خصوصی ..... Private Machine Learning

یک‌به‌یک ..... Injective

یک‌نواختی ..... Monotonicity

Abstract

The Differential Privacy framework, serving as the gold standard for data protection, provides a strong mathematical guarantee ensuring that the output of algorithms does not exhibit meaningful sensitivity to the presence or absence of a specific individual's data. However, the standard centralized model of this framework requires the aggregation of raw data by an intermediary entity. The Local Differential Privacy (LDP) framework was developed to eliminate this reliance; in this model, data is perturbed with noise on the user's side prior to aggregation. Nevertheless, resolving the trust issue leads to another fundamental challenge: a severe degradation in statistical utility, such that the sample complexity increases from $\mathcal{O}(n)$ in the centralized model to approximately $\mathcal{O}(n^2)$ in the local model.

The central question is whether this substantial drop in accuracy constitutes an inherent limitation or stems from the shortcomings of current algorithms. To address this question, this thesis presents an analytical survey of the leading literature in this domain. Two complementary approaches are studied and integrated: first, the statistical minimax theory framework based on the works of Duchi, Jordan, and Wainwright, which establishes lower bounds on estimation error using tools such as Assouad's Lemma and Fano's Inequality; second, the information geometry framework based on the works of Asoodeh et al., which interprets local mechanisms as *information contraction channels*.

This dissertation demonstrates that the Local Differential Privacy constraint is equivalent to the contraction of f-divergences, and exploits this equivalence to derive precise contraction coefficients for the Chi-squared, Kullback–Leibler, and Hockey-Stick divergences. Combining these coefficients with classical tools from statistical decision theory yields minimax lower bounds in a unified and coherent manner, and

explicitly establishes that the degradation of the convergence rate to $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ is an unavoidable information-theoretic and geometric barrier. This framework provides a rigorous benchmark for evaluating the optimality of future algorithms in the domain of data privacy.

Keywords: Local Differential Privacy, Statistical Minimax Theory, f-Divergences, Strong Data Processing Inequalities, Privacy–Accuracy Trade-off

Sharif University of Technology

Department of Mathematics

M.Sc. Thesis

Privacy-Preserving Machine Learning

By:

Firoozeh Abrishami

Supervisor:

Dr. Javad Ebrahimi Boroujeni

March 2026