



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

اسفند ۱۴۰۴



به نام خدا
دانشگاه صنعتی شریف
دانشکده علوم ریاضی

پایان نامه کارشناسی ارشد

این پایان نامه به عنوان تحقق 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 محرمانگی تفاضلی موضعی و کاربردهای آماری آن

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

این جانب فیروزه ابریشمی اظهار می دارم:

۱. متن و نتایج علمی ارائه شده در این پایان نامه اصیل بوده و زیر نظر 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 فرمول ها، توابع کتابخانه ای، نرم افزارها، سخت افزارها و مواردی که قابلیت ثبت اختراع دارد) متعلق به دانشگاه صنعتی شریف است. هیچ شخصیت حقیقی یا حقوقی بدون کسب اجازه از دانشگاه صنعتی شریف حق فروش و ادعای مالکیت مادی یا معنوی بر آن یا ثبت اختراع از آن را ندارد. همچنین، کلیه حقوق مربوط به چاپ، تکثیر، نسخه برداری، ترجمه، اقتباس و نظائر آن در محیط های مختلف اعم از الکترونیکی، مجازی یا فیزیکی برای دانشگاه صنعتی شریف محفوظ است. نقل مطلب با ذکر ماخذ بلامانع است.

نگارنده: فیروزه ابریشمی

تاریخ:

امضاء:

استاد راهنما: دکتر جواد ابراهیمی بروجنی

تاریخ:

امضاء:

چکیده

کلیدواژه‌ها:

اول

فهرست مطالب

۲	۱ مقدمه
۲	۱-۱ اهمیت داده‌ها و ضرورت حفظ حریم خصوصی
۴	۱-۱-۱ محرمانگی تفاضلی
۷	۲-۱-۱ محرمانگی تفاضلی موضعی (α -LDP)
۸	۲-۱ کارهای پیشین و مرور ادبیات
۸	۱-۲-۱ مکانیزم‌های بنیادین و مثال‌های عملی
۱۰	۲-۲-۱ چالش بزرگ‌نمایی نویز و هزینه سودمندی
۱۱	۳-۱ اهداف پژوهش
۱۱	۴-۱ ساختار پایان‌نامه
۱۲	۲ پیش‌نیازها
۱۲	۱-۲ محرمانگی تفاضلی متمرکز (CDP)
۱۲	۱-۱-۲ مدل اعتماد و تعریف رسمی
۱۴	۲-۱-۲ مکانیزم‌های پایه در CDP
۱۵	۳-۱-۲ خواص کلیدی محرمانگی تفاضلی
۱۵	۴-۱-۲ محدودیت مدل متمرکز
۱۶	۲-۲ محرمانگی تفاضلی موضعی
۱۶	۱-۲-۲ مقدمه و گذار از مدل متمرکز

۱۶	۲-۲-۲ تعاریف رسمی و مدل‌های محاسباتی
۱۹	۳-۲-۲ مکانیزم‌های پایه در α -LDP
۲۲	۴-۲-۲ چالش سودمندی در مدل موضعی
۲۲	۳-۲ f -واگرایی‌ها
۲۳	۱-۳-۲ تعریف f -واگرایی
۲۳	۲-۳-۲ نمونه‌های مهم f -واگرایی
۲۴	۳-۳-۲ ارتباط f -واگرایی‌ها با یکدیگر
۲۴	۴-۲ مبانی آماری و نظریه اطلاعات
۲۵	۱-۴-۲ معیارهای فاصله اطلاعاتی
۲۵	۲-۴-۲ ریسک مینیماکس
۲۶	۵-۲ آزمون فرض آماری و روش تقلیل
۲۶	۱-۵-۲ آزمون فرض دودویی
۲۷	۲-۵-۲ تقلیل تخمین به آزمون (روش بسته‌بندی)
۲۷	۳-۵-۲ نامساوی‌های کران پایین
۲۹	۳ تحلیل‌های مبتنی بر انقباض و نرخ‌های مینیماکس
۲۹	۱-۳ مقدمه
۲۹	۲-۳ محرمانگی به عنوان انقباض اطلاعاتی
۳۱	۱-۲-۳ انقباض در فاصله واریانس کل
۳۱	۳-۳ تحلیل نرخ‌های مینیماکس با استفاده از انقباض
۳۲	۴-۳ مطالعه موردی: تخمین میانگین
۳۲	۵-۳ محدودیت‌های تحلیل کلاسیک
۳۴	۴ هم‌ارزی α -LDP و انقباض E_γ -واگرایی
۳۴	۱-۴ مقدمه و انگیزه

۳۵	۲-۴ معرفی E_γ - واگرایی
۳۵	۱-۲-۴ خواص هندسی
۳۵	۳-۴ قضیه هم‌ارزی اصلی
۳۷	۴-۴ بهبود کران‌های انقباض
۳۷	۵-۴ تعمیم به محرمانگی تقریبی $((\alpha, \delta)$ -LDP)
۳۸	۶-۴ کاربرد در تخمین توزیع گسسته
۳۸	۷-۴ انقباض قوی برای خانواده‌ی f - واگرایی‌ها
۳۸	۱-۷-۴ کران دقیق برای واگرایی کای-دو (χ^2)
۳۹	۲-۷-۴ تعمیم به سایر واگرایی‌ها
۴۰	۸-۴ نامساوی ون‌تریز خصوصی (Private van Trees Inequality)
۴۰	۹-۴ کاربردهای نوین و بهبود نرخ‌ها

۴۲ ۵ نتیجه‌گیری

۴۳ مراجع

۴۵ واژه‌نامه

۴۷ آ مطالب تکمیلی

فهرست جداول

فهرست تصاویر

- ۱-۲ مدل محرمانگی تفاضلی متمرکز با یک متصدی مورد اعتماد. ۱۳
- ۲-۲ مدل محرمانگی تفاضلی موضعی (α -LDP). نویز به صورت موضعی روی دستگاه کاربر
اضافه می شود. ۱۷

فصل ۱

مقدمه

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

۱-۱ اهمیت داده‌ها و ضرورت حفظ حریم خصوصی

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

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

^۱Big Data

^۲Machine Learning

^۳Data Analytics

^۴Privacy

^۵Sensitive 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 [۱۴].

- **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 [۱۳].

- **داده‌های جستجوی AOL:** در سال ۲۰۰۶، شرکت AOL تاریخچه‌ی جستجوی هزاران کاربر خود را منتشر کرد. اگرچه نام کاربران حذف شده بود، اما تحلیل محتوای جستجوها منجر به شناسایی هویت افراد شد (از جمله پرونده مشهور تلما آرنولد) که نشان داد حتی خودِ داده‌ها نیز می‌توانند به عنوان شناسه عمل کنند [۴].

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

⁶Utility

⁷Explicit Identifiers

⁸Anonymization

⁹Re-identification Attacks

¹⁰Auxiliary Information

¹¹Quasi-identifier

¹²Netflix Prize Data

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

۱-۱-۱ محرمانگی تفاضلی

در پاسخ به چالش‌های امنیتی و ناکارآمدی روش‌های سنتی گمنام‌سازی، دُورک و همکاران در سال ۲۰۰۶ مفهوم محرمانگی تفاضلی^{۱۳} را معرفی کردند [۱۰]. این چارچوب ریاضی دقیق، به جای تمرکز بر ویژگی‌های ظاهری داده‌ها (مانند حذف نام‌ها)، بر فرایند تولید خروجی تمرکز دارد و تضمین می‌کند که حضور یا عدم حضور یک فرد خاص در پایگاه‌داده، تأثیر ناچیزی بر خروجی نهایی الگوریتم داشته باشد.

پیش از آنکه به تعاریف صوری و ریاضی بپردازیم، ضروری است که درک عمیقی از چیستی محرمانگی و تمایز بنیادین آن با مفاهیم امنیتی کلاسیک پیدا کنیم. بسیاری از سوتفاهم‌ها در این حوزه ناشی از تمایز ندادن دو مفهوم امنیت داده (که قلمرو رمزنگاری^{۱۴} است) و محرمانگی داده (که هدف ماست) می‌باشد. رمزنگاری اساساً سازوکاری برای کنترل دسترسی^{۱۵} است و تضمین می‌کند که تنها افراد مجاز می‌توانند داده‌ها را ببینند؛ اما در برابر نشت اطلاعات از خروجی‌های مجاز سکوت می‌کند. تصور کنید یک پایگاه‌داده‌ی حساس پزشکی کاملاً رمزنگاری شده باشد و پژوهشگری مجاز، نتیجه‌ی یک تحلیل آماری ساده (مانند میانگین حقوق یا نرخ یک بیماری) را منتشر کند. رمزنگاری هیچ محافظتی در برابر استنتاج‌های ثانویه ارائه نمی‌دهد و مهاجم می‌تواند با ترکیب این خروجی مجاز با دانش پیشین^{۱۶} خود، اطلاعات خصوصی افراد را بازسازی کند. بنابراین، رمزنگاری شرط لازم است، اما برای حفظ محرمانگی کافی نیست؛ چرا که 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 پایگاه‌داده‌ی همسایه^{۱۸} (مانند D و 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 محرمانگی بتواند کاری کند که مهاجم با مشاهده‌ی خروجی، نتواند تشخیص دهد که این خروجی از کدام جهان آمده است، آنگاه حریم خصوصی کاربر x حفظ شده است.

برای رسیدن به این هدف، ما از الگوریتم‌های تصادفی^{۲۰} بهره می‌بریم. این الگوریتم‌ها با تزریق نویز کنترل‌شده، توزیع خروجی‌ها را بین دو مجموعه داده‌ی 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 برمی‌گرداند $(f(D) = \frac{1}{n} \sum_{i=1}^n x_i)$. مهاجم می‌تواند دو پرس‌وجو انجام دهد:

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

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

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

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

¹⁸Neighboring Datasets

¹⁹Indistinguishable

²⁰Randomized Algorithms

²¹Deterministic

²²Differencing Attack

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

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

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

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

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

این رابطه اثبات می‌کند که هیچ سطح محدودی از محرمانگی ($\varepsilon < \infty$) با توابع قطعی غیرثابت قابل دستیابی نیست. بنابراین، همان‌طور که در ادبیات موضوع تأکید شده است [۱۱]، 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 ε -DP در محاسبه‌ی آماره‌های توصیفی است. سازمان‌های آماری (مانند اداره سرشماری آمریکا) از این روش برای انتشار میانگین درآمد، سن یا جمعیت مناطق استفاده می‌کنند، بدون آنکه داده‌های فردی شهروندان به خطر بیفتد.
- **انتشار هیستوگرام^{۲۹}:** بسیاری از تحلیل‌ها نیازمند دانستن توزیع داده‌ها هستند. ε -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 به‌خاطر سپاری^{۳۱} داده‌های آموزشی وجود دارد. با استفاده از الگوریتم‌هایی نظیر DP-SGD، می‌توان مدل‌هایی آموزشی 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 ϵ -DP برای جمع‌آوری آمارهای رفتاری (مانند پربازدیدترین وبسایت‌ها یا مکان‌های پرتدد) استفاده می‌کنند تا بدون ردیابی لحظه‌ای کاربران، کیفیت خدمات خود را بهبود بخشند (مانند مکانیزم RAPPOR در گوگل کروم).

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

۲-۱-۱ محرمانگی تفاضلی موضعی (α -LDP)

اگرچه محرمانگی تفاضلی متمرکز (CDP) استاندارد طلایی حفاظت از داده‌ها محسوب می‌شود، اما پاشنه‌ی آشیل آن در فرضیه‌ی وجود یک متصدی مورد اعتماد^{۳۲} نهفته است که به تمام داده‌های خام دسترسی دارد. این مدل در دنیای واقعی با چالش‌های امنیتی و حقوقی جدی روبروست؛ چرا که تجربه نشان داده است اعتماد کامل به سرورهای مرکزی، حتی در صورت مدیریت توسط نهادهای بزرگ فناوری، همواره در معرض تهدید است. یکی از این خطرات، نفوذهای خارجی و سرقت انبوه داده‌هاست؛ به طوری که حتی پیشرفته‌ترین دیوارهای آتش^{۳۳} نیز در برابر حملات پیچیده آسیب‌پذیرند. در چنین شرایطی، اگر داده‌ها به صورت خام ذخیره شده باشند، نشت اطلاعاتی مانند آنچه در واقعه‌ی Equifax رخ داد، مکانیزم‌های محرمانگی تفاضلی مرکزی را عملاً بی‌فایده می‌کند؛ زیرا مهاجم با دور زدن مکانیزم، مستقیماً به مخزن داده‌های حساس دست می‌یابد [۱۵].

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

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 محرمانگی تفاضلی موضعی (LDP) است که در آن فرآیند خصوصی‌سازی پیش از خروج داده از دستگاه کاربر انجام می‌شود.

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

۲-۱ کارهای پیشین و مرور ادبیات

۱-۲-۱ مکانیزم‌های بنیادین و مثال‌های عملی

برای درک عمیق‌تر چگونگی عملکرد α -LDP در عمل، ابتدا به بررسی مکانیزم بنیادین این حوزه می‌پردازیم و سپس نحوه‌ی تکامل آن را در پروتکل‌های صنعتی مدرن بررسی می‌کنیم.

• پاسخ تصادفی: سنگ بنای α -LDP

پایه‌ای‌ترین مکانیزم در ادبیات α -LDP، «پاسخ تصادفی»^{۳۵} است که نخستین بار در سال ۱۹۶۵ توسط وارنر برای انجام نظرسنجی‌های حساس (نظیر مصرف مواد مخدر یا عقاید سیاسی خاص) معرفی شد [۱۷]. هدف این مکانیزم، ایجاد یک «انکارپذیری باورپذیر»^{۳۶} برای پاسخ‌دهنده است. سازوکار کلاسیک این روش برای یک پرسش با پاسخ «بله/خیر» به صورت زیر است: فرض کنید از کاربر i خواسته می‌شود که ویژگی حساس $X_i \in \{0, 1\}$ را گزارش کند. کاربر به جای پاسخ مستقیم، طبق دستورالعمل زیر عمل می‌کند:

۱. یک سکه را پرتاب می‌کند. (می‌تواند سکه غیرمنصفانه^{۳۷} باشد)

³⁵Randomized Response (RR)

³⁶Plausible Deniability

³⁷Unfair

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

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

در این سناریو، حتی اگر سرور پاسخ «بله» را دریافت کند، با قطعیت نمی‌داند که آیا کاربر واقعاً دارای ویژگی X بوده است (شیر آمده) یا صرفاً به دلیل تصادف (خط آمدن سکه‌ی اول و شیر آمدن سکه‌ی دوم) این پاسخ را ارسال کرده است. با این حال، از آنجایی که احتمالات سکه‌ها مشخص است، سرور می‌تواند با جمع‌آوری تعداد زیادی از پاسخ‌ها (n بسیار بزرگ)، اثر نویز را به صورت آماری حذف کرده و توزیع واقعی جامعه را با خطا تخمین بزند.

• گذار به مقیاس صنعتی: چالش‌ها و راه‌حل‌ها

اما پاسخ تصادفی تنها راهکاری برای ایجاد محرمانگی در داده‌های دودویی است، و در کاربردهای مدرن با چالش دامنه‌ی بسیار بزرگ^{۳۸} روبروست. شرکت‌های بزرگ فناوری نیاز دارند داده‌هایی نظیر «آدرس‌های اینترنتی بازدید شده» یا «کلمات جدید تایپ‌شده» را جمع‌آوری کنند که دامنه‌ی آن‌ها (\mathcal{X}) می‌تواند شامل میلیون‌ها حالت باشد. اعمال مستقیم RR در این حالات منجر به نویز بسیار زیاد و کاهش شدید سودمندی می‌شود. در ادامه، راهکارهای اتخاذ شده توسط بزرگ‌ترین شرکت‌های فناوری را مرور می‌کنیم:

– **گوگل و پروتکل RAPPOR:** در سال ۲۰۱۴، گوگل برای جمع‌آوری آمار تنظیمات مرورگر کروم و شناسایی بدافزارها، پروتکل RAPPOR^{۳۹} را معرفی کرد [۱۶]. چالش اصلی گوگل، جمع‌آوری رشته‌های متنی^{۴۰} بود. راه‌حل آن‌ها ترکیب پاسخ تصادفی با فیلترهای بلوم^{۴۱} 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. راه‌حل اپل استفاده از تکنیک‌های مبتنی بر طرح‌ریزی^{۴۲} و تبدیل‌های ریاضی مانند تبدیل هادامارد^{۴۳}

³⁸High-Dimensional Domain

³⁹Randomized Aggregatable Privacy-Preserving Ordinal Response

⁴⁰String

⁴¹Bloom Filters

⁴²Sketching

⁴³Hadamard Transform

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 [۱۶].

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

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

۱-۲-۲ چالش بزرگ‌نمایی نویز و هزینه سودمندی

یکی از مهم‌ترین مباحث در ادبیات α -LDP که انگیزه‌ی اصلی پژوهش حاضر نیز می‌باشد، مسئله‌ی هزینه‌ی سودمندی است. بیایید تفاوت ریاضی دو مدل را بررسی کنیم. در مدل متمرکز، متصدی ابتدا آمار واقعی را محاسبه می‌کند و سپس یک بار نویز به آن اضافه می‌کند. از آن‌جا که نویز تنها یک بار اضافه می‌شود، خطا با افزایش تعداد کاربران (n) به سرعت و با نرخ $O(1/n)$ کاهش می‌یابد. اما در مدل موضعی، هر کاربر باید به صورت مستقل نویز اضافه کند. وقتی میانگین این داده‌های نویزدار محاسبه می‌شود، طبق قانون اعداد بزرگ، واریانس نویزها با هم جمع می‌شود. ریاضیات نشان می‌دهد که خطای نهایی در این حالت با نرخ بسیار کندتر یعنی $O(1/\sqrt{n})$ کاهش می‌یابد [۸].

این تفاوت چشم‌گیر به این معناست که اگر در مدل متمرکز با هزار داده به دقت مطلوب می‌رسیدیم، در مدل موضعی ممکن است به یک میلیون داده نیاز داشته باشیم.

انواع پروتکل‌های موضعی

پروتکل‌های α -LDP را می‌توان بر اساس نحوه‌ی ارتباط کاربران با سرور به دو دسته‌ی اصلی تقسیم کرد:

۱. **پروتکل‌های غیرتعاملی^{۴۴}:** در این حالت، هر کاربر i مکانیزم M را مستقل اجرا می‌کند و پیام خود را یک‌بار برای همیشه به سرور می‌فرستد. هیچ ارتباطی بین کاربران یا بازخوردی از سرور وجود

⁴⁴Non-interactive

ندارد. مزیت این روش سادگی پیاده‌سازی و مقیاس‌پذیری بالا است.

۲. پروتکل‌های تعاملی^{۴۵}: در این روش، کاربران به نوبت با سرور ارتباط برقرار می‌کنند. کاربر بعدی می‌تواند قبل از ارسال داده‌ی خود، خلاصه‌ای از داده‌های کاربران قبلی را مشاهده کند و نویز خود را هوشمندانه‌تر تنظیم نماید. اگرچه این روش از نظر تئوری می‌تواند دقت را در مسائل پیچیده بهبود بخشد، اما به دلیل تأخیر شبکه و پیچیدگی هماهنگی، در سیستم‌های واقعی کم‌تر استفاده می‌شود [۱۲].

در فصل‌های آینده، ما عمدتاً بر تحلیل حدود پایین و نرخ‌های مینیماکس در این پروتکل‌ها تمرکز خواهیم کرد و نشان خواهیم داد که چگونه ابزارهای مدرن نظریه اطلاعات می‌توانند درک ما را از محدودیت‌های ذاتی این سیستم‌ها دگرگون کنند.

۳-۱ اهداف پژوهش

۴-۱ ساختار پایان‌نامه

⁴⁵Interactive

فصل ۲

پیش‌نیازها

۱-۲ محرمانگی تفاضلی متمرکز (CDP)

مفهوم محرمانگی تفاضلی^۱ یا به اختصار ϵ -DP، اولین بار توسط دژرک و همکاران [۱۰] 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 مجموعه داده است.

۱-۱-۲ مدل اعتماد و تعریف رسمی

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

برای تعریف رسمی محرمانگی تفاضلی، مفاهیم الگوریتم (مکانیزم) تصادفی، فاصله‌ی بین دو

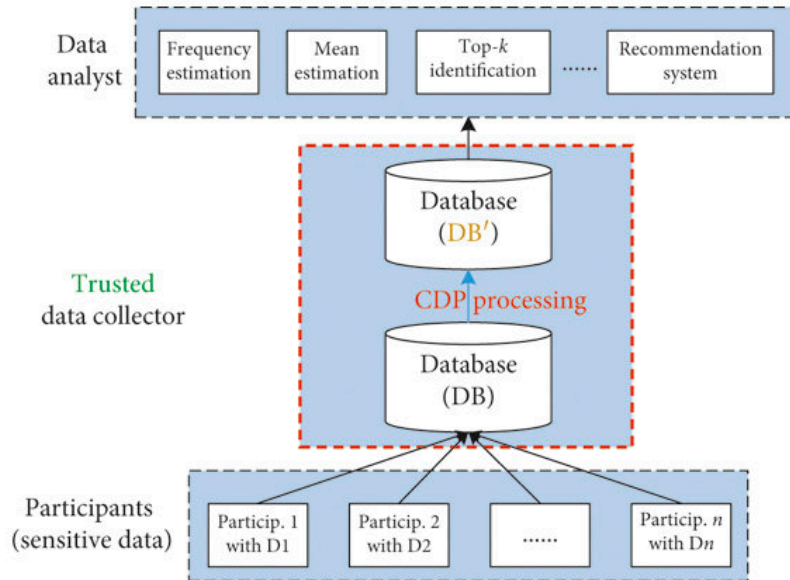
¹Differential Privacy

²Centralized

³Trusted Curator

⁴Randomized Mechanism

⁵Query



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

پایگاه داده^۶ و سپس همسایگی^۷ را تعریف می‌کنیم.

تعریف ۱-۲ (الگوریتم تصادفی) طبق تعریف کتاب اضافه شود. عکسش توهمین فولدر هست. همچنین تعریف \mathcal{X} و پایگاه داده هم اضافه شود.

تعریف ۲-۲ (فاصله بین دو پایگاه داده) نرم ℓ_1 یک پایگاه داده D به صورت $\|D\|_1$ نمایش و به صورت زیر تعریف می‌شود:

$$\|D\|_1 = \sum_{i=1}^{|\mathcal{X}|} |x_i|$$

فاصله ℓ_1 بین دو پایگاه داده D_1 و D_2 برابر است با $\|D_1 - D_2\|_1$.

تعریف ۳-۲ (پایگاه داده‌های همسایه) دو پایگاه داده D_1 و D_2 را همسایه^۸ می‌گوییم (و با $D_1 \sim D_2$ نشان می‌دهیم) اگر $\|D_1 - D_2\|_1 \leq 1$.

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

⁶Database

⁷Neighboring

⁸Adjacent

تعریف ۴-۲ (ϵ -محرم‌انگی تفاضلی (ϵ -DP)) یک مکانیزم تصادفی \mathcal{M} ، تعریف ϵ -محرم‌انگی تفاضلی را برآورده می‌سازد، اگر برای هر دو مجموعه داده‌ی همسایه‌ی \mathcal{D}_1 و \mathcal{D}_2 و برای هر زیرمجموعه S از خروجی‌های ممکن ($\text{Range}(\mathcal{M})$)، داشته باشیم:

$$\Pr[\mathcal{M}(\mathcal{D}_1) \in S] \leq \exp(\epsilon) \cdot \Pr[\mathcal{M}(\mathcal{D}_2) \in S] \quad (۱-۲)$$

گاهی اوقات، یک تعریف انعطاف‌پذیرتر به نام (ϵ, δ) -DP نیز استفاده می‌شود که اجازه‌ی یک احتمال شکست کوچک δ را می‌دهد:

$$\Pr[\mathcal{M}(\mathcal{D}_1) \in S] \leq \exp(\epsilon) \cdot \Pr[\mathcal{M}(\mathcal{D}_2) \in S] + \delta \quad (۲-۲)$$

۲-۱-۲ مکانیزم‌های پایه در CDP

برای دستیابی به ϵ -DP، باید به پاسخ دقیق پرس‌وجو «نویز»^۹ اضافه کنیم. میزان نویز به حساسیت^{۱۰} پرس‌وجو بستگی دارد.

تعریف ۵-۲ (حساسیت سراسری) برای یک تابع f ، حساسیت سراسری ℓ_1 ($\Delta_1 f$) و ℓ_2 ($\Delta_2 f$) به صورت زیر تعریف می‌شوند:

$$\Delta_1 f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1 \quad (۳-۲)$$

$$\Delta_2 f = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_2 \quad (۴-۲)$$

سه مکانیزم اساسی برای دستیابی به CDP عبارتند از:

- مکانیزم لاپلاس^{۱۱}: برای توابع عددی، با افزودن نویز از توزیع لاپلاس متناسب با حساسیت ℓ_1 ، می‌توان به ϵ -DP دست یافت:

$$\mathcal{M}(\mathcal{D}) = f(\mathcal{D}) + \text{Lap}\left(\frac{\Delta_1 f}{\epsilon}\right) \quad (۵-۲)$$

- مکانیزم گوسی^{۱۲}: این مکانیزم اغلب زمانی استفاده می‌شود که حساسیت ℓ_2 تابع کمتر از حساسیت ℓ_1 باشد. در اینجا نویز از توزیع نرمال (گوسی) افزوده می‌شود:

$$\mathcal{M}(\mathcal{D}) = f(\mathcal{D}) + \mathcal{N}(0, \sigma^2) \quad (۶-۲)$$

^۹Noise

^{۱۰}Sensitivity

^{۱۱}Laplace

^{۱۲}Gaussian

که در آن $\sigma \geq \sqrt{2 \ln(1/25/\delta)} \cdot \frac{\Delta f}{\epsilon}$ است. برخلاف مکانیزم لاپلاس، این مکانیزم تنها (ϵ, δ) -DP را (با $\delta > 0$) تضمین می‌کند.

- مکانیزم نمایشی^{۱۳}: برای خروجی‌های غیر عددی (دسته‌ای)، از یک «تابع امتیاز» $q(D, r)$ استفاده می‌شود. این مکانیزم خروجی r را با احتمالی متناسب با امتیاز آن برمی‌گرداند:

$$\Pr[\mathcal{M}(D) = r] \propto \exp\left(\frac{\epsilon \cdot q(D, r)}{2\Delta q}\right) \quad (7-2)$$

۳-۱-۲ خواص کلیدی محرمانگی تفاضلی

قدرت چارچوب DP در خواص ترکیبی آن نهفته است:

- مصونیت در برابر پس‌پردازش^{۱۴}: انجام هرگونه محاسبات بر روی خروجی یک مکانیزم ϵ -DP (بدون دسترسی مجدد به داده‌های اصلی)، نمی‌تواند سطح محرمانگی را کاهش دهد.
- ترکیب‌پذیری^{۱۵}: اگر چندین مکانیزم ϵ -DP را اجرا کنیم، بودجه‌های محرمانگی جمع می‌شوند.

– ترکیب‌پذیری پایه‌ای: اجرای k مکانیزم ϵ_i -DP منجر به $\sum \epsilon_i$ -DP می‌شود.

– ترکیب‌پذیری پیشرفته: با پذیرش یک δ کوچک، می‌توان نشان داد که بودجه کل با نرخ \sqrt{k} رشد می‌کند (نه k).

- محرمانگی گروهی^{۱۶}: محرمانگی تفاضلی به طور طبیعی برای گروه‌هایی از افراد نیز صادق است. اگر دو پایگاه داده در k رکورد با هم متفاوت باشند، تضمین محرمانگی به صورت $k\epsilon$ -DP برقرار خواهد بود. این یعنی با افزایش اندازه گروه، تضمین محرمانگی به صورت خطی تضعیف می‌شود.

۴-۱-۲ محدودیت مدل متمرکز

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

¹³Exponential

¹⁴Post-processing Immunity

¹⁵Composition

¹⁶Group Privacy

۲-۲ محرمانگی تفاضلی موضعی

در فصل گذشته، مبانی نظری حریم خصوصی متمرکز (CDP) و ابزارهای آماری لازم برای تحلیل آن را مرور کردیم. در این فصل، به طور اختصاصی به چارچوب محرمانگی تفاضلی موضعی^{۱۷} (α -LDP) می‌پردازیم. این مدل، که امروزه در سیستم‌های توزیع‌شده و جمع‌آوری داده‌های بزرگ‌مقیاس کاربرد فراوان دارد، پارادایم اعتماد را از «سرور مرکزی» به «کاربر نهایی» تغییر می‌دهد.

۱-۲-۲ مقدمه و گذار از مدل متمرکز

همان‌طور که در بخش ۱-۲ دیدیم، مدل متمرکز نیازمند وجود یک متصدی مورد اعتماد^{۱۸} است که به داده‌های خام دسترسی داشته باشد. اگرچه این مدل دقت آماری بالایی را فراهم می‌کند، اما در دنیای واقعی با چالش‌های امنیتی و حقوقی جدی روبروست:

- **نقطه شکست مرکزی^{۱۹}:** سرور مرکزی هدف جذابی برای مهاجمان است. نشت اطلاعات از سرور (چه بر اثر هک و چه بر اثر خطای انسانی) حریم خصوصی تمام کاربران را به خطر می‌اندازد.
- **عدم اعتماد کاربران:** در بسیاری از کاربردها (مانند جمع‌آوری داده‌های پزشکی یا تاریخچه مرورگر)، کاربران تمایلی ندارند داده‌های حساس خود را حتی به یک سرور «مطمئن» بسپارند.

در پاسخ به این چالش‌ها، مدل محرمانگی تفاضلی موضعی مطرح شد. در α -LDP، فرآیند خصوصی‌سازی (افزودن نویز) به سمت کلاینت (کاربر) منتقل می‌شود. به این معنا که داده‌ها قبل از ترک دستگاه کاربر، نویزدار می‌شوند و سرور تنها به داده‌های بی‌نام و نویزدار دسترسی دارد (شکل ۲-۲). این رویکرد توسط شرکت‌های بزرگ فناوری برای جمع‌آوری داده‌های تله‌متری پذیرفته شده است. برای مثال، گوگل از مکانیزم RAPPOR در مرورگر کروم، و اپل و مایکروسافت از روش‌های مشابهی برای جمع‌آوری داده‌های آماری از سیستم‌عامل‌های خود استفاده می‌کنند.

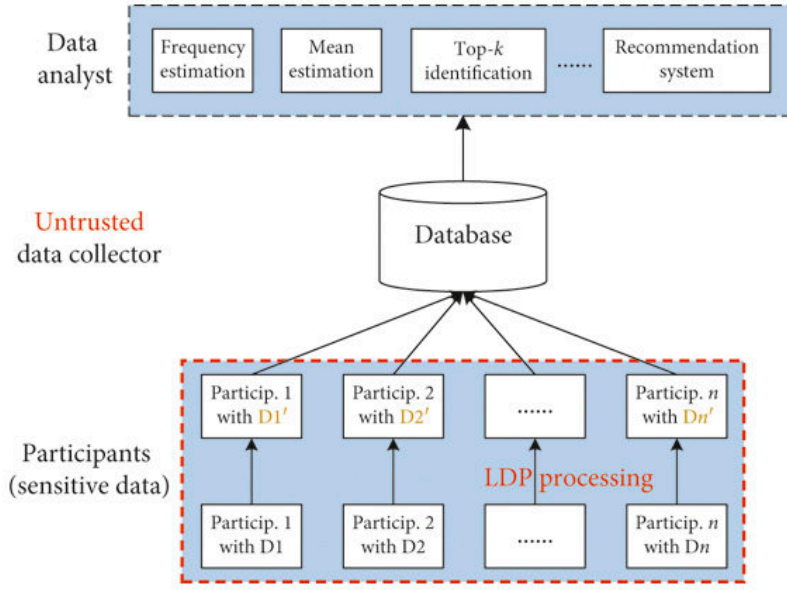
۲-۲-۲ تعاریف رسمی و مدل‌های محاسباتی

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.

¹⁷Local Differential Privacy (LDP)

¹⁸Trusted Curator

¹⁹Single Point of Failure



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

تعریف α -LDP

هسته‌ی اصلی این مدل، «تصادفی‌ساز موضعی» است.

تعریف ۲-۶ (تصادفی‌ساز موضعی^{۲۰}) یک مکانیزم تصادفی $M: \mathcal{X} \rightarrow \mathcal{Z}$ را یک تصادفی‌ساز موضعی می‌نامیم که ورودی $x \in \mathcal{X}$ را می‌گیرد و خروجی $z \in \mathcal{Z}$ را بر اساس توزیع احتمال شرطی $Q(z|x)$ تولید می‌کند.

شرط محرمانگی در اینجا تضمین می‌کند که با مشاهده‌ی خروجی z ، تمایز قائل 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 ۲-۷ (α -محرمانگی تفاضلی موضعی) یک مکانیزم M دارای α -محرمانگی تفاضلی موضعی (α -LDP) است اگر برای تمام جفت ورودی‌های $x, x' \in \mathcal{X}$ و هر زیرمجموعه از خروجی‌ها $S \subseteq \mathcal{Z}$ داشته باشیم:

$$\sup_S \frac{\Pr[M(x) \in S]}{\Pr[M(x') \in S]} \leq e^\alpha \quad (۸-۲)$$

(نکته: در متون آماری مانند [۸] 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:

$$\sup_{x, x' \in \mathcal{X}} D_\infty(Q(\cdot|x) || Q(\cdot|x')) \leq \alpha \quad (9-2)$$

تعمیم‌ها و خواص

علاوه بر تعریف استاندارد α -LDP (معادله ۸-۲)، دو مفهوم دیگر نیز در تحلیل‌های نظری و طراحی مکانیزم‌ها اهمیت دارند: محرمانگی تقریبی و خاصیت ترکیب.

تعریف ۸-۲ ((α, δ) -محرمانگی تفاضلی موضعی) یک مکانیزم تصادفی \mathcal{M} دارای محرمانگی تفاضلی موضعی تقریبی یا (α, δ) -LDP است اگر برای هر دو ورودی $x, x' \in \mathcal{X}$ و هر زیرمجموعه خروجی $S \subseteq \mathcal{Z}$ داشته باشیم:

$$\Pr[\mathcal{M}(x) \in S] \leq e^\alpha \cdot \Pr[\mathcal{M}(x') \in S] + \delta \quad (10-2)$$

این تعریف (که در [۱۶] 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 فصل قبل معرفی شد) دارد.

قضیه ۱-۲ (ترکیب ترتیبی^{۲۱}) اگر یک کاربر در k مرحله‌ی مختلف در پروتکل‌های $\mathcal{M}_1, \dots, \mathcal{M}_k$ شرکت کند که هر کدام به ترتیب دارای بودجه‌ی حریم خصوصی α_i باشند، آنگاه کل فرآیند دارای 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 [۸] پروتکل‌های موضعی را به دو دسته تقسیم می‌کنند:

۱. **پروتکل‌های غیرتعاملی^{۲۲}**: در این حالت، خروجی هر کاربر Z_i تنها به ورودی خودش X_i وابسته است و مستقل از داده‌ها یا خروجی‌های سایر کاربران تولید می‌شود.

$$Z_i = \mathcal{M}_i(X_i) \quad (11-2)$$

²²Non-interactive

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

۲. پروتکل‌های تعاملی (ترتیبی)^{۲۳}: 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، کانال ارتباطی Q_i می‌تواند به صورت پویا بر اساس تاریخچه تغییر کند:

$$Z_i \sim Q_i(\cdot | X_i, Z_1, \dots, Z_{i-1}) \quad (۱۲-۲)$$

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

۳-۲-۲ مکانیزم‌های پایه در α -LDP

در این بخش، مکانیزم‌های بنیادین را معرفی می‌کنیم که برای تحقق محرمانگی 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»^{۲۴} است که دهه‌ها پیش از تعریف رسمی α -LDP توسط وارنر [۱۷] برای نظرسنجی‌های حساس معرفی شد. فرض کنید دامنه ورودی دودویی باشد ($\mathcal{X} = \{0, 1\}$).

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

$$\Pr[z = x] = p, \quad \Pr[z \neq x] = 1 - p \quad (۱۳-۲)$$

برای اینکه این مکانیزم شرط α -LDP را برآورده کند، طبق تعریف ۷-۲ باید نسبت احتمالات حداکثر e^α باشد:

$$\frac{\Pr[z = 1 | x = 1]}{\Pr[z = 1 | x = 0]} = \frac{p}{1 - p} \leq e^\alpha \quad (۱۴-۲)$$

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

$$p = \frac{e^\alpha}{1 + e^\alpha} \quad (۱۵-۲)$$

^{۲۳}Sequential/Interactive

^{۲۴}Randomized Response (RR)

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

$$\text{Var}[\hat{x}] = \frac{e^\alpha}{(e^\alpha - 1)^2} \quad (16-2)$$

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

پاسخ تصادفی تعمیم‌یافته (GRR)

زمانی که دامنه ورودی شامل $k > 2$ عنصر باشد ($\mathcal{X} = \{1, \dots, k\}$)، از نسخه تعمیم‌یافته پاسخ تصادفی^{۲۵} استفاده می‌شود [۱۲]. این روش تعمیم مستقیم RR برای دامنه‌های گسسته است.

در این مکانیزم، برای ورودی x :

$$\Pr[\mathcal{M}(x) = z] = \begin{cases} p & \text{if } z = x \\ q & \text{if } z \neq x \end{cases} \quad (17-2)$$

از آنجا که مجموع احتمالات باید ۱ باشد، داریم $p + (k - 1)q = 1$. 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:

$$p = \frac{e^\alpha}{e^\alpha + k - 1}, \quad q = \frac{1}{e^\alpha + k - 1} \quad (18-2)$$

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.

مکانیزم‌های مبتنی بر کدگذاری یگانی (UE)

برای غلبه بر مشکل کاهش دقت GRR در دامنه‌های بزرگ، خانواده‌ای از مکانیزم‌ها تحت عنوان «کدگذاری یگانی»^{۲۶} توسعه یافته‌اند. این رویکرد اساس پروتکل مشهور RAPOR گوگل را تشکیل می‌دهد [۱۶].

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

²⁵Generalized Randomized Response (GRR)

²⁶Unary Encoding (UE)

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:

$$\Pr[z_i = 1] = \begin{cases} p & \text{if } v_i = 1 \\ q & \text{if } v_i = 0 \end{cases} \quad (19-2)$$

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:

$$p = \frac{e^{\alpha/2}}{e^{\alpha/2} + 1}, \quad q = \frac{1}{e^{\alpha/2} + 1} \quad (20-2)$$

مزیت SUE این است که واریانس تخمین برای هر آیت مستقل از تعداد کل آیت‌ها (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 [۱۶] نشان دادند که برای تخمین فراوانی در دامنه‌های بزرگ، نیازی به متقارن بودن مکانیزم نیست. در روش OUE^{۲۷}، پارامترها به گونه‌ای تنظیم می‌شوند که اطلاعات بیت‌های ۱ (سیگنال اصلی) با بیشترین دقت حفظ شود ($p = 1/2$) و نویز روی بیت‌های ۰ (که تعدادشان زیاد است) کنترل شود:

$$p = \frac{1}{2}, \quad q = \frac{1}{e^{\alpha} + 1} \quad (21-2)$$

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.

²⁷Optimized Unary Encoding

مکانیزم لاپلاس موضعی

برای داده‌های عددی (مثلاً $x \in [-1, 1]$)، 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 $\mathcal{X} = [-1, 1]$ باشد، حساسیت برابر است با:

$$\Delta = \max_{x, x'} |x - x'| = |1 - (-1)| = 2 \quad (22-2)$$

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

$$\mathcal{M}_{Lap}(x) = x + \eta, \quad \eta \sim \text{Lap}\left(\frac{2}{\alpha}\right) \quad (23-2)$$

نکته مهمی که دوجی و همکاران [۸] به آن اشاره کرده‌اند این است که برخلاف مدل متمرکز، مکانیزم لاپلاس در مدل موضعی برای ابعاد بالا ($d > 1$) زیر-بهینه^{۲۸} است و نرخ خطای آن بدتر از مکانیزم‌های پیشرفته‌تر (مانند نمونه‌برداری هاپیرکیوب) 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^{۲۹} 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 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 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.

۳-۲-۲ f -واگرایی‌ها

در بخش‌های قبلی، ما مکانیزم‌های محرمانگی 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

²⁸Sub-optimal

²⁹Utility

این «شبهات» یا «فاصله» بین توزیع‌های احتمالی را معرفی می‌کنیم. این ابزار، خانواده‌ی f -واگرایی‌ها^{۳۰} 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 [۶] و به طور همزمان توسط علی و سیلوی [۱] 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 ۹-۲ (f -واگرایی) فرض کنید P و Q دو توزیع احتمال باشند به طوری که P نسبت به Q مطلقاً پیوسته^{۳۱} باشد. فرض کنید p و q توابع چگالی احتمال (یا توابع جرم احتمال) آن‌ها باشند. برای هر تابع محذب $f: (0, \infty) \rightarrow \mathbb{R}$ که $f(1) = 0$ باشد، f -واگرایی P از Q به صورت زیر تعریف می‌شود:

$$D_f(P||Q) = \int q(x) f\left(\frac{p(x)}{q(x)}\right) dx \quad (24-2)$$

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

$$D_f(P||Q) = \sum_{x \in \mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) \quad (25-2)$$

تابع f «ژنراتور» (تولیدکننده) واگرایی نامیده می‌شود و با انتخاب f ‌های متفاوت، می‌توان معیارهای فاصله یا 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^{۳۲}: این معیار که به آن آنتروپی نسبی^{۳۳} نیز گفته می‌شود، با انتخاب تابع $f(t) = t \log t$ به دست می‌آید:

$$D_{KL}(P||Q) = \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right) \quad (26-2)$$

^{۳۰} f -divergences

^{۳۱} Absolutely Continuous

^{۳۲} Kullback-Leibler (KL) Divergence

^{۳۳} Relative Entropy

- فاصله‌ی واریانس کل^{۳۴}: فاصله‌ی TV (که اغلب با $\Delta(P, Q)$ یا d_{TV} نشان داده می‌شود) ارتباط نزدیکی با f -واگرایی دارد و با انتخاب $f(t) = \frac{1}{t}|t - 1|$ حاصل می‌شود:

$$d_{TV}(P, Q) = \frac{1}{2} \sum_x |p(x) - q(x)| \quad (27-2)$$

- واگرایی کای-دو^{۳۵}: این معیار آماری با انتخاب $f(t) = (t - 1)^2$ به دست می‌آید:

$$\chi^2(P||Q) = \sum_x \frac{(p(x) - q(x))^2}{q(x)} \quad (28-2)$$

- فاصله‌ی هلینجر (مربع)^{۳۶}: این فاصله با انتخاب $f(t) = (\sqrt{t} - 1)^2$ (یا معادل آن $f(t) = \frac{1}{4}(\sqrt{t} - 1)^2$) حاصل می‌شود:

$$H^2(P, Q) = \sum_x (\sqrt{p(x)} - \sqrt{q(x)})^2 \quad (29-2)$$

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^{۳۷} است که ارتباط بین واگرایی KL و فاصله‌ی TV را نشان می‌دهد:

$$d_{TV}(P, Q)^2 \leq \frac{1}{4} D_{KL}(P||Q) \quad (30-2)$$

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 ۸-۲ تعریف شد) مستقیماً منجر به ایجاد یک کران بالا بر روی f -واگرایی‌های مختلف بین توزیع‌های خروجی می‌شود.

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 خطا در حضور

³⁴Total Variation (TV) Distance

³⁵Chi-Squared (χ^2) Divergence

³⁶Squared-Hellinger Distance

³⁷Pinsker's Inequality

محدودیت‌های محرمانگی استفاده می‌شود. این تعاریف و قضایا عمدتاً بر اساس چارچوب ارائه‌شده در [۸] تدوین شده‌اند.

۲-۴-۱ 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 ۲-۱۰ (واگرایی کولبک-لایبلر) واگرایی کولبک-لایبلر (KL) بین دو توزیع P و Q به صورت زیر تعریف می‌شود:

$$D_{KL}(P||Q) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} d\mu(x) \quad (۳۱-۲)$$

تعریف ۲-۱۱ (فاصله‌ی واریانس کل) فاصله‌ی واریانس کل^{۳۸} بین دو توزیع P و Q به صورت زیر تعریف می‌شود:

$$\|P - Q\|_{TV} = \sup_{S \in \sigma(\mathcal{X})} |P(S) - Q(S)| = \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)| d\mu(x) \quad (۳۲-۲)$$

تعریف ۲-۱۲ (اطلاعات متقابل) اگر X و V دو متغیر تصادفی باشند، اطلاعات متقابل^{۳۹} بین آن‌ها به صورت امید ریاضی واگرایی KL بین توزیع شرطی و توزیع حاشیه‌ای تعریف می‌شود:

$$I(X; V) = D_{KL}(P_{X,V} || P_X \otimes P_V) = \mathbb{E}_V [D_{KL}(P_{X|V} || P_X)] \quad (۳۳-۲)$$

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 $\theta(P)$ 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).

نرخ مینیماکس^{۴۰}، کمترین خطای ممکن است که یک تخمین‌گر در بدترین سناریو (بدترین توزیع P در کلاس \mathcal{P}) متحمل می‌شود.

³⁸Total Variation Distance

³⁹Mutual Information

⁴⁰Minimax Rate

تعریف ۲-۱۳ (نرخ مینیماکس) برای یک کلاس از توزیع‌ها \mathcal{P} و پارامتر θ ، نرخ مینیماکس \mathfrak{M}_n به صورت زیر تعریف می‌شود:

$$\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho) = \inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\Phi(\rho(\hat{\theta}(Z^n), \theta(P)))] \quad (۲-۳۴)$$

که در آن اینفیمم روی تمام تخمین‌گرهای ممکن $\hat{\theta}$ گرفته می‌شود.

در حالتی که محدودیت محرمانگی تفاضلی موضعی با پارامتر α وجود داشته باشد، نرخ مینیماکس خصوصی (α -Private Minimax Rate) با در نظر گرفتن اینفیمم روی تمام مکانیزم‌های کانال Q که شرط α -LDP را برآورده می‌کنند، تعریف می‌شود [۸].

۲-۵ 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^{۴۱} تقلیل دهیم. ایده اصلی این است: اگر بتوانیم بین چند مقدار گسسته از پارامتر با دقت بالا تمایز قائل شویم، قطعاً نمی‌توانیم پارامتر را در فضای پیوسته با خطای کم تخمین بزنیم.

۲-۵-۱ 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 می‌شود:

$$P_{err}(\psi) = \Pr_{H_0}(\psi(Z) = 1) + \Pr_{H_1}(\psi(Z) = 0) \quad (۲-۳۵)$$

⁴¹Hypothesis Testing

لم نیمن-پیرسون^{۴۲} نشان می‌دهد که کمترین خطای ممکن برای هر آزمون دودویی، مستقیماً با فاصله‌ی واریانس کل (d_{TV}) بین دو توزیع ارتباط دارد:

$$\inf_{\psi} P_{err}(\psi) = 1 - \|P_0 - P_1\|_{TV} \quad (۳۶-۲)$$

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

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

برای استفاده از ابزارهای آزمون فرض در مسئله‌ی تخمین نرخ مینیماکس (معادله ۲-۳۴)، از تکنیک گسسته‌سازی فضای پارامتر Θ استفاده می‌کنیم. این روش شامل مراحل زیر است:

۱. ساخت مجموعه‌ی بسته‌بندی^{۴۳}: مجموعه‌ای متناهی از پارامترها $\Theta \subset \mathcal{V} = \{\theta_1, \dots, \theta_M\}$ را انتخاب می‌کنیم به طوری که از یکدیگر فاصله‌ی معناداری داشته باشند. به طور دقیق‌تر، اگر ρ متریک خطا باشد، برای هر $i \neq j$ باید داشته باشیم $\rho(\theta_i, \theta_j) \geq 2\delta$.

۲. تعریف مسئله‌ی آزمون: فرض می‌کنیم طبیعت^{۴۴} 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 خطاها: اگر یک تخمین‌گر $\hat{\theta}$ وجود داشته باشد که خطای تخمین آن با احتمال بالا کمتر از δ باشد، می‌توانیم از آن برای حل مسئله‌ی آزمون فرض استفاده کنیم (با انتخاب نزدیک‌ترین θ_i به $\hat{\theta}$). بنابراین، کران پایین روی خطای آزمون فرض، یک کران پایین برای خطای تخمین ایجاد می‌کند:

$$\mathfrak{M}_n(\theta(\mathcal{P})) \geq \Phi(\delta) \cdot \inf_{\psi} \Pr(\psi(Z^n) \neq V) \quad (۳۷-۲)$$

۳-۵-۲ نامساوی‌های کران پایین

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

⁴²Neyman-Pearson Lemma

⁴³Packing Set

⁴⁴Nature

قضیه ۲-۲ (نامساوی لو کم^{۴۵}) این روش برای آزمون بین دو توزیع P_1 و P_2 استفاده می‌شود. کمینه احتمال خطا با استفاده از فاصله‌ی واریانس کل (رابطه ۲-۳۲) کران‌دار می‌شود:

$$\inf_{\psi} \Pr(\psi(Z^n) \neq V) \geq \frac{1}{4} (1 - \|P_1^n - P_2^n\|_{TV}) \quad (۳۸-۲)$$

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

قضیه ۳-۲ (نامساوی فانو^{۴۶}) 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:

$$\inf_{\psi} \Pr(\psi(Z^n) \neq V) \geq 1 - \frac{I(Z^n; V) + \log 2}{\log |\mathcal{V}|} \quad (۳۹-۲)$$

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

لم ۴-۲ (لم اسود^{۴۷}) 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 [۸] استفاده شده است، کران پایین را بر اساس فاصله‌ی واریانس کل توزیع‌های مخلوط حاشیه‌ای بیان می‌کند:

$$\mathfrak{M}_n(\theta(\mathcal{P})) \geq \delta \sum_{j=1}^d [1 - \|M_{+j}^n - M_{-j}^n\|_{TV}] \quad (۴۰-۲)$$

که در آن M_{+j}^n و M_{-j}^n توزیع‌های حاشیه‌ای مخلوط روی مقادیر $+1$ و -1 در بُعد 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 استاندارد می‌پردازیم که توسط دوچی و همکاران [۸] 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 $\mathcal{M}(\cdot|x)$ و $\mathcal{M}(\cdot|x')$ 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 [۸].

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

^۱Contraction Operators

^۲Data Processing Inequality

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

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

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

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

به طور دقیق‌تر، اگر $\alpha \leq 1$ باشد، این کران به صورت $O(\alpha^2)$ رفتار می‌کند [۸].

اثبات. برای اثبات دقیق این قضیه، از تعریف واگرایی KL شروع می‌کنیم. فرض کنید $q(z|x)$ و $q(z|x')$ چگالی‌های احتمال خروجی باشند. طبق تعریف α -LDP 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 α -LDP بر روی نسبت p/q ، جمله درجه اول صفر می‌شود و جمله درجه دوم ضریب $(e^\alpha - 1)^2$ را تولید می‌کند. جزئیات کامل این محاسبات در لم ۱ مقاله [۸] 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 α -LDP 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 [۷] یا واگرایی E_γ [۲] نیز بیان کرد که در فصل بعد به آن می‌پردازیم.

³Strong Data Processing Inequality (SDPI)

۳-۲-۱ 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 [۸]:

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

$$\|\mathcal{M}(\cdot|x) - \mathcal{M}(\cdot|x')\|_{TV} \leq \min\{1, e^\alpha - 1\} \cdot \|x - x'\|. \quad (۵-۳)$$

(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 [۸] و بعدها با جزئیات بیشتر در [۹] 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 لم 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 [۵].

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:

$$I(V; Z^n) \leq \sum_{i=1}^n I(V; Z_i) \leq n \cdot \alpha^2 \cdot C \quad (۶-۳)$$

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.

^۴Le Cam

^۵Lower 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 خطا» (MSE) در نظر می‌گیریم.

قضیه ۳-۳ (کران پایین تخمین میانگین) برای هر مکانیزم α -LDP و هر تخمین‌گر $\hat{\mu}$ ، ماکسیم خطای مورد انتظار با رابطه زیر محدود می‌شود [۸]:

$$\inf_{\hat{\mu}, \mathcal{M}} \sup_P \mathbb{E}_{\|\hat{\mu} - \mu\|^2} \geq \Omega\left(\frac{d}{n \min\{\alpha, \alpha^2\}}\right) \quad (۷-۳)$$

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

$$\|\mathcal{M}(\cdot|x) - \mathcal{M}(\cdot|x')\|_{TV}^2 \leq \frac{1}{4} D_{KL}(\mathcal{M}(\cdot|x) \parallel \mathcal{M}(\cdot|x')) \leq O(\alpha^2) \quad (۸-۳)$$

بنابراین فاصله TV حداکثر 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 $1/\alpha^2$ افزایش یابد، که هزینه‌ی بسیار سنگین‌تری نسبت به مدل متمرکز (که متناسب با $1/\varepsilon$ است) دارد.

۵-۳ محدودیت‌های تحلیل کلاسیک

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

همان‌طور که در رابطه (۱-۳) دیدیم، کران انقباض KL با ضریب $(e^\alpha - 1)^2$ رشد می‌کند. زمانی که α بزرگ باشد، این کران به سرعت به بی‌نهایت میل 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 وزن‌دهی می‌کند. برای رفع این مشکل و به دست آوردن تحلیل‌های دقیق‌تر^۶ که در تمام بازه‌های α معتبر باشند، نیازمند

^۶Tight

معیار هندسی متفاوتی هستیم. این نیاز، انگیزه اصلی معرفی واگرایی‌های جدید مانند f - واگرایی‌های خاص (نظیر E_γ) است [۲، ۳] که در فصل آینده به تفصیل به آن خواهیم پرداخت.

فصل ۴

هم‌ارزی α -LDP و انقباض E_γ - واگرایی

۴-۱ مقدمه و انگیزه

در فصل پیشین، دیدیم که چگونه دوجی و همکاران [۸] از واگرایی کولبک-لایبلر (KL) برای تحلیل محرمانگی تفاضلی موضعی استفاده کردند. اگرچه کران‌های آن‌ها برای رژیم‌های محرمانگی 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 متریک «بومی» برای تعریف α -LDP نیست. تعریف α -LDP (معادله ۲-۸) مبتنی بر نسبت احتمالات است، در حالی که KL مبتنی بر لگاریتم نسبت‌هاست. این ناهمخوانی 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 $\gamma = e^\alpha$) است. سپس از این هم‌ارزی برای استخراج کران‌های انقباض دقیق^۱ برای سایر واگرایی‌ها استفاده خواهیم کرد که نتایج دوجی را بهبود می‌بخشند [۲].

^۱Tight

۲-۴ معرفی E_γ -واگرایی

E_γ -واگرایی یکی از اعضای کمتر شناخته شده‌ی خانواده f -واگرایی‌هاست که در نظریه اطلاعات برای مقایسه نسبت درست‌نمایی توزیع‌ها کاربرد دارد.

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

$$E_\gamma(P||Q) = \sup_{\mathcal{S} \in \sigma(\mathcal{X})} (P(\mathcal{S}) - \gamma Q(\mathcal{S})) \quad (۱-۴)$$

این تعریف را می‌توان به صورت بسته‌ی زیر نیز نوشت:

$$E_\gamma(P||Q) = \int_{\mathcal{X}} \max\{0, p(x) - \gamma q(x)\} d\mu(x) \quad (۲-۴)$$

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 محرمانگی ایده‌آل می‌کند:

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

$$E_1(P||Q) = \sup_{\mathcal{S}} (P(\mathcal{S}) - Q(\mathcal{S})) = \|P - Q\|_{TV} \quad (۳-۴)$$

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 مرتبط می‌کند.

۳-۴ قضیه هم‌ارزی اصلی

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

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

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

اثبات. اثبات را در دو جهت انجام می‌دهیم.

جهت اول (\Rightarrow): فرض کنید \mathcal{M} خاصیت α -LDP دارد. طبق تعریف ۲-۷، برای هر زیرمجموعه خروجی $\mathcal{S} \subseteq \mathcal{Z}$ و هر x, x' داریم:

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

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 $\mathcal{S} = \emptyset$ مقدار حداقل صفر است)، پس حتماً برابر صفر است.

جهت دوم (\Leftarrow): فرض کنید $E_{e^\alpha}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) = 0$. 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 بلافاصله 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 α -LDP است.

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

۴-۴ بهبود کران‌های انقباض

در فصل ۳ دیدیم که دوجی [۸] کران زیر را برای انقباض KL ارائه کرد:

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

حال با استفاده از چارچوب E_γ ، می‌توانیم کران‌های بسیار دقیق‌تری استخراج کنیم. آسوده و همکاران [۲] 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.

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

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

برای مقادیر کوچک α (رژیم محرمانگی بالا)، این کران به $\alpha^2/2$ میل می‌کند که ۴ برابر کوچکتر (بهتر) از کران دوجی است.

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

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

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

این بهبود ضریب ثابت (از ۴ به ۰/۵) در تحلیل‌های مینیماکس بسیار حیاتی است و نشان می‌دهد که «اندازه نمونه موثر» واقعی می‌تواند تا ۸ برابر بهتر از چیزی باشد که آنالیزهای قبلی نشان می‌دادند.

۵-۴ تعمیم به محرمانگی تقریبی $((\alpha, \delta)$ -LDP)

یکی دیگر از قدرت‌های چارچوب E_γ ، توانایی آن در توصیف ساده‌ی محرمانگی 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 (α, δ) -LDP دقیقاً معادل محدود کردن مقدار 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.

۴-۶ کاربرد در تخمین توزیع گسسته

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 $(e^\alpha - 1)^2$ به صورت نمایی رشد می‌کند و نشان می‌دهد که دقت می‌تواند بسیار سریع‌تر از پیش‌بینی‌های قبلی بهبود یابد.

۴-۷ 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 [۲]، نشان داده شده است که مکانیزم‌های موضعی خاصیت «انقباض قوی» را برای طیف وسیعی از واگرایی‌ها به ارمغان می‌آورند.

۴-۷-۱ 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.

قضیه ۴-۴ (انقباض χ^2) فرض کنید \mathcal{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:

$$\chi^2(\mathcal{M}P || \mathcal{M}Q) \leq \eta_\alpha \cdot \chi^2(P || Q) \quad (15-4)$$

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 مجانبی: برای مقادیر کوچک α (رژیم محرمانگی 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 دوجی (فصل ۳) ضریبی از مرتبه $O(\alpha^2)$ داشتند، اما ضریب $1/4$ در اینجا نشان‌دهنده 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.

۴-۷-۲ تعمیم به سایر واگرایی‌ها

نویسندگان در [۳] 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 مجذور H^2 و واگرایی KL) نیز صادق است.

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

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

$$H^2(\mathcal{M}P, \mathcal{M}Q) \leq \eta_\alpha \cdot H^2(P, Q) \quad (19-4)$$

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

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

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 $\pi(\theta)$ است.

نامساوی ون‌تریز (van Trees) ابزاری کلاسیک برای کران‌دار کردن خطای بیزی بر اساس «اطلاعات فیشر» است. آسوده و ژانگ [۳] نسخه‌ی خصوصی‌شده‌ی این نامساوی را ارائه کرده‌اند که ابزاری نوین در جعبه‌ابزار تحلیل محرمانگی محسوب می‌شود.

قضیه ۴-۶ (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^۲ دارای کران پایین زیر است:

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

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

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

که در آن $\mathcal{I}_{\text{orig}}$ 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 $\eta_\alpha \approx \alpha^2/4$ تضعیف می‌شود.» این نتیجه، اثبات حدود پایین برای مسائل تخمین پارامتر را بسیار ساده می‌کند. به جای درگیر شدن با لم‌های پیچیده‌ی اسود یا فانو، کافیت اطلاعات فیشر مسئله‌ی اصلی را محاسبه کنیم و در ضریب η_α ضرب کنیم.

۹-۴ کاربردهای نوین و بهبود نرخ‌ها

استفاده از کران‌های انقباض قوی (بخش ۴-۷) و نامساوی ون‌تریز خصوصی (بخش ۴-۸) 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:

$$R_{opt} \asymp \left(\frac{1}{n\alpha^2} \right)^{\frac{2\beta}{2\beta+1}} \quad (۲۲-۴)$$

اگرچه مرتبه‌ی کلی نرخ همگرایی مشابه نتایج دوچی است، اما ضرایب ثابت بهبود یافته‌اند و مهم‌تر از آن، اثبات با استفاده از انقباض χ^2 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 محرمانگی تفاضلی است.

فصل ۵

نتیجه گیری

Bibliography

- [1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):131–142, 1966.
- [2] Shahab Asoodeh, Maryam Aliakbarpour, and Flavio P. Calmon. Local differential privacy is equivalent to contraction of an f -divergence. In *2021 IEEE International Symposium on Information Theory (ISIT)*, page 545–550. IEEE Press, 2021.
- [3] Shahab Asoodeh and Huanyu Zhang. Contraction of locally differentially private mechanisms. *IEEE Journal on Selected Areas in Information Theory*, 5:385–395, 2024.
- [4] Michael Barbaro and Tom Zeller. A face is exposed for aol searcher no. 4417749. *New York Times*, 01 2006.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.

- [9] 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.
- [10] Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.
- [11] 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.
- [12] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.*, 17(1):492–542, January 2016.
- [13] 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.
- [14] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.
- [15] 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.
- [16] 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.
- [17] 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

ب

پ

پرس و جو Query

پایگاه داده Database

ت

ج

ح

خ

د

داده Data

دودویی Binary

ر

ز

س

ش

ص

غ

ف

ق

ک

گ

ل

م

مجموعه Set.....
 متصدی مورد اعتماد Trusted Curator.....
 مکانیزم تصادفی Randomized Mechanism.....
 محرمانگی تفاضلی Differential Privacy.....
 مطلقاً پیوسته Absolutely Continuous.....

و

واگرایی Divergence.....

هـ

همسایه Ajacent.....
 همسایگی Adjacency.....

ن

ی

پیوست آ

مطالب تکمیلی

Abstract

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

Keywords: Thesis, Typesetting, Template, X_YTeX Persian



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