



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

کلیدواژه‌ها:

اول

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 چالش سودمندی و موازنۀ دقت-محرمانگی	۱۲
۱۳	۲-۱-۳ نگاهی آماری به LDP: چارچوب مینیماکس و حدود بنيادین	۱۳
۱۴	۲-۱-۴ دسته‌بندی پروتکل‌های موضعی: تعاملی و غیرتعاملی	۱۴
۱۵	۳-۱ بیان مسئله و اهداف پژوهش	۱۵
۱۶	۳-۱-۱ رویکرد تحلیل: f-واگرایی‌ها به عنوان زبان مشترک	۱۶
۱۶	۳-۱-۲ اهداف و ساختار پژوهش	۱۶
۱۷	۴-۱ ساختار پایان‌نامه	۱۷
۱۹	۲ پیش‌نیازها	۲
۱۹	۱-۱ محرمانگی تفاضلی مرکز (CDP)	۱۹
۱۹	۱-۱-۱ مدل اعتماد و تعریف رسمی	۱۹
۲۲	۱-۱-۲ تفسیر پارامترهای محرمانگی	۲۲

۲۲	۳-۱-۲ تعاریف معادل و صورت بندی‌های جایگزین
۲۲	۴-۱-۲ مکانیزم‌های پایه
۲۵	۵-۱-۲ مکانیزم‌های بنیادی محرمانگی تفاضلی
۲۹	۶-۱-۲ ترکیب‌پذیری(Composition)
۳۰	۷-۱-۲ محرمانگی گروهی(Group Privacy)
۳۱	۸-۱-۲ محدودیت مدل متمرکز
۳۱	۲-۲ محرمانگی تفاضلی موضعی
۳۱	۱-۲-۲ مقدمه و گذار از مدل متمرکز
۳۲	۲-۲-۲ تعاریف رسمی و مدل‌های محاسباتی
۳۴	۳-۲-۲ مکانیزم‌های پایه در LDP
۳۷	۴-۲-۲ چالش سودمندی در مدل موضعی
۳۸	f	۳-۲-۲ واگرایی‌ها
۳۸	۱-۳-۲ تعریف f-واگرایی
۳۸	۲-۳-۲ نمونه‌های مهم f-واگرایی
۳۹	۳-۳-۲ ارتباط f-واگرایی‌ها با یک دیگر
۴۰	۴-۲ مبانی آماری و نظریه اطلاعات
۴۰	۱-۴-۲ معیارهای فاصله اطلاعاتی
۴۰	۲-۴-۲ ریسک مینیماکس
۴۱	۵-۲ آزمون فرض آماری و روش تقلیل
۴۱	۱-۵-۲ آزمون فرض دودویی
۴۲	۲-۵-۲ تقلیل تخمین به آزمون (روش بسته‌بندی)
۴۳	۳-۵-۲ نامساوی‌های کران پایین
۴۴	۳ تحلیل‌های مبتنی بر انقباض و نرخ‌های مینیماکس
۴۴	۱-۳ مقدمه

۴۴	۲-۳ محرمانگی به عنوان انقباض اطلاعاتی
۴۶	۱-۲-۳ انقباض در فاصله واریانس کل
۴۶	۳-۳ تحلیل نرخ‌های مینیماکس با استفاده از انقباض
۴۷	۴-۳ مطالعه موردی: تخمین میانگین
۴۷	۵-۳ محدودیت‌های تحلیل کلاسیک
۴۹	۴ همارزی LDP و انقباض E_γ-واگرایی
۴۹	۱-۴ مقدمه و انگیزه
۵۰	۲-۴ معرفی E_γ -واگرایی
۵۰	۱-۲-۴ خواص هندسی
۵۰	۳-۴ قضیه همارزی اصلی
۵۲	۴-۴ بهبود کران‌های انقباض
۵۲	۵-۴ تعمیم به محرمانگی تقریبی $((\alpha, \delta)\text{-LDP})$
۵۳	۶-۴ کاربرد در تخمین توزیع گستته
۵۳	۷-۴ انقباض قوی برای خانواده‌ی f -واگرایی‌ها
۵۳	۱-۷-۴ کران دقیق برای واگرایی کای-دو (χ^2)
۵۴	۲-۷-۴ تعمیم به سایر واگرایی‌ها
۵۵	۴-۴ نامساوی ون‌تریز خصوصی (Private van Trees Inequality)
۵۵	۹-۴ کاربردهای نوین و بهبود نرخ‌ها
۵۷	۵ نتیجه‌گیری
۵۸	مراجع
۶۰	واژه‌نامه
۶۲	آ مطالب تکمیلی

فهرست جداول

فهرست تصاویر

۲-۱	مدل محramانگی تفاضلی متمرکز با یک متصدی مورد اعتماد.	۲۰
۲-۲	مدل محramانگی تفاضلی موضعی (α -LDP). نویز به صورت موضعی روی دستگاه کاربر اضافه می شود.	۳۲

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 [۱۶].

• **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 محتواهای 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 [۵].

این شواهد تجربی و نظری نشان می‌دهند که تعاریف هیوریستیک و روش‌های موردنی (مانند حذف ستون‌ها یا مخدوش‌سازی ساده) نمی‌توانند تضمین امنیتی پایداری ارائه دهنند. 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 [۱۱]. 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 پایگاهداده‌ی همسایه‌ی \mathcal{D} و \mathcal{D}' چنان باشند که $f(\mathcal{D}') \neq f(\mathcal{D})$ ، آن‌گاه تکیه‌گاه^{۲۵} توزیع‌های خروجی کاملاً مجزا خواهد بود. در نتیجه، واگرایی کولبک-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 [۱۲]، برای شکستن این وابستگی قطعی و ایجاد ابهام آماری، تصادفی‌سازی^{۲۷} در فرآیند مکانیزم الزامی است.

کاربردهای محرمانگی تفاضلی

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

- **تخمین میانگین و مجموع^{۲۸}**: اساسی‌ترین کاربرد DP در محاسبه‌ی آماره‌های توصیفی است. سازمان‌های آماری (مانند اداره سرشماری آمریکا) از این روش برای انتشار میانگین درآمد، سن یا جمعیت مناطق استفاده می‌کنند، بدون آنکه داده‌های فردی شهروندان به خطر 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^[۳۴].

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^{۳۵} 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 کنجهکاو 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.

۱-۲ کارهای پیشین و مرور ادبیات

۱-۲-۱ آغازگرها: 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 [۱۱]. 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 پایگاه داده 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:

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

³⁵Statistical Disclosure Control

³⁶Differential Privacy

که در آن ϵ پارامتری کلیدی به نام بودجه محرمانگی^{۳۷} است و δ احتمال شکست ناچیز مکانیزم را نشان می‌دهد [۱۲]. برای درک شهودی این مفهوم، می‌توان ϵ را به عنوان یک «پیچ تنظیم» برای کنترل توازن میان امنیت و مطلوبیت داده‌ها در نظر گرفت. این پارامتر تعیین می‌کند که خروجی مکانیزم تا چه حد اجازه دارد بین دو جهان موازی (جهانی با حضور داده‌ی شما و جهانی بدون آن) تمایز قائل شود:

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

همان‌طور که در بخش قبل توضیح داده شد، پیاده‌سازی این تعریف نیازمند یک پیش‌فرض قوی است: وجود یک متصدی مورد اعتماد که تمام داده‌های خام را جمع‌آوری کرده و نویز را به صورت مرکزی اعمال کند. اما دیدیم که این مدل دارای نقطه ضعف‌هایی است. حذف این فرض و انتقال اعتماد از سرور به کاربر، منجر به شکل‌گیری مفهوم محرمانگی تفاضلی موضعی^{۳۸} (LDP) شد. اگرچه اصطلاح LDP و صورت‌بندی مدرن آن در سال‌های اخیر توسط پژوهشگرانی نظری کاسیسی‌سوانatan و دیگران تدوین شد [۱۴]، اما ریشه‌های عملی آن به دهه‌ها قبل باز می‌گردد.

تعریف ۱-۱ (محرمانگی تفاضلی موضعی (α, δ) -LDP) یک مکانیزم تصادفی $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{Z}$ (تصادفی‌ساز موضعی) شرط «محرمانگی تفاضلی موضعی تقریبی» یا (α, δ) -LDP را برآورده می‌کند، اگر برای تمام جفت ورودی‌های ممکن $x, x' \in \mathcal{X}$ و هر زیرمجموعه‌ی خروجی $\mathcal{S} \subseteq \mathcal{Z}$ ، رابطه زیربرقرار 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 توزیع‌های خروجی را کنترل می‌کند.
- δ ، احتمال ناچیز شکست مکانیزم در برقراری شرط محرمانگی است.

³⁷Privacy Budget

³⁸Local Differential Privacy

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

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 [۱۹]. 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 n 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 تخمین بزند. به زبان ریاضی مدرن، اگر احتمال گزارش پاسخ واقعی p باشد، نسبت احتمال خروجی‌ها برای دو ورودی متفاوت x و x' به صورت زیر محدود می‌شود:

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

این رابطه دقیقاً منطبق بر تعریف α -LDP است و نشان می‌دهد که بودجه محروم‌انگی 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:

³⁹Randomized Response (RR)

⁴⁰Unfair

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

- اگر $1 \rightarrow p$ (پاسخ تقریباً همیشه راست)، آنگاه $\infty \rightarrow \alpha$ می‌شود. یعنی داده‌ها دقیق هستند اما هیچ محرمانگی 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 محرمانگی در داده‌های دودویی است، و در کاربردهای مدرن با چالش دامنه‌ی بسیار بزرگ^{۴۱} روبروست. شرکت‌های بزرگ فناوری نیاز دارند داده‌هایی نظری «آدرس‌های اینترنتی بازدید شده» یا «کلمات جدید تایپ شده» را جمع‌آوری کنند که دامنه‌ی آنها^{۴۲} (\mathcal{X}) می‌تواند شامل میلیون‌ها حالت باشد. اعمال مستقیم RR در این حالات منجر به نویز بسیار زیاد و کاهش شدید سودمندی می‌شود. در ادامه، راهکارهای اتخاذ شده توسط بزرگ‌ترین شرکت‌های فناوری را مورر می‌کنیم:

- گوگل و پروتکل RAPPOR:** در سال ۲۰۱۴، گوگل برای جمع‌آوری آمار تنظیمات مرورگر کروم و شناسایی بدافزارها، پروتکل RAPPOR^{۴۳} را معرفی کرد [۱۸]. چالش اصلی گوگل، جمع‌آوری رشته‌های متنه^{۴۴} بود. راه حل آن‌ها ترکیب پاسخ تصادفی با فیلترهای بلوم^{۴۵} بود. در این روش، داده‌ی ورودی ابتدا به یک بردار بیتی (با استفاده از توابع درهم‌سان) نگاشت می‌شود و سپس پاسخ تصادفی روی تک‌تک بیت‌های این فیلتر اعمال می‌گردد. این معماری به گوگل اجازه داد تا بدون دانستن ورودی دقیق هر کاربر، الگوهای پرتکرار و ناهنجاری‌ها را در مقیاس میلیونی شناسایی کند.

- اپل و جمع‌آوری داده‌های دایره‌لغات:** شرکت اپل از LDP برای بهبود کیبورد QuickType، شناسایی ایموجی‌های پرطرفدار و داده‌های سلامت در سیستم‌عامل‌های iOS و macOS استفاده می‌کند. مسئله‌ی اپل، مخابره‌ی کارآمد داده‌ها با حفظ حریم خصوصی بود. راه حل اپل استفاده از تکنیک‌های مبتنی بر طرح ریزی^{۴۶} و تبدیل‌های ریاضی مانند تبدیل هادامارد^{۴۷} 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 [۱۸].

⁴¹High-Dimensional Domain

⁴²Randomized Aggregatable Privacy-Preserving Ordinal Response

⁴³String

⁴⁴Bloom Filters

⁴⁵Sketching

⁴⁶Hadamard Transform

• مایکروسافت و داده‌های تله‌متري: مایکروسافت برای جمع‌آوری داده‌های تله‌متري ویندوز (مانند مدت زمان استفاده از برنامه‌ها) با چالش تخمين هیستوگرام‌های پیوسته روبرو بود. آن‌ها از مکانیزم‌هایی نظیر نمونه‌برداری هیستوگرام و روش‌های تکرارکننده برای بازسازی توزیع داده‌ها استفاده کردند. تمرکز اصلی در این‌جا، ایجاد تعادل بین دقت آماری در جمع‌آوری داده‌های سیستمی و عدم امکان بازناسایی رفتار یک کاربر خاص در طول زمان است.

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

۲-۲-۱ چالش سودمندی و موازنۀ دقت-محترمانگی

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^{۴۷} داده‌ها یا همان دقت تحلیل‌های آماری است. این پدیده تحت عنوان موازنۀ محترمانگی-دقت^{۴۸} شناخته می‌شود.

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

• در مدل متمرکز (CDP): نویز تنها «یک‌بار» و پس از تجمع داده‌ها به نتیجه نهایی اضافه می‌شود. از آن‌جا که مجموع (یا میانگین) داده‌ها حساسیت کمی دارد، مقدار نویز معمولاً^{۴۹} 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 انباست نویز باعث می‌شود که نسبت سیگنال به نویز^{۴۹} 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 تخمين فراوانی یک بیماری نادر باشد، نویز اضافه شده توسط مکانیزم‌های LDP ممکن است سیگنال اصلی را کاملاً بپوشاند.

⁴⁷Utility

⁴⁸Privacy-Accuracy Trade-off

⁴⁹Signal-to-Noise Ratio (SNR)

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 مینیماکس» فراهم کرد.

۳-۲-۱ نگاهی آماری به LDP: چارچوب مینیماکس و حدود بنیادین

پاسخ به پرسش بالا، مسیر پژوهش‌های این حوزه را به سمت نظریه مینیماکس آماری^{۵۰} تغییر داد. نقطه عطف این تحول، سلسله مقالات جریان‌ساز دوچی، جردن و وین‌رایت^{۵۱} [۹، ۱۰] بود. آن‌ها با صورتی بندی مسئله در قالب نظریه اطلاعات، نشان دادند که هزینه محرومگی در مدل موضعی بسیار سنگین و غیرقابل اجتناب است.

در تحلیل مینیماکس، هدف یافتن «ریسک مینیماکس» (\mathfrak{M}_n) است؛ یعنی کمترین خطایی که «بهترین تخمین‌گر ممکن» در «بدترین توزیع داده‌ی ممکن» مرتكب می‌شود. دوچی و همکاران با استفاده از ابزارهایی نظری نامساوی فانو^{۵۲} و لم اسود^{۵۳} (که در فصل سوم به تفصیل بررسی خواهند شد)، ثابت کردند که برای مسائل پایه‌ای نظری تخمین میانگین یا چگالی احتمال، نرخ هم‌گرایی خطا در مدل LDP رفتاری متفاوت با مدل متمرکز دارد.

به طور مشخص، برای n کاربر و بودجه محرومگی α ، کران پایین خطا (\mathcal{E}) به صورت مجانی از رابطه‌ی زیر پیروی می‌کند:

$$\mathcal{E}_{LDP} \asymp \frac{1}{\sqrt{n\alpha^2}} \quad \text{در حالی که} \quad \mathcal{E}_{CDP} \asymp \frac{1}{n\varepsilon}$$

این نتیجه که به «قانون مقیاس‌دهی کانونی»^{۵۴} معروف است، 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 کندر $1/\sqrt{n}$ 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

⁵⁰Statistical Minimax Theory

⁵¹Duchi, Jordan, and Wainwright

⁵²Fano's Inequality

⁵³Assouad's Lemma

⁵⁴Canonical Scaling Law

حجم داده‌ها باید با ضریب $\alpha^2 / 1$ افزایش یابد.

پس از استقرار این چارچوب نظری، تمرکز جامعه علمی بر طراحی «mekanizm‌های بهینه ترتیب مقدماتی»⁵⁵ قرار گرفت که بتوانند به این کران‌های نظری دست یابند. از جمله مهم‌ترین این تلاش‌ها می‌توان به معرفی «mekanizm‌های پله‌ای»⁵⁶ توسط کایروز و همکاران [۱۳] و توسعه پروتکل‌های پیشرفت‌های نظری UE (کدگذاری یگانی)⁵⁷ و OLH (هشینگ محلی بهینه)⁵⁸ توسط وانگ و همکاران [۱۸] اشاره کرد. این روش‌ها تلاش می‌کنند با بهینه‌سازی ساختار نویز و استفاده از تکنیک‌های فشرده‌سازی اطلاعات، فاصله بین عملکرد عملی و حدود نظری مینیماکس را به حداقل برسانند.

۴-۲-۱ دسته‌بندی پروتکل‌های موضوعی: تعاملی و غیرتعاملی

به عنوان آخرین مبحث در مرور ادبیات موضوع، لازم است به دسته‌بندی پروتکل‌های α -LDP بر اساس «معماری ارتباطی» اشاره کنیم. پژوهش‌های انجام شده در این حوزه، مکانیزم‌ها را به دو دسته‌ی کلی تقسیم می‌کنند:

۱. پروتکل‌های غیرتعاملی⁵⁹: در این حالت، تمام کاربران به صورت همزمان و مستقل عمل می‌کنند. هر کاربر i مکانیزم M را تنها بر اساس داده‌ی خودش X_i اجرا کرده و پیام Z_i را به سرور می‌فرستد. هیچ ارتباطی بین کاربران وجود ندارد و سرور نیز هیچ بازخوردی به کاربران نمی‌دهد. به دلیل سادگی پیاده‌سازی و مقیاس‌پذیری بالا، اکثر پروتکل‌های صنعتی (مانند RAPPOR گوگل یا سیستم‌های اپل) در این دسته قرار می‌گیرند.

۲. پروتکل‌های تعاملی⁶⁰: در این روش، کاربران به صورت متوالی با سرور ارتباط برقرار می‌کنند. کاربر i می‌تواند قبل از ارسال داده‌ی خود، خلاصه‌ای از داده‌های کاربران قبلی (Z_1, \dots, Z_{i-1}) را از سرور دریافت کند و نویز خود را هوشمندانه‌تر تنظیم نماید. اگرچه به نظر می‌رسد این آزادی عمل باید دقیق 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 مینیماکس را بهبود نمی‌بخشند [۱۳, ۹].

⁵⁵Order-optimal Mechanisms

⁵⁶Staircase Mechanisms

⁵⁷Unary Encoding

⁵⁸Optimal Local Hashing

⁵⁹Non-interactive / Simultaneous

⁶⁰Interactive / Sequential

در این پایان نامه، چارچوب نظری ارائه شده به گونه‌ای است که نتایج (بهویژه کران‌های پایین مینیماکس) برای هر دو کلاس پروتکل‌های تعاملی و غیرتعاملی صادق هستند. ما نشان خواهیم داد که محدودیت‌های ذاتی مدل LDP، مستقل از معماری ارتباطی شبکه عمل می‌کنند و افزودن تعامل، لزوماً راه گریزی از این محدودیت‌های بنیادین فراهم نمی‌کند. ابزارهای ریاضیاتی قدرتمندی که امکان چنین تحلیل یک‌پارچه‌ای را فراهم می‌سازند، در بخش بعدی معرفی خواهند شد.

۳-۱ بیان مسئله و اهداف پژوهش

محرمانگی تفاضلی موضعی (LDP) به عنوان یک حوزه‌ی میان‌رشته‌ای، محل تلاقی «علوم کامپیوتر» (با تمرکز بر طراحی الگوریتم و امنیت) و «آمار ریاضی» (با تمرکز بر نظریه تخمین و مینیماکس) است. این 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 مینیماکس که توسط دوچی و همکاران [۹] پایه‌گذاری شده، نشان می‌دهد که اعمال محدودیت محرمانگی منجر به کاهش نرخ هم‌گرایی در تخمین‌های آماری می‌شود. با این حال، اثبات این نتایج در مقالات مختلف با ابزارهای متفاوتی صورت گرفته است که منجر به نوعی «اثر برج بابل» در ادبیات علمی شده است. برای مثال:

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

این تنوع در ابزارها، درک هندسی و شهودی از ماهیت «اتلاف اطلاعات» را برای پژوهشگران دشوار می‌سازد. پژوهشگری که قصد ورود به این حوزه را دارد، با مجموعه‌ای از تکنیک‌های اثباتی مجزا روبرو می‌شود که ارتباط درونی آن‌ها شفاف نیست. مسئله‌ی اصلی که این پایان‌نامه بر آن تمرکز دارد، فقدان یک چارچوب یک‌پارچه و منسجم در ادبیات (بهویژه منابع فارسی) است که بتواند این ابزارهای به‌ظاهر متفاوت را زیر یک چتر واحد گردآوری و تحلیل کند.

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^{۶۱} بازنویسی و تفسیر شوند. f -واگرایی‌ها (معرفی شده توسط سیسار^{۶۲})، کلاسی جامع از معیارهای فاصله بین دو توزیع احتمال P و Q هستند:

$$D_f(P\|Q) = \mathbb{E}_Q \left[f \left(\frac{dP}{dQ} \right) \right] \quad (5-1)$$

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

انتخاب این رویکرد به ما اجازه می‌دهد تا «محرمانگی» را نه صرفاً به عنوان یک ویژگی الگوریتمی، بلکه به عنوان یک محدودیت هندسی تفسیر کنیم. در این دیدگاه، هر مکانیزم LDP مانند یک «کanal انقباضی»^{۶۳} عمل می‌کند که فاصله‌ی بین توزیع‌های ورودی را فشرده می‌سازد. هدف ما در این پژوهش، بررسی این است که چگونه ادبیات پیشرو، مفهوم «ضریب انقباض»^{۶۴} را برای f -واگرایی‌های مختلف محاسبه کرده و از آن برای استخراج کران‌های مینیماکس استفاده می‌کنند. این دیدگاه هندسی، پلی میان اثبات‌های پراکنده ایجاد می‌کند و نشان می‌دهد که انتخاب f مناسب، تابعی از هندسه‌ی مسئله است.

۱-۳-۲ اهداف و ساختار پژوهش

هدف نهایی این پایان‌نامه، ارائه‌ی یک «بازخوانی تحلیلی و آموزشی» از نظریه مینیماکس تحت محدودیت محرمانگی است تا شکاف میان مفاهیم علوم کامپیوتر و آمار پر شود. اهداف مشخص این پژوهش عبارتند از:

۱. یکپارچه‌سازی مبانی نظری: گردآوری و بازتعریف قضایای بنیادین (مانند نامساوی‌های پردازش داده قوی) با استفاده از نمادگذاری واحد و چارچوب f -واگرایی، به‌گونه‌ای که خواننده بتواند ارتباط ریاضی بین معیارهای مختلف (KL, TV, χ^2) را به وضوح مشاهده کند.

۲. تبیین روش‌های کران‌پایین: تشریح دقیق و گام‌به‌گام متدهای کلاسیک آماری نظری «متد دو نقطه‌ای لوکام»^{۶۵} و «نامساوی فانو»^{۶۶} در بستر محرمانگی. هدف این است که نشان دهیم چگونه می‌توان

⁶¹ f -divergences

⁶² Csiszár

⁶³ Contraction Channel

⁶⁴ Contraction Coefficient

⁶⁵ Le Cam's Method

⁶⁶ Fano's Inequality

مسئله‌ی پیچیده‌ی «تخمین» را به مسئله‌ی ساده‌تر «آزمون فرض» تقلیل داد و از هندسه‌ی f -واگرایی برای حل آن استفاده کرد.

۳. تفسیر هندسی رفتار مکانیزم‌ها: تحلیل «رفتار انقباضی»^{۶۷} مکانیزم‌ها با هدف محاسبه‌ی ضرباب انقباض برای f -واگرایی‌های مختلف. ما نشان می‌دهیم که چگونه مکانیزم‌های محربمانگی (مانند پاسخ تصادفی) باعث کاهش فاصله بین توزیع‌ها می‌شوند و این کاهش چگونه به طور مستقیم بر خطای تخمین تأثیر می‌گذارد.

۴. مرور تحلیلی و بازخوانی ادبیات موضوع: انجام مروری جامع و ساختاریافته بر کارهای انجام شده در حوزه‌ی تخمین مینیماکس تحت محدودیت α -LDP و تحلیل دقیق نتایج موجود با استفاده از ابزار f -واگرایی‌ها، به منظور یکپارچه‌سازی اثبات‌های پراکنده و شفاف‌سازی مسیرهای استدلالی برای پژوهشگران این حوزه.

۴-۱ ساختار پایان‌نامه

ساختار ادامه‌ی این پایان‌نامه به شرح زیر سازمان‌دهی شده است:

• فصل دوم: مفاهیم اولیه و تعاریف در این فصل، بستر ریاضیاتی پژوهش بنا می‌شود. ابتدا چارچوب‌های محربمانگی تفاضلی مت مرکز (CDP) و موضعی (LDP) به صورت دقیق تعریف شده و مکانیزم‌های پایه‌ی این حوزه (نظیر پاسخ تصادفی وارنر و مکانیزم لاپلاس) به همراه چالش‌های سودمندی آن‌ها بررسی می‌گرددند. سپس خانواده‌ی f -واگرایی‌ها و در بخش آخر مفاهیم بنیادین مورد نیاز آماری معرفی می‌شوند.

• فصل سوم: نرخ‌های مینیماکس و حدود پایین این فصل هسته‌ی اصلی تحلیل‌های آماری پایان‌نامه را تشکیل می‌دهد. در این بخش، چارچوب تخمین مینیماکس معرفی شده و با استفاده از ابزار f -واگرایی‌ها، متدهای کلاسیک (نظیر روش لوکام و نامساوی فانو) بازنویسی می‌شوند. هدف نهایی این فصل، اثبات کران‌های پایین برای خطای تخمین در مسائل مختلف تحت قید α -LDP با استفاده از نامساوی‌های پردازش داده است.

• فصل چهارم: همارزی LDP با انقباض و کاربردهای آماری f -واگرایی‌ها در این فصل، روابط و همارزی‌های موجود میان f -واگرایی‌های مختلف (مانند χ^2 و KL) در بستر محربمانگی موضعی

⁶⁷Contraction Behavior

بررسی می‌شود. نشان خواهیم داد که در رژیم‌های محروم‌نگی بالا (High Privacy)، این معیارها رفتاری مشابه از خود نشان می‌دهند و نتایج به دست آمده با یک معیار، قابل تعمیم به سایرین است.

- **فصل پنجم: نتیجه‌گیری و پیشنهادها** در نهایت، در فصل آخر ضمن مرور دستاوردهای کلیدی پژوهش، به جمع‌بندی مباحث پرداخته و پیشنهادهایی برای پژوهش‌های آتی در این حوزه ارائه خواهیم داد.

فصل ۲

پیش‌نیازها

۱-۲ محرمانگی تفاضلی مت مرکز (CDP)

مفهوم محرمانگی تفاضلی^۱ یا به اختصار DP، اولین بار توسط دُورک و همکاران^[۱] 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 (مثلًاً 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

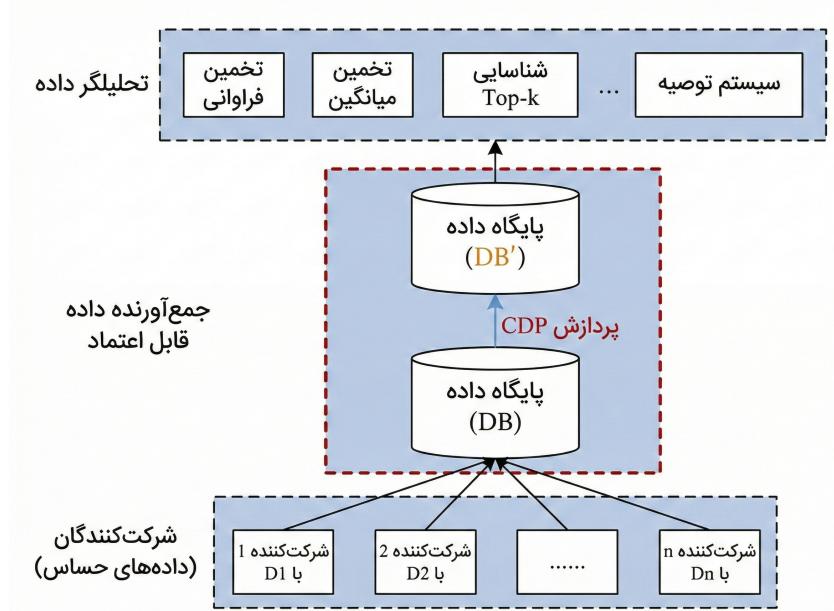
¹Differential Privacy

²Centralized

³Trusted Curator

⁴Randomized Mechanism

⁵Query



شکل ۱-۲: مدل محترمانگی تفاضلی متبرکز با یک متصلی مورد اعتماد.

داده‌ها^۶ می‌نامیم و آن را با \mathcal{X} نمایش می‌دهیم. یک پایگاه داده‌ی D مجموعه‌ای از رکوردها است که اعضای آن از \mathcal{X} انتخاب شده‌اند. در ادبیات محترمانگی تفاضلی، پایگاه داده معمولاً به صورت یک بردار (هیستوگرام) $\mathcal{X} \in \mathbb{N}^{|\mathcal{X}|}$ نمایش داده می‌شود که در آن هر مولفه x_i نشان‌دهنده‌ی تعداد تکرار عنصر i – ام از \mathcal{X} در پایگاه داده است.

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

$$D = (3, 1)$$

تعریف ۲-۲ (الگوریتم تصادفی) یک الگوریتم (mekanizm) تصادفی M تابعی است که دامنه‌ی آن مجموعه‌ی تمام پایگاه داده‌های ممکن و برآ آن مجموعه‌ی خروجی‌های ممکن R است. برخلاف الگوریتم‌های قطعی، خروجی M برای یک ورودی ثابت D ، یک متغیر تصادفی است. به عبارت دیگر، M یک توزیع احتمال روی 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 پرتاپ 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 منهن) 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}_2 به \mathcal{D}_1 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 $(1, 0, 0, 3, 3) = \mathcal{D}_2$ و $(3, 1, 0, 0, 3) = \mathcal{D}_1$ باشند (یعنی در پایگاهداده دوم، یک نفر بیکار حذف شده است). فاصله‌ی ℓ_1 آن‌ها برابر است با:

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

تعريف ۴-۲ (پایگاهداده‌های همسایه) دو پایگاهداده‌ی \mathcal{D}_1 و \mathcal{D}_2 را همسایه^۷ 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 (مانند مثال بالا).

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:

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

در این نامساوی، پارامتر $\varepsilon \geq 0$ را بودجه‌ی محترمانگی^۸ 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 ϵ -DP-تقریبی یا (ϵ, δ) -DP در بسیاری از موارد (مانند مکانیزم گوسی)، اوضاعی شرط ϵ -DP خالص ممکن نیست. در این شرایط، از تعریف انعطاف‌پذیرتری به نام (ϵ, δ) -DP استفاده می‌شود که اجازه‌ی یک احتمال شکست کوچک δ را می‌دهد:

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

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 ϵ -DP 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 ϵ -DP را به صورت‌های همارز دیگری بیان کرد. در ادامه دو دیدگاه مهم « نقطه‌ای » و « واگرایی » را بررسی می‌کنیم.

۱. دیدگاه نقطه‌ای و چگالی احتمال (لم همارزی):

اگرچه تعریف اصلی بر روی زیرمجموعه‌ها بنا شده است، اما لم زیر نشان می‌دهد که کنترل نسبت احتمالات در تک‌تک نقاط برای برقراری شرط کافی است.

لم ۱-۲ (همارزی نقطه‌ای) یک مکانیزم M شرط ϵ -DP را بآورده 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 $z \in \mathcal{R}$

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

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

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

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 $\mathcal{S} \subseteq \mathcal{R}$:

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

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

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

۲. تعریف مبتنی بر واگرایی ماکزیمم:

از دیدگاه نظریه اطلاعات، 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:

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

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 $f : \mathcal{X}^n \rightarrow \mathcal{R}^k$ 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 محروم‌گی 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 «تک تک 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. جزئیات این انتخاب و تأثیر آن بر خطای نهایی، در بخش‌های آینده و پس از معرفی این مکانیزم‌ها به تفصیل بررسی خواهد شد.

۱-۲-۵ مکانیزم‌های بنیادی محترمانگی تفاضلی

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

مکانیزم لاپلاس

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

قضیه‌ی ۲-۲ (محترمانگی مکانیزم لاپلاس) فرض کنید $f : \mathcal{X}^n \rightarrow \mathcal{R}^k$ 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 $(\frac{\Delta_1 f}{\varepsilon}, Y_i) \stackrel{i.i.d.}{\sim} \text{Lap}(\frac{\Delta_1 f}{\varepsilon})$ 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}، پرس‌وجوهای شمارشی^{۱۴}، این نوع توابع بلوک‌های سازنده‌ی بسیاری از تحلیل‌های آماری و داده‌کاوی (مانند هیستوگرام‌ها) هستند [۱۲].

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

¹⁴Counting Queries

حالت برداری (چند پرس و جو): فرض کنید لیستی از k پرس و جوی شمارشی (f_1, \dots, f_k) داریم (یک پرس و جوی برداری). بدون داشتن اطلاعات اضافی درباره ارتباط پرس و جوها، در بدترین حالت یک فرد مشخص می‌تواند در تمام k شمارش تأثیر بگذارد (مثالاً فردی که تمام k ویژگی مورد نظر را دارد). بنابراین حساسیت ℓ_1 کل بردار برابر با مجموع تغییرات، یعنی k خواهد بود ($\Delta_1 f = k$). در این حالت برای دست‌یابی به $-DP-\varepsilon$ ، باید به هر پاسخ نویزی با مقیاس ε/k اضافه کنیم. این مثال نشان می‌دهد که چگونه افزایش تعداد پرس و جوها می‌تواند حساسیت و در نتیجه نویز را افزایش دهد.

مکانیزم گوسی

زمانی که حساسیت ℓ_2 تابع بسیار کمتر از حساسیت ℓ_1 باشد (مثالاً در پرس و جوهای برداری)، مکانیزم گوسی ترجیح داده می‌شود.

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

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

$$\mathcal{M}_{\text{Gauss}}(\mathcal{D}) = f(\mathcal{D}) + \mathcal{N}(0, \sigma^2 \mathbf{I}_k) \quad (12-2)$$

اگر $(\varepsilon, 0) \in \mathcal{E}$ باشد، با انتخاب انحراف معیار σ به صورت زیر، این مکانیزم شرط $(\varepsilon, \delta) - DP$ را برآورده می‌کند:

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

اثبات. [طرح کلی] اثبات دقیق این قضیه نیازمند تحلیل «متغیر تصادفی زیان محرمانگی»¹⁵ است که جزئیات کامل آن در [A 12, Appendix A] موجود است. در اینجا ایده اصلی اثبات را بیان می‌کنیم:

۱. بررسی نسبت چگالی‌ها:

برخلاف مکانیزم لاپلاس که در آن نسبت چگالی‌های احتمال همواره با e^ε کران دار است، در توزیع گوسی نسبت $\frac{p(z)}{q(z)}$ می‌تواند برای مقادیر z خیلی بزرگ، به بی‌نهایت میل کند. بنابراین محرمانگی خالص ($\delta = 0$) غیرممکن است.

¹⁵Privacy Loss Random Variable

۲. تحلیل دمی:

با این حال، احتمال اینکه خروجی مکانیزم در ناحیه‌ای بیفتد که نسبت چگالی‌ها بسیار بزرگ است، بسیار ناچیز است. با استفاده از «کران‌های دمی»^{۱۶} توزیع نرمال، می‌توان نشان داد که احتمال رخداد این نواحی خطر، با δ محدود می‌شود.

۳. شرط انحراف معیار:

محاسبات نشان می‌دهد که اگر واریانس نویز (σ^2) به اندازه کافی بزرگ باشد (رباطه صورت قضیه)، با احتمال حداقل $\delta - 1$ ، نسبت چگالی‌ها کمتر از e^ϵ باقی می‌ماند.

□

مثال ۶-۲ (انتشار آماره‌های چندگانه و اثر ابعاد) فرض کنید یک بیمارستان می‌خواهد میانگین d ویژگی حیاتی مختلف (مانند فشار خون، قند، کلسیتروول و ...) را برای بیمارانش منتشر کند. داده‌های هر بیمار را می‌توان به صورت یک بردار $v \in \mathbb{R}^d$ (پس از نرمال‌سازی) در نظر گرفت. اگر یک بیمار پرونده‌اش را تغییر دهد (یا حذف کند)، در بدترین حالت تمام d ویژگی او می‌توانند از ۰ به ۱ تغییر کنند.

تحلیل حساسیت:

- حساسیت ℓ_1 : مجموع قدر مطلق تغییرات برابر است با $d = |\mathbf{v} - \mathbf{0}| = \sum_{i=1}^d |v_i - 0|$.
- حساسیت ℓ_2 : جذر مجموع مربعات تغییرات برابر است با $\sqrt{d} = \sqrt{\sum_{i=1}^d (v_i - 0)^2}$.
- مقایسه نویز: اگر تعداد ویژگی‌ها زیاد باشد (مثالاً $d = 100$):
- مکانیزم لاپلاس باید نویزی متناسب با $d = 100$ اضافه کند تا $DP = e^d$ حفظ شود. این حجم از نویز عملاً داده‌ها را بی‌فایده می‌کند.
- مکانیزم گوسی نویزی متناسب با $\sqrt{100} = 10$ اضافه می‌کند (با پذیرش یک δ ناچیز).

این کاهش 10 برابری نویز (و در حالت کلی کاهش با ضریب \sqrt{d}) دلیل اصلی استفاده از مکانیزم گوسی در کاربردهای پیشرفته نظری یادگیری عمیق با محرمانگی تفاضلی (DP-SGD) است که در آن گرادیان‌ها بردارهای بسیار بزرگ هستند.

¹⁶Tail Bounds

مکانیزم نمایی

مکانیزم‌های قبلی برای خروجی‌های عددی بودند. اگر خروجی یک «عضو» از یک مجموعه باشد (مثلًاً «بهترین» رنگ، یا «پرطرفدارترین» کاندیدا)، از مکانیزم نمایی استفاده می‌شود. این مکانیزم بر اساس یک تابع امتیاز^{۱۷} $q(\mathcal{D}, r)$ کار می‌کند که میزان «خوبی» خروجی r را برای داده‌های \mathcal{D} 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 (محرمانگی مکانیزم نمایی) مکانیزم نمایی، یک خروجی r از مجموعه ممکن \mathcal{R} را با احتمالی متناسب با امتیاز آن انتخاب می‌کند:

$$\Pr[\mathcal{M}_{\text{Exp}}(\mathcal{D}) = r] = \frac{\exp\left(\frac{\varepsilon \cdot q(\mathcal{D}, r)}{2\Delta q}\right)}{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon \cdot q(\mathcal{D}, r')}{2\Delta q}\right)} \quad (14-2)$$

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:

$$\frac{\Pr[\mathcal{M}(\mathcal{D}_1) = r]}{\Pr[\mathcal{M}(\mathcal{D}_2) = r]} = \frac{\exp\left(\frac{\varepsilon q(\mathcal{D}_1, r)}{2\Delta q}\right)}{\exp\left(\frac{\varepsilon q(\mathcal{D}_2, r)}{2\Delta q}\right)} \cdot \frac{\sum_{r'} \exp\left(\frac{\varepsilon q(\mathcal{D}_2, r')}{2\Delta q}\right)}{\sum_{r'} \exp\left(\frac{\varepsilon q(\mathcal{D}_1, r')}{2\Delta q}\right)}$$

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

مثال ۷-۲ (رأی‌گیری محبوب‌ترین کالا) فرض کنید می‌خواهیم از بین چند کالا، محبوب‌ترین آن‌ها را بر اساس آرا انتخاب کنیم. تابع امتیاز q را تعداد آرای هر کالا تعریف می‌کنیم (با حساسیت ۱). مکانیزم نمایی به کالاهای پرطرفدار شناسن انتخاب بسیار بالایی می‌دهد، اما برای حفظ محرمانگی، احتمال انتخاب کالاهای کم‌طرفدار را نیز صفر نمی‌کند. این احتمالات غیرصفر باعث می‌شود مهاجم نتواند با قطعیت بگوید که آیا انتخاب نهایی واقعاً بیشترین رأی را داشته یا خیر.

۶-۱-۲ ترکیب‌پذیری (Composition)

در کاربردهای واقعی، معمولاً چندین پرس‌وجو روی یک پایگاه‌داده اجرا می‌شود. قضایای ترکیب‌پذیری نشان می‌دهند که چگونه بودجه‌ی محرمانگی با افزایش تعداد پرس‌وجوها انباسته می‌شود.

¹⁷Score Function

قضیه‌ی ۵-۲ (ترکیب‌پذیری ساده^{۱۸}) اگر k مکانیزم M_1, \dots, M_k به ترتیب دارای بودجه‌های $\varepsilon_1, \dots, \varepsilon_k$ باشند، اجرای متوالی آن‌ها روی یک پایگاهداده واحد، تضمین $-DP-(\sum_i \varepsilon_i)$ را فراهم می‌کند [۱۲].

این کران خطی $(\sum_i \varepsilon_i)$ در بسیاری موارد بدینانه است. «قضیه ترکیب پیشرفته»^{۱۹} نشان می‌دهد که با پذیرش اندکی احتمال شکست (δ') اضافه، انباست بودجه بسیار کنتر (با نرخ \sqrt{k}) رشد می‌کند [۱۲].

قضیه‌ی ۶-۲ (ترکیب‌پذیری پیشرفته) برای هر $\delta' > 0$ ، 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. این نتیجه اساس بسیاری از الگوریتم‌های یادگیری ماشین خصوصی ($DP-SGD$) است [۱].

۷-۱-۲ محرمانگی گروهی (Group Privacy)

محرمانگی تفاضلی نه تنها از یک فرد، بلکه به صورت خودکار از گروه‌های کوچک نیز محافظت می‌کند [۱۲].

قضیه‌ی ۷-۲ (محرمانگی گروهی) 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:

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

این خاصیت نشان می‌دهد که با بزرگ شدن گروه (k)، تضمین محرمانگی به صورت نمایی تضعیف می‌شود ($e^{k\varepsilon}$). این یک ویژگی مطلوب است، زیرا پنهان کردن ویژگی‌های یک گروه بزرگ (مثلاً « تمام زنان ایرانی ») نباید ممکن باشد، در حالی که پنهان کردن اطلاعات یک خانواده کوچک ($k = 4$) همچنان ممکن است.

¹⁹ Advanced Composition

۸-۱-۲ محدودیت مدل مت默کز

با وجود تمام مزايا، مدل CDP يك نقطه‌ي ضعف اساسی دارد: نياز به يك متصدی کاملاً مورد اعتماد. در بسياري از سناريوهای دنيا واقعي (مانند جمع‌آوري داده از گوشی‌های هوشمند)، کاربران به سرور مرکزي اعتماد ندارند. اين عدم اعتماد، ما را به سمت مدل جايگزين، يعني «محرمانگي تفاضلي موضعی» سوق می‌دهد. در بخش بعد با محرمانگي تفاضلي موضعی و تعاريف و قضایای اساسی آن آشنا خواهيم شد.

۲-۲ محرمانگي تفاضلي موضعی

در بخش قبل، مبانی نظری حریم خصوصی مت默کز (CDP) و ابزارهای آماری لازم برای تحلیل آن را مرور کردیم. در این فصل، به طور اختصاصی به چارچوب محرمانگي تفاضلي موضعی^{۲۰} می‌پردازیم. این مدل، که امروزه در سیستم‌های توزیع شده و جمع‌آوري داده‌های بزرگ مقیاس کاربرد فراوان دارد، پارادایم اعتماد را از «سرور مرکزی» به «کاربر نهايی» تغيير می‌دهد.

۱-۲-۲ مقدمه و گذار از مدل مت默کز

همان‌طور که در بخش ۱-۲ ديدیم، مدل مت默کز نيازنده وجود يك متصدی مورد اعتماد^{۲۱} است که به داده‌های خام دسترسی داشته باشد. اگرچه اين مدل دقت آماری بالايی را فراهم می‌کند، اما در دنيا واقعي با چالش‌های امنیتی و حقوقی جدی روبروست:

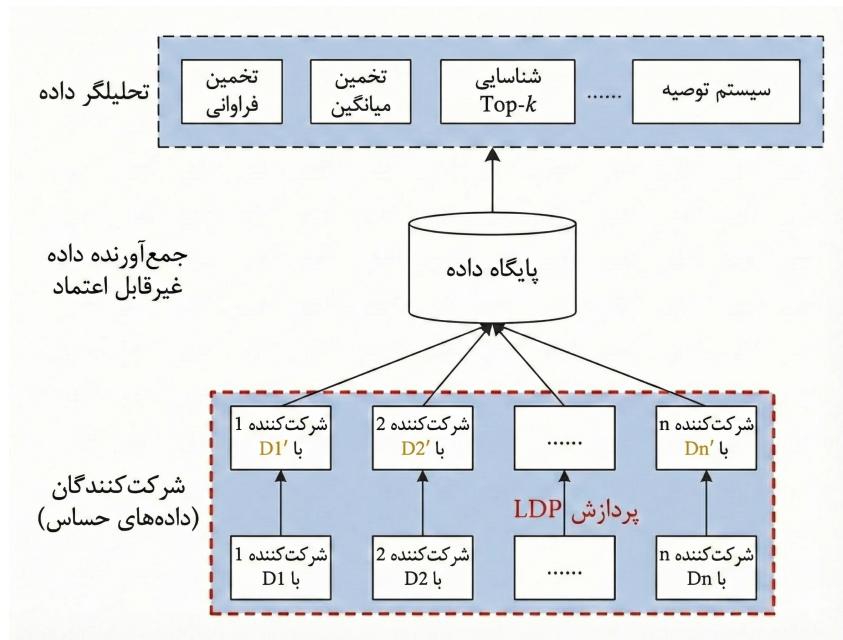
- **نقطه شکست مرکزی^{۲۲}:** سرور مرکزی هدف جذابی برای مهاجمان است. نشت اطلاعات از سرور (چه بر اثر هک و چه بر اثر خطای انساني) حریم خصوصی تمام کاربران را به خطر می‌اندازد.
- **عدم اعتماد کاربران:** در بسياري از کاربردها (مانند جمع‌آوري داده‌های پزشکي یا تاریخچه مرورگر)، کاربران تمایلي ندارند داده‌های حساس خود را حتی به يك سرور «مطمئن» بسپارند.

در پاسخ به اين چالش‌ها، مدل محرمانگي تفاضلي موضعی مطرح شد. در LDP، فرآيند خصوصی‌سازی (افزون نويز) به سمت کلاینت (کاربر) منتقل می‌شود. به اين معنا که داده‌ها قبل از ترک دستگاه کاربر، نويزدار می‌شوند و سرور تنها به داده‌های بي‌نام و نويزدار دسترسی دارد (شکل ۲-۲).

²⁰Local Differential Privacy (LDP)

²¹Trusted Curator

²²Single Point of Failure



شکل ۲-۲: مدل محترمانگی تفاضلی موضعی (α -LDP). نویز به صورت موضعی روی دستگاه کاربر اضافه می‌شود.

این رویکرد توسط شرکت‌های بزرگ فناوری برای جمع‌آوری داده‌های تله‌متري پذیرفته شده است. برای مثال، گوگل از مکانیزم RAPPOR در مرورگر کروم، و اپل و مایکروسافت از روش‌های مشابهی برای جمع‌آوری داده‌های آماری از سیستم عامل‌های خود استفاده می‌کنند.

۲-۲-۲ تعاریف رسمی و مدل‌های محاسباتی

در مدل موضعی، ما n کاربر داریم که هر کدام یک داده‌ی خصوصی X_i از دامنه \mathcal{X} در اختیار دارند. هر کاربر به طور مستقل یک الگوریتم تصادفی (مکانیزم) را اجرا می‌کند و خروجی Z_i را منتشر می‌کند.

تعريف LDP

هسته‌ی اصلی این مدل، «تصادفی‌ساز موضعی» است.

تعريف ۹-۲ (تصادفی‌ساز موضعی^{۲۳}) یک مکانیزم تصادفی $Z \rightarrow M : \mathcal{X}$ را یک تصادفی‌ساز موضعی می‌نامیم که ورودی $x \in \mathcal{X}$ را می‌گیرد و خروجی $Z \in z \in \mathcal{Z}$ را بر اساس توزیع احتمال شرطی $(z|x)$ تولید می‌کند.

شرط محترمانگی در اینجا تضمین می‌کند که با مشاهده‌ی خروجی z ، تمایز قابل شدن بین هر دو ورودی

اولیه x و x' دشوار باشد. تفاوت کلیدی این تعریف با مدل متمرکز در این است که در CDP ما دو پایگاه داده‌ی همسایه را مقایسه می‌کردیم، اما در اینجا هر دو مقدار ورودی ممکن مقایسه می‌شوند.

تعریف ۱۰-۲ (α -محرمانگی تفاضلی موضعی) یک مکانیزم M دارای α -محرمانگی تفاضلی موضعی است اگر برای تمام جفت ورودی‌های \mathcal{X} و هر زیرمجموعه از خروجی‌ها $\mathcal{Z} \subseteq \mathcal{S}$ داشته باشیم:

$$\sup_{\mathcal{S}} \frac{\Pr[\mathcal{M}(x) \in \mathcal{S}]}{\Pr[\mathcal{M}(x') \in \mathcal{S}]} \leq e^{\alpha} \quad (17-2)$$

(نکته: در متون آماری مانند [۹] معمولاً از پارامتر α به جای ϵ برای نمایش بودجه حریم خصوصی موضعی استفاده می‌شود تا تمایز آن با مدل متمرکز مشخص باشد. ما نیز در این فصل و فصول بعدی از این نمادگذاری پیروی می‌کنیم).

این تعریف معادل شرط زیر بر روی واگرایی ماکزیمم (D_∞) بین توزیع‌های شرطی است:

$$\sup_{x, x' \in \mathcal{X}} D_\infty(Q(\cdot|x)||Q(\cdot|x')) \leq \alpha \quad (18-2)$$

تعمیم‌ها و خواص

علاوه بر تعریف استاندارد α -LDP (معادله ۱۷-۲)، دو مفهوم دیگر نیز در تحلیل‌های نظری و طراحی مکانیزم‌ها اهمیت دارند: محرمانگی تقریبی و خاصیت ترکیب.

تعریف ۱۱-۲ ((α, δ) -محرمانگی تفاضلی موضعی) یک مکانیزم تصادفی M دارای α -محرمانگی تفاضلی موضعی تقریبی یا (α, δ) -LDP است اگر برای هر دو ورودی $\mathcal{X} \in \mathcal{X}, x, x' \in \mathcal{X}$ و هر زیرمجموعه خروجی $\mathcal{Z} \subseteq \mathcal{S}$ داشته باشیم:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq e^\alpha \cdot \Pr[\mathcal{M}(x') \in \mathcal{S}] + \delta \quad (19-2)$$

این تعریف (که در [۱۱] نیز بررسی شده است)، اجازه‌ی یک احتمال شکست کوچک δ را می‌دهد. اهمیت نظری این تعریف در آن است که ارتباط مستقیمی با واگرایی E_γ (که در بخش قبل معرفی شد) دارد.

قضیه ۸-۲ (ترکیب ترتیبی^{۲۴}) اگر یک کاربر در k مرحله‌ی مختلف در پروتکل‌های M_1, \dots, M_k شرکت کند که هر کدام به ترتیب دارای بودجه‌ی حریم خصوصی α_i باشند، آنگاه کل فرآیند دارای α -محرمانگی تفاضلی موضعی با بودجه‌ی $\sum_{i=1}^k \alpha_i$ خواهد بود. این خاصیت در تحلیل پروتکل‌های تعاملی (بخش بعد) که در آن خروجی‌های بعدی به خروجی‌های قبلی وابسته هستند، نقش بنیادین دارد.

پروتکل‌های تعاملی و غیرتعاملی

یکی از جنبه‌های مهم در تحلیل نرخ‌های مینیماکس (که در فصل بعد به آن می‌پردازیم)، نحوه‌ی تعامل کاربران با سرور است. دوچی و همکاران [۹] پروتکل‌های موضعی را به دو دسته تقسیم می‌کنند:

۱. **پروتکل‌های غیرتعاملی**^{۲۰}: در این حالت، خروجی‌ی هر کاربر Z_i تنها به ورودی خودش X_i وابسته است و مستقل از داده‌ها یا خروجی‌های سایر کاربران تولید می‌شود.

$$Z_i = \mathcal{M}_i(X_i) \quad (20-2)$$

این مدل ساده‌ترین و رایج‌ترین شکل پیاده‌سازی α -LDP است.

۲. **پروتکل‌های تعاملی (ترتیبی)**^{۲۶}: در این حالت، مکانیزم کاربر i می‌تواند به خروجی‌های منتشر شده توسط کاربران قبلی (Z_1, \dots, Z_{i-1}) وابسته باشد. به عبارت دیگر، کanal ارتباطی Q_i می‌تواند به صورت پویا بر اساس تاریخچه تغییر کند:

$$Z_i \sim Q_i(\cdot | X_i, Z_1, \dots, Z_{i-1}) \quad (21-2)$$

این مدل به الگوریتم‌های تطبیقی اجازه می‌دهد تا دقیق تخمین را بهبود بخشد. با این حال، همان‌طور که در فصل بعد خواهیم دید، حتی با وجود تعامل، محدودیت‌های بنیادی f -واگرایی همچنان مانع از کاهش چشمگیر نرخ خطأ می‌شوند.

۳-۲-۲ مکانیزم‌های پایه در LDP

در این بخش، مکانیزم‌های بنیادین را معرفی می‌کنیم که برای تحقق محترمانگی تفاضلی موضعی استفاده می‌شوند. این مکانیزم‌ها بلوک‌های سازنده‌ی پروتکل‌های پیچیده‌تر هستند و بسته به نوع داده (دودویی، دسته‌ای یا عددی) و اندازه دامنه انتخاب می‌شوند.

پاسخ تصادفی دودویی (RR)

پایه‌ای‌ترین و کلاسیک‌ترین مکانیزم در مدل موضعی، «پاسخ تصادفی»^{۲۷} است که داده‌ها پیش از تعریف رسمی α -LDP-توسط وارنر [۱۹] برای نظرسنجی‌های حساس معرفی شد. فرض کنید دامنه ورودی دودویی باشد ($\mathcal{X} = \{0, 1\}$).

²⁵Non-interactive

²⁶Sequential/Interactive

²⁷Randomized Response (RR)

مکانیزم \mathcal{M}_{RR} با ورودی $\{0, 1\}$ ، خروجی $\{0, 1\}$ را طبق احتمالات زیر تولید می‌کند:

$$\Pr[z = x] = p, \quad \Pr[z \neq x] = 1 - p \quad (22-2)$$

برای اینکه این مکانیزم شرط α -LDP را برآورده کند، طبق تعریف ۲-۱۰ باید نسبت احتمالات حداکثر e^α باشد:

$$\frac{\Pr[z = 1|x = 1]}{\Pr[z = 1|x = 0]} = \frac{p}{1-p} \leq e^\alpha \quad (23-2)$$

بنابراین، برای دستیابی به کمترین خطأ، پارامتر احتمال p به صورت زیر تنظیم می‌شود:

$$p = \frac{e^\alpha}{1 + e^\alpha} \quad (24-2)$$

در این حالت، واریانس تخمین‌گر حاصل از این مکانیزم برابر است با:

$$\text{Var}[\hat{x}] = \frac{e^\alpha}{(e^\alpha - 1)^2} \quad (25-2)$$

این رابطه نشان می‌دهد که برای α ‌های کوچک، واریانس به سرعت (تقریباً با نرخ $1/\alpha^2$) افزایش می‌یابد که نشان‌دهنده‌ی هزینه بالای حریم خصوصی در دقت تخمین است.

پاسخ تصادفی تعمیم‌یافته (GRR)

زمانی که دامنه ورودی شامل $2 > k$ عنصر باشد ($\mathcal{X} = \{1, \dots, k\}$)، از نسخه تعمیم‌یافته پاسخ تصادفی ۲۸ استفاده می‌شود [۱۳]. این روش تعمیم مستقیم RR برای دامنه‌های گسته است.

در این مکانیزم، برای ورودی x :

$$\Pr[\mathcal{M}(x) = z] = \begin{cases} p & \text{if } z = x \\ q & \text{if } z \neq x \end{cases} \quad (26-2)$$

از آنجا که مجموع احتمالات باید ۱ باشد، داریم $1 = p + (k-1)q$. همچنین شرط α -LDP ایجاب می‌کند که $e^\alpha \leq \frac{p}{q}$. با حل این دستگاه معادلات، مقادیر بهینه p و q به صورت زیر به دست می‌آیند:

$$p = \frac{e^\alpha}{e^\alpha + k - 1}, \quad q = \frac{1}{e^\alpha + k - 1} \quad (27-2)$$

این مکانیزم برای دامنه‌های کوچک (k کوچک) بسیار کارآمد است و واریانس آن بهینه است. اما با افزایش k ، دقت آن به شدت کاهش می‌یابد زیرا احتمال گزارش پاسخ صحیح (p) با افزایش k کاهش می‌یابد و به سمت صفر میل می‌کند. بنابراین برای دامنه‌های بزرگ (مانند لغات یک دیکشنری)، GRR گزینه مناسبی نیست.

²⁸Generalized Randomized Response (GRR)

مکانیزم‌های مبتنی بر کدگذاری یگانی (UE)

برای غلبه بر مشکل کاهش دقت GRR در دامنه‌های بزرگ، خانواده‌ای از مکانیزم‌ها تحت عنوان «کدگذاری یگانی»^{۲۹} توسعه یافته‌اند. این رویکرد اساس پروتکل مشهور RAPPOR گوگل را تشکیل می‌دهد [۱۸].

در این روش، فرآیند خصوصی‌سازی طی دو مرحله انجام می‌شود:

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

۲. **اختلال (Perturbation)**: هر بیت این بردار به صورت مستقل با استفاده از یک مکانیزم باینری معکوس می‌شود.

اگر v_i بیت i -ام بردار کدگذاری شده باشد، خروجی z_i به صورت زیر تولید می‌شود:

$$\Pr[z_i = 1] = \begin{cases} p & \text{if } v_i = 1 \\ q & \text{if } v_i = 0 \end{cases} \quad (28-2)$$

بر اساس انتخاب مقادیر p و q ، دو مکانیزم مهم در این خانواده تعریف می‌شوند:

- **کدگذاری یگانی متقارن (SUE)**: که به آن Basic RAPPOR نیز گفته می‌شود. در این حالت، احتمالات تغییر بیت به گونه‌ای انتخاب می‌شوند که مکانیزم متقارن باشد ($p + q = 1$). مقادیر بهینه عبارتند از:

$$p = \frac{e^{\alpha/2}}{e^{\alpha/2} + 1}, \quad q = \frac{1}{e^{\alpha/2} + 1} \quad (29-2)$$

مزیت SUE این است که واریانس تخمین برای هر آیتم مستقل از تعداد کل آیتم‌ها (k) است، اما به دلیل استفاده از نیمی از بودجه حریم خصوصی ($\alpha/2$) برای هر بیت، خطای آن همچنان قابل توجه است.

- **کدگذاری یگانی بهینه (OUE)**: وانگ و همکاران [۱۸] نشان دادند که برای تخمین فراوانی در دامنه‌های بزرگ، نیازی به متقارن بودن مکانیزم نیست. در روش OUE^{۳۰}، پارامترها به گونه‌ای تنظیم می‌شوند که اطلاعات بیت‌های ۱ (سیگنال اصلی) با بیشترین دقت حفظ شود ($p = 1/2$) و نیز روی بیت‌های ۰ (که تعدادشان زیاد است) کنترل شود:

$$p = \frac{1}{2}, \quad q = \frac{1}{e^\alpha + 1} \quad (30-2)$$

²⁹Unary Encoding (UE)

³⁰Optimized Unary Encoding

تحلیل‌های نظری و تجربی نشان می‌دهند که OUE برای α -های متوسط و بزرگ، واریانس کمتری نسبت به GRR و SUE دارد و استاندارد فعلی برای جمع‌آوری داده‌های دسته‌ای بزرگ مقیاس است.

مکانیزم لاپلاس موضعی

برای داده‌های عددی (مثالاً $x \in [1, -1]$)، استفاده از مکانیزم لاپلاس که در مدل متمرکز محبوب است، در مدل موضعی نیز ممکن است اما با چالش‌هایی همراه است.

حساسیت سراسری (Δ) در مدل موضعی برابر با قطر دامنه است، زیرا هر دو ورودی $x, x' \in \mathcal{X}$ باید از نظر مکانیزم غیرقابل تمایز باشند. اگر دامنه ورودی $[1, -1] = \mathcal{X}$ باشد، حساسیت برابر است با:

$$\Delta = \max_{x, x'} |x - x'| = |1 - (-1)| = 2 \quad (31-2)$$

بنابراین، مکانیزم لاپلاس موضعی خروجی را به صورت زیر تولید می‌کند:

$$\mathcal{M}_{Lap}(x) = x + \eta, \quad \eta \sim \text{Lap}\left(\frac{2}{\alpha}\right) \quad (32-2)$$

نکته مهمی که دوچی و همکاران [۹] به آن اشاره کرده‌اند این است که برخلاف مدل متمرکز، مکانیزم لاپلاس در مدل موضعی برای ابعاد بالا ($d > 1$) زیر-بهینه^{۳۱} است و نرخ خطای آن بدتر از مکانیزم‌های پیشرفته‌تر (مانند نمونه‌برداری هایپرکیوب) است که در فصل‌های آینده به آن‌ها اشاره خواهیم کرد.

۴-۲-۲ چالش سودمندی در مدل موضعی

بهای عدم اعتماد به سرور، کاهش شدید سودمندی^{۳۲} آماری است. از آنجایی که نویز به داده‌ی هر فرد به صورت مستقل اضافه می‌شود، خطای تجمعی در مدل α -LDP بسیار بیشتر از مدل متمرکز CDP است.

برای رسیدن به سطح دقت مشابه، مدل موضعی معمولاً^{۳۳} به تعداد کاربران n بسیار بیشتری نیاز دارد. به طور کلی، در حالی که خطای مکانیزم‌های CDP اغلب با $O(1/n)$ کاهش می‌یابد، خطای مکانیزم‌های α -LDP معمولاً^{۳۴} با $O(1/\sqrt{n})$ کاهش می‌یابد. این کاهش در «اندازه نمونه مؤثر» یکی از موضوعات اصلی است که در فصل آینده با استفاده از ابزارهای f -واگرایی آن را اثبات خواهیم کرد.

³¹Sub-optimal

³²Utility

۳-۲ f-واگرایی‌ها

در بخش‌های قبلی، ما مکانیزم‌های محترمانگی تفاضلی را به عنوان روش‌هایی برای ایجاد «شباهت آماری» بین خروجی‌های دو پایگاه داده‌ی همسایه معرفی کردیم. در این بخش، ما ابزار ریاضیاتی اصلی برای سنجش این «شباهت» یا «فاصله» بین توزیع‌های احتمالی را معرفی می‌کنیم. این ابزار، خانواده f -واگرایی‌ها^{۳۳} است که بسیاری از معیارهای رایج فاصله‌ی آماری را به عنوان حالت‌های خاص خود در بر می‌گیرد.

۱-۳-۲ تعریف f -واگرایی

مفهوم f -واگرایی اولین بار توسط سیسر [۴] و به طور همزمان توسط علی و سیلوی [۲] معرفی شد. این معیار، یک روش عمومی برای اندازه‌گیری تفاوت بین دو توزیع احتمال P و Q (تعریف شده بر روی یک فضای یکسان) ارائه می‌دهد.

تعریف ۱۲-۲ (f -واگرایی) فرض کنید P و Q دو توزیع احتمال باشند به طوری که P نسبت به Q مطلقاً پیوسته^{۳۴} باشد. فرض کنید p و q توابع چگالی احتمال (یا توابع جرم احتمال) آن‌ها باشند. برای هر تابع محدب $\mathbb{R} \rightarrow (0, \infty)$ که $f(0) = 0$ باشد، f -واگرایی P از Q به صورت زیر تعریف می‌شود:

$$D_f(P||Q) = \int q(x)f\left(\frac{p(x)}{q(x)}\right) dx \quad (33-2)$$

در حالت گسسته، این تعریف به صورت حاصل جمع زیر در می‌آید:

$$D_f(P||Q) = \sum_{x \in \mathcal{X}} q(x)f\left(\frac{p(x)}{q(x)}\right) \quad (34-2)$$

تابع f «ژنراتور» (تولیدکننده) واگرایی نامیده می‌شود و با انتخاب f ‌های متفاوت، می‌توان معیارهای فاصله‌ی واگرایی متفاوتی را به دست آورد. شرط $f(1) = 0$ تضمین می‌کند که اگر دو توزیع یکسان باشند ($P = Q$)، واگرایی آن‌ها صفر خواهد بود.

۲-۳-۲ نمونه‌های مهم f -واگرایی

بسیاری از معیارهای معروف در آمار و نظریه اطلاعات، حالت‌های خاصی از f -واگرایی هستند:

- **واگرایی کولبک-لایبلر^{۳۵}**: این معیار که به آن آنتروپی نسبی^{۳۶} نیز گفته می‌شود، با انتخاب تابع

³³ f -divergences

³⁴ Absolutely Continuous

³⁵ Kullback-Leibler (KL) Divergence

³⁶ Relative Entropy

• فاصله‌ی واریانس کل^{۳۷}: فاصله‌ی TV (که اغلب با $\Delta(P, Q)$ یا d_{TV} نشان داده می‌شود) ارتباط نزدیکی با f -واگرایی دارد و با انتخاب $f(t) = \frac{1}{2}|t - 1|$ حاصل می‌شود:

$$D_{KL}(P||Q) = \sum_x p(x) \log \left(\frac{p(x)}{q(x)} \right) \quad (35-2)$$

- فاصله‌ی واریانس کل^{۳۸}: فاصله‌ی TV (که اغلب با $\Delta(P, Q)$ یا d_{TV} نشان داده می‌شود) ارتباط نزدیکی با f -واگرایی دارد و با انتخاب $f(t) = \frac{1}{2}|t - 1|$ حاصل می‌شود:

$$d_{TV}(P, Q) = \frac{1}{2} \sum_x |p(x) - q(x)| \quad (36-2)$$

- واگرایی کای-دو^{۳۹}: این معیار آماری با انتخاب $f(t) = (t - 1)^2$ به دست می‌آید:

$$\chi^2(P||Q) = \sum_x \frac{(p(x) - q(x))^2}{q(x)} \quad (37-2)$$

- فاصله‌ی هلینجر (مربع)^{۴۰}: این فاصله با انتخاب $f(t) = (\sqrt{t} - 1)^2$ (یا معادل آن $= f(t) = (\sqrt{t} - 1)^2$) حاصل می‌شود:

$$H^2(P, Q) = \sum_x (\sqrt{p(x)} - \sqrt{q(x)})^2 \quad (38-2)$$

۳-۳-۲ ارتباط f -واگرایی‌ها با یکدیگر

این واگرایی‌ها مستقل از هم نیستند و روابط ریاضی مهمی بین آنها برقرار است. یکی از مشهورترین این روابط، نامساوی پینسکر^{۴۱} است که ارتباط بین واگرایی KL و فاصله‌ی TV را نشان می‌دهد:

$$d_{TV}(P, Q) \leq \frac{1}{\sqrt{2}} D_{KL}(P||Q) \quad (39-2)$$

این نامساوی‌ها در تحلیل‌های حریم خصوصی بسیار کاربردی هستند، زیرا به ما اجازه می‌دهند که با داشتن یک کران (حد) بر روی یک معیار واگرایی، بتوانیم کرانی برای سایر معیارها نیز به دست آوریم.

در فصل بعدی، ما به تفصیل بررسی خواهیم کرد که چگونه تضمین α -LDP (که در معادله ۱۷-۲ تعریف شد) مستقیماً منجر به ایجاد یک کران بالا بر روی f -واگرایی‌های مختلف بین توزیع‌های خروجی می‌شود.

³⁷Total Variation (TV) Distance

³⁸Chi-Squared (χ^2) Divergence

³⁹Squared-Hellinger Distance

⁴⁰Pinsker's Inequality

۴-۲ مبانی آماری و نظریه اطلاعات

در بخش‌های پیشین، ابزارهای سنجش فاصله بین توزیع‌ها (مانند f -واگرایی‌ها) را معرفی کردیم. در این بخش، به معرفی چارچوب آماری می‌پردازیم که در آن از این ابزارها برای تحلیل حدود پایین خطأ در حضور محدودیت‌های محترمانگی استفاده می‌شود. این تعاریف و قضایا عمده‌تاً بر اساس چارچوب ارائه شده در [۹] تدوین شده‌اند.

۱-۴-۲ معیارهای فاصله اطلاعاتی

برای دو توزیع احتمال P و Q که روی فضای \mathcal{X} تعریف شده‌اند و نسبت به یک اندازه‌ی پایه μ مطلقاً پیوسته هستند (با توابع چگالی p و q)، معیارهای زیر را تعریف می‌کنیم:

تعریف ۱۳-۲ (واگرایی کولبک-لایبلر) واگرایی کولبک-لایبلر (KL) بین دو توزیع P و Q به صورت زیر تعریف می‌شود:

$$D_{KL}(P||Q) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} d\mu(x) \quad (40-2)$$

تعریف ۱۴-۲ (فاصله‌ی واریانس کل) فاصله‌ی واریانس کل^{۴۱} بین دو توزیع P و Q به صورت زیر تعریف می‌شود:

$$\|P - Q\|_{TV} = \sup_{S \in \sigma(\mathcal{X})} |P(S) - Q(S)| = \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)| d\mu(x) \quad (41-2)$$

تعریف ۱۵-۲ (اطلاعات متقابل) اگر X و V دو متغیر تصادفی باشند، اطلاعات متقابل^{۴۲} بین آن‌ها به صورت امید ریاضی واگرایی KL بین توزیع شرطی و توزیع حاشیه‌ای تعریف می‌شود:

$$I(X; V) = D_{KL}(P_{X,V}||P_X \otimes P_V) = \mathbb{E}_V [D_{KL}(P_{X|V}||P_X)] \quad (42-2)$$

این معیار نقش کلیدی در نامساوی فانو (که در ادامه می‌آید) ایفا می‌کند.

۲-۴-۲ ریسک مینیماکس

در نظریه تصمیم آماری، هدف تخمین یک پارامتر $(P)\theta$ از یک توزیع ناشناخته $P \in \mathcal{P}$ است. اگر $\hat{\theta}$ یک تخمین‌گر باشد که تابعی از داده‌های مشاهده شده (مانند Z_1, \dots, Z_n) است، کیفیت آن با استفاده از یک تابع زیان صعودی $\Phi \circ \rho$ سنجیده می‌شود (که ρ یک شبهمتر روی فضای پارامتر است).

⁴¹Total Variation Distance

⁴²Mutual Information

نرخ مینیماکس^{۴۳}، کمترین خطای ممکنی است که یک تخمین‌گر در بدترین سناریو (بدترین توزیع P در کلاس \mathcal{P}) متحمل می‌شود.

تعریف ۲-۱۶ (نرخ مینیماکس) برای یک کلاس از توزیع‌ها \mathcal{P} و پارامتر θ ، نرخ مینیماکس \mathfrak{M}_n به صورت زیر تعریف می‌شود:

$$\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho) = \inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\Phi(\rho(\hat{\theta}(Z^n), \theta(P)))] \quad (43-2)$$

که در آن اینفیم روى تمام تخمین‌گرهای ممکن $\hat{\theta}$ گرفته می‌شود.

در حالتی که محدودیت محرومانگی تفاضلی موضعی با پارامتر α وجود داشته باشد، نرخ مینیماکس خصوصی (α -Private Minimax Rate) با در نظر گرفتن اینفیم روی تمام مکانیزم‌های کانال Q که شرط α -LDP را برآورده می‌کنند، تعریف می‌شود [۹].

۵-۲ آزمون فرض آماری و روش تقلیل

برای اثبات حدود پایین نرخ‌های مینیماکس، روش استاندارد این است که مسئله‌ی تخمین پارامتر را به یک مسئله‌ی آزمون فرض^{۴۴} تقلیل دهیم. ایده اصلی این است: اگر نتوانیم بین چند مقدار گسته از پارامتر با دقت بالا تمایز قائل شویم، قطعاً نمی‌توانیم پارامتر را در فضای پیوسته با خطای کم تخمین بزنیم.

۱-۵-۲ آزمون فرض دودویی

ساده‌ترین حالت آزمون فرض، تصمیم‌گیری بین دو توزیع احتمال P_0 و P_1 است. فرض کنید داده‌ی مشاهده شده Z از یکی از این دو توزیع تولید شده است. ما دو فرض داریم:

• فرض صفر (H_0): $Z \sim P_0$

• فرض مقابل (H_1): $Z \sim P_1$

یک آزمون (یا تابع تست) $\psi : \mathcal{Z} \rightarrow \{0, 1\}$ تابعی است که بر اساس داده‌ی مشاهده شده، حدس می‌زند کدام فرض صحیح است. خطای این آزمون به صورت مجموع احتمال خطای نوع اول و دوم تعریف

⁴³Minimax Rate

⁴⁴Hypothesis Testing

می‌شود:

$$P_{err}(\psi) = \Pr_{H_0}(\psi(Z) = 1) + \Pr_{H_1}(\psi(Z) = 0) \quad (44-2)$$

لم نیمن-پیرسون⁴⁵ نشان می‌دهد که کمترین خطای ممکن برای هر آزمون دودویی، مستقیماً با فاصلهٔ واریانس کل (d_{TV}) بین دو توزیع ارتباط دارد:

$$\inf_{\psi} P_{err}(\psi) = 1 - \|P_0 - P_1\|_{TV} \quad (45-2)$$

این رابطه نشان می‌دهد که هرچه دو توزیع P_0 و P_1 به هم شبیه‌تر باشند (فاصلهٔ TV کمتر)، احتمال خطای بیشتر شده و به ۱ (حدس تصادفی) نزدیک‌تر می‌شود. در فضای α -LDP، نویز اضافه شده باعث کاهش شدید فاصلهٔ TV و در نتیجه افزایش خطای آزمون می‌شود.

۲-۵-۲ تقلیل تخمین به آزمون (روش بسته‌بندی)

برای استفاده از ابزارهای آزمون فرض در مسئلهٔ تخمین نرخ مینیماکس (معادله ۴۳-۲)، از تکنیک گسسته‌سازی فضای پارامتر Θ استفاده می‌کنیم. این روش شامل مراحل زیر است:

۱. ساخت مجموعهٔ بسته‌بندی⁴⁶: مجموعه‌ای متناهی از پارامترها $\Theta \subset \{\theta_1, \dots, \theta_M\}$ را انتخاب می‌کنیم به طوری که از یکدیگر فاصلهٔ معناداری داشته باشند. به طور دقیق‌تر، اگر ρ متریک خطا باشد، برای هر $j \neq i$ باید داشته باشیم $\rho(\theta_i, \theta_j) \geq 2\delta$.

۲. تعریف مسئلهٔ آزمون: فرض می‌کنیم طبیعت⁴⁷ یک اندیس V را به صورت تصادفی و یکنواخت از مجموعه $\{1, \dots, M\}$ انتخاب می‌کند و داده‌ها بر اساس توزیع P_{θ_V} تولید می‌شوند. هدف، یافتن V بر اساس داده‌های مشاهده شده است.

۳. ارتباط خطاهای تخمین: اگر یک تخمین‌گر $\hat{\theta}$ وجود داشته باشد که خطای تخمین آن با احتمال بالا کمتر از δ باشد، می‌توانیم از آن برای حل مسئلهٔ آزمون فرض استفاده کنیم (با انتخاب نزدیک‌ترین i به $\hat{\theta}$). بنابراین، کران پایین روی خطای آزمون فرض، یک کران پایین برای خطای تخمین ایجاد می‌کند:

$$\mathfrak{M}_n(\theta(\mathcal{P})) \geq \Phi(\delta) \cdot \inf_{\psi} \Pr_{\psi}(\psi(Z^n) \neq V) \quad (46-2)$$

⁴⁵Neyman-Pearson Lemma

⁴⁶Packing Set

⁴⁷Nature

۳-۵-۲ نامساوی‌های کران پایین

برای اثبات کران‌های پایین، سه روش اصلی که بر پایه f -واگرایی‌ها بنا شده‌اند را معرفی می‌کنیم:

قضیه‌ی ۹-۲ (نامساوی لو کم^{۴۸}) این روش برای آزمون بین دو توزیع P_1 و P_2 استفاده می‌شود. کمینه احتمال خطأ با استفاده از فاصله‌ی واریانس کل (رابطه‌ی ۴۱-۲) کران‌دار می‌شود:

$$\inf_{\psi} \Pr(\psi(Z^n) \neq V) \geq \frac{1}{2} (1 - \|P_1^n - P_2^n\|_{TV}) \quad (47-2)$$

این روش زمانی مفید است که مسئله را به تشخیص بین دو حالت ساده تقلیل دهیم.

قضیه‌ی ۱۰-۲ (نامساوی فانو^{۴۹}) زمانی که پارامتر مورد نظر متعلق به مجموعه‌ای بزرگتر \mathcal{V} باشد (تعداد فرضیه‌ها $|\mathcal{V}| > 2$)، نامساوی فانو کران پایین قوی‌تری ارائه می‌دهد که مبتنی بر اطلاعات متقابل است:

$$\inf_{\psi} \Pr(\psi(Z^n) \neq V) \geq 1 - \frac{I(Z^n; V) + \log 2}{\log |\mathcal{V}|} \quad (48-2)$$

که در آن V متغیر تصادفی یکنواخت روی مجموعه اندیس‌ها \mathcal{V} است.

لم ۱۱-۲ (لم اسود^{۵۰}) این لم مسئله تخمین را به چندین آزمون فرض دودویی مستقل روی مختصات یک ابرمکعب $\{1, -1\}^d$ تبدیل می‌کند. نسخه دقیق‌تر آن که در [۴] استفاده شده است، کران پایین را بر اساس فاصله‌ی واریانس کل توزیع‌های مخلوط حاشیه‌ای بیان می‌کند:

$$\mathfrak{M}_n(\theta(\mathcal{P})) \geq \delta \sum_{j=1}^d [1 - \|M_{+j}^n - M_{-j}^n\|_{TV}] \quad (49-2)$$

که در آن M_{+j}^n و M_{-j}^n توزیع‌های حاشیه‌ای مخلوط روی مقادیر $+1$ و -1 در بعد j -ام هستند.

فصل ۳

تحلیل‌های مبتنی بر انقباض و نرخ‌های مینیماکس

۱-۳ مقدمه

در فصل پیشین، تعاریف پایه محرمانگی تفاضلی موضعی (LDP) و مکانیزم‌های ابتدایی آن را بررسی کردیم. همان‌طور که دیدیم، چالش اصلی در مدل موضعی، کاهش شدید نسبت سیگنال به نویز است. برای تحلیل دقیق این پدیده و یافتن حدود نهایی دقت آماری، نیازمند ابزارهای قوی‌تری هستیم.

در این فصل، به بررسی چارچوب نظری استانداردی می‌پردازیم که توسط دوچی و همکاران [۹] توسعه داده شده است. ایده مرکزی این چارچوب، نگاه به مکانیزم‌های محرمانگی به عنوان «عملگرهای انقباضی»^۱ است. به بیان شهودی، اعمال شرط α -LDP باعث می‌شود که توزیع‌های خروجی $(x|M)$ و $(x'|M)$ بسیار به یکدیگر شبیه شوند، حتی اگر ورودی‌های x و x' کاملاً متفاوت باشند.

ما نشان خواهیم داد که چگونه می‌توان این شباهت اجباری را با استفاده از نامساوی‌های پردازش داده و f -واگرایی‌ها (به‌ویژه واگرایی کولبک-لایبلر) مدل‌سازی کرد و از آن برای اثبات نرخ‌های مینیماکس در مسائل تخمین آماری استفاده نمود [۹].

۲-۳ محرمانگی به عنوان انقباض اطلاعاتی

یکی از ویژگی‌های بنیادین نظریه اطلاعات، «نامساوی پردازش داده»^۲ است که بیان می‌کند پردازش روی داده‌ها (بدون دسترسی به منبع اصلی) نمی‌تواند اطلاعات متقابل را افزایش دهد. در زمینه محرمانگی، ما

¹Contraction Operators

²Data Processing Inequality

با نسخه قوی‌تری از این مفهوم سروکار داریم که به آن «نامساوی قوی پردازش داده»^۳ می‌گویند [۴].

فرض کنید M یک مکانیزم $\text{LDP}-\alpha$ -بашد. هدف ما یافتن کرانی برای واگرایی بین توزیع‌های خروجی بر حسب واگرایی ورودی‌هاست. دوچی و همکاران نشان دادند که مکانیزم‌های موضعی باعث انقباض شدید در واگرایی KL می‌شوند.

قضیه ۱-۳ (انقباض KL در مکانیزم‌های موضعی) فرض کنید M یک مکانیزم $\text{LDP}-\alpha$ -باشد. برای هر دو ورودی $x, x' \in \mathcal{X}$, واگرایی کولبک-لایلر بین توزیع‌های خروجی متناظر $M(\cdot|x)$ و $M(\cdot|x')$ با رابطه زیر محدود می‌شود:

$$D_{KL}(M(\cdot|x) \| M(\cdot|x')) \leq 4(e^\alpha - 1)^2 \quad (1-3)$$

به طور دقیق‌تر، اگر $1 \leq \alpha$ باشد، این کران به صورت $O(\alpha^2)$ رفتار می‌کند [۹].

اثبات. برای اثبات دقیق این قضیه، از تعریف واگرایی KL شروع می‌کنیم. فرض کنید $(z|x)$ و $(z|x')$ چگالی‌های احتمال خروجی باشند. طبق تعریف $\text{LDP}-\alpha$ -می‌دانیم که برای هر $z \in \mathcal{Z}$:

$$e^{-\alpha} \leq \frac{q(z|x)}{q(z|x')} \leq e^\alpha \quad (2-3)$$

این شرط تضمین می‌کند که نسبت درست‌نمایی‌ها حول عدد ۱ محدود است. با بسط تیلور تابع $\log t$ حول $t = 1$ و استفاده از خواص تحدب، می‌توان نشان داد که:

$$D_{KL}(P \| Q) = \int p(z) \log \frac{p(z)}{q(z)} dz \quad (3-3)$$

$$\leq \int p(z) \left(\frac{p(z)}{q(z)} - 1 + \frac{1}{2} \left(\frac{p(z)}{q(z)} - 1 \right)^2 \right) dz \quad (4-3)$$

با اعمال کران‌های $\text{LDP}-\alpha$ -بر روی نسبت p/q ، جمله درجه اول صفر می‌شود و جمله درجه دوم ضریب $(1 - e^\alpha)^2$ را تولید می‌کند. جزئیات کامل این محاسبات در لم ۱ مقاله [۹] آمده است. نکته کلیدی این است که برای α کوچک، فاصله KL به صورت مربعی با α کاهش می‌یابد. \square

این قضیه ابزار بسیار قدرتمندی است. به جای اینکه مستقیماً با تعریف دشوار $\text{LDP}-\alpha$ -کار کنیم، می‌توانیم از این کران ساده در نامساوی‌هایی مثل فانو استفاده کنیم. همچنین مطالعات جدیدتر نشان داده‌اند که این انقباض را می‌توان با استفاده از معیارهای دیگری نظیر اطلاعات متقابل [۸] یا واگرایی E_γ [۳] نیز بیان کرد که در فصل بعد به آن می‌پردازیم.

³Strong Data Processing Inequality (SDPI)

۱-۲-۳ انقباض در فاصله واریانس کل

علاوه بر KL، کران مشابهی برای فاصله واریانس کل (TV) نیز ارائه شده است که در استفاده از روش «لم لو کم^۴» کاربرد دارد [۹]:

قضیه‌ی ۲-۳ (انقباض TV) تحت شرایط مشابه، برای هر مکانیزم α -LDP:

$$\|\mathcal{M}(\cdot|x) - \mathcal{M}(\cdot|x')\|_{TV} \leq \min\{1, e^\alpha - 1\} \cdot \|x - x'\|. \quad (5-3)$$

(در اینجا $\|x - x'\|$ نشان‌دهنده فاصله همینگ یا متریک مجزا روی ورودی است). برای α کوچک، این رابطه بیان می‌کند که فاصله آماری خروجی‌ها نمی‌تواند بیشتر از $O(\alpha)$ باشد.

۳-۳ تحلیل نرخ‌های مینیماکس با استفاده از انقباض

حال که ابزار انقباض را در اختیار داریم، می‌توانیم استراتژی کلی اثبات حدود پایین^۵ در مدل موضعی را صورت‌بندی کنیم. این استراتژی که توسط دوچی [۹] و بعدها با جزئیات بیشتر در [۱۰] بسط داده شد، شامل سه گام است:

۱. تقلیل به آزمون فرض: تبدیل مسئله تخمین پارامتر θ به مسئله تشخیص اندیس V در یک مجموعه متناهی (استفاده از LM فانو یا اسود).

۲. کران‌دار کردن اطلاعات متقابل: استفاده از خاصیت انقباض α -LDP برای محدود کردن اطلاعاتی که نمونه‌های مشاهده شده Z_1, \dots, Z_n درباره اندیس V می‌دهند [۶].

۳. محاسبه ریسک نهایی: ترکیب نتایج برای رسیدن به کران پایین خطای تخمین.

مهم‌ترین گام، گام دوم است. طبق نامساوی قوی پردازش داده برای مدل موضعی، داریم:

$$I(V; Z^n) \leq \sum_{i=1}^n I(V; Z_i) \leq n \cdot \alpha^2 \cdot C \quad (6-3)$$

که در آن C ثابتی است که به هندسه مسئله بستگی دارد. این رابطه نشان می‌دهد که اطلاعات موثر با نرخ $n\alpha^2$ رشد می‌کند، نه n . این همان دلیلی است که «اندازه نمونه موثر» در مدل موضعی برابر با $n\alpha^2$ در نظر گرفته می‌شود.

⁴Le Cam

⁵Lower Bounds

۴-۳ مطالعه موردی: تخمین میانگین

برای نمایش قدرت این چارچوب، مسئله کلاسیک تخمین میانگین را در نظر می‌گیریم. فرض کنید هر کاربر i برداری $X_i \in [-1, 1]^d$ دارد و هدف تخمین میانگین جامعه $\mu = \mathbb{E}_X$ است. معیار خطا را «میانگین مربعات خطای مدل» (MSE) در نظر می‌گیریم.

قضیه ۳-۳ (کران پایین تخمین میانگین) برای هر مکانیزم $LDP-\alpha$ -گرایی^۶، ماکسیمم خطای موردنظر با رابطه زیر محدود می‌شود [۹]:

$$\inf_{\hat{\mu}, \mathcal{M}} \sup_P \mathbb{E}_{\|\hat{\mu} - \mu\|^2} \geq \Omega \left(\frac{d}{n \min\{\alpha, \alpha^2\}} \right) \quad (7-3)$$

تحلیل اثبات: برای اثبات این کران، از لم اسود استفاده می‌کنیم. فضای پارامتر را به صورت یک ابرمکعب $\{-1, 1\}^d$ گسترش می‌کنیم. طبق لم اسود، خطای مجموع فاصله‌های TV بین توزیع‌های شرطی مرتبط است. با استفاده از قضیه انقباض ۱-۳ و نامساوی پینسکر، می‌دانیم که:

$$\|\mathcal{M}(\cdot|x) - \mathcal{M}(\cdot|x')\|_{TV}^2 \leq \frac{1}{\alpha} D_{KL}(\mathcal{M}(\cdot|x) || \mathcal{M}(\cdot|x')) \leq O(\alpha^2) \quad (8-3)$$

بنابراین فاصله TV حداقل از مرتبه α است. با جایگذاری این مقدار در لم اسود، کران پایین $\frac{1}{n\alpha^2}$ حاصل می‌شود.

این نتیجه نشان می‌دهد که برای رسیدن به خطای کم در مدل موضعی، تعداد داده‌ها باید متناسب با α^2/α افزایش یابد، که هزینه‌ی بسیار سنگین‌تری نسبت به مدل متغیر (که متناسب با $1/\varepsilon$ است) دارد.

۵-۳ محدودیت‌های تحلیل کلاسیک

با وجود موفقیت چارچوب دوچی در اثبات نرخ‌های مینیماکس بهینه برای α ‌های کوچک (رزیم محترمانگی بالا)، این روش در رزیم α ‌های بزرگ (محترمانگی پایین) دچار ضعف است.

همان‌طور که در رابطه (۱-۳) دیدیم، کران انقباض KL با ضریب $(1 - e^\alpha)^2$ رشد می‌کند. زمانی که α بزرگ باشد، این کران به سرعت به بینهایت می‌کند و اطلاعاتی فراتر از کران بدیهی به ما نمی‌دهد. این در حالی است که به طور شهودی، حتی با α بزرگ، مکانیزم همچنان باید مقداری انقباض ایجاد کند.

این محدودیت ناشی از ذات واگرایی KL است که رفتار دنباله‌های توزیع را با حساسیت زیادی وزن دهدی می‌کند. برای رفع این مشکل و به دست آوردن تحلیل‌های دقیق‌تر^۹ که در تمام بازه‌های α معتبر باشند، نیازمند

⁶Tight

معیار هندسی متفاوتی هستیم. این نیاز، انگیزه اصلی معرفی واگرایی‌های جدید مانند f – واگرایی‌های خاص (نظیر E_γ) است [۳، ۴] که در فصل آینده به تفصیل به آن خواهیم پرداخت.

فصل ۴

همارزی LDP و انقباض E_γ -واگرایی

۱-۴ مقدمه و انگیزه

در فصل پیشین، دیدیم که چگونه دوچی و همکاران [۴] از واگرایی کولبک-لایلر (KL) برای تحلیل محترمانگی تفاضلی موضعی استفاده کردند. اگرچه کرانهای آنها برای رژیم‌های محترمانگی بالا (α کوچک) بسیار کارآمد هستند، اما در رژیم‌های α متوسط و بزرگ، دقت خود را از دست می‌دهند.

مشکل اصلی در آنجاست که واگرایی KL متريک «بومی» برای تعريف α -LDP می‌باشد. تعريف α -LDP (معادله ۱۷-۲) مبنی بر نسبت احتمالات است، در حالی که KL مبنی بر لگاریتم نسبت‌هاست. اين نامخوانی باعث می‌شود که در تبدیل شرایط α -LDP به کرانهای KL، اطلاعاتی از دست برود (lossy conversion).

در این فصل، نشان می‌دهیم که یک معیار واگرایی دیگر به نام E_γ -واگرایی وجود دارد که دقیقاً ساختار هندسی α -LDP را تسخیر می‌کند. ما ثابت خواهیم کرد که شرط α -LDP دقیقاً معادل صفر شدن E_γ -واگرایی (برای $e^\alpha = \gamma$) است. سپس از این همارزی برای استخراج کرانهای انقباض دقیق^۱ برای سایر واگرایی‌ها استفاده خواهیم کرد که نتایج دوچی را بهبود می‌بخشند [۳].

¹Tight

۲-۴ معرفی E_γ -واگرایی

-واگرایی یکی از اعضای کمتر شناخته شدهی خانواده f -واگرایی هاست که در نظریه اطلاعات برای مقایسه نسبت درست نمایی توزیع‌ها کاربرد دارد.

تعريف ۱-۴ (-واگرایی) فرض کنید P و Q دو توزیع احتمال باشند و $1 \geq \gamma \geq 0$ یک عدد حقیقی باشد.
-واگرایی بین P و Q به صورت زیر تعریف می‌شود:

$$E_\gamma(P||Q) = \sup_{\mathcal{S} \in \sigma(\mathcal{X})} (P(\mathcal{S}) - \gamma Q(\mathcal{S})) \quad (1-4)$$

این تعریف را می‌توان به صورت بسطهی زیر نیز نوشت:

$$E_\gamma(P||Q) = \int_{\mathcal{X}} \max\{\cdot, p(x) - \gamma q(x)\} d\mu(x) \quad (2-4)$$

که در آن p و q توابع چگالی احتمال هستند.

۱-۲-۴ خواص هندسی

این واگرایی خواص جالبی دارد که آن را برای تحلیل محروم‌نگی ایده‌آل می‌کند:

- ارتباط با فاصله واریانس کل: اگر $\gamma = 1$ باشد، داریم:

$$E_1(P||Q) = \sup_{\mathcal{S}} (P(\mathcal{S}) - Q(\mathcal{S})) = \|P - Q\|_{TV} \quad (3-4)$$

بنابراین E_γ تعمیمی از فاصله TV است.

- غیرمنفی بودن: همواره $E_\gamma(P||Q) \geq 0$ نیست. در واقع، اگر نسبت $p(x)/q(x)$ همواره کمتر از γ باشد، این مقدار صفر می‌شود. دقیقاً همین ویژگی است که آن را به α -LDP-مرتبه می‌کند.

۳-۴ قضیه همارزی اصلی

اکنون به مهم‌ترین نتیجه‌ی نظری این پایان‌نامه می‌رسیم: اثبات اینکه α -LDP-چیزی جز محدودیت بر روی -واگرایی نیست.

قضیه‌ی ۱-۴ (همارزی- LDP - α) یک مکانیزم \mathcal{M} در شرط E_γ می‌کند اگر و تنها اگر برای تمام جفت ورودی‌های $x, x' \in \mathcal{X}$

$$E_{e^\alpha}(\mathcal{M}(\cdot|x) \parallel \mathcal{M}(\cdot|x')) = 0 \quad (4-4)$$

اثبات. اثبات را در دو جهت انجام می‌دهیم.

جهت اول (\Rightarrow): فرض کنید \mathcal{M} خاصیت LDP - α دارد. طبق تعریف ۱۰-۲، برای هر زیرمجموعه خروجی $\mathcal{S} \subseteq \mathcal{Z}$ و هر $x, x' \in \mathcal{S}$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq e^\alpha \Pr[\mathcal{M}(x') \in \mathcal{S}] \quad (5-4)$$

این نامساوی را می‌توان به صورت زیر بازنویسی کرد:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] - e^\alpha \Pr[\mathcal{M}(x') \in \mathcal{S}] \leq 0 \quad (6-4)$$

از آنجایی که این رابطه برای تمام \mathcal{S} ‌ها برقرار است، سوپریمم آن نیز باید کوچکتر یا مساوی صفر باشد. اما طبق تعریف E_γ در معادله ۱-۴، این سوپریمم دقیقاً همان E_{e^α} است. چون E_γ نمی‌تواند منفی باشد (با انتخاب $\emptyset = \mathcal{S}$ مقدار حداقل صفر است)، پس حتماً برابر صفر است.

جهت دوم (\Leftarrow): فرض کنید $E_{e^\alpha}(\mathcal{M}(\cdot|x) \parallel \mathcal{M}(\cdot|x')) = 0$. طبق تعریف سوپریمم، برای هر مجموعه دلخواه \mathcal{S} :

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] - e^\alpha \Pr[\mathcal{M}(x') \in \mathcal{S}] \leq 0 \quad (7-4)$$

که بلافارسله نتیجه می‌دهد:

$$\frac{\Pr[\mathcal{M}(x) \in \mathcal{S}]}{\Pr[\mathcal{M}(x') \in \mathcal{S}]} \leq e^\alpha \quad (8-4)$$

□ این دقیقاً همان تعریف LDP - α است.

این قضیه ساده اما بنیادین، یک تفسیر هندسی دقیق از محرمانگی ارائه می‌دهد: LDP - α یعنی توزیع‌های خروجی چنان به هم نزدیک باشند که هیچ بخشی از دامنه نتواند نسبت درست‌نمایی بیشتر از e^α ایجاد کند.

۴-۴ بهود کران‌های انقباض

در فصل ۳ دیدیم که دوچی [۹] کران زیر را برای انقباض KL ارائه کرد:

$$D_{KL}(\mathcal{M}(\cdot|x) \parallel \mathcal{M}(\cdot|x')) \leq 4(e^\alpha - 1)^2 \quad (9-4)$$

حال با استفاده از چارچوب E_γ ، می‌توانیم کران‌های بسیار دقیق‌تری استخراج کنیم. آسوده و همکاران [۳] نشان داده‌اند که اگر شرط $E_{e^\alpha} = 0$ برقرار باشد، می‌توان کران‌های انقباض برای سایر f -واگرایی‌ها را از طریق بهینه‌سازی محدب به دست آورد.

قضیه‌ی ۲-۴ (کران دقیق انقباض KL) اگر \mathcal{M} یک مکانیزم α -LDP باشد، آنگاه:

$$D_{KL}(\mathcal{M}(\cdot|x) \parallel \mathcal{M}(\cdot|x')) \leq \frac{e^\alpha - 1}{e^\alpha + 1} \cdot (e^\alpha - 1) \quad (10-4)$$

برای مقادیر کوچک α (رژیم محروم‌گی بالا)، این کران به $\alpha^2/2$ میل می‌کند که 4 برابر کوچکتر (بهتر) از کران دوچی است.

تحلیل مقایسه‌ای: باید رفتار دو کران را در $\alpha \rightarrow 0$ بررسی کنیم:

- کران دوچی: $4(e^\alpha - 1)^2 \approx 4\alpha^2$

- کران مبتنی بر E_γ : $\frac{e^\alpha - 1}{e^\alpha + 1} \approx \tanh(\alpha/2) \approx \alpha/2 \approx \frac{\alpha}{2} \cdot \alpha = \frac{\alpha^2}{2}$

این بهود ضریب ثابت (از 4 به 0.5) در تحلیل‌های مینیماکس بسیار حیاتی است و نشان می‌دهد که «اندازه نمونه موثر» واقعی می‌تواند تا 8 برابر بهتر از چیزی باشد که آنالیزهای قبلی نشان می‌دادند.

۵-۴ تعمیم به محروم‌گی تقریبی (α, δ) -LDP

یکی دیگر از قدرت‌های چارچوب E_γ ، توانایی آن در توصیف ساده‌ی محروم‌گی تقریبی است. یادآوری می‌کنیم که (α, δ) -LDP شرط زیر را دارد:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq e^\alpha \Pr[\mathcal{M}(x') \in \mathcal{S}] + \delta \quad (11-4)$$

با بازنویسی این رابطه داریم:

$$\sup_{\mathcal{S}} (\Pr[\mathcal{M}(x) \in \mathcal{S}] - e^\alpha \Pr[\mathcal{M}(x') \in \mathcal{S}]) \leq \delta \quad (12-4)$$

که دقیقاً معادل است با:

$$E_{e^\alpha}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) \leq \delta \quad (13-4)$$

نتیجه‌ی ۳-۴ محرمانگی تقریبی $LDP_{\alpha, \delta}$ - E_{e^α} -واگرایی توزیع‌های خروجی به مقدار δ است. این نتیجه نشان می‌دهد که E_γ -واگرایی طبیعی‌ترین زبان برای صحبت درباره محرمانگی تفاضلی (چه خالص و چه تقریبی) است.

۶-۴ کاربرد در تخمین توزیع گستته

برای نشان دادن کاربرد عملی این نتایج، مسئله تخمین توزیع احتمال روی یک دامنه k -تایی را در نظر بگیرید. با استفاده از تکنیک‌های انقباض E_γ ، می‌توان نشان داد که نرخ مینیماکس برای این مسئله تحت شرط α -LDP برابر است با:

$$\mathfrak{M}_n \asymp \frac{k}{n(e^\alpha - 1)^2} \quad (14-4)$$

در حالی که استفاده از تکنیک‌های کلاسیک (دوچی)، جمله‌ای به صورت $\frac{k}{n\alpha^2}$ را پیشنهاد می‌کرد. تفاوت این دو عبارت در رژیم α بزرگ (محرمانگی کم) آشکار می‌شود؛ جایی که $(1 - e^\alpha)^2$ به صورت نمایی رشد می‌کند و نشان می‌دهد که دقت می‌تواند بسیار سریع‌تر از پیش‌بینی‌های قبلی بهبود یابد.

۷-۴ انقباض قوی برای خانواده‌ی f -واگرایی‌ها

تا اینجا دیدیم که شرط α -LDP معادل صفر شدن E_{e^α} -واگرایی است. یک پرسش طبیعی و بسیار مهم این است: آیا این شرط بر روی سایر معیارهای فاصله (مثل χ^2 یا هلینجر) نیز انقباض ایجاد می‌کند؟ پاسخ مثبت است. در مقاله‌ی اخیر آسوده و ژانگ [۴]، نشان داده شده است که مکانیزم‌های موضعی خاصیت «انقباض قوی» را برای طیف وسیعی از واگرایی‌ها به ارمغان می‌آورند.

۱-۷-۴ کران دقیق برای واگرایی کای-دو (χ^2)

یکی از مهم‌ترین نتایج این پژوهش، ارائه‌ی یک ضریب انقباض دقیق برای واگرایی χ^2 است. اهمیت این واگرایی در آن است که کار با آن در محاسبات واریانس و کران‌های مینیماکس بسیار ساده‌تر از KL است.

قضیه‌ی ۴-۴ (انقباض χ^2) فرض کنید M یک مکانیزم α -LDP باشد. برای هر دو توزیع ورودی P و Q ، واگرایی کای-دو بین توزیع‌های خروجی با رابطه زیر محدود می‌شود:

$$\chi^2(MP||MQ) \leq \eta_\alpha \cdot \chi^2(P||Q) \quad (15-4)$$

که در آن η_α ضریب انقباض بهینه است و برابر است با:

$$\eta_\alpha = \left(\frac{e^\alpha - 1}{e^\alpha + 1} \right)^2 \quad (16-4)$$

تحلیل مجانی: برای مقادیر کوچک α (رژیم محترمانگی بالا)، داریم:

$$\eta_\alpha \approx \left(\frac{1 + \alpha - 1}{1 + \alpha + 1} \right)^2 \approx \left(\frac{\alpha}{2} \right)^2 = \frac{\alpha^2}{4} \quad (17-4)$$

این نتیجه بسیار قابل توجه است. یادآوری می‌کنیم که کران‌های کلاسیک دوچی (فصل ۳) ضریبی از مرتبه $O(\alpha^2)$ داشتند، اما ضریب $1/4$ در اینجا نشان‌دهنده‌ی یک انقباض بسیار شدیدتر است. این ضریب دقیقاً با ضریب انقباض «پاسخ تصادفی دودویی» برای واریانس مطابقت دارد و نشان می‌دهد که این کران برای کل کلاس مکانیزم‌های α -LDP «تایت» (Tight) است.

۲-۷-۴ تعمیم به سایر واگرایی‌ها

نویسنده‌گان در [۴] نشان داده‌اند که این ضریب انقباض η_α تنها مختص χ^2 نیست، بلکه برای خانواده‌ای از واگرایی‌ها که خاصیت «تحدب مشترک» دارند (شامل فاصله هلینجر مجدور H^2 و واگرایی KL) نیز صادق است.

نتیجه‌ی ۵-۴ برای هر مکانیزم α -LDP، کران‌های زیر برقرار هستند:

$$D_{KL}(MP||MQ) \leq \eta_\alpha \cdot D_{KL}(P||Q) \quad (18-4)$$

$$H^2(MP, MQ) \leq \eta_\alpha \cdot H^2(P, Q) \quad (19-4)$$

این یکسان‌سازی ضرایب انقباض، تحلیل مکانیزم‌های پیچیده را بسیار ساده می‌کند؛ زیرا کافیست فقط ضریب η_α را محاسبه کنیم.

۸-۴ نامساوی ونتریز خصوصی (Inequality Private van Trees)

اکثر تحلیل‌های موجود در ادبیات α -LDP (مانند کارهای دوچی)، بر روی «ریسک مینیماکس» (بدترین حالت) تمرکز دارند. اما در بسیاری از کاربردهای مدرن، ما به تحلیل‌های بیزی (Bayesian) علاقه‌مندیم، جایی که پارامتر مجهول θ دارای یک توزیع پیشین $(\theta) \pi$ است.

نامساوی ونتریز (van Trees) ابزاری کلاسیک برای کران دار کردن خطای بیزی بر اساس «اطلاعات فیشر» است. آسوده و ژانگ^۴ [۴] نسخه‌ی خصوصی شده‌ی این نامساوی را ارائه کردند که ابزاری نوین در جعبه‌ابزار تحلیل محروم‌گی محسوب می‌شود.

قضیه‌ی ۶-۴ (نامساوی ونتریز موضعی) فرض کنید می‌خواهیم پارامتر θ را از روی مشاهدات Z^n که خروجی یک مکانیزم α -LDP-هستند، تخمین بزنیم. اگر $\hat{\theta}$ هر تخمین‌گر دلخواهی باشد، آنگاه میانگین مربعات خطای بیزی^۵ دارای کران پایین زیر است:

$$\mathbb{E}_{(\hat{\theta}-\theta)^2} \geq \frac{1}{\mathbb{E}_{\mathcal{I}(\theta)} + \mathcal{I}_{prior}(\pi)} \quad (20-4)$$

نکته‌ی کلیدی اینجاست که در نسخه خصوصی، اطلاعات فیشر مشاهدات $((\mathcal{I}(\theta))$ با ضریب انقباض تضعیف می‌شود:

$$\mathcal{I}_{priv}(\theta) \leq \eta_\alpha \cdot \mathcal{I}_{orig}(\theta) \quad (21-4)$$

که در آن \mathcal{I}_{orig} اطلاعات فیشر داده‌های خام است.

تفسیر: این نامساوی به زبان ساده می‌گوید: «در دنیای α -LDP، هر بیت اطلاعات فیشر که از داده‌ها می‌گیرید، به اندازه‌ی $\alpha^2/4 \approx \eta_\alpha$ تضعیف می‌شود.» این نتیجه، اثبات حدود پایین برای مسائل تخمین پارامتر را بسیار ساده می‌کند. به جای درگیر شدن بالمهای پیچیده‌ی اسود یا فانو، کافیست اطلاعات فیشر مسئله‌ی اصلی را محاسبه کنیم و در ضریب η ضرب کنیم.

۹-۴ کاربردهای نوین و بهبود نرخ‌ها

استفاده از کران‌های انقباض قوی (بخش ۴-۴^۵) و نامساوی ونتریز خصوصی (بخش ۴-۴^۶) منجر به بهبود نتایج در مسائل کلاسیک می‌شود.

²Bayesian Mean Square Error

به عنوان مثال، در مسئله‌ی تخمین چگالی غیرپارامتری برای کلاس توزیع‌های هموار (کلاس هولدر با پارامتر β)، استفاده از این ابزارهای جدید نشان می‌دهد که نرخ خطای بهینه دقیقاً برابر است با:

$$R_{opt} \asymp \left(\frac{1}{n\alpha^2} \right)^{\frac{1-\beta}{1+\beta}} \quad (22-4)$$

اگرچه مرتبه‌ی کلی نرخ همگرایی مشابه نتایج دوچی است، اما ضرایب ثابت بهبود یافته‌اند و مهم‌تر از آن، اثبات با استفاده از انقباض^۲ χ بسیار کوتاه‌تر و مستقیم‌تر از روش‌های مبتنی بر KL است. این امر نشان‌دهنده‌ی برتری رویکرد مبتنی بر E_γ و انقباض قوی در تحلیل سیستم‌های محترمانگی تفاضلی است.

فصل ۵

نتیجه‌گیری

Bibliography

- [1] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS ’16*, page 308–318, New York, NY, USA, 2016. Association for Computing Machinery.
- [2] 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.
- [3] 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.
- [4] Shahab Asoodeh and Huanyu Zhang. Contraction of locally differentially private mechanisms. *IEEE Journal on Selected Areas in Information Theory*, 5:385–395, 2024.
- [5] Michael Barbaro and Tom Zeller. A face is exposed for aol searcher no. 4417749. *New York Times*, 01 2006.
- [6] Leighton Pate Barnes, Wei Ning Chen, and Ayfer Özgür. Fisher information under local differential privacy. *IEEE Journal on Selected Areas in Information Theory*, 1:645–659, 2020.
- [7] 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.
- [8] Paul Cuff and Lanqing Yu. Differential privacy as a mutual information constraint. In *Proceedings of the ACM Conference on Computer and Communications Security*, pages 43–54, 2016.

- [9] 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.
- [10] 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.
- [11] Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.
- [12] 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.
- [13] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.*, 17(1):492–542, January 2016.
- [14] 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.
- [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] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.
- [17] 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.
- [18] 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.
- [19] 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 Set

متصدی مورد اعتماد Trusted Curator

مکانیزم تصادفی Randomized Mechanism

محرمانگی تفاضلی Differential Privacy

مطلقاً پیوسته Absolutely Continuous

و

Divergence واگرایی

هـ

Ajacent همسایه

Adjacency همسایگی

ن**ى**

پیوست آ

مطالب تكميلی

Abstract

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

Keywords: Thesis, Typesetting, Template, `XEPersian`



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