



دانشگاه صنعتی شریف
دانشکده علوم ریاضی

پایان نامه کارشناسی ارشد
ریاضی کاربردی

تحلیل نظریه اطلاعاتی محرمانگی تفاضلی موضعی و کاربردهای آماری آن

نگارش

فیروزه ابریشمی

استاد راهنما

دکتر جواد ابراهیمی بروجنی

۱۴۰۴ اسفند

الله اعلم

به نام خدا

دانشگاه صنعتی شریف

دانشکده علوم ریاضی

پایان نامه کارشناسی ارشد

این پایان نامه به عنوان تحقیق بخشی از شرایط دریافت درجه کارشناسی ارشد است.

عنوان: تحلیل نظریه اطلاعاتی محرمانگی تفاضلی موضعی و کاربردهای آماری آن

نگارش: فیروزه ابریشمی

کمیته ممتحنین

استاد راهنما: دکتر جواد ابراهیمی امضاء:

بروجنی

استاد مشاور: استاد مشاور امضاء:

استاد مدعو: استاد ممتحن امضاء:

تاریخ:



اظهارنامه

(اصالت متن و محتوای پایان نامه کارشناسی ارشد)

عنوان پایان نامه: تحلیل نظریه اطلاعاتی محترمانگی تفاضلی موضعی و کاربردهای آماری آن

استاد مشاور: استاد مشاور

استاد راهنما: دکتر جواد ابراهیمی بروجنی

این جانب فیروزه ابریشمی اظهار می دارد:

۱. متن و نتایج علمی ارائه شده در این پایان نامه اصیل بوده و زیرنظر استادان نام برده شده در بالا تهیه شده است.
۲. متن پایان نامه به این صورت در هیچ جای دیگری منتشر نشده است.
۳. متن و نتایج مندرج در این پایان نامه، حاصل تحقیقات این جانب به عنوان دانشجوی کارشناسی ارشد دانشگاه صنعتی شریف است.
۴. کلیه مطالبی که از منابع دیگر در این پایان نامه مورد استفاده قرار گرفته، با ذکر مرجع مشخص شده است.

نگارنده: فیروزه ابریشمی

تاریخ:

امضاء:

نتایج تحقیقات مندرج در این پایان نامه 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 اختراق 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 مأخذ بلامانع است.

نگارنده: فیروزه ابریشمی

تاریخ:

امضاء:

استاد راهنما: دکتر جواد ابراهیمی بروجنی

تاریخ:

امضاء:

چکیده تامین امنیت مسئله‌ایست که از دیرباز انسان‌ها را به خود مشغول کرده است. در عصر اطلاعات و ارتباطات نیز این مسئله به شکلی متفاوت اهمیت خود را نشان می‌دهد.

کلیدواژه‌ها:

فهرست مطالب

۱	مقدمه	۱
۲	۱-۱ اهمیت داده‌ها و ضرورت حفظ حریم خصوصی	۲
۴	۱-۱-۱ محرمانگی تفاضلی (Differential Privacy)	۴
۷	۱-۱-۲ محرمانگی تفاضلی موضعی (Local Differential Privacy)	۷
۸	۲-۱ کارهای پیشین و مورادیات	۸
۸	۲-۱-۱ آغازگرها: از پیمایش‌های آماری تا تعریف مدرن محرمانگی	۸
۱۲	۲-۱-۲ دسته‌بندی پروتکل‌های موضعی: تعاملی و غیرتعاملی	۱۲
۱۲	۳-۱ چالش سودمندی و موازنۀ دقت-محرمانگی	۱۲
۱۳	۴-۱ نگاهی آماری به LDP: چارچوب مینی‌مکس و حدود بنیادین	۱۳
۱۴	۱-۲ بیان مسئله و اهداف پژوهش	۱۴
۱۵	۱-۳-۱ رویکرد تحلیل: f -واگرایی‌ها به عنوان زبان مشترک	۱۵
۱۶	۱-۳-۲ رویکرد تحلیل: چارچوب انقباض f -واگرایی‌ها	۱۶
۱۶	۱-۳-۳ اهداف و ساختار پژوهش	۱۶
۱۷	۱-۴ ساختار پایان‌نامه	۱۷
۱۹	۲ پیش‌نیازها	۱۹
۱۹	۱-۱ محرمانگی تفاضلی مرکز (CDP)	۱۹
۱۹	۱-۱-۲ مدل اعتماد و تعریف رسمی	۱۹

۲۲	۲-۱-۲ تفسیر پارامترهای محramانگی
۲۲	۳-۱-۲ تعاریف معادل و صورت بندی های جایگزین
۲۳	۴-۱-۲ مکانیزم های پایه
۲۵	۵-۱-۲ مکانیزم های بنیادی محramانگی تفاضلی
۲۹	۶-۱-۲ ترکیب پذیری
۲۹	۷-۱-۲ محramانگی گروهی
۳۰	۸-۱-۲ محدودیت مدل مرکز
۳۰	f	۲-۲ واگرایی ها
۳۰	۱-۲-۲ تعریف رسمی در فضای اندازه پذیر
۳۲	۲-۲-۲ نمونه های مهم و توابع مولد
۳۴	۳-۲-۲ خواص بنیادین و روابط بین f-واگرایی ها
۳۶	۳-۲ مبانی آماری و کران های اطلاعاتی
۳۷	۱-۳-۲ نظریه تصمیم و ریسک مینی مکس
۳۷	۲-۳-۲ تقلیل به آزمون فرض (روش بسته بندی)
۳۸	۳-۳-۲ نامساوی های کران پایین
۴۰	۳ محramانگی تفاضلی موضعی
۴۰	۱-۳ مقدمه
۴۰	۲-۳ تعاریف رسمی و مدل های محاسباتی
۴۱	۱-۲-۳ تعریف ریاضی LDP
۴۲	۲-۲-۳ محramانگی تقریبی
۴۲	۳-۳ پروتکل های تعاملی و خواص ترکیب
۴۳	۱-۳-۳ پروتکل های غیر تعاملی
۴۳	۲-۳-۳ پروتکل های تعاملی (ترتیبی)
۴۴	۳-۳-۳ قضیه ترکیب ترتیبی

۴۴	۴-۳ مکانیزم‌های پایه در LDP
۴۵	۱-۴-۳ پاسخ تصادفی دودویی (RR)
۴۶	۲-۴-۳ پاسخ تصادفی تعیین‌یافته (GRR)
۴۷	۳-۴-۳ مکانیزم‌های مبتنی بر کدگذاری یگانی (UE)
۴۹	۴-۴-۳ تحلیل مقایسه‌ای: چرا GRR در ابعاد بالا شکست می‌خورد؟
۴۹	۵-۴-۳ مکانیزم لاپلاس موضعی
۵۱	۶-۴-۳ مکانیزم تخمین میانگین دوچی (۱-بیتی)
۵۲	۵-۳ چالش سودمندی و هزینه عدم اعتماد
۵۲	۱-۵-۳ تعریف مسئله: تخمین میانگین دودویی
۵۳	۲-۵-۳ تحلیل در مدل مرکزی (CDP)
۵۴	۳-۵-۳ تحلیل در مدل موضعی (LDP)
۵۴	۴-۵-۳ نتیجه‌گیری: شکاف کارایی
۵۶	۴ تحلیل مینی‌مکس و هندسه اطلاعاتی در LDP
۵۶	۱-۴ مقدمه
۵۷	۲-۴ محترمانگی به عنوان انقباض
۵۹	۱-۲-۴ تفسیر رژیم‌های محترمانگی
۶۰	۴-۳ نامساوی‌های پردازش داده قوی (SDPI)
۶۰	۱-۳-۴ کران انقباض برای واگرایی کای-دو (χ^2)
۶۱	۲-۳-۴ مزیت استفاده از χ^2 نسبت به KL
۶۲	۴-۴ اثبات نرخ مینی‌مکس برای تخمین میانگین
۶۲	۱-۴-۴ تعریف مسئله
۶۴	۲-۴-۴ بحث و تفسیر
۶۵	۵ چارچوب انقباض E_7 -واگرایی

۶۵	۱-۵ مقدمه
۶۶	۲-۵ همارزی E_γ -واگرایی LDP
۶۷	۱-۲-۵ تفسیر بر اساس ضریب انقباض
۶۷	۳-۵ کران‌های انقباض برای f -واگرایی‌های عمومی
۶۷	۱-۳-۵ قضیه انقباض بهبودیافته برای χ^2
۶۸	۲-۳-۵ مقایسه با کران‌های دوچی (فصل ۴)
۶۹	۴-۵ کاربرد در نرخ‌های مینی‌مکس و مقایسه نهایی
۶۹	۱-۴-۵ بازبینی مسئله تخمین میانگین
۷۰	۵-۵ نتیجه‌گیری فصل
۷۲	مراجع
۷۵	واژه‌نامه
۷۷	آ مطالب تكميلی

فهرست جداول

فهرست تصاویر

- ۱-۲ مدل محترمانگی تفاضلی مت مرکز (CDP) با یک مت صدی مورد اعتماد.

۱-۳ گذار از مدل مت مرکز به موضوعی؛ نویز به صورت محلی (Local) روی دستگاه کاربر اضافه می شود.

۲-۳ (الف) مدل گرافیکی نمایش دهنده روابط استقلال شرطی بین داده های خصوصی $\{X_i\}_{i=1}^n$ و متغیر های مشاهده شده $\{Z_i\}_{i=1}^n$ در پروتکل های تعاملی؛ پیکان های افقی نشان دهنده وابستگی Z_i به تاریخچه های پیشین هستند. (ب) مدل گرافیکی ساده تر در پروتکل های غیر تعاملی که در آن خروجی ها به شرط ورودی ها از یکدیگر مستقل هستند.

فصل ۱

مقدمه

در این بخش به توضیح مسئله‌ی محترمانگی 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 نظری 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

¹Big Data

²Machine Learning

³Data Analytics

⁴Privacy

⁵Sensitive 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 خود داده‌ها نیز می‌توانند به عنوان شناسه عمل کنند [۴].

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

⁶Utility

⁷Explicit Identifiers

⁸Anonymization

⁹Re-identification Attacks

¹⁰Auxiliary Information

¹¹Quasi-identifier

¹²Netflix Prize Data

دانش پس زمینه‌ی پیش‌بینی نشده‌ای 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، زمینه‌ی را برای ظهور مفهوم محرمانگی تفاضلی فراهم کرد که در بخش‌های آتی به تفصیل به آن خواهیم پرداخت.

۱-۱-۱ محرمانگی تفاضلی (Differential Privacy)

در پاسخ به چالش‌های امنیتی و ناکارآمدی روش‌های سنتی گمنام‌سازی، دُورک 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 بین این 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 خود نتیجه‌ی تحلیل، حامل اطلاعات است.

در پاسخ به این چالش، دُورک 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

¹³Differential Privacy

¹⁴Cryptography

¹⁵Access Control

¹⁶Auxiliary Knowledge

¹⁷Semantic Security

فرد به طور چشمگیری افزایش یابد.

برای مدل‌سازی این مفهوم، می‌توانیم از استعاره‌ی جهان‌های موازی استفاده کنیم. دو پایگاهداده همسایه^{۱۸} (\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^{۱۹} باشند. اگر مکانیزم محروم‌گی 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^{۲۰} بهره می‌بریم. این الگوریتم‌ها با تزریق نویز کنترل شده، 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^{۲۱} ممکن نیست. برای درک بهتر، یک حمله‌ی تفاضلی^{۲۲} کلاسیک را در نظر بگیرید. فرض کنید f تابعی قطعی است که میانگین درآمد n فرد در پایگاهداده را برمی‌گردد:

$$f(\mathcal{D}) = \frac{1}{n} \sum_{i=1}^n x_i$$

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

۲. میانگین درآمد همان افراد، به جز فرد هدف 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 بلاfacile دو جهان موازی را از هم تمایز می‌کند.

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

¹⁸Neighboring Datasets

¹⁹Indistinguishable

²⁰Randomized Algorithms

²¹Deterministic

²²Differencing Attack

بدینانه 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 دامنه \mathcal{X} یکبهیک^{۲۳} (یا حتی در بازه‌ای مشخص وارونپذیر) باشد، محرمانگی به طور کامل از بین می‌رود؛ زیرا مهاجم با مشاهده خروجی z ، می‌تواند ورودی را به صورت $(z^{-1} = g^{-1}(z))$ بازیابی کند. حتی اگر g کاملاً وارونپذیر نباشد، مشاهده 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} تک نقطه‌ای (تابع دلتای دیراک) است. اگر دو پایگاهداده 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-لایلر^{۲۶} بین آن‌ها بی‌نهایت می‌شود:

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

این رابطه اثبات می‌کند که هیچ سطح محدودی از محرمانگی ($\infty < \epsilon$) با توابع قطعی 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 محاسبه 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 بیفتند.

- **انتشار هیستوگرام^{۲۹}**: بسیاری از تحلیل‌ها نیازمند دانستن توزیع داده‌ها هستند. DP اجازه می‌دهد

²³Injective

²⁴Probability Mass

²⁵Support

²⁶Kullback-Leibler Divergence

²⁷Randomization

²⁸Mean and Sum Estimation

²⁹Histogram 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 کم جمعیت می‌شود.

• **یادگیری ماشین خصوصی^{۳۰}**: در آموزش مدل‌های عمیق، خطر به خاطر سپاری^{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 بخشد (مانند مکانیزم RAPPOR در گوگل کروم).

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

۲-۱-۱ محرمانگی تفاضلی موضعی (Local Differential Privacy)

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 انبوی 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 عملأً 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 می‌زنند. از سوی دیگر، محدودیت‌های حقوقی و احصاریه‌های قضایی نیز متصدی را ملزم به افشاء اطلاعات می‌کند. در تمام این سناریوها، مدل متمرکز با یک نقطه شکست مرکزی^{۳۴} روبروست.

³⁰Private Machine Learning

³¹Memorization

³²Trusted Curator

³³Firewalls

³⁴Single 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 محترمانگی 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 کنجه‌کار و نه حتی دولت‌ها، هرگز داده‌ی واقعی کاربر را مشاهده نکند. سرور تنها نسخه‌هایی مخدوش و نویزدار از داده‌ها را دریافت می‌کند که به تنها‌ی بی‌معنی هستند، اما در تجمعی با تعداد زیادی داده‌ی دیگر، الگوهای آماری دقیق را آشکار می‌سازند. این رویکرد، خطر نقض حریم خصوصی را بسیار کنترل می‌کند.

۲-۱ کارهای پیشین و مرور ادبیات

۱-۲-۱ آغازگرهای آماری: از پیمایش‌های آماری تا تعریف مدرن محترمانگی

اگرچه نگرانی پیرامون محترمانگی داده‌ها قدمتی به اندازه خود آمار دارد، اما فرمول‌بندی ریاضی دقیق آن دستاورد قرن بیست و یکم است. ادبیات کلاسیک این حوزه با تلاش برای کنترل افشاری آماری^{۳۵} آغاز شد، اما ناکارآمدی روش‌های مبتنی بر گمنامسازی در برابر دانش پس‌زمینه مهاجم، نیاز به یک تعریف معنایی قوی‌تر را ایجاد کرد.

نقطه عطف این تحول، معرفی مفهوم محترمانگی تفاضلی^{۳۶} توسط دُورک 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 M دارای شرایط $\text{DP}-\epsilon$ 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 $\mathcal{S} \subseteq \text{Range}(M)$ رابطه زیر برقرار باشد:

$$\Pr[M(\mathcal{D}) \in \mathcal{S}] \leq e^{\epsilon} \cdot \Pr[M(\mathcal{D}') \in \mathcal{S}] + \delta \quad (1-1)$$

که در آن ϵ پارامتری کلیدی به نام بودجه محترمانگی^{۳۷} است و δ احتمال شکست ناچیز مکانیزم را نشان

³⁵Statistical Disclosure Control

³⁶Differential Privacy

³⁷Privacy Budget

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

- **مقادیر کوچک ϵ (محرمانگی قوی):** زمانی که $0 \rightarrow \epsilon$ ، توزیع های خروجی برای دو پایگاه داده همسایه تقریباً بر هم منطبق می شوند. در این حالت، مکانیزم مجبور است نویز بسیار زیادی به پاسخ اضافه کند تا تفاوت ها را بپوشاند. در نتیجه، مهاجم تقریباً هیچ توانی برای تشخیص حضور فرد ندارد، اما در مقابل، دقت آماری خروجی کاهش می یابد.

- **مقادیر بزرگ ϵ (محرمانگی ضعیف):** با افزایش ϵ ، مکانیزم آزادی عمل بیشتری دارد تا خروجی های متمایزتری تولید کند (نویز کمتر). این امر دقت تحلیل را افزایش می دهد، اما همزمان ریسک بازشناسایی فرد و نشت اطلاعات خصوصی نیز به صورت نمایی بالا می رود.

همان طور که در بخش قبل توضیح داده شد، پیاده سازی این تعریف نیازمند یک پیش فرض قوی است: وجود یک متصدی مورد اعتماد که تمام داده های خام را جمع آوری کرده و نویز را به صورت مرکزی اعمال کند. اما دیدیم که این مدل دارای نقطه ضعف هایی است. حذف این فرض و انتقال اعتماد از سرور به کاربر، منجر به شکل گیری مفهوم محرمانگی تفاضلی موضعی^{۳۸} (LDP) شد. اگرچه اصطلاح LDP و صورت بندی مدرن آن در سال های اخیر توسط پژوهشگرانی نظری کاسی یسوانتان و دیگران تدوین شد [۱۳]، اما ریشه های عملی آن به دهه ها قبل باز می گردد.

تعریف ۱-۱ (محرمانگی تفاضلی موضعی α -LDP) یک مکانیزم تصادفی $\mathcal{Z} \rightarrow \mathcal{X}$ (تصادفی ساز موضعی) شرط «محرمانگی تفاضلی موضعی تقریبی» یا (α, δ) -LDP را برآورده می کند، اگر برای تمام جفت ورودی های ممکن \mathcal{X} و هر زیر مجموعه خروجی $\mathcal{Z} \subseteq \mathcal{S}$ ، رابطه زیر برقرار باشد:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq e^\alpha \cdot \Pr[\mathcal{M}(x') \in \mathcal{S}] + \delta \quad (2-1)$$

در این تعریف:

- α ، بودجه محرمانگی است که میزان شباهت توزیع های خروجی را کنترل می کند.
 - 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 می گردد.

³⁸Local Differential Privacy

مثال‌هایی از مکانیزم و کاربرد LDP

در واقع، ساده‌ترین و نخستین نمونه از یک مکانیزم LDP، روش «پاسخ تصادفی»^{۳۹} 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^{۴۰} خواسته می‌شود که ویژگی حساس $\{X_i \in \{0, 1\}\}$ را گزارش کند. کاربر به جای پاسخ مستقیم، طبق دستورالعمل زیر عمل می‌کند:

۱. یک سکه را پرتاب می‌کند. (می‌تواند سکه غیرمنصفانه^{۴۰} 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 خطأ تخمين 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:

$$\frac{\Pr[\mathcal{M}(x) = z]}{\Pr[\mathcal{M}(x') = z]} \leq \frac{p}{1-p} \quad (3-1)$$

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

$$e^\alpha = \frac{p}{1-p} \quad \Rightarrow \quad \alpha = \ln\left(\frac{p}{1-p}\right) \quad (4-1)$$

این فرمول، تفسیر شهودی LDP را کامل می‌کند:

- اگر $0.5 \approx p$ (سکه کاملاً تصادفی)، آنگاه $0 \approx \alpha$ می‌شود. یعنی خروجی هیچ اطلاعاتی از ورودی ندارد (محروم‌نگی کامل، اما بدون فایده آماری).

³⁹Randomized Response (RR)

⁴⁰Unfair

- اگر $1 \rightarrow p$ (پاسخ تقریباً همیشه راست)، آنگاه $\infty \rightarrow \alpha$ می‌شود. یعنی داده‌ها دقیق هستند اما هیچ محروم‌انگی وجود ندارد.

بنابراین، کار وارنر را می‌توان سنگ بنای 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 محروم‌انگی در داده‌های دودویی است، و در کاربردهای مدرن با چالش دامنه‌ی بسیار بزرگ⁴¹ روبروست. شرکت‌های بزرگ فناوری نیاز دارند داده‌هایی نظری «آدرس‌های اینترنتی بازدید شده» یا «کلمات جدید تایپ شده» را جمع‌آوری کنند که دامنه‌ی آن‌ها^(۲) 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 RAPPOR⁴² را معرفی کرد [۲۰]. چالش اصلی گوگل، جمع‌آوری رشته‌های متنی⁴³ بود. راه حل آن‌ها ترکیب پاسخ تصادفی با فیلترهای بلوم⁴⁴ 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 پرطریفدار و داده‌های سلامت در سیستم‌عامل‌های iOS و macOS استفاده می‌کند. مسئله‌ی اپل، مخابره‌ی کارآمد داده‌ها با حفظ حریم خصوصی بود. راه حل اپل استفاده از تکنیک‌های مبتنی بر طرح‌ریزی⁴⁵ و تبدیل‌های ریاضی مانند تبدیل هادامارد⁴⁶ 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 تله‌متري:** مایکروسافت برای جمع‌آوری داده‌های تله‌متري ویندوز (مانند مدت زمان استفاده از برنامه‌ها) با چالش تخمین هیستوگرام‌های پیوسته روبرو بود. آن‌ها از مکانیزم‌هایی نظری نمونه‌برداری هیستوگرام و روش‌های تکرارکننده برای بازسازی توزیع داده‌ها

⁴¹High-Dimensional Domain

⁴²Randomized Aggregatable Privacy-Preserving Ordinal Response

⁴³String

⁴⁴Bloom Filters

⁴⁵Sketching

⁴⁶Hadamard 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 موضوعی (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.

۲-۲-۱ دسته‌بندی پروتکل‌های موضوعی: 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 کاربرⁿ مکانیزم M را تنها بر اساس داده‌ی خودشⁱ, X_i اجرا کرده و پیامⁱ, 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.

کاربرⁿ می‌تواند قبل از ارسال داده‌ی خود، خلاصه‌ای از داده‌های کاربران قبلی (Z_1, \dots, 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 [۱۱، ۶].

۳-۲-۱ چالش سودمندی و موازنۀ دقت-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 موازنۀ محرمانگی-دقت^{۵۰} شناخته می‌شود.

⁴⁷Non-interactive / Simultaneous

⁴⁸Interactive / Sequential

⁴⁹Utility

⁵⁰Privacy-Accuracy Trade-off

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

- در مدل متمرکز (CDP): نویز تنها «یکبار» و پس از تجمعی 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^{۵۱} 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 محروم‌گی را فراهم کنند؟» و «حد نهایی این دقت کجاست؟» این پرسش‌ها زمینه را برای ورود تئوری‌های پیشرفته‌تر نظری «تخمین مینی‌مکس» فراهم کرد.

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

پاسخ به پرسش بالا، مسیر پژوهش‌های این حوزه را به سمت نظریه مینی‌مکس آماری^{۵۲} تغییر داد. نقطه عطف این تحول، سلسله مقالات جریان‌ساز دوچی، جردن و وین‌رایت^{۵۳} بود [۶، ۷]. آن‌ها با صورت‌بندی مسئله در قالب نظریه اطلاعات، نشان دادند که هزینه محروم‌گی در مدل موضعی بسیار سنگین و غیرقابل اجتناب است.

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

⁵¹Signal-to-Noise Ratio (SNR)

⁵²Statistical Minimax Theory

⁵³Duchi, Jordan, and Wainwright

از ابزارهایی نظری 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 بینایین با مدل متمرکز دارد.

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

$$\mathfrak{M}_{\text{LDP}} \asymp \frac{1}{n\alpha^2} \quad , \quad \mathfrak{M}_{\text{CDP}} \asymp \frac{1}{n^2\epsilon^2} \quad (5-1)$$

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 $1/n$ کاهش می‌یابد، خطای مدل موضعی با سرعت بسیار کندتر $1/\sqrt{n}$ کم می‌شود.

۲. اندازه نمونه مؤثر: ضریب α^2 نشان می‌دهد که هر نمونه داده‌ی خصوصی‌سازی شده، عملأً حاوی اطلاعاتی معادل با α^2 نمونه داده‌ی خام است (برای $1 < \alpha$). این یعنی برای جبران نویز α -LDP، حجم داده‌ها باید با ضریب $1/\alpha^2$ افزایش یابد.

پس از استقرار این چارچوب نظری، مرکز 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 نظری 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.

۳-۱ بیان مسئله و اهداف پژوهش

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

⁵⁴Fano's Inequality

⁵⁵Assouad's Lemma

⁵⁶Optimal Mechanisms

⁵⁷Staircase Mechanisms

⁵⁸Unary Encoding (UE)

⁵⁹Optimal Local Hashing (OLH)

ماهیت دوگانه، اگرچه باعث غنای ادبیات موضوع شده، اما منجر به پراکندگی قابل توجهی 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 محروم‌نگی 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 کولبک-لایبلر (KL) و نامساوی فانو استفاده کردند.
- در مسائل آزمون فرض ساده، اغلب از فاصله‌ی تغییرات کل (TV) و لم لوکام بهره گرفته شده است.
- در تحلیل‌های اخیرتر، فاصله‌ی کای-دو (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: f -واگرایی‌ها به عنوان زبان مشترک

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

$$D_f(P\|Q) = \mathbb{E}_Q \left[f \left(\frac{dP}{dQ} \right) \right] \quad (6-1)$$

که در آن f یک تابع محدب با ویژگی $f(1) = 0$ است.

انتخاب این رویکرد به ما اجازه می‌دهد تا «محروم‌نگی» را نه صرفاً به عنوان یک ویژگی الگوریتمی، بلکه به عنوان یک محدودیت هندسی تفسیر کنیم. در این دیدگاه، هر مکانیزم LDP مانند یک «کanal 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»⁶³ را برای f -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.

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

برای غلبه بر چالش پراکنده‌ی مفاهیم و ارائه یک تحلیل منسجم، این پایان‌نامه بر مرور و تشریح رویکرد نوینی تمرکز دارد که در سال‌های اخیر توسط γ -آسوده، علی‌اکبری و کالمون⁶⁴ [۲] و γ -آسوده و ژانگ⁶⁵ [۳] توسعه یافته است. این پژوهشگران نشان داده‌اند که قیود محرمانگی تفضیلی موضعی (LDP) را می‌توان به صورت دقیق و ریاضی معادل با «خواص انقباضی»⁶⁶ خانواده f -واگرایی‌ها فرمول‌بندی کرد.

به‌طور مشخص، در این رویکرد اثبات می‌شود که شرط LDP معادل است با کران‌دار بودن ضریب انقباض برای «واگرایی‌هاکی-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 کanal 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.

۳-۳-۱ 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 تفضیلی موضعی دقیقاً معادل با انقباض در واگرایی‌هاکی-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 دقیق⁶⁵ برای ضریب

⁶³Contraction Coefficient

⁶⁴Contraction Properties

⁶⁵Sharp Bounds

انقباض مکانیزم‌های LDP تحت واگرایی‌های کای-دو و کولبک-لایبلر. این بخش با تمرکز بر نتایج [۲]، نشان می‌دهد که چگونه می‌توان «نامساوی‌های پردازش داده قوی»^{۶۶} را برای محیط‌های خصوصی استخراج کرد.

۳. یک پارچه‌سازی 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.

۴-۱ 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 کولبک-لایبلر 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 لی کم^{۶۷} و لم فانو^{۶۸} بررسی می‌شوند. بخش پایانی این فصل به معرفی مدل کلاسیک محترمانگی تفاضلی متمرکز (CDP) و ویژگی‌های آن اختصاص دارد.

فصل سوم به معرفی دقیق مدل محترمانگی 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 محترمانگی

⁶⁶Strong Data Processing Inequalities (SDPI)

⁶⁷Le Cam's Lemma

⁶⁸Fano's Lemma

⁶⁹Randomized Response

⁷⁰Minimax Risk

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

در فصل پنجم، چارچوب تحلیل خود را با استفاده از واگرایی‌های \mathcal{F} تعمیم می‌دهیم. در این فصل ارتباط میان محروم‌نگی تفاضلی موضعی و مفهوم ضریب انقباض⁷¹ 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.

⁷¹Contraction Coefficient

فصل ۲

پیش‌نیازها

۱-۲ محرمانگی تفاضلی مت مرکز (CDP)

مفهوم محرمانگی تفاضلی^۱ یا به اختصار DP، اولین بار توسط دُرک و همکاران [A] 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.

۱-۱-۲ مدل اعتماد و تعریف رسمی

در مدل مت مرکز^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 ۱-۲ را ببینید). متصدی، مجموعه‌داده‌ی کامل D را در اختیار دارد. وظیفه‌ی متصدی این است که با اجرای یک مکانیزم تصادفی^۴ M بر روی مجموعه‌داده‌ی D ، نتایجی (مثالاً پاسخ به یک پرس‌وجو^۵) را به صورت عمومی منتشر کند، به طوری که اطلاعات حساس افراد فاش نشود.

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

¹Differential Privacy

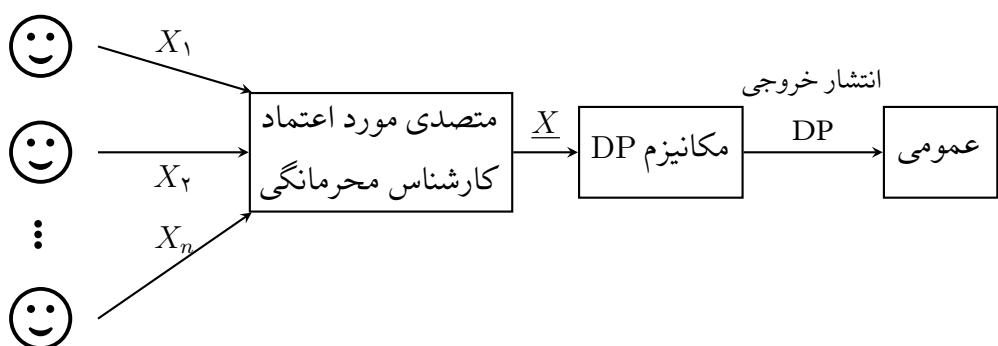
²Centralized

³Trusted Curator

⁴Randomized Mechanism

⁵Query

افراد (داده‌ها)



شکل ۱-۲: مدل محیانگی، تفاضلی، متمرکز (CDP) یا یک متصلی مورد اعتماد.

داده‌ها^۹ می‌نامیم و آن را با \mathcal{X} نمایش می‌دهیم. یک پایگاهداده‌ی \mathcal{D} مجموعه‌ای از رکوردهاست که اعضای آن از \mathcal{X} انتخاب شده‌اند. در ادبیات محترمانگی 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 i است.

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

$$\mathcal{D} = (\mathfrak{r}, \mathfrak{l})$$

تعريف ۲-۲ (الگوریتم تصادفی) یک الگوریتم (mekanizm) تصادفی M تابعی است که دامنه‌ی آن مجموعه‌ی تمام پایگاهداده‌های ممکن و برد آن مجموعه‌ی خروجی‌های ممکن \mathcal{R} است. برخلاف الگوریتم‌های قطعی، خروجی M برای یک ورودی ثابت D ، یک متغیر تصادفی است. به عبارت دیگر، 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.

⁶Data Universe

تعريف ۳-۲ (فاصله و نرم‌های ℓ_p) برای سنجش میزان تفاوت دو پایگاهداده، از مفهوم نرم ℓ_p استفاده می‌شود. در حالت کلی برای $1 \geq p$ ، فاصله‌ی ℓ_p بین دو پایگاهداده‌ی \mathcal{D}_1 و \mathcal{D}_2 (با بردارهای تکرار متناظر) به صورت زیر تعریف می‌گردد:

$$\|\mathcal{D}_1 - \mathcal{D}_2\|_p = \left(\sum_{i=1}^{|\mathcal{X}|} |x_{1,i} - x_{2,i}|^p \right)^{1/p} \quad (1-2)$$

در ادبیات محترمانگی تفاضلی، نرم ℓ_1 (یا فاصله‌ی منتهن) 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}_1 به \mathcal{D}_2 تبدیل شود:

$$\|\mathcal{D}_1 - \mathcal{D}_2\|_1 = \sum_{i=1}^{|\mathcal{X}|} |x_{1,i} - x_{2,i}|$$

مثال ۳-۲ فرض کنید $(1, 3, 0) = \mathcal{D}_2$ باشند (یعنی در پایگاهداده دوم، یک نفر بیکار حذف شده است). فاصله‌ی ℓ_1 آن‌ها برابر است با:

$$\|\mathcal{D}_1 - \mathcal{D}_2\|_1 = |3 - 3| + |1 - 0| = 1$$

تعريف ۴-۲ (پایگاهداده‌های همسایه) دو پایگاهداده‌ی \mathcal{D}_1 و \mathcal{D}_2 را همسایه^۵ می‌گوییم (و با $\mathcal{D}_1 \sim \mathcal{D}_2$ نشان می‌دهیم) اگر فاصله‌ی ℓ_1 آن‌ها حداقل ۱ باشد:

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

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

ایده‌ی اصلی محترمانگی 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.

تعريف ۵-۲ (ε -محترمانگی تفاضلی (ε -DP)) یک مکانیزم تصادفی M با دامنه \mathcal{X}^n و برد \mathcal{R} ، ویژگی ε -محترمانگی 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 σ -جبر برد تعریف شده باشد)، داشته باشیم:

$$\mathbb{P}[M(\mathcal{D}_1) \in \mathcal{S}] \leq \exp(\varepsilon) \cdot \mathbb{P}[M(\mathcal{D}_2) \in \mathcal{S}] \quad (2-2)$$

در این نامساوی، پارامتر $\varepsilon \geq 0$ را «بودجه‌ی محترمانگی»^۶ می‌نامیم.

⁷Neighboring / Adjacent

⁸Privacy Budget

۲-۱-۲ تفسیر پارامترهای محربانگی

پارامتر ϵ نقش کنترل کننده‌ی توازن میان «محربانگی» و «سودمندی» را ایفا می‌کند:

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

تعريف ۶-۲ DP -تقریبی یا (δ, ϵ) در بسیاری از موارد (مانند مکانیزم گوسی)، اوضاعی شرط ϵ خالص ممکن نیست. در این شرایط، از تعریف انعطاف‌پذیرتری به نام DP - ϵ استفاده می‌شود که اجازه‌ی یک احتمال شکست کوچک δ را می‌دهد:

$$\mathbb{P}[\mathcal{M}(\mathcal{D}_1) \in \mathcal{S}] \leq \exp(\epsilon) \cdot \mathbb{P}[\mathcal{M}(\mathcal{D}_2) \in \mathcal{S}] + \delta \quad (3-2)$$

پارامتر $[0, 1] \in \delta$ را معمولاً «احتمال شکست»^۹ یا احتمال نشت اطلاعات می‌نامند. تفسیر شهودی این است که مکانیزم با احتمال حداقل $\delta - 1$ ، تضمین DP - ϵ را رعایت می‌کند. در کاربردهای عملی، مقدار δ باید بسیار ناچیز (کمتر از معکوس چند جمله‌ای اندازه‌ی داده‌ها، مثلاً $1/n < \delta$) در نظر گرفته شود.

۳-۱-۲ تعاریف معادل و صورت‌بندی‌های جایگزین

برای تسهیل تحلیل‌های ریاضی و درک عمیق‌تر، می‌توان تعریف اصلی DP - ϵ 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 ۱-۲ (همارزی نقطه‌ای) یک مکانیزم M شرط DP - ϵ را برآورده می‌کند اگر و تنها اگر برای تمام همسایه‌های $D_2 \sim D_1$ شرایط زیر برقرار باشد:

⁹Failure Probability

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

$$\frac{\mathbb{P}[\mathcal{M}(\mathcal{D}_1) = z]}{\mathbb{P}[\mathcal{M}(\mathcal{D}_2) = z]} \leq e^\varepsilon \quad (4-2)$$

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

$$p(z) \leq e^\varepsilon \cdot q(z) \quad (5-2)$$

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

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

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

نکته: این همارزی تنها برای $\delta = \varepsilon, \delta$ -DP صادق است. برای بررسی نقطه‌به‌ نقطه کافی نیست و شرط باید روی زیرمجموعه‌ها چک شود.

۲. تعریف مبتنی بر واگرایی ماکزیمم:

از دیدگاه نظریه اطلاعات، محترمانگی تفاضلی خالص محدودیتی بر روی «واگرایی ماکزیمم»^{۱۰} بین توزیع‌های خروجی است. واگرایی ماکزیمم به صورت $D_\infty(P||Q) = \sup_S \ln \frac{P(S)}{Q(S)}$ تعریف می‌شود. بنابراین تعریف ۵-۲ معادل است با:

$$\sup_{\mathcal{D}_1 \sim \mathcal{D}_2} D_\infty(\mathcal{M}(\mathcal{D}_1) \parallel \mathcal{M}(\mathcal{D}_2)) \leq \varepsilon \quad (6-2)$$

در بخش‌های بعدی با تعریف دقیق واگرایی آشنا خواهیم شد.

۴-۱-۲ مکانیزم‌های پایه

برای دست‌یابی به محترمانگی تفاضلی، باید به پاسخ دقیق پرس‌وجو «نویز»^{۱۱} اضافه کنیم. میزان نویز به حساسیت^{۱۲} پرس‌وجو بستگی دارد.

¹⁰Max Divergence

¹¹Noise

¹²Sensitivity

تعريف ۷-۲ (حساسیت سراسری ℓ_p) برای هر تابع پرس و جوی $f : \mathcal{X}^n \rightarrow \mathbb{R}^k$ که خروجی برداری دارد، «حساسیت سراسری ℓ_p »^{۱۳} 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 (تعريف ۲-۱) داریم:

$$\Delta_p f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_p \quad (7-2)$$

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:

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

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

$$\Delta_1 f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1 \quad (8-2)$$

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:

$$\Delta_2 f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_2 \quad (9-2)$$

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:

$$\Delta_\infty f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_\infty \quad (10-2)$$

مثال ۴-۲ 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 $(0, \dots, +1, \dots, -1, \dots, 0)$. حال حساسیت‌ها را محاسبه می‌کنیم:

¹³ ℓ_p -Global Sensitivity

• حساسیت ℓ_1 : مجموع قدر مطلق تغییرات: $2 = |1| + |-1|$.

• حساسیت ℓ_2 : جذر مجموع مربعات: $\sqrt{2} = \sqrt{1^2 + (-1)^2}$.

• حساسیت ℓ_∞ : ماکزیمم قدر مطلق تغییرات: $1 = \max(|1|, |-1|, 0)$.

این مثال به وضوح رابطه $\Delta_\infty f \leq \Delta_2 f \leq \Delta_1 f$ 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.

۵-۱-۲ مکانیزم‌های بنیادی محترمانگی تفاضلی

در این بخش، سه مکانیزم اصلی را که بلوک‌های سازنده 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 $h(z) = \frac{1}{2b} \exp\left(-\frac{|z-\mu|}{b}\right)$ است.

قضیه‌ی ۲-۲ (محترمانگی مکانیزم لاپلاس) فرض کنید $f : \mathcal{X}^n \rightarrow \mathbb{R}^k$ یک تابع پرس‌وجو با حساسیت سراسری $\Delta_1 f$ باشد. مکانیزم لاپلاس که خروجی آن به صورت زیر تعریف می‌شود:

$$\mathcal{M}_{\text{Lap}}(\mathcal{D}) = f(\mathcal{D}) + (Y_1, \dots, Y_k) \quad (11-2)$$

که در آن $(\frac{\Delta_1 f}{\varepsilon}, Y_i)_{i=1}^{i.i.d.} \sim \text{Lap}(\frac{\Delta_1 f}{\varepsilon})$ را برآورده می‌کند.

اثبات. فرض کنید $\mathcal{D}_1 \sim \mathcal{D}_2$ دو پایگاه داده‌ی همسایه باشند و خروجی تابع f یک بردار k -بعدی باشد. نویز لاپلاس به هر مؤلفه به صورت مستقل اضافه می‌شود، بنابراین تابع چگالی احتمال توأم برابر با حاصل ضرب چگالی‌های مؤلفه‌هاست. با فرض $\frac{\Delta_1 f}{\varepsilon} = b$ ، نسبت چگالی احتمال را برای یک بردار خروجی دلخواه

z بررسی می‌کنیم:

$$\begin{aligned} \frac{p(z|\mathcal{D}_1)}{p(z|\mathcal{D}_2)} &= \frac{\prod_{i=1}^k \frac{1}{\sqrt{b}} \exp\left(-\frac{|z_i - f(\mathcal{D}_1)_i|}{b}\right)}{\prod_{i=1}^k \frac{1}{\sqrt{b}} \exp\left(-\frac{|z_i - f(\mathcal{D}_2)_i|}{b}\right)} \\ &= \prod_{i=1}^k \exp\left(\frac{|z_i - f(\mathcal{D}_2)_i| - |z_i - f(\mathcal{D}_1)_i|}{b}\right) \\ &= \exp\left(\frac{1}{b} \sum_{i=1}^k (|z_i - f(\mathcal{D}_2)_i| - |z_i - f(\mathcal{D}_1)_i|)\right) \end{aligned}$$

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

$$|z_i - f(\mathcal{D}_2)_i| - |z_i - f(\mathcal{D}_1)_i| \leq |f(\mathcal{D}_1)_i - f(\mathcal{D}_2)_i|$$

با اعمال این نامساوی در مجموع توان نمایی:

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

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 $\|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1 \leq \Delta_1 f$. بنابراین:

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

و حکم ثابت می‌شود. \square

مثال ۲-۵ (پرس‌وجوهای شمارشی)^{۱۴}، پرس‌وجوهایی شمارشی 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 حداقل ۱ واحد تغییر می‌دهد. بنابراین طبق قضیه ۲-۲، با افزودن نویز با مقیاس $\varepsilon/1$ (یعنی $(1/\varepsilon)$ Lap($1/\varepsilon$)) به پاسخ واقعی، محترمانگی $-DP(\varepsilon, 0)$ تضمین می‌شود.

حالت برداری (چند پرس‌وجو): فرض کنید لیستی از k پرس‌وجوی شمارشی $(f_1, \dots, f_k) = f$ داریم (یک پرس‌وجوی برداری). بدون داشتن اطلاعات اضافی درباره‌ی ارتباط پرس‌وجوها، در بدترین حالت

¹⁴Counting Queries

یک فرد مشخص می‌تواند در تمام k شمارش تأثیر بگذارد. بنابراین حساسیت ℓ_1 کل بردار برابر با مجموع تغییرات، یعنی k خواهد بود ($f = \Delta_1 f = k$). در این حالت برای دست‌یابی به ϵ -DP، باید به هر پاسخ نویزی با مقیاس ϵ/k اضافه کنیم.

مکانیزم گوسی

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

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

$$\mathcal{M}_{\text{Gauss}}(\mathcal{D}) = f(\mathcal{D}) + \mathcal{N}(\cdot, \sigma^2 \mathbf{I}_k) \quad (12-2)$$

اگر $\epsilon \in (0, 1)$ باشد، با انتخاب انحراف معیار σ به صورت زیر، این مکانیزم شرط (ϵ, δ) -DP را برآورده می‌کند:

$$\sigma \geq \sqrt{2 \ln(1/25/\delta)} \cdot \frac{\Delta_2 f}{\epsilon} \quad (13-2)$$

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 لایپلاس، در توزیع گوسی نسبت $\frac{p(z)}{q(z)}$ کران دار 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 $v \in \mathbb{R}^d$ در نظر گرفت. اگر یک بیمار پرونده‌اش را تغییر دهد، در بدترین حالت تمام d ویژگی او می‌توانند از 0 به 1 تغییر کنند.

تحلیل حساسیت:

• حساسیت ℓ_1 : مجموع قدر مطلق تغییرات برابر است با $d = \sum_{i=1}^d |v_i - 0|$

• حساسیت ℓ_2 : جذر مجموع مربعات تغییرات برابر است با $\sqrt{d} = \sqrt{\sum_{i=1}^d (v_i - 0)^2}$

¹⁵Privacy Loss Random Variable

مقایسه نویز: اگر تعداد ویژگی‌ها زیاد باشد (مثلًا $d = 100$):

- مکانیزم لاپلاس باید نویزی متناسب با $d = 100$ اضافه کند.

- مکانیزم گوسی نویزی متناسب با $\sqrt{100} = 10$ اضافه می‌کند.

این کاهش چشمگیر نویز (با ضریب \sqrt{d}) دلیل اصلی استفاده از مکانیزم گوسی در الگوریتم‌هایی نظریه یادگیری عمیق با محترمانگی 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^{۱۶} $q(\mathcal{D}, r)$ 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.

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

$$\mathbb{P}[\mathcal{M}_{\text{Exp}}(\mathcal{D}) = r] = \frac{\exp\left(\frac{\varepsilon \cdot q(\mathcal{D}, r)}{\gamma \Delta q}\right)}{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon \cdot q(\mathcal{D}, r')}{\gamma \Delta q}\right)} \quad (4-2)$$

این مکانیزم شرط $DP-\varepsilon$ -را برآورده می‌کند.

اثبات. فرض کنید $\mathcal{D}_1 \sim \mathcal{D}_2$. نسبت احتمالات برای یک خروجی ثابت r عبارت است از:

$$\frac{\mathbb{P}[\mathcal{M}(\mathcal{D}_1) = r]}{\mathbb{P}[\mathcal{M}(\mathcal{D}_2) = r]} = \frac{\exp\left(\frac{\varepsilon q(\mathcal{D}_1, r)}{\gamma \Delta q}\right)}{\exp\left(\frac{\varepsilon q(\mathcal{D}_2, r)}{\gamma \Delta q}\right)} \cdot \frac{\sum_{r'} \exp\left(\frac{\varepsilon q(\mathcal{D}_2, r')}{\gamma \Delta q}\right)}{\sum_{r'} \exp\left(\frac{\varepsilon q(\mathcal{D}_1, r')}{\gamma \Delta q}\right)}$$

ترم اول (صورت کسرها) با استفاده از خاصیت حساسیت q حداقل $e^{\varepsilon/2}$ است. ترم دوم (نسبت مخرج‌ها) نیز با استدلالی مشابه حداقل $e^{\varepsilon/2}$ خواهد بود. حاصل ضرب این دو مقدار حداقل e^ε می‌شود. \square

¹⁶Score Function

۶-۱-۲ ترکیب‌پذیری

در کاربردهای واقعی، معمولاً چندین پرس‌وجو روی یک پایگاه‌داده اجرا می‌شود. قضایای ترکیب‌پذیری^{۱۷} نشان می‌دهند که چگونه بودجه‌ی محرمانگی انباسته می‌شود.

قضیه‌ی ۵-۲ (ترکیب‌پذیری ساده) اگر k مکانیزم M_1, \dots, M_k به ترتیب دارای بودجه‌های $\varepsilon_1, \dots, \varepsilon_k$ باشند، اجرای متوالی آن‌ها روی یک پایگاه‌داده‌ی واحد، تضمین $-DP(\sum \varepsilon_i)$ را فراهم می‌کند [۹].

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

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

$$\varepsilon' \approx \varepsilon \sqrt{2k \ln(1/\delta')} + k\varepsilon(e^\varepsilon - 1) \quad (15-2)$$

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

۷-۱-۲ محرمانگی گروهی

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

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

$$\mathbb{P}[\mathcal{M}(\mathcal{D}_1) \in S] \leq e^{k\varepsilon} \mathbb{P}[\mathcal{M}(\mathcal{D}_2) \in S] \quad (16-2)$$

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

¹⁷Composition

۸-۱-۲ محدودیت مدل متتمرکز

با وجود تمام مزايا، مدل CDP يك نقطه‌ي ضعف اساسی دارد: نياز به يك متصدی کاملاً مورد اعتماد. در بسياری از سناريوهای دنيای واقعی (مانند جمع‌آوري داده از گوشی‌های هوشمند)، کاربران به سرور مرکزی اعتماد ندارند. اين عدم اعتماد، ما را به سمت مدل جايگزین، یعنی «محرمانگی تفاضلی موضوعی» سوق می‌دهد. در فصل بعد با محرمانگی تفاضلی موضوعی و تعريف و قضایای اساسی آن آشنا خواهيم شد.

۲-۲ f-واگرایی‌ها

در بخش‌های پيشين، مکانيزم‌های محرمانگی تفاضلی را ابزاری برای ايجاد «شباخت آماری» بين خروجی‌های دو پایگاهداده همسایه معرفی کردیم. برای کمی‌سازی دقیق اين شباخت و اثبات کران‌های پايان در فصل‌های آينده، نيازمند معیاري هستيم که فاصله ميان توزيع‌های احتمالي را در يك چارچوب عمومی بسنجد. اين ابزار، خانواده‌ی f-واگرایی‌ها^{۱۹} است که نخستین بار توسط سizar [۵] و على و سيلوي [۱] معرفی شد.

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. فرض کنید (Ω, \mathcal{F}) يك فضای اندازه‌پذیر^{۲۱} باشد و P و Q دو «اندازه احتمال»^{۲۲} تعریف شده روی اين فضا باشنند.

پيش از تعریف واگرایی، مفهوم «پيوستگی مطلق» که شرط وجود چگالی نسبت است را يادآوري می‌کنیم.

تعريف ۲-۹ (پيوستگی مطلق^{۲۳}) می‌گوییم اندازه P نسبت به Q مطلقاً پيوسته است و می‌نویسیم $Q \ll P$ ، اگر برای هر مجموعه اندازه‌پذیر $\mathcal{F} \in A \in \mathcal{F}$:

$$Q(A) = 0 \implies P(A) = 0 \quad (17-2)$$

اين شرط تضمین می‌کند که هر رویدادی که تحت توزيع Q غیرممکن باشد، تحت P نیز غیرممکن است.

¹⁹ f-divergences

²⁰ Measure Theory

²¹ Measurable Space

²² Probability Measures

اگر $Q \ll P$ باشد، بنابر قضیه رادون-نیکودیم^{۲۴}، تابعی اندازه‌پذیر و غیرمنفی روی Ω وجود دارد که «مشتق رادون-نیکودیم» P نسبت به Q نامیده می‌شود و با $\frac{dP}{dQ}$ نمایش داده می‌شود. این مشتق نقش همان نسبت درست‌نمایی^{۲۵} را در حالت کلی ایفا می‌کند.

تعريف ۱۰-۲ (f -واگرایی) فرض کنید P و Q دو اندازه احتمال روی (Ω, \mathcal{F}) باشند به طوری که $\ll P$ Q . برای هر تابع محدب^{۲۶} $\mathbb{R} \rightarrow (0, \infty)$: f با این شرط که $f(1) = 0$ ، f -واگرایی^{۲۷} P از Q به صورت امید ریاضی تابع f بر روی مشتق رادون-نیکودیم (تحت اندازه Q) تعریف می‌شود:

$$D_f(P\|Q) \triangleq \int_{\Omega} f\left(\frac{dP}{dQ}(\omega)\right) dQ(\omega) \quad (18-2)$$

یا به بیانی دیگر با استفاده از نماد امید ریاضی:

$$D_f(P\|Q) = \mathbb{E}_Q \left[f\left(\frac{dP}{dQ}\right) \right] \quad (19-2)$$

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^{۲۸} و شرط محدب بودن f ، می‌توان نشان داد که واگرایی همواره نامنفی است:

$$D_f(P\|Q) = \mathbb{E}_Q \left[f\left(\frac{dP}{dQ}\right) \right] \geq f\left(\mathbb{E}_Q \left[\frac{dP}{dQ} \right]\right) = f(1) = 0 \quad (20-2)$$

تساوی $0 = D_f(P\|Q)$ برقرار است اگر و تنها اگر^{۲۹} $P = Q$ (در صورت اکیداً محدب بودن f).

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 رادون-نیکودیم به نسبت معمولی چگالی‌ها $\frac{p(x)}{q(x)}$ تبدیل می‌شود و تعریف انتگرالی بالا به فرم‌های آشنای زیر تقلیل می‌یابد:

$$D_f(P\|Q) = \int_{\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) dx \quad \text{یا} \quad \sum_{x \in \mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) \quad (21-2)$$

²⁴Radon-Nikodym Theorem

²⁵Likelihood Ratio

²⁶Convex

²⁷Jensen's Inequality

تفاوت واگرایی با متر ریاضی: اگرچه در ادبیات موضوع، گاهی با اغماس از واژه‌ی «فاصله»^{۲۸} 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^{۲۹} ($d(P, Q) = d(Q, P)$) و نامساوی مثلث^{۳۰} ($d(P, R) \leq d(P, Q) + d(Q, R)$) را ارضاء کند.

بسیاری از واگرایی‌های معرفی شده در این بخش، به ویژه واگرایی کولبک-لایبلر (بخش ۱۲-۲)، این دو شرط را نقض می‌کنند. عدم تقارن در واگرایی‌ها دارای تفسیر آماری مهمی است؛ به عنوان مثال $KL(P\|Q)$ 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، f -واگرایی‌ها همچنان خواص توپولوژیکی و هندسی قدرتمندی دارند که در فصل‌های آتی برای استخراج کران‌های پایین مینی‌مکس از آن‌ها بهره خواهیم برد.

۲-۲-۲ 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.

تعريف ۱۱-۲ (تغییرات کل) فاصله تغییرات کل^{۳۱}، 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 $|f(t) - 1|$ است.

$$TV(P, Q) = \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)| dx \quad (22-2)$$

$$= \sup_{A \subseteq \mathcal{X}} |P(A) - Q(A)| \quad (23-2)$$

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

تعريف ۱۲-۲ (کولبک-لایبلر) واگرایی کولبک-لایبلر^{۳۲}، معروف‌ترین واگرایی در نظریه اطلاعات که آنتروپی نسبی^{۳۳} نیز نامیده می‌شود. این معیار نامتقارن است و نقش اساسی در فشرده‌سازی داده‌ها و استنتاج

²⁸Distance

²⁹Symmetry

³⁰Triangle Inequality

³¹Total Variation (TV)

³²Kullback-Leibler (KL)

³³Relative Entropy

بیزی دارد. تابع مولد آن $f(t) = t \ln t$ است.

$$\text{KL}(P\|Q) = \int_{\mathcal{X}} p(x) \ln \left(\frac{p(x)}{q(x)} \right) dx \quad (24-2)$$

تعریف ۱۳-۲ (اطلاعات متقابل) اگر X و V دو متغیر تصادفی باشند، اطلاعات متقابل^{۳۴} بین آن‌ها به صورت امید ریاضی واگرایی KL بین توزیع شرطی و توزیع حاشیه‌ای تعریف می‌شود:

$$I(X; V) = \text{KL}(P_{X,V}\|P_X \otimes P_V) = \mathbb{E}_V [\text{KL}(P_{X|V}\|P_X)] \quad (25-2)$$

این معیار نقش کلیدی در نامساوی فانو و تحلیل کانال‌های اطلاعاتی ایفا می‌کند [۱۵].

تعریف ۱۴-۲ (کای-دو) واگرایی کای-دو^{۳۵}، 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^۲ $f(t) = (t - 1)^2$ است.

$$\chi^2(P\|Q) = \int_{\mathcal{X}} \frac{(p(x) - q(x))^2}{q(x)} dx \quad (26-2)$$

تعریف ۱۵-۲ (هلینگر-دو) فاصله هلینگر-دو^{۳۶}، تابع مولد آن^۲ $f(t) = (\sqrt{t} - 1)^2$ است. این فاصله به دلیل خواص ریاضی خوش‌رفتار (مانند متریک و کران‌دار بودن بین ۰ و ۲)، در نظریه برآورد^{۳۷} و استخراج کران‌های مینی‌مکس (مانند روش *Le Cam*) کاربرد فراوان دارد.

$$H^2(P, Q) = \int_{\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx \quad (27-2)$$

تعریف ۱۶-۲ (E_γ یا هاکی-استیک) واگرایی E_γ یا واگرایی هاکی-استیک^{۳۸}، ابزاری کلیدی در تحلیل‌های مدرن حریم خصوصی و آزمون‌های فرضیه است. تابع مولد آن برای پارامتر $\gamma \geq 0$ به صورت $f_\gamma(t) = [t - \gamma]_+ = \max \{0, t - \gamma\}$ است.

تعریف و صورت‌های معادل: تعریف اصلی بر اساس انگرال جرم اضافی نسبت درست‌نمایی است، اما صورت‌های معادل زیر بیان عملیاتی تری ارائه می‌دهند:

$$E_\gamma(P\|Q) = \int_{\mathcal{X}} \max \{0, p(x) - \gamma q(x)\} dx \quad (\tilde{28}-2)$$

$$= \sup_{A \subseteq \mathcal{X}} (P(A) - \gamma Q(A)) \quad (28-2\text{ب})$$

$$= P(Z > \gamma) - \gamma Q(Z > \gamma) \quad \left(\text{که } Z = \frac{p(x)}{q(x)} \right) \quad (28-2\text{ج})$$

³⁴Mutual Information

³⁵Chi-Squared (χ^2)

³⁶Squared Hellinger

³⁷Estimation Theory

³⁸Hockey-Stick Divergence

رابطه (۲-۲۸ ب) 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 وردو^[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 $E_\gamma(P\|Q)$ 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 محترمانگی تفاضلی تقریبی ($DP(\delta, \varepsilon)$) دقیقاً معادل است با اینکه برای هر دو دیتابیس همسایه، واگرایی هاکی-استیک خروجی‌ها از مقدار δ تجاوز نکند:

$$E_{e^\varepsilon}(P\|Q) \leq \delta \quad (29-2)$$

تعريف ۱۷-۲ (رنی) واگرایی رنی^{۳۹}، با پارامتر $(1, \infty)$ ، به صورت زیر تعریف می‌شود:

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \ln \left(\int_{\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} dx \right) \quad (30-2)$$

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

تعريف ۱۸-۲ (ماکزیمم (D_∞)) این واگرایی متناظر با بدترین نسبت درست‌نمایی نقطه‌ای است و به عنوان حد واگرایی رنی به دست می‌آید:

$$D_\infty(P\|Q) = \lim_{\alpha \rightarrow \infty} D_\alpha(P\|Q) = \sup_{x \in \mathcal{X}} \ln \left(\frac{p(x)}{q(x)} \right) \quad (31-2)$$

کاربرد در حریم خصوصی: شرط $DP(\varepsilon)$ (خالص) دقیقاً معادل است با کران دار بودن این واگرایی توسط بودجه حریم خصوصی: $D_\infty(P\|Q) \leq \varepsilon$.

۳-۲-۲ خواص بنیادین و روابط بین f -واگرایی‌ها

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

³⁹Rényi Divergence

نامساوی پردازش داده (DPI)

مهم‌ترین ویژگی f -واگرایی‌ها در نظریه اطلاعات، خاصیت یکنواختی^{۴۰} 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 ۸-۲ (نامساوی پردازش داده^{۴۱}) فرض کنید P و Q دو توزیع احتمال روی فضای \mathcal{X} باشند و $\mathcal{X} \rightarrow \kappa$ یک هسته‌ی احتمالاتی (کانال مارکوف)^{۴۲} باشد که داده‌ها را از فضای \mathcal{X} به κ نگاشت می‌کند. اگر P_κ و Q_κ توزیع‌های خروجی پس از اعمال کرنل باشند، آنگاه برای هر f -واگرایی داریم:

$$D_f(P_\kappa \| Q_\kappa) \leq D_f(P \| Q) \quad (۳۲-۲)$$

تفسیر در حريم 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^{۴۳} روی آن خروجی نیز قادر به بهبود توان تشخیص خود نخواهد بود. به عبارت دیگر، اطلاعات (و حريم خصوصی) با پردازش بیشتر، «خلق» یا «تخرب» نمی‌شود.

تحدب مشترک

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

قضیه‌ی ۹-۲ (تحدب مشترک^{۴۴}) نگاشت $(P, Q) \mapsto D_f(P \| Q)$ یک تابع محدب مشترک است. یعنی برای هر $\lambda \in [0, 1]$ و توزیع‌های P_1, P_2, Q_1, Q_2

$$D_f(\lambda P_1 + (1 - \lambda)P_2 \| \lambda Q_1 + (1 - \lambda)Q_2) \leq \lambda D_f(P_1 \| Q_1) + (1 - \lambda)D_f(P_2 \| Q_2) \quad (۳۳-۲)$$

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.

⁴⁰Monotonicity

⁴²Markov Kernel / Probability Kernel

⁴³Post-processing

- نامساوی پینسکر^{۴۵}: 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_1 را تضمین می‌کند:

$$\|P - Q\|_{TV} \leq \sqrt{\frac{1}{2} \text{KL}(P\|Q)} \quad (34-2)$$

- واگرایی $E_\gamma(P\|Q)$ (چوب هاکی) و ارتباط با TV: واگرایی E_γ که به صورت $\int(dP - \gamma dQ)^+$ تعریف می‌شود، تعمیمی از فاصله تغییرات کل است. به طور مشخص، در نقطه $\gamma = 1$ این دو معیار برابر هستند:

$$E_1(P\|Q) = \|P - Q\|_{TV} = \frac{1}{2} \|P - Q\|_1 \quad (35-2)$$

این ویژگی نشان می‌دهد که E_γ طیفی از فواصل را می‌سازد که از هندسه محض ($\gamma = 1$) شروع شده و به معیارهای اطلاعاتی می‌رسد.

- نمایش انتگرالی f -واگرایی‌ها^{۴۶}: 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 f -واگرایی‌ها عمل می‌کند. هر f -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:

$$D_f(P\|Q) = \int_1^\infty (f''(\gamma) E_\gamma(P\|Q) + \gamma^{-1} f''(\gamma^{-1}) E_\gamma(Q\|P)) d\gamma \quad (36-2)$$

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 $f''(\gamma)$ (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.

۳-۲ مبانی آماری و کران‌های اطلاعاتی

ابزارهای سنجش فاصله بین توزیع‌ها (مانند f -واگرایی‌ها) را در بخش ۲-۲ معرفی کردیم. در این بخش، به معرفی چارچوب «نظریه تصمیم آماری»^{۴۷} می‌پردازیم. هدف ما ایجاد پلی مستحکم میان مفاهیم انتزاعی فاصله و مسئله‌ی عملیاتی تخمین پارامتر است. تعاریف و لمحه‌ای ارائه شده در این بخش، ابزار اصلی برای اثبات کران‌های پایین مینی‌مکس در فصل‌های بعد خواهد بود.

⁴⁵Pinsker's Inequality

⁴⁶Integral Representation of f-Divergences

⁴⁷Statistical Decision Theory

۱-۳-۲ نظریه تصمیم و ریسک مینی‌مکس

فرض کنید مدل آماری ما شامل فضای نمونه \mathcal{X} و خانواده‌ای از توزیع‌های احتمال $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ باشد که توسط فضای پارامتر Θ اندیس‌گذاری شده‌اند. داده‌های مشاهده شده $X^n = (X_1, \dots, X_n) \in \mathcal{X}^n$ متغیرهای تصادفی مستقلی هستند که از توزیع ناشناخته P_θ نمونه‌برداری شده‌اند.

یک تخمین‌گر^{۴۸}، تابعی اندازه‌پذیر به صورت $\hat{\theta} : \mathcal{X}^n \rightarrow \hat{\Theta}$ است که بر اساس مشاهدات، تقریبی از پارامتر واقعی ارائه می‌دهد. برای سنجش کیفیت این تقریب، از یک متریک (یا شبهمتریک) $\rho : \Theta \times \Theta \rightarrow \mathbb{R}_{\geq 0}$ بر روی فضای پارامتر استفاده می‌کنیم.تابع زیان معمولاً تابعی صعودی از این متریک است:

$$L(\hat{\theta}(X^n), \theta) = \Phi(\rho(\hat{\theta}(X^n), \theta)) \quad (37-2)$$

که در آن $\Phi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ تابعی صعودی و غیرمنفی است (مانند t^2 برای خطای میانگین مربعات).

«ریسک»^{۴۹} یک تخمین‌گر در نقطه θ ، برابر با امید ریاضی زیان تحت توزیع P_θ تعریف می‌شود:

$$R(\hat{\theta}, \theta) = \mathbb{E}_{X^n \sim P_\theta^n} [\Phi(\rho(\hat{\theta}(X^n), \theta))] \quad (38-2)$$

در تحلیل مینی‌مکس، ما بدترین عملکرد ممکن یک تخمین‌گر را در سراسر فضای پارامتر در نظر می‌گیریم.

تعریف ۱۹-۲ (ریسک مینی‌مکس) ریسک مینی‌مکس $\mathfrak{M}_n(\Theta)$ برای خانواده \mathcal{P} و متریک ρ ، برابر است با کمینه‌ی بیشینه‌ی ریسک:

$$\mathfrak{M}_n(\Theta, \Phi \circ \rho) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{P_\theta^n} [\Phi(\rho(\hat{\theta}(X^n), \theta))] \quad (39-2)$$

که در آن اینفیم بر روی تمامی تخمین‌گرهای اندازه‌پذیر $\hat{\theta}$ گرفته می‌شود و سوپریم روی تمام پارامترهای عضو خانواده Θ .

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

محاسبه مستقیم مقدار (۳۹-۲) دشوار است. استراتژی استاندارد ما برای غلبه بر این دشواری، «تبديل فضای پیوسته پارامتر به یک مجموعه گسسته» است. به عبارت دیگر، مسئله‌ی «تخمین» را به مسئله‌ی «آزمون فرض چندگزینه‌ای» تقلیل می‌دهیم.

⁴⁸Estimator

⁴⁹Risk

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 ۲۰-۲ (مجموعه بسته‌بندی^{۵۰}) مجموعه‌ی متناهی $\Theta \subset \{\theta_1, \dots, \theta_M\}$ یک $\mathcal{V} = \{2\delta, \dots, \delta\}$ -بسته‌بندی برای فضای (Θ, ρ) نامیده می‌شود اگر اعضای آن حداقل به اندازه 2δ از یک دیگر فاصله داشته باشند:

$$\min_{i \neq j} \rho(\theta_i, \theta_j) \geq 2\delta \quad (40-2)$$

نکته کلیدی در این است که اگر تخمین‌گر $\hat{\theta}$ به پارامتر واقعی θ_v نزدیک باشد ($\hat{\theta} < \theta_v$), آنگاه به دلیل نامساوی مثلث، فاصله آن از سایر نقاط شبکه بیش از δ خواهد بود. بنابراین، وقوع خطای آزمون فرضیه، مستلزم وقوع خطای تخمین بزرگ است. با استفاده از مفاهیم فاصله که در بخش ۲-۲ دیدیم و ترکیب آن با نامساوی مارکوف، به قضیه‌ی زیر می‌رسیم:

قضیه‌ی ۱۰-۲ (کران پایین عمومی) فرض کنید $\{\theta_1, \dots, \theta_M\}$ یک \mathcal{V} -بسته‌بندی برای Θ باشد. آنگاه ریسک مینی‌مکس به خطای آزمون فرض محدود می‌شود:

$$\mathfrak{M}_n(\Theta) \geq \Phi(\delta) \cdot \inf_{\hat{\theta}} \sup_{\theta \in \mathcal{V}} \Pr_{P_{\theta}^n} \left(\rho(\hat{\theta}, \theta) \geq \delta \right) \quad (41-2)$$

$$\geq \Phi(\delta) \cdot \inf_{\psi} \bar{P}_{err}(\psi) \quad (42-2)$$

که در آن $v \neq \psi$ میانگین احتمال خطای آزمون روی فرضیه‌های گسته است و ψ آزمون‌گری است که سعی در بازیافت اندیس v دارد.

۳-۳-۲ 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 f -واگرایی‌ها بنا شده‌اند عبارتند از:

قضیه‌ی ۱۱-۲ (نامساوی لو کم^{۵۱}) 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.

$$\inf_{\psi} \bar{P}_{err}(\psi) \geq \frac{1}{2} (1 - \|P_1^n - P_2^n\|_{TV}) \quad (43-2)$$

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 ۱۲-۲ (نامساوی فانو^{۵۲}) زمانی که پارامتر مورد نظر متعلق به مجموعه‌ای بزرگ‌تر \mathcal{V} باشد ($M = |\mathcal{V}| > 2$)، نامساوی فانو کران پایین قوی‌تری ارائه می‌دهد:

$$\inf_{\psi} \bar{P}_{err}(\psi) \geq 1 - \frac{I(X^n; V) + \log 2}{\log M} \quad (44-2)$$

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

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

لم ۱۳-۲ (لم اسود^{۵۳}) این لم ابزاری قدرتمند برای فضاهای پارامتر با ابعاد بالا (مانند $\{1, \dots, d\}$) است. قدرت لم اسود در شکستن یک مسئله d -بعدی دشوار به d مسئله ۱-بعدی مستقل است.

$$\mathfrak{M}_n(\Theta) \geq \frac{\delta}{2} \sum_{j=1}^d \left[1 - \|M_{+j}^n - M_{-j}^n\|_{TV} \right] \quad (45-2)$$

که در آن M_{+j}^n و M_{-j}^n توزیع‌های مخلوط حاشیه‌ای هستند (میانگین توزیع‌هایی که بیت زام آن‌ها به ترتیب $+1$ و -1 است). اگر در هر بعد تمایز قائل شدن سخت باشد، مجموع خطاهای با تعداد ابعاد d جمع شده و کران محکمی می‌سازد.

فصل ۳

محرمانگی تفاضلی موضعی

۱-۳ مقدمه

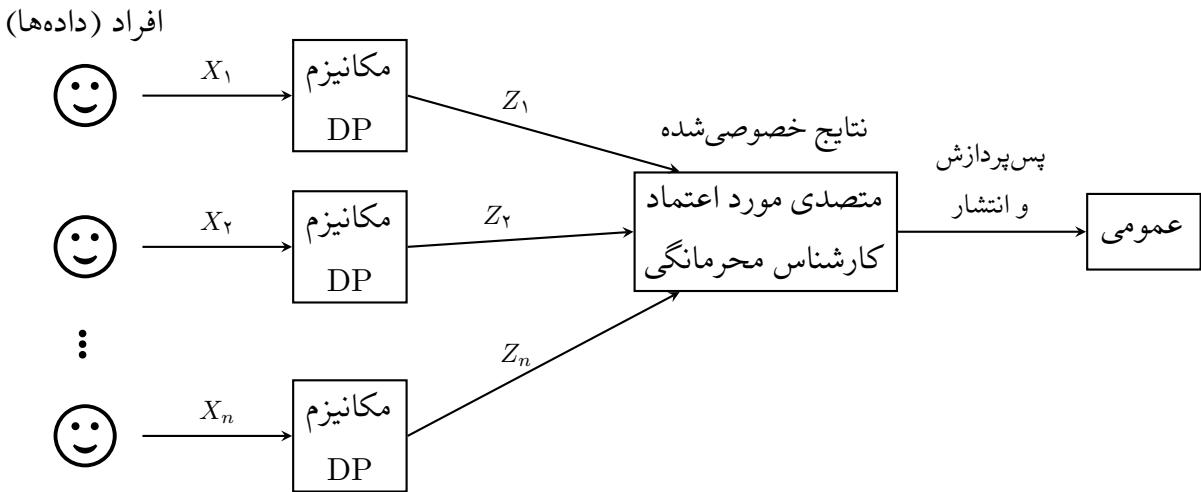
در فصل پیشین (۲)، مفاهیم بنیادی نظریه اندازه و مدل محرمانگی تفاضلی مت مرکز (CDP) را بررسی کردیم. همان‌طور که در بخش ۱-۲ مشاهده شد، مدل مت مرکز (CDP) بر فرض وجود یک متصلی مورد اعتماد استوار است که به داده‌های خام تمامی کاربران دسترسی دارد ($M : \mathcal{X}^n \rightarrow \mathcal{R}$). اگرچه این مدل دقیق آماری بالایی را فراهم می‌کند، اما ذخیره‌سازی مت مرکز داده‌ها یک «نقطه شکست مرکزی» ایجاد می‌کند؛ به این معنا که نفوذ به سرور یا خیانت متصلی، حریم خصوصی تمامی کاربران را به خطر می‌اندازد. برای مثال در بسیاری از کاربردهای مدرن، مانند جمع‌آوری داده‌های تله‌منtri مروگرها 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»^۱ 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 M_i را روی داده‌ی خود اجرا کرده و تنها خروجی نویزدار Z_i را منتشر می‌کند (شکل ۱-۳).

۲-۳ تعاریف رسمی و مدل‌های محاسباتی

در مدل موضعی، مجموعه‌ای از n کاربر وجود دارند که هر کدام داده‌ای خصوصی $X_i \in \mathcal{X}$ در اختیار دارند. برخلاف مدل مت مرکز که شرط محرمانگی روی «پایگاه داده‌های همسایه» تعریف می‌شد، در این جا

¹Local Differential Privacy



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

شرط محترمانگی باید برای «هر جفت ورودی ممکن» در دامنه برقرار باشد تا تمایز قائل شدن بین مقادیر مختلف ورودی برای مهاجم دشوار گردد.

تعريف ۱-۳ (تصادفی‌ساز موضعی^۲) یک مکانیزم تصادفی $\mathcal{Z} \rightarrow \mathcal{M} : \mathcal{X}$ را یک تصادفی‌ساز موضعی می‌نامیم که ورودی $x \in \mathcal{X}$ را دریافت کرده و خروجی $z \in \mathcal{Z}$ را بر اساس توزیع احتمال شرطی $Q(z|x)$ تولید می‌کند.

۱-۲-۳ تعریف ریاضی LDP

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

تعريف ۲-۳ (α -LDP) یک مکانیزم تصادفی \mathcal{M} ، α -LDP است، اگر برای تمام جفت ورودی‌های $x, x' \in \mathcal{X}$ و برای هر رویداد خروجی $\mathcal{Z} \subseteq \mathcal{S}$ (در σ -جبر برد) داشته باشیم:

$$\sup_{x, x' \in \mathcal{X}} \sup_{\mathcal{S} \subseteq \mathcal{Z}} \frac{\mathbb{P}[\mathcal{M}(x) \in \mathcal{S}]}{\mathbb{P}[\mathcal{M}(x') \in \mathcal{S}]} \leq e^\alpha \quad (1-3)$$

نکته: در متون آماری این حوزه (مانند [۶])، معمولاً از پارامتر α برای بودجه‌ی محترمانگی 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 (۱-۳) دقیقاً معادل است با:

$$\sup_{x, x' \in \mathcal{X}} D_\infty(Q(\cdot|x) \| Q(\cdot|x')) \leq \alpha \quad (2-3)$$

این رابطه نشان می‌دهد که α -LDP محدودیتی سخت‌گیرانه بر روی «نسبت درست‌نمایی»^۳ توزیع‌های خروجی اعمال می‌کند و تضمین می‌دهد که مشاهده‌ی خروجی z ، اطلاعات اندکی درباره‌ی ورودی x افشا می‌کند.

۲-۲-۳ محرمانگی تقریبی

مشابه مدل متمرکز، در برخی کاربردها نیاز است که تعریف α -LDP را تضعیف کنیم تا اجازه‌ی یک احتمال شکست ناچیز^۴ داده شود. این حالت معمولاً زمانی رخ می‌دهد که دامنه یا برد مکانیزم نامتناهی باشد (مانند مکانیزم گوسی).

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

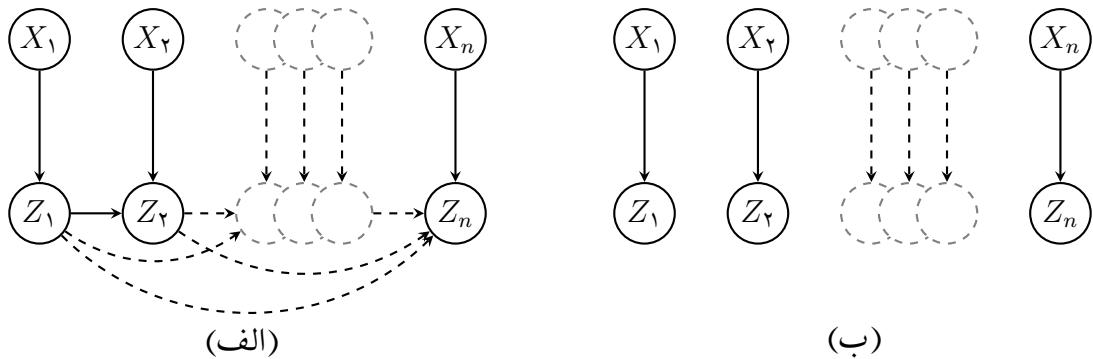
$$\mathbb{P}[M(x) \in Z] \leq e^\alpha \cdot \mathbb{P}[M(x') \in Z] + \delta \quad (3-3)$$

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

۳-۳ پروتکل‌های تعاملی و خواص ترکیب

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

³Likelihood Ratio



شکل ۲-۳: (الف) مدل گرافیکی نمایش دهنده روابط استقلال شرطی بین داده های خصوصی $\{X_i\}_{i=1}^n$ و متغیر های مشاهده شده $\{Z_i\}_{i=1}^n$ در پروتکل های تعاملی؛ پیکان های افقی نشان دهنده واپستگی Z_i به تاریخچه پیشین هستند. (ب) مدل گرافیکی ساده تر در پروتکل های غیر تعاملی که در آن خروجی ها به شرط ورودی ها از یک دیگر مستقل هستند.

۱-۳-۳ پروتکل های غیر تعاملی

در پروتکل های غیر تعاملی^۴، تمام کاربران $n = 1, \dots, i = i$ مکانیزم های خود را به صورت همزمان و مستقل از یک دیگر اجرا می کنند. اگر Z_i خروجی کاربر i -ام باشد، توزیع آن تنها به داده خصوصی X_i وابسته است و هیچ واپستگی ای به خروجی سایر کاربران ندارد. به بیان ریاضی، توزیع مشترک خروجی ها به صورت حاصل ضرب توزیع های حاشیه ای فاکتور می شود:

$$\mathbb{P}(Z_1, \dots, Z_n | X_1, \dots, X_n) = \prod_{i=1}^n Q_i(Z_i | X_i) \quad (4-3)$$

بسیاری از پیاده سازی های صنعتی فعلی، از جمله RAPPOR گوگل [۱۰]، در این دسته قرار می گیرند.

۲-۳-۳ پروتکل های تعاملی (ترتیبی)

در پروتکل های تعاملی^۵، کاربران به نوبت داده های خود را ارسال می کنند و مکانیزم کاربر i -ام می تواند به خروجی های مشاهده شده از کاربران پیشین (Z_1, \dots, Z_{i-1}) وابسته باشد. این مدل، آزادی عمل بیشتری را برای طراحی الگوریتم های تطبیقی فراهم می کند.

از دیدگاه آنالیز ریاضی، این فرآیند با استفاده از «کرنل های احتمالاتی»^۶ مدل سازی می شود. فرض کنید (Z_1, \dots, Z_{i-1}) بردار خروجی های پیشین باشد که σ -فیلد \mathcal{F}_{i-1} را تولید می کند. مکانیزم کاربر

⁴Non-interactive

⁵Interactive / Sequential

⁶Probability Kernels

i-ام، یک کرنل احتمالاتی Q_i است که خروجی $Z_i \in Z$ را مشروط بر دادهای خصوصی X_i و تاریخچه‌ی عمومی $Z_{1:i-1}$ تولید می‌کند:

$$Z_i \sim Q_i(dz_i|x_i, z_{1:i-1}) \quad (5-3)$$

شرط اساسی محترمانگی در اینجا این است که با شرط‌سازی 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 کسب شده 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 محترمانگی با یک‌دیگر جمع می‌شوند. قضیه‌ی زیر، کران بالای محترمانگی را برای یک پروتکل ترتیبی بیان می‌کند [۶].

قضیه‌ی ۱-۳ (ترکیب ترتیبی)^۵ فرض کنید در یک پروتکل تعاملی، برای هر کاربر $\{1, \dots, n\}$ و به ازای هر تاریخچه‌ی ممکن $z^{i-1} \in Z_{1:i-1}$ ، 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 محترمانگی تفاضلی موضعی با بودجه‌ی مجموع است:

$$\alpha_{total} = \sum_{i=1}^n \alpha_i \quad (6-3)$$

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 x'^n هستند، نسبت احتمال توأم به حاصل ضرب 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 حاصل ضرب با $e^{\sum \alpha_i}$ کران‌دار خواهد بود. \square

۴-۳ 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 اراضی شرط محترمانگی، «خطای تخمین» (که

معمولًاً 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 نااریب»^۸ معکوس استفاده می‌شود.

۱-۴-۳ پاسخ تصادفی دودویی (RR)

پایه‌ای ترین مکانیزم α -LDP، پاسخ تصادفی^۹ (RR) برای دامنه‌ی دودویی $\{0, 1\}^{\mathcal{X}}$ است. این مکانیزم را می‌توان به صورت یک کانال متقارن دودویی مدل‌سازی کرد که ورودی x را با احتمال p حفظ کده و با احتمال $p - 1$ قرینه می‌کند:

$$\mathbb{P}[y = z|x] = \begin{cases} p & \text{if } z = x \\ 1 - p & \text{if } z \neq x \end{cases} \quad (7-3)$$

اثبات α -LDP: برای اینکه این مکانیزم شرط α -LDP را براورد کند، طبق تعریف^{۱۰} 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 ($p - 1$) باشد:

$$\sup_{y \in \{0, 1\}} \frac{\mathbb{P}[y|x]}{\mathbb{P}[y|x']} = \frac{p}{1-p} \leq e^\alpha \quad (8-3)$$

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

$$p = \frac{e^\alpha}{e^\alpha + 1} \quad (9-3)$$

تحلیل واریانس: برای تحلیل دقیق خطای ابتدا تخمین‌گر نااریب را استخراج می‌کنیم. اگر $y \in \{0, 1\}$ باشد، هدف یافتن تابعی $(\hat{f}(y))$ است که $x = \mathbb{E}_{\hat{f}(y)}$ باشد.

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

$$\hat{x} = \frac{y - (1 - p)}{2p - 1} \quad (10-3)$$

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

$$\text{Var}[\hat{x}] = \frac{e^\alpha}{(e^\alpha - 1)^2} \quad (11-3)$$

⁸Unbiased Estimator

⁹Randomized Response

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

$$\mathbb{E}_y = p \cdot x + (1 - p)(1 - x) = (2p - 1)x + (1 - p)$$

با جایگذاری در معادله تخمین‌گر:

$$\mathbb{E}_{\hat{x}} = \frac{\mathbb{E}_y - (1 - p)}{2p - 1} = \frac{(2p - 1)x}{2p - 1} = x$$

برای محاسبه واریانس، چون y یک متغیر بُرنولی است، واریانس آن $(1 - p)$ می‌شود. با اعمال خواص واریانس ($\text{Var}[aY + b] = a^2 \text{Var}[Y]$) داریم:

$$\text{Var}[\hat{x}] = \frac{\text{Var}[y]}{(2p - 1)^2} = \frac{p(1 - p)}{(2p - 1)^2}$$

حال با جایگذاری $\frac{e^\alpha}{e^\alpha + 1} = p$ در رابطه بالا:

$$\text{Var}[\hat{x}] = \frac{\frac{e^\alpha}{(e^\alpha + 1)^2}}{\left(\frac{e^\alpha}{e^\alpha + 1} - 1\right)^2} = \frac{\frac{e^\alpha}{(e^\alpha + 1)^2}}{\left(\frac{e^\alpha - 1}{e^\alpha + 1}\right)^2} = \frac{e^\alpha}{(e^\alpha - 1)^2}$$

□

۲-۴-۳ پاسخ تصادفی تعمیم‌یافته (GRR)

پاسخ تصادفی تعمیم‌یافته^{۱۰}، برای دامنه‌های گسسته با $k > 2$ عنصر ($\mathcal{X} = \{1, \dots, k\}$ ، به عنوان تعمیم مستقیم RR معرفی می‌شود. این مکانیزم را می‌توان با یک «ماتریس گذار»^{۱۱} تصادفی توصیف کرد.

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

$$Q = \begin{pmatrix} p & q & \dots & q \\ q & p & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p \end{pmatrix}_{k \times k} \quad (12-3)$$

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

¹⁰Generalized Randomized Response

¹¹Transition Matrix

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

$$p = \frac{e^\alpha}{e^\alpha + k - 1}, \quad q = \frac{1}{e^\alpha + k - 1} \quad (13-3)$$

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

$$\text{Var}_{GRR} = \frac{p(1-p)}{(p-q)^2} = \frac{(e^\alpha)(k-1) + (k-1)^2}{(e^\alpha - 1)^2} \approx O(k) \quad (14-3)$$

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 یگانی (UE)

برای غلبه بر مشکل کاهش دقت GRR در دامنه‌های بزرگ، خانواده‌ای از مکانیزم‌ها تحت عنوان «کدگذاری یگانی»^{۱۲} توسعه یافته‌اند. این رویکرد اساس پروتکل مشهور RAPPOR گوگل را تشکیل می‌دهد [۱۰، ۲۰].

تعريف ۵-۳ (کدگذاری یگانی) در این روش، فرآیند خصوصی‌سازی طی دو مرحله انجام می‌شود:

۱. **کدگذاری قطعی:** ورودی $\{1, \dots, k\}^k$ به یک بردار بیتی $v \in \{0, 1\}^k$ تبدیل می‌شود که تنها در موقعیت x برابر با ۱ و در سایر جاهای ۰ است (*One-hot encoding*).

۲. **اختلال**^{۱۳} مستقل: هر بیت این بردار به صورت مستقل با استفاده از یک مکانیزم بازتری معکوس می‌شود. اگر v_i بیت i -ام بردار کدگذاری شده باشد، خروجی z_i به صورت زیر تولید می‌شود:

$$\mathbb{P}[z_i = 1] = \begin{cases} p & \text{if } v_i = 1 \\ q & \text{if } v_i = 0 \end{cases} \quad (15-3)$$

مثال ۱-۳ فرض کنید دامنه شامل ۴ آیتم باشد ($\mathcal{X} = \{1, 2, 3, 4\}$) و ورودی کاربر $x = 2$ باشد.

۱. بردار کدگذاری شده: $v = [0, 1, 0, 0]$

¹²Unary Encoding (UE)

¹³Perturbation

۲. اعمال نویز: هر بیت مستقل پرتاپ می‌شود. ممکن است خروجی نهایی $[0, 1, 0] = z$ شود (بیت سوم از \cdot به \cdot تغییر کرده است).

اثبات α -LDP: برای بررسی شرط α -LDP، نسبت احتمال خروجی برداری $(z_1, \dots, z_k) = z$ را برای دو ورودی متمایز x و x' محاسبه می‌کنیم. بردارهای متناظر v و v' تنها در دو موقعیت تفاوت دارند: موقعیت x (که $v_x = 1, v'_{x'} = 0$) و موقعیت x' (که $v_x = 0, v'_{x'} = 1$). 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:

$$\frac{\mathbb{P}[z|x]}{\mathbb{P}[z|x']} = \frac{\mathbb{P}[z_x|v_x=1] \cdot \mathbb{P}[z_{x'}|v_{x'}=0]}{\mathbb{P}[z_x|v'_x=0] \cdot \mathbb{P}[z_{x'}|v'_{x'}=1]} \quad (16-3)$$

بیشینه‌ی این کسر زمانی رخ می‌دهد که صورت کسر ماکزیمم و مخرج مینیمم شود؛ یعنی زمانی که $z_x = 1$ (حفظ بیت \cdot) و $z_{x'} = 0$ (حفظ بیت \cdot) باشد. در این حالت:

$$\frac{\mathbb{P}[z|x]}{\mathbb{P}[z|x']} \leq \frac{p}{q} \cdot \frac{1-q}{1-p} = \frac{p(1-q)}{q(1-p)} \quad (17-3)$$

بنابراین شرط α -LDP معادل است با:

$$\alpha = \ln \left(\frac{p(1-q)}{q(1-p)} \right) \quad (18-3)$$

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:

$$\text{Var}_{UE} = \frac{q(1-q)}{(p-q)^2} \quad (19-3)$$

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

دو استراتژی اصلی برای تنظیم p و q بر اساس رابطه (18-3) وجود دارد:

۱. **کدگذاری یگانی متقارن (SUE):** در این روش $1 = p + q$ در نظر گرفته می‌شود. با جایگذاری در شرط α -LDP، مقادیر بهینه عبارتند از $p = \frac{1}{e^{\alpha/2} + 1}$ و $q = \frac{e^{\alpha/2}}{e^{\alpha/2} + 1}$. واریانس در این حالت برابر است با:

$$\text{Var}_{SUE} = \frac{e^{\alpha/2}}{(e^{\alpha/2} - 1)^2} \quad (20-3)$$

۲. **کدگذاری یگانی بهینه (OUE):** وانگ و همکاران [۲۰] نشان دادند که برای کمینه‌سازی واریانس در دامنه‌های بزرگ، باید اطلاعات بیت‌های ۱ (سیگنال) حفظ شود ($p = 1/2$) و نویز روی بیت‌های \cdot (که اکثر بردار را تشکیل می‌دهند) کنترل شود ($q = \frac{1}{e^\alpha + 1}$). واریانس حاصل برابر است با:

$$\text{Var}_{OUE} = \frac{4e^\alpha}{(e^\alpha - 1)^2} \quad (21-3)$$

۴-۴-۳ 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 است.^k

مثال ۲-۳ (ناکارآمدی GRR در دامنه‌های بزرگ) فرض کنید می‌خواهیم کلمات پرکاربرد را از یک لغت‌نامه با $k = 100,000$ کلمه استخراج کنیم.

- در مکانیزم GRR ، طبق رابطه (۱۴-۳)، واریانس تقریباً با k رشد می‌کند:

$$\text{Var}_{GRR} \approx \frac{k}{(e^\alpha - 1)^2}$$

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:

$$\text{Var}_{OUE} = \frac{4e^\alpha}{(e^\alpha - 1)^2}$$

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^{۱۴}» 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.

۵-۴-۳ 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 $\mathbb{R} \subset \mathcal{X}$ باشد. بدون کاستن از کلیت^{۱۵}، فرض می‌کنیم داده‌ها با یک تبدیل خطی به بازه‌ی $[-1, 1]$ نگاشت شده‌اند. در مدل موضعی، شرط محترمانگی باید برای هر جفت ورودی $x, x' \in \mathcal{X}$ برابر باشد. بنابراین «حساسیت سراسری» (Δ) برابر با قطر دامنه است. در این صورت Δ ، بیشترین مقدار ممکن را خواهد داشت:

$$\Delta = \sup_{x, x' \in [-1, 1]} |x - x'| = |1 - (-1)| = 2 \quad (۲۲-۳)$$

¹⁴Optimized Local Hashing (OLH)

¹⁵Without Loss of Generality

تعريف ۶-۳ (مکانیزم لاپلاس موضعی) مکانیزم M_{Lap} ورودی نرمالایز شده $[1, -1] \in x$ را دریافت کرده و خروجی z را طبق رابطه زیر تولید می‌کند:

$$z = x + \eta, \quad \eta \sim \text{Lap} \left(\frac{\Delta}{\alpha} \right) = \text{Lap} \left(\frac{2}{\alpha} \right) \quad (23-3)$$

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

$$f(z|x) = \frac{\alpha}{4} \exp \left(-\frac{\alpha|z-x|}{2} \right) \quad (24-3)$$

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

$$\frac{f(z|x)}{f(z|x')} = \frac{\exp \left(-\frac{\alpha}{2}|z-x| \right)}{\exp \left(-\frac{\alpha}{2}|z-x'| \right)} \quad (25-3)$$

$$= \exp \left(\frac{\alpha}{2}(|z-x'| - |z-x|) \right) \quad (26-3)$$

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

$$|z-x'| - |z-x| \leq |(z-x') - (z-x)| = |x-x'| \leq 2$$

بنابراین نسبت احتمال با $\exp \left(\frac{\alpha}{2} \cdot 2 \right) = e^\alpha$ کران دار می‌شود.

مثال ۳-۳ فرض کنید می‌خواهیم دمای بدن یک بیمار را گزارش کنیم. اگر دامنه تغییرات دما $[42, 35]$ درجه باشد، طول بازه $7 = 35 - 42$ است. اگر بدون نرمال‌سازی از لاپلاس استفاده کنیم، باید نویزی متناسب با $\frac{7}{\alpha}$ اضافه کنیم. اما در روش استاندارد، ابتدا دما را به $[1, -1]$ نگاشت می‌کنیم (که حساسیت ۲ شود)، نویز با مقیاس $\frac{2}{\alpha}$ اضافه می‌کنیم و در سمت سرور مجدد نتیجه را به مقیاس اصلی بر می‌گردانیم. حال اگر بخواهیم دمای بدن یک بیمار را که به بازه $[1, -1]$ نرمالایز شده است، با بودجه $\alpha = 1$ منتشر کنیم. اگر مقدار واقعی $0.5 = x$ باشد:

- مقیاس نویز برابر است با $2 = \frac{2}{1} \cdot b$.

- یک نمونه تصادفی ممکن است $z = -0.5 + 0.5 = 0$ باشد.

- واریانس خطای برابر است با $b^2 = 2^2 = 4$. این واریانس برای یک مقدار در بازه $[1, -1]$ بسیار زیاد است و نشان می‌دهد که LDP برای داده‌های عددی تک بعدی خطای زیادی تحمیل می‌کند مگر اینکه n (تعداد کاربران) بسیار زیاد باشد.

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

همان طور که در مثال قبل دیدیم، مکانیزم لاپلاس به دلیل دامنهٔ نامتناهی خروجی، واریانس قابل توجهی را تحمیل می‌کند. برای حل این مشکل در داده‌های کران‌دار (مثلاً $[1, -1]$)، دوچی و همکاران [۵] یک مکانیزم بهینه ارائه کردند که خروجی را به شدت کوانتیده (باینری) می‌کند.

این مکانیزم که به «مکانیزم ۱-بیتی» نیز شهرت دارد، ورودی پیوسته $x \in [-1, 1]$ را به یک متغیر تصادفی $z \in \{-B, +B\}$ تبدیل می‌کند، به طوری که ناریب باقی بماند ($\mathbb{E}_z = x$).

الگوریتم ۱ مکانیزم ۱-بیتی دوچی

۱: ورودی: داده کاربر $x \in [-1, 1]$ و بودجه محروم‌گی α .

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

$$u = \begin{cases} 1 & \text{با احتمال } \frac{1+x}{2} \\ -1 & \text{با احتمال } \frac{1-x}{2} \end{cases} \quad (27-3)$$

۳: گام ۲ (اختلال): مقدار u را با مکانیزم RR مخدوش کنید تا z' به دست آید (با احتمال $\frac{e^\alpha}{e^\alpha + 1}$ خود u و با احتمال مکمل، قرینه‌ی آن).

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

$$z = z' \cdot \frac{e^\alpha + 1}{e^\alpha - 1} \quad (28-3)$$

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 مینیماکس برای تخمین میانگین بهینه است. واریانس آن برابر است با:

$$\text{Var}_{Duchi} \leqslant \left(\frac{e^\alpha + 1}{e^\alpha - 1} \right)^2 \approx O\left(\frac{1}{\alpha^2}\right) \quad (29-3)$$

مزیت اصلی این روش نسبت به لاپلاس، کران‌دار بودن قطعی خروجی است که در تحلیل‌های بدترین حالت^{۱۶} اهمیت زیادی دارد.

¹⁶Worst-case

چالش ابعاد بالا (نفرین ابعاد)

چرا در ابعاد بالا ($d > 1$) از مکانیزم لاپلاس استفاده نمی‌شود؟ فرض کنید ورودی کاربر یک بردار $x \in \mathbb{R}^d$ باشد که در توب واحد اقلیدسی قرار دارد ($\|x\|_2 \leq 1$).

مشکل بینایدین این است که مکانیزم لاپلاس متکی بر حساسیت در فضای ℓ_1 است، در حالی که هندسه فضای برداری اقلیدسی منطبق بر ℓ_2 می‌باشد. برای پوشش دادن توب واحد ℓ_2 با نویز لاپلاس، باید حساسیت ℓ_1 را در بدترین حالت در نظر بگیریم. می‌دانیم برای هر بردار با نرم اقلیدسی ۱، نرم ℓ_1 می‌تواند تا \sqrt{d} رشد کند. بنابراین قطر دامنه در متر ℓ_1 برابر است با:

$$\Delta_{\ell_1} = \sup_{x, x' \in B_{\ell_1}} \|x - x'\|_1 \leq \sqrt{d} \cdot \sup \|x - x'\|_2 = 2\sqrt{d} \quad (30-3)$$

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

- واریانس خطأ در هر بعد: $2 \times \left(\frac{\sqrt{d}}{\alpha}\right)^2 = \frac{8d}{\alpha^2}$

- خطای میانگین مربعات کل 17 برای d بعد: $d \times \frac{8d}{\alpha^2} = \frac{8d^3}{\alpha^2}$

این نرخ رشد ($O(d^2)$) برای خطأ، بسیار ناکارآمد است. دوچی و همکاران [۶] ثابت کردند که حد پایین نظری مینی‌مکس برای این مسئله ($O(d)$) است. به همین دلیل، در ابعاد بالا از مکانیزم‌های پیشرفته‌تری مانند «توزیع برنولی چندبعدی» یا «نمونه‌برداری هایپرکیوب» استفاده می‌شود که با هندسه فضای سازگارترند.

مقایسه با مدل مرکزی: در مثال ۲-۶ فصل قبل دیدیم که در مدل مرکزی، 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 موضوعی، این پدیده بسیار شدیدتر است. در اینجا نه تنها نویز با d رشد می‌کند، بلکه به دلیل عدم مرکز و جمع شدن خطاهای تک‌تک کاربران، خطای نهایی (MSE) با d^2 افزایش می‌یابد. به همین دلیل، راهکارهای ساده‌ی مدل مرکز (مانند افزودن نویز به هر بعد) در مدل موضوعی تقریباً بلااستفاده هستند.

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

¹⁷Mean Squared Error (MSE)

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.

۱-۵-۳ تعریف مسئله: تخمین میانگین دودویی

فرض کنید n کاربر وجود دارند و هر کاربر i دارای یک بیت خصوصی $\{0, 1\} \in X_i$ است. هدف تخمین‌گر، محاسبه میانگین واقعی جامعه است:

$$p = \frac{1}{n} \sum_{i=1}^n X_i \quad (31-3)$$

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

$$\text{MSE}(\hat{p}) = \mathbb{E}[(\hat{p} - p)^2] = \text{Var}[\hat{p}] + (\text{Bias}[\hat{p}])^2 \quad (32-3)$$

ما در هر دو مدل از تخمین‌گرهای ناریب ($Bias = 0$) استفاده می‌کنیم، بنابراین خطأ صرفاً ناشی از واریانس نویز تزریق شده است.

۲-۵-۳ تحلیل در مدل متمرکز (CDP)

در مدل متمرکز با بودجه محرمانگی ε -DP، متصدی به تمام X_i ‌ها دسترسی دارد. او ابتدا مقدار دقیق مجموع $\sum X_i$ را محاسبه می‌کند. چون تغییر یک بیت حداقل مجموع را ۱ واحد تغییر می‌دهد، حساسیت سراسری برابر با $1/\Delta = 1/\varepsilon$ است. طبق مکانیزم لاپلاس، نویزی با مقیاس $1/\varepsilon$ به مجموع اضافه شده و سپس بر n تقسیم می‌شود تا میانگین به دست آید:

$$\hat{p}_{CDP} = \frac{1}{n} \left(\sum_{i=1}^n X_i + \eta \right), \quad \eta \sim \text{Lap}(1/\varepsilon) \quad (33-3)$$

خطای این تخمین‌گر برابر است با:

$$\text{MSE}_{CDP} = \text{Var}\left[\frac{\eta}{n}\right] = \frac{1}{n^2} \text{Var}[\eta] = \frac{1}{n^2} \cdot \frac{2}{\varepsilon^2} = \mathcal{O}\left(\frac{1}{n^2 \varepsilon^2}\right) \quad (34-3)$$

این رابطه نشان می‌دهد که در مدل متمرکز، خطأ با سرعت $1/n$ به سمت صفر میل می‌کند (یا انحراف معیار با سرعت $1/n$).

۳-۵-۳ تحلیل در مدل موضعی (LDP)

در مدل موضعی با محدودیت LDP، هیچ کس به X_i های 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

۴-۳، واریانس تخمین‌گر هر کاربر برای α های کوچک ($1 < \alpha$) تقریباً برابر است با:

$$\text{Var}[\hat{X}_i] \approx \frac{1}{\alpha^2} \quad (35-3)$$

متصدی برای تخمین میانگین کل، میانگین گزارش‌های دریافتی را محاسبه می‌کند: $\hat{p}_{LDP} = \frac{1}{n} \sum_{i=1}^n \hat{X}_i$. از آنجا که نویز کاربران مستقل از یکدیگر است، واریانس مجموع برابر با مجموع واریانس‌هاست:

$$\text{MSE}_{LDP} = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n \hat{X}_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[\hat{X}_i] = \frac{1}{n^2} \cdot n \cdot \mathcal{O}\left(\frac{1}{\alpha^2}\right) = \mathcal{O}\left(\frac{1}{n\alpha^2}\right) \quad (36-3)$$

در اینجا خطأ با سرعت $1/\sqrt{n}$ به سمت صفر میل می‌کند.

۴-۵-۳ نتیجه‌گیری: شکاف کارایی

با مقایسه روابط (۳۶-۳) و (۳۶-۴)، تفاوت بین این دو مدل آشکار می‌شود (با فرض ثابت بودن بودجه‌های محرمانگی $\alpha \approx \varepsilon$):

- مدل متتمرکز: نرخ همگرایی خطأ $\mathcal{O}(1/n)$ است.

- مدل موضعی: نرخ همگرایی خطأ $\mathcal{O}(1/\sqrt{n})$ است.

این تفاوت در نرخ همگرایی پیامد بسیار مهمی در حجم نمونه^{۱۸} مورد نیاز دارد. برای رسیدن به یک دقیق ثابت مشخص (مثلاً خطای τ ، تعداد کاربران مورد نیاز در هر مدل عبارت است از:

$$n_{CDP} \propto \frac{1}{\tau}, \quad n_{LDP} \propto \frac{1}{\tau^2} \quad (37-3)$$

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

$$n_{LDP} \approx \mathcal{O}(n_{CDP}^2) \quad (38-3)$$

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

¹⁸Sample Complexity

۱,۰۰۰ کاربر به دقت مطلوبی برسیم، در مدل موضعی برای همان دقت به $1,000,000$ کاربر نیاز خواهیم داشت [۱۲، ۶].

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 بخشد؟»

فصل ۴

تحلیل مینی‌مکس و هندسه اطلاعاتی در LDP

۱-۴ مقدمه

در فصل پیشین (۳)، چارچوب محترمانگی تفاضلی موضعی (α -LDP) را به عنوان جایگزینی برای مدل متتمرکز معرفی کردیم. دیدیم که مکانیزم‌های استاندارد مانند «پاسخ تصادفی تعیین‌یافته» (GRR) و «کدگذاری‌های یکانی» (UE)، اگرچه امنیت داده‌ها را تضمین می‌کنند، اما بهای سنگینی را از نظر دقت آماری تحمل می‌کنند. به طور خاص، دیدیم که در پسیاری از مسائل تخمین میانگین، واریانس خطأ از مرتبه $O(1/n)$ در مدل متتمرکز به $O(1/n\epsilon^2)$ در مدل موضعی افزایش می‌یابد (که در اینجا n تعداد کاربران است).

این مشاهده، یک پرسش بنیادین را ایجاد می‌کند: آیا این کاهش دقت، ناشی از ضعف در طراحی مکانیزم‌های موجود (مانند RR) است، یا یک محدودیت ذاتی و غیرقابل اجتناب در طبیعت محترمانگی موضعی است؟ به بیان دیگر، آیا می‌توان مکانیزم هوشمندانه‌تری طراحی کرد که هم شرط α -LDP را ارضاء کند و هم به نرخ خطای مدل متتمرکز نزدیک شود؟

هدف اصلی این فصل، پاسخ به این پرسش با استفاده از ابزارهای قدرتمند «نظریه مینی‌مکس» است. ما نشان خواهیم داد که هزینه پرداختی در مدل موضعی (معمولًاً ضریبی از \sqrt{n} در نرخ همگرایی)، یک مانع اطلاعاتی بنیادین است و هیچ الگوریتمی نمی‌تواند از آن عبور کند.

برای اثبات این ادعا، ما از زیرساخت‌های ریاضی که در فصل ۲ بنا نهادیم (شامل f -واگرایی‌ها، لم LDP، نامساوی فانو و لم اسود) بهره خواهیم برد. با این حال، استفاده مستقیم از این ابزارها در محیط

ممکن نیست. نوآوری اصلی که در این فصل بررسی می‌کنیم، چارچوب «محرمانگی به عنوان انقباض»^۱ است که عمدتاً توسط دوچی، جردن و وین رایت [۶، ۷] توسعه یافته است.

ایده مرکزی این است که مکانیزم‌های LDP مانند یک فیلتر اطلاعاتی عمل می‌کنند که فاصله آماری بین توزیع‌های ورودی را «منقبض» کرده و تشخیص آنها را دشوار می‌سازند. ما در این فصل:

۱. ابتدا خاصیت انقباضی مکانیزم‌های α -LDP را بر حسب واگرایی کولبک-لایبلر (KL) و فاصله تغییرات کل (TV) فرمولبندی می‌کنیم.

۲. سپس این خاصیت را با نامساوی‌های پردازش داده^۲ ترکیب کرده تا کران‌های پایین مینی‌مکس را برای مسائل کلاسیک آماری استخراج کنیم.

۳. در نهایت، نشان می‌دهیم که مکانیزم‌های معرفی شده در فصل ۳، در واقع «بهینه مینی‌مکس» هستند و به بهترین نرخ ممکن دست می‌یابند.

۲-۴ محرمانگی به عنوان انقباض

همان‌طور که در فصل پیش‌نیازها (۲-۲) بحث شد، f -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»^۳ عمل می‌کنند. به عبارت دیگر، شرط α -LDP باعث می‌شود که توزیع‌های خروجی مکانیزم برای هر دو ورودی دلخواه، به یک دیگر بسیار نزدیک شوند و در نتیجه واگرایی بین آنها محدود شود.

در این بخش، ما لم اساسی دوچی و همکاران [۵] 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 کران دار می‌شود.

قضیه‌ی ۱-۴ (کران انقباض قوی برای LDP) فرض کنید $M : \mathcal{X} \rightarrow \mathcal{Z}$ 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 $(M_i(S) = \int M(S|x)dP_i(x)$)، آنگاه نامساوی زیر برقرار است:

$$D_{KL}(M_1 \| M_2) + D_{KL}(M_2 \| M_1) \leq (e^\alpha - 1)^{\frac{1}{\alpha}} \| P_1 - P_2 \|_{TV}^{\frac{2}{\alpha}} \quad (1-4)$$

¹Privacy as Contraction

²Data Processing Inequalities

³Contraction Operators

اثبات. برای اثبات این قضیه، از خاصیت تحدب مشترک f -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 آن محاسبه می‌کنیم.

گام اول: تقلیل به ورودی‌های نقطه‌ای. می‌دانیم که نگاشت $(P, Q) \mapsto D_{KL}(P\|Q)$ محدب مشترک است (بخش ۲-۱). همچنین فاصله تغییرات کل $\|\cdot\|_{TV}$ نیز محدب است. بیشینه واگرایی زمانی رخ می‌دهد که توزیع‌های P_1 و P_2 کمترین همپوشانی را داشته باشند (متعامد باشند). با توجه به خطی بودن مکانیزم M روی مخلوط‌های احتمالی، می‌توان نشان داد که کران بالا با در نظر گرفتن توزیع‌های مرکزی روی تک نقطه‌ها (توزیع‌های دیراک) به دست می‌آید [۶]. فرض کنید $x, x' \in \mathcal{X}$ دو نقطه دلخواه باشند و $(\cdot|x)$ و $(\cdot|x')$ توزیع‌های خروجی متناظر با آن‌ها باشند. اگر بتوانیم نشان دهیم که برای هر x, x' :

$$D_{KL}(Q(\cdot|x)\|Q(\cdot|x')) + D_{KL}(Q(\cdot|x')\|Q(\cdot|x)) \leq (e^\alpha - 1)^2 \quad (2-4)$$

آنگاه با توجه به این‌که برای ورودی‌های نقطه‌ای $1 = \|\delta_x - \delta_{x'}\|_{TV}$ است، حکم برای حالت کلی با استفاده از نامساوی ینسن و تحدب نتیجه می‌شود.

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:

$$e^{-\alpha} \leq \frac{q(z|x)}{q(z|x')} \leq e^\alpha \quad (3-4)$$

که در آن $q(\cdot|x)$ تابع چگالی (یا جرم) احتمال مکانیزم است. عبارت سمت چپ نامساوی (۳-۴) را بسط می‌دهیم (که همان واگرایی متقارن یا واگرایی جفریز است):

$$D_{KL}(Q_x\|Q_{x'}) + D_{KL}(Q_{x'}\|Q_x) = \int_{\mathcal{Z}} (q(z|x) - q(z|x')) \log \left(\frac{q(z|x)}{q(z|x')} \right) \nu(dz) \quad (4-4)$$

از شرط α -LDP، می‌دانیم که نسبت درست‌نمایی‌ها کراندار است. به طور خاص، چون نسبت $q(z|x)/q(z|x')$ در بازه $[e^{-\alpha}, e^\alpha]$ قرار دارد، لگاریتم آن در بازه $[-\alpha, \alpha]$ است. با این حال، برای به دست آوردن ضریب دقیق $(e^\alpha - 1)^2$ ، از یک نامساوی دقیق‌تر استفاده می‌کنیم. طبق لم ۱ در [۶]، اگر نسبت دو چگالی q/p در بازه $[e^{-\alpha}, e^\alpha]$ باشد، آنگاه:

$$D_{KL}(p\|q) \leq \frac{1}{2}(e^\alpha - 1) \|p - q\|_1 \quad (5-4)$$

با جمع دو طرف برای واگرایی متقارن و استفاده از این‌که $\|p - q\|_1 = 2\|P - Q\|_{TV}$ ، به رابطه دقیق‌تر می‌رسیم. اما یک روش مستقیم‌تر استفاده از کران زیر است که برای هر دو توزیع P, Q که شرط نسبت محدود را دارند برقرار است:

$$\int (p(z) - q(z)) \log \frac{p(z)}{q(z)} dz \leq (e^\alpha - 1) \int |p(z) - q(z)| dz = (e^\alpha - 1) \|Q_x - Q_{x'}\|_1 \quad (6-4)$$

همچنین می‌دانیم که تحت شرط α -LDP، فاصله L_1 خروجی‌ها نیز محدود است:

$$\|Q_x - Q_{x'}\|_1 \leq e^\alpha - 1 \quad (7-4)$$

(این نامساوی از آن‌جا ناشی می‌شود که جرم احتمال نمی‌تواند سریع‌تر از ضریب e^α جابجا شود). با ترکیب این دو نتیجه:

$$\begin{aligned} D_{KL}(Q_x \| Q_{x'}) + D_{KL}(Q_{x'} \| Q_x) &\leq (e^\alpha - 1) \|Q_x - Q_{x'}\|_1 \\ &\leq (e^\alpha - 1)^2 \end{aligned} \quad (8-4)$$

این اثبات برای حالت ورودی‌های نقطه‌ای کامل می‌شود. تعمیم به توزیع‌های کلی P_1, P_2 از طریق نامساوی پردازش داده (DPI) و تحدب مشترک حاصل می‌شود که در آن ضریب $\|P_1 - P_2\|_{TV}^2$ ظاهر می‌گردد. \square

۴-۲-۱ تفسیر رژیم‌های محروم‌گی

نامساوی (۱-۴) پیامدهای مهمی برای نرخ‌های همگرایی آماری دارد. رفتار ضریب $(e^\alpha - 1)^2$ در دو رژیم حدی قابل توجه است:

- **رژیم محروم‌گی بالا ($1 \leq \alpha$):** در این حالت، می‌توان از بسط تیلور استفاده کرد: $\alpha \approx e^\alpha - 1$. بنابراین کران انقباض به صورت زیر در می‌آید:

$$D_{KL}(M_1 \| M_2) \lesssim \alpha^2 \|P_1 - P_2\|_{TV}^2 \quad (9-4)$$

این نتیجه نشان می‌دهد که میزان اطلاعات فیشر و واگرایی‌ها با توان دوم α کاهش می‌یابند. این پدیده دلیل اصلی نرخ همگرایی کندر در مدل موضوعی (نرخ $1/\alpha^2$) نسبت به مدل متمرکز است.

- **رژیم محروم‌گی پایین ($\alpha \rightarrow \infty$):** در این حالت، e^α به سرعت رشد می‌کند و کران بی‌نهایت می‌شود. این منطقی است، زیرا وقتی α بسیار بزرگ است، مکانیزم α -LDP تقریباً محدودیتی اعمال نمی‌کند و اجازه می‌دهد توزیع‌های خروجی کاملاً متمایز باشند (عدم انقباض).

این خاصیت انقباضی، ابزار اصلی ما در فصل بعد برای اثبات کران‌های پایین مینی‌مکس خواهد بود، جایی که نشان می‌دهیم تمایز بین پارامترهای مدل حتی با وجود داده‌های زیاد، به دلیل این انقباض دشوار باقی می‌ماند.

۳-۴ نامساوی‌های پردازش داده قوی (SDPI)

در بخش قبل دیدیم که مکانیزم‌های α -LDP باعث کاهش فاصله بین توزیع‌ها می‌شوند. این مفهوم را می‌توان به صورت صوری‌تر در قالب «نامساوی پردازش داده قوی»^۴ بیان کرد. در حالی که نامساوی $D_f(\mathcal{M}(P)\|\mathcal{M}(Q))$ پردازش داده استاندارد (DPI) تنها بیان می‌کند که واگرایی افزایش نمی‌یابد (\leqslant)، نسخه قوی آن تضمین می‌کند که واگرایی اکیداً کاهش می‌یابد.

تعريف ۱-۴ (ضریب انقباض Dobrushin) برای یک کانال یا مکانیزم \mathcal{M} و یک واگرایی f ، ضریب انقباض $(\eta_f(\mathcal{M}))$ به صورت زیر تعریف می‌شود:

$$\eta_f(\mathcal{M}) = \sup_{P \neq Q} \frac{D_f(\mathcal{M}(P)\|\mathcal{M}(Q))}{D_f(P\|Q)} \quad (10-4)$$

اگر $\eta_f(\mathcal{M}) < 1$ باشد، می‌گوییم مکانیزم دارای خاصیت SDPI است.

برای مکانیزم‌های α -LDP، می‌توان نشان داد که این ضریب همواره کمتر از یک است، که نشان‌دهنده اتلاف حتمی اطلاعات در فرآیند خصوصی‌سازی است. با این حال، در تحلیل‌های مینی‌مکس، ما اغلب به کران‌هایی نیاز داریم که واگرایی خروجی را مستقیماً به فاصله تغییرات کل (TV) ورودی مرتبط کنند، زیرا فاصله TV متر طبیعی روی فضای پارامترها در بسیاری از مسائل است.

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

اگرچه واگرایی KL (بخش قبل) ابزار استانداردی در نظریه اطلاعات است، اما کار با واگرایی χ^2 در مسائل α -LDP اغلب ساده‌تر است. دلیل این امر رفتار چندجمله‌ای χ^2 است که برخلاف لگاریتم در KL، با عملیات جمع و انتگرال‌گیری سازگاری بیشتری دارد.

قضیه زیر که برگرفته از تحلیل‌های دوچی و همکاران [۶] است، کران انقباض را برای واگرایی χ^2 بیان می‌کند.

قضیه ۲-۴ (انقباض χ^2 در LDP) فرض کنید \mathcal{M} یک مکانیزم α -LDP باشد. برای هر دو توزیع P_1 و P_2 روی \mathcal{X} ، اگر M_1 و M_2 توزیع‌های خروجی باشند، داریم:

$$D_{\chi^2}(M_1\|M_2) \leq (e^\alpha + 1)(e^\alpha - 1)^{\frac{1}{2}} \|P_1 - P_2\|_{TV}^2 \quad (11-4)$$

⁴Strong Data Processing Inequality (SDPI)

اثبات. اثبات این قضیه نیز بر پایه تحدب مشترک بنا شده است. تابع $D_{\chi^2}(P\|Q) = \int \frac{(dP - dQ)^2}{dQ}$ نسبت به جفت (P, Q) محدب است. مشابه اثبات لم ۱-۴، ماکسیمم انقباض روی نقاط حدی دامنه (توزیع‌های دیراک) رخ می‌دهد. برای دو ورودی x, x' ، و خروجی z ، با استفاده از شرط α -LDP و کران‌دار بودن نسبت چگالی‌ها، می‌توان نشان داد که واریانس نسبت درست‌نمایی کنترل می‌شود. عبارت $(1 - e^\alpha)^2$ ناشی از فاصله L_1 توزیع‌هاست و ضریب $(e^\alpha + 1)$ از نرمال کردن مخرج در تعریف χ^2 حاصل می‌شود. جزئیات کامل جبری این اثبات در [۶] (لم ۳) آمده است. \square

۲-۳-۴ مزیت استفاده از χ^2 نسبت به KL

در ادامه فصل و هنگام اثبات کران‌های پایین برای تخمین چگالی و میانگین، ما عمدتاً از واگرایی χ^2 استفاده خواهیم کرد. دلایل این انتخاب عبارتند از:

۱. رفتار حول صفر: واگرایی KL رفتار نامتقارنی دارد و بسط تیلور آن شامل جملات لگاریتمی است. در مقابل، D_{χ^2} به صورت محلی شبیه یک فرم مربعی (فاصله اقلیدسی وزن‌دار) عمل می‌کند. از آنجا که مکانیزم‌های α -LDP توزیع‌ها را بسیار به هم نزدیک می‌کنند (رژیم محلی)، تقریب مرتبه دوم χ^2 بسیار دقیق و کارآمد است.

۲. سادگی در مخلوط‌ها: در روش‌هایی مانند لم اسود یا فانو، ما نیاز به محاسبه واگرایی بین «مخلوطی از توزیع‌ها» $\bar{P} = \mathbb{E}_{P_\theta}(P)$ داریم. به دلیل ساختار انتگرالی $D_{\chi^2}(P\|Q) = \int \frac{P^2}{Q}$ ، محاسبه کران برای مخلوط‌ها ساده‌تر از KL است که لگاریتم در داخل انتگرال دارد (نامساوی ینسن در KL جهت نامطلوبی دارد، اما در χ^2 می‌توان راحت‌تر کران بالا پیدا کرد).

۳. رابطه با خطای میانگین مربعات: در مسائل تخمین پارامتر، هدف معمولاً کمینه‌سازی خطای MSE است. واگرایی χ^2 ارتباط مستقیمی با واریانس تخمین‌گرها دارد و به طور طبیعی با کران‌های کرامر-رائو و اطلاعات فیشر مرتبط می‌شود.

با در دست داشتن این ابزارها (انقباض KL و χ^2)، اکنون آماده‌ایم تا در بخش‌های بعدی نرخ‌های مینی‌مکس را برای مسائل خاص استخراج کنیم.

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.

۱-۴-۴ تعریف مسئله

فرض کنید داده‌های ورودی $\theta = \mathbb{E}_{X_i} \in \{-1, 1\}^d$ بردارهای تصادفی مستقل با میانگین $X_1, \dots, X_n \in \{-1, 1\}^d$ 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) می‌سنجدیم. هدف یافتن نرخ مینیمکس زیر است:

$$\mathfrak{M}_n(\theta, \mathcal{M}) = \inf_{\mathcal{M}, \hat{\theta}} \sup_{\theta \in [-1, 1]^d} \mathbb{E} \left[\|\hat{\theta}(Z_1, \dots, Z_n) - \theta\|_2^2 \right] \quad (12-4)$$

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

قضیه‌ی ۳-۴ (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:

$$\inf_{\hat{\theta}} \sup_P \mathbb{E} \left[\|\hat{\theta} - \theta(P)\|_2^2 \right] \geq c \cdot \frac{d}{n\alpha^2} \quad (13-4)$$

که $c > 0$ یک ثابت عددی مطلق است.

اثبات. برای اثبات این قضیه، از روش «لم اسود»^۵ (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 $\{-1, 1\}^d = \mathcal{V}$ 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:

$$\mathbb{E}_{P_v}[(X)_j] = v_j \Delta \quad (14-4)$$

^۵ Assouad's Lemma

که $\Delta \in [0, 1)$ 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 بازیزی در هر مؤلفه کران دار می‌شود:

$$\max_{v \in \mathcal{V}} \mathbb{E} \|\hat{\theta} - \theta_v\|_2^2 \geq \frac{d}{2} \cdot \Delta^2 \cdot \min_{v, v': H(v, v')=1} (1 - \|M_v^n - M_{v'}^n\|_{TV}) \quad (15-4)$$

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 کران دار 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_1, \dots, Z_n) :

$$\|M_v^n - M_{v'}^n\|_{TV}^2 \leq \frac{1}{2} D_{KL}(M_v^n \| M_{v'}^n) = \frac{n}{2} D_{KL}(M_v \| M_{v'}) \quad (16-4)$$

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

$$D_{KL}(M_v \| M_{v'}) \leq (e^\alpha - 1)^2 \|P_v - P_{v'}\|_{TV}^2 \quad (17-4)$$

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

$$\|P_v - P_{v'}\|_{TV} = \frac{(1 + \Delta) - (1 - \Delta)}{2} = \Delta \quad (18-4)$$

با ترکیب این روابط و فرض $1 \leq \alpha \leq e^\alpha - 1$ (که نتیجه می‌دهد $\alpha^2 \approx \alpha$):

$$\|M_v^n - M_{v'}^n\|_{TV}^2 \leq \frac{n}{2} \alpha^2 \Delta^2 \quad (19-4)$$

بنابراین:

$$\|M_v^n - M_{v'}^n\|_{TV} \leq \alpha \Delta \sqrt{\frac{n}{2}} \quad (20-4)$$

۴. تنظیم پارامتر Δ : برای این‌که کران پایین در رابطه (۱۵-۴) 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:

$$\alpha\Delta\sqrt{\frac{n}{2}} = \frac{1}{2} \implies \Delta = \frac{1}{\alpha\sqrt{2n}} \quad (21-4)$$

با جایگذاری این مقدار Δ در رابطه لم اسود:

$$\begin{aligned} \sup_{\theta} \mathbb{E}\|\hat{\theta} - \theta\|_2^2 &\geqslant \frac{d}{2} \cdot \Delta^2 \cdot \left(1 - \frac{1}{2}\right) \\ &= \frac{d}{4} \left(\frac{1}{2n\alpha^2}\right) \\ &= \frac{d}{\lambda n\alpha^2} \end{aligned} \quad (22-4)$$

این اثبات نشان می‌دهد که کران پایین از مرتبه $(\frac{d}{n\alpha^2})\Omega$ است. \square

۲-۴-۴ بحث و تفسیر

نتیجه به دست آمده در قضیه ۳-۴ دارای پیام‌های مهمی برای طراحی سیستم‌های خصوصی است:

۱. هزینه محرومانگی: در مقایسه با تخمین میانگین در حالت غیرخصوصی (یا متمرکز) که نرخ خطای $\frac{d}{n}$ است، در حالت α -LDP خطای با ضریب $\frac{1}{\alpha}$ افزایش می‌یابد. این یعنی برای جبران نویز اضافه شده توسط محرومانگی، حجم داده‌ها (n) باید متناسب با مربع بودجه محرومانگی افزایش یابد.

۲. وابستگی به ابعاد: خطای 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 ۳-۶ دیدیم که مکانیزم تخمین میانگین دوچی دقیقاً واریانسی از مرتبه $\frac{d \log d}{n\alpha^2}$ (برای روش‌های تنک) یا $\frac{d}{n\alpha^2}$ 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 محرومانگی نهفته است.

فصل ۵

چارچوب انقباض E_γ -واگرایی

۱-۵ مقدمه

در فصل قبل (۴)، ما از واگرایی‌های کلاسیک مانند کولبک-لایلر (KL) و کای-دو (χ^2) برای تحلیل مکانیزم‌های LDP استفاده کردیم. دیدیم که شرط α -LDP باعث می‌شود که واگرایی بین توزیع‌های خروجی با ضریبی از $(1 - e^\alpha)^2$ محدود شود (بخش ۲-۴). این کران‌ها، اگرچه برای اثبات نرخ‌های همگرایی مجانبی (وقتی $n \rightarrow \infty$ و $\alpha \rightarrow 0$) بسیار مفید هستند، اما دو محدودیت اساسی دارند:

۱. عدم دقیقیت در رژیم‌های میانی: کران‌های مبتنی بر KL معمولاً برای مقادیر بزرگ α (رژیم محروم‌گی پایین) بسیار شل^۱ هستند و رفتار دقیق مکانیزم را بازتاب نمی‌دهند.

۲. عدم همارزی: رابطه بین LDP و واگرایی KL یک طرفه است؛ یعنی LDP انقباض KL را نتیجه می‌دهد، اما انقباض KL لزوماً تضمین‌کننده LDP نیست.

در این فصل، ما به سراغ ابزاری می‌رویم که از نظر هندسی و جبری، «همزبان» با تعریف α -LDP است. بر اساس پژوهش‌های آسوده و همکاران [۲]، نشان می‌دهیم که واگرایی E_γ -واگرایی^۲، ابزار طبیعی برای تحلیل LDP است. هدف اصلی این فصل اثبات این حقیقت است که α -LDP معادل با انقباض کامل (صفر شدن) این واگرایی برای یک مقدار مشخص γ است.

¹Loose

² E_γ -Divergence

۲-۵ همارزی LDP و E_γ -واگرایی

در این بخش، قضیه بنیادی همارزی را بیان می‌کنیم. این قضیه نشان می‌دهد که تعریف LDP که معمولاً به صورت "کران دار بودن نسبت احتمالات" بیان می‌شود، دقیقاً معادل با "صفر شدن E_γ -واگرایی" برای $\gamma = e^\alpha$ است.

قضیه ۱-۵ (همارزی LDP و انتباخت E_γ) یک مکانیزم تصادفی $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{Z}$ در شرط LDP صدق می‌کند اگر و تنها اگر برای هر جفت ورودی $x, x' \in \mathcal{X}$ ، مقدار E_γ -واگرایی بین توزیع‌های خروجی با پارامتر $\gamma = e^\alpha$ برابر صفر باشد. به زبان ریاضی:

$$\mathcal{M} \in \alpha\text{-LDP} \iff \sup_{x, x' \in \mathcal{X}} E_{e^\alpha}(\mathcal{M}(\cdot|x) \| \mathcal{M}(\cdot|x')) = 0 \quad (1-5)$$

اثبات. اثبات این همارزی مستقیماً از تعاریف و ویژگی‌های انتگرال حاصل می‌شود.

جهت اول (\Rightarrow): فرض کنید \mathcal{M} یک مکانیزم α -LDP باشد. طبق تعریف α -LDP (رابطه ۲۸-۲)، برای هر x, x' و هر مجموعه خروجی $S \subseteq \mathcal{Z}$ داریم:

$$\Pr[\mathcal{M}(x) \in S] \leq e^\alpha \Pr[\mathcal{M}(x') \in S] \quad (2-5)$$

اگر $P = \mathcal{M}(\cdot|x')$ و $Q = \mathcal{M}(\cdot|x)$ بازنویسی می‌شود. این نامساوی به صورت $P(S) \leq e^\alpha Q(S)$ باشند، بنابراین:

$$P(S) - e^\alpha Q(S) \leq 0 \quad (3-5)$$

با گرفتن سوپریمم روی تمام مجموعه‌های S (طبق تعریف E_γ در معادله ۲۸-۲):

$$E_{e^\alpha}(P \| Q) = \sup_S (P(S) - e^\alpha Q(S)) \leq 0 \quad (4-5)$$

از آنجا که واگرایی همیشه نامنفی است، نتیجه می‌گیریم $E_{e^\alpha}(P \| Q) = 0$.

جهت دوم (\Leftarrow): فرض کنید برای تمام x, x' داشته باشیم $E_{e^\alpha}(\mathcal{M}(\cdot|x) \| \mathcal{M}(\cdot|x')) = 0$. از تعریف انتگرالی (۲۸-۲) استفاده می‌کنیم (فرض می‌کنیم چگالی‌های p و q وجود دارند):

$$\int_{\mathcal{Z}} (p(z|x) - e^\alpha p(z|x'))^+ \nu(dz) = 0 \quad (5-5)$$

از آنجا که عبارت داخل پرانتز $(\cdot)^+$ همواره نامنفی است، انتگرال آن تنها زمانی صفر می‌شود که خود تابع تقریباً همه‌جا (a.e.) صفر باشد. یعنی برای تمام $z \in \mathcal{Z}$:

$$(p(z|x) - e^\alpha p(z|x'))^+ = 0 \implies p(z|x) - e^\alpha p(z|x') \leq 0 \quad (6-5)$$

که بازآرایی آن منجر به رابطه زیر می‌شود:

$$\frac{p(z|x)}{p(z|x')} \leq e^\alpha \quad (7-5)$$

□ این دقیقاً همان تعریف شرط α -LDP است.

۱-۲-۵ تفسیر بر اساس ضریب انقباض

در ادبیات موضوع [۲]، این نتیجه گاهی با استفاده از مفهوم «ضریب انقباض» بیان می‌شود. ضریب انقباض واگرایی برای مکانیزم M به صورت زیر است:

$$\eta_{E_{e^\alpha}}(M) = \sup_{P \neq Q} \frac{E_{e^\alpha}(MP\|MQ)}{E_{e^\alpha}(P\|Q)} \quad (8-5)$$

قضیه ۱-۵ بیان می‌کند که مکانیزم M خصوصی است اگر صورت این کسر برای هر ورودی دیراک (نقطه‌ای) صفر شود. به بیان دیگر، α -LDP معادل است با وضعیتی که مکانیزم تمامی E_{e^α} -واگرایی موجود در ورودی را «از بین برد» و به صفر برساند. این نگاه، E_γ را از سایر واگرایی‌ها (مانند KL که تنها کاهش می‌یابند اما صفر نمی‌شوند) متمایز می‌کند.

۳-۵ کران‌های انقباض برای f -واگرایی‌های عمومی

با اثبات همارزی میان شرط α -LDP و صفر شدن E_γ -واگرایی (قضیه ۱-۵)، اکنون ابزاری قدرتمند برای بازنویسی کران‌های انقباض در اختیار داریم. استراتژی اصلی در این بخش مبتنی بر این واقعیت است که هر f -واگرایی محدب را می‌توان به صورت انتگرالی از E_γ -واگرایی‌ها نوشت. بنابراین، اگر ما کنترلی روی E_γ داشته باشیم (که به واسطه α -LDP داریم)، می‌توانیم آن را به سایر واگرایی‌ها تعمیم دهیم.

۱-۳-۵ قضیه انقباض بهبودیافته برای χ^2

یکی از مهم‌ترین نتایج این چارچوب جدید، بهبود کران‌های انقباض برای واگرایی کای-دو (χ^2) است. همان‌طور که در فصل ۴ (قضیه ۴-۲) دیدیم، تحلیل‌های کلاسیک دوچی و همکاران کرانی با ضریب $(1 - (e^\alpha + 1)(e^\alpha))$ ارائه می‌دادند. در اینجا، با استفاده از تکنیک‌های مقاله ۲۰۲۴ آسوده و ژانگ [۳]، کران بسیار دقیق‌تری (تنگ‌تر) را بیان می‌کنیم.

قضیه‌ی ۲-۵ (کران دقیق انقباض χ) فرض کنید M یک مکانیزم $LDP-\alpha$ -باشد. برای هر جفت توزیع ورودی P_1 و P_2 ، و اگرایی χ بین توزیع‌های خروجی $M_1 = M(P_1)$ و $M_2 = M(P_2)$ به صورت زیر کران دار می‌شود:

$$D_{\chi^2}(M_1 \| M_2) \leq (e^\alpha - 1)^{\frac{1}{2}} \|P_1 - P_2\|_{TV}^{\frac{1}{2}} \quad (9-5)$$

همچنین، یک کران کلی‌تر برای هر f -واگرایی به صورت زیر برقرار است (قضیه ۳ در [؟]):

$$D_f(M_1 \| M_2) \leq \eta_f(\mathcal{M}) \cdot D_f(P_1 \| P_2) \quad (10-5)$$

که در آن ضریب انقباض برای مکانیزم‌های باینری بهینه برابر است با $\eta_{\chi^2} = \tanh\left(\frac{\alpha}{2}\right)$.

اثبات. اثبات دقیق این قضیه نیازمند تحلیل مقادیر ویژه عملگرهای مارکوف است که در [؟] آمده است. ایده کلی این است که چون $E_{e^\alpha}(M_1 \| M_2) = 0$ است، جرم احتمال در «دنباله» توزیع‌ها (جایی که نسبت p/q بزرگ است) کاملاً حذف می‌شود. با استفاده از نامساوی‌های پردازش داده قوی (SDPI) که مستقیماً از همارزی E_γ استخراج شده‌اند، می‌توان نشان داد که ضریب $(e^\alpha + 1)$ که در تحلیل‌های قبلی ظاهر می‌شد، ناشی از تقریب‌های نادقيق بوده و قابل حذف است. \square

۲-۳-۵ مقایسه با کران‌های دوچی (فصل ۴)

برای درک اهمیت قضیه ۲-۵، بیایید ضریب به دست آمده را با نتیجه دوچی (رابطه ۱۱-۴ در فصل ۴) مقایسه کنیم.

• کران دوچی (Duchi et al. 2013) :

$$D_{\chi^2}(M_1 \| M_2) \leq \underbrace{(e^\alpha + 1)(e^\alpha - 1)^{\frac{1}{2}}}_{\approx e^{\frac{3}{2}\alpha}} \|P_1 - P_2\|_{TV}^{\frac{1}{2}}$$

این کران در رژیم محروم‌نگی پایین (α بزرگ) با سرعت $e^{3\alpha}$ رشد می‌کند.

• کران بهبودیافته (Asoodeh et al. 2024) :

$$D_{\chi^2}(M_1 \| M_2) \leq \underbrace{(e^\alpha - 1)^{\frac{1}{2}}}_{\approx e^{\frac{1}{2}\alpha}} \|P_1 - P_2\|_{TV}^{\frac{1}{2}}$$

این کران با سرعت $e^{2\alpha}$ رشد می‌کند.

تحلیل بهبود: تفاوت این دو کران در ضریب $(1 + e^\alpha)$ است.

۱. در رژیم $\alpha \rightarrow 0$ (محرمانگی بالا)، $1 + e^\alpha \rightarrow 1$ ، بنابراین کران جدید تا ضریب ۲ بهتر است (که در تحلیل‌های مجانبی تأثیر کمی دارد).

۲. اما در رژیم $\alpha > 0$ (که در کاربردهای عملی بسیار رایج است)، حذف عامل نمایی e^α از ضریب، بهبودی چشمگیر است. این نشان می‌دهد که تحلیل‌های مبتنی بر E_γ ، رفتار مکانیزم را در تمام بازه‌های α بسیار دقیق‌تر توصیف می‌کنند و از برآوردهای بیش از حد بدینانه (Over-pessimistic) جلوگیری می‌کنند.

این کران‌های دقیق‌تر، مستقیماً منجر به استخراج نرخ‌های مینی‌مکس دقیق‌تر در مسائل تخمین توزیع و آزمون فرض می‌شوند که در ادامه فصل به آن‌ها خواهیم پرداخت.

۴-۵ کاربرد در نرخ‌های مینی‌مکس و مقایسه نهایی

در بخش‌های پیشین، ما ابزارهای قدرتمندی را برای تحلیل مکانیزم‌های α -LDP توسعه دادیم. اکنون زمان آن رسیده است که این ابزارها را در عمل به کار بگیریم. در این بخش، نشان می‌دهیم که چگونه استفاده از چارچوب انقباض E_γ و کران‌های دقیق χ ، منجر به اثبات نرخ‌های مینی‌مکس دقیق‌تر (تنگ‌تر) نسبت به روش‌های کلاسیک فصل ۴ می‌شود.

۴-۵-۱ بازبینی مسئله تخمین میانگین

مسئله تخمین میانگین^d $[-1, 1] \in \theta$ را که در بخش ۴-۴ معرفی شد، در نظر بگیرید. در آن‌جا با استفاده از لم آسواد و کران انقباض کلاسیک دوچی، به نرخ زیر رسیدیم:

$$M_n(\alpha\text{-LDP}) \geq c \cdot \frac{d}{n\alpha^2} \quad (11-5)$$

نکته کلیدی در اثبات آن بخش، استفاده از کران انقباضی بود که شامل ضریب $(1 + e^\alpha)$ می‌شد. حال باید بینیم با ابزارهای جدید چه تغییری در این تحلیل ایجاد می‌شود.

طبق روش لم آسواد (رابطه ۱۵-۴)، بخش اصلی اثبات، کران دار کردن مجموع احتمال‌های خطأ در تشخیص بین دو توزیع همسایه P_v و $P_{v'}$ است. دیدیم که این خطأ با فاصله تغییرات کل توزیع‌های خروجی

که $\|M_v^n - M_{v'}^n\|_{TV}$ مرتبط است. با استفاده از نامساوی‌های پردازش داده قوی (SDPI) و کران دقیق χ^2 در قضیه ۲-۵ ثابت کردیم، داریم:

$$\begin{aligned}\|M_v - M_{v'}\|_{TV}^2 &\leq \frac{1}{4} D_{\chi^2}(M_v \| M_{v'}) \\ &\leq \frac{1}{4} \eta_{\chi^2}(\mathcal{M}) \cdot D_{\chi^2}(P_v \| P_{v'})\end{aligned}\quad (12-5)$$

که در آن $\eta_{\chi^2}(\mathcal{M})$ ضریب انقباض دقیق مکانیزم است. بر اساس نتایج آسوده و همکاران [؟]، برای مکانیزم‌های بهینه باینری، این ضریب دقیقاً برابر است با:

$$\eta_{\chi^2}^*(\alpha) = \tanh^2\left(\frac{\alpha}{2}\right) = \left(\frac{e^{\alpha/2} - e^{-\alpha/2}}{e^{\alpha/2} + e^{-\alpha/2}}\right)^2 = \left(\frac{e^\alpha - 1}{e^\alpha + 1}\right)^2 \quad (13-5)$$

این نتیجه بسیار دقیق‌تر از کران‌های پیشین است. باید رفتار این ضریب را در دو رژیم حدی بررسی کنیم:

۱. رژیم محramانگی بالا ($\alpha \rightarrow 0^+$):

$$\eta_{\chi^2}^*(\alpha) \approx \left(\frac{\alpha}{2}\right)^2 = \frac{\alpha^2}{4}$$

این نتیجه با تحلیل‌های فصل ۴ (که α^2 بود) سازگار است. این نشان می‌دهد که تحلیل‌های دوچی در رژیم α کوچک، مجانباً بهینه بوده‌اند.

۲. رژیم محramانگی پایین ($\alpha \rightarrow \infty$): در این حالت، $1 \rightarrow \eta_{\chi^2}^*(\alpha)$. این رفتار بسیار منطقی است؛ وقتی محramانگی وجود ندارد، انقباضی رخ نمی‌دهد. اما کران‌های کلاسیک فصل ۴ در این حالت به بی‌نهایت میل می‌کردند ($e^{3\alpha} \approx$) که از نظر فیزیکی بی‌معنی بود. چارچوب جدید به درستی پیش‌بینی می‌کند که با افزایش بودجه محramانگی، اطلاعات از دست رفته به صفر میل می‌کند.

با جایگذاری این ضریب دقیق در لم آسوده، نرخ مینی‌مکس به صورت دقیق‌تری بیان می‌شود:

$$\mathfrak{M}_n \geq \frac{d}{n \cdot \tanh^2(\alpha/2)} \quad (14-5)$$

این فرمول نه تنها برای α ‌های کوچک، بلکه برای تمام طیف پارامترهای محramانگی معتبر است.

۵-۵ نتیجه‌گیری فصل

در این فصل، ما یک گام بنیادین فراتر از تحلیل‌های استاندارد LDP برداشتیم. هدف ما گذار از «تقریب‌های جبری» به «تعاریف هندسی دقیق» بود. دستاوردهای اصلی این فصل را می‌توان در سه محور خلاصه کرد:

• همارزی ذاتی: ما نشان دادیم که LDP صرفاً یک محدودیت روی نسبت‌های احتمال نیست، بلکه دقیقاً معادل با صفر شدن E_γ -واگرایی (برای $e^\alpha = \gamma$) است. این یعنی E_γ ، متريک طبیعی و ذاتی فضای α -LDP است، همان‌طور که فاصله اقلیدسی متريک طبیعی فضای برداری \mathbb{R}^n است.

• انقباض دقیق: با استفاده از تجزیه طیفی f -واگرایی‌ها بر حسب E_γ ، توانستیم نشان دهیم که مکانیزم‌های خصوصی چقدر اطلاعات را «منقبض» می‌کنند. ما ثابت کردیم که بسیاری از ضرایب نمایی e^α که در مقالات پیشین (مانند [۶]) وجود داشتند، ناشی از ابزار نامناسب تحلیل (استفاده از KL به جای E_γ) بوده‌اند و قابل حذف هستند.

• وحدت رویه: این چارچوب جدید، یک روش واحد برای تحلیل انواع مسائل (تخمین، آزمون فرض، و یادگیری) فراهم می‌کند. به جای استفاده از تکنیک‌های مجزا برای هر واگرایی، ما تنها رفتار E_γ را بررسی می‌کنیم و نتایج را به سایر معیارها تعمیم می‌دهیم.

به این ترتیب، فصل ۵ نه تنها نتایج فصل‌های قبل را تأیید می‌کند، بلکه پایه‌ای مستحکم برای طراحی مکانیزم‌های بهینه در آینده فراهم می‌آورد. این دیدگاه هندسی به ما می‌گوید که یک مکانیزم α -LDP بهینه، مکانیزمی است که دقیقاً روی مرز انقباض E_{e^α} حرکت کند.

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

ز

ب

س

پ

ش

پرس‌وجو Query
پایگاه‌داده Database

ص

ت

غ

ج

ف

ح

ق

خ

ک

د

گ

داده Data
دودوبی Binary

ل

ر

م

مجموعه واگرایی Set

متصدی مورد اعتماد Trusted Curator

مکانیزم تصادفی Randomized Mechanism

محترمانگی تفاضلی Differential Privacy

مطلقاً پیوسته Absolutely Continuous

و

Divergence واگرایی

هـ

Ajacent همسایه

Adjacency همسایگی

ن

ی

پیوست آ

مطالب تكميلی

Abstract

We present a standard template for typesetting theses in Persian. The template is based on the XE_PA_HPersian package for the L_AT_EX typesetting system. This write-up shows a sample usage of this template.

Keywords: Thesis, Typesetting, Template, XE_PA_HPersian



Sharif University of Technology

Department of Mathematics

M.Sc. Thesis

Information-Theoretic Analysis of Local Differential Privacy and its Statistical Applications

By:

Firoozeh Abrishami

Supervisor:

Dr. Javad Ebrahimi Boroujeni

March 2026