



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 تامین امنیت مسئله ایست که از دیرباز انسان ها را به خود مشغول کرده است. در عصر اطلاعات و ارتباطات نیز این مسئله به شکلی متفاوت اهمیت خود را نشان می دهد.

کلیدواژه ها:

فهرست مطالب

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

۲۲	۲-۱-۲ تفسیر پارامترهای محرمانگی
۲۲	۳-۱-۲ تعاریف معادل و صورت‌بندی‌های جایگزین
۲۳	۴-۱-۲ مکانیزم‌های پایه
۲۵	۵-۱-۲ مکانیزم‌های بنیادی محرمانگی تفاضلی
۲۹	۶-۱-۲ ترکیب‌پذیری
۲۹	۷-۱-۲ محرمانگی گروهی
۳۰	۸-۱-۲ محدودیت مدل متمرکز
۳۰	۲-۲ f -واگرایی‌ها
۳۰	۱-۲-۲ تعریف رسمی در فضای اندازه‌پذیر
۳۱	۲-۲-۲ تعریف f -واگرایی
۳۳	۳-۲-۲ نمونه‌های مهم و توابع مولد
۳۶	۴-۲-۲ خواص بنیادین و روابط بین f -واگرایی‌ها
۳۸	۳-۲ مبانی آماری و کران‌های اطلاعاتی
۳۸	۱-۳-۲ نظریه تصمیم و ریسک مینی‌مکس
۳۹	۲-۳-۲ تقلیل به آزمون فرض (روش بسته‌بندی)
۴۲	۳-۳-۲ نامساوی‌های کران پایین

۳ محرمانگی تفاضلی موضعی

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

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

۶۱	تحلیل مینی مکس و هندسه اطلاعاتی در LDP ۴
۶۱ مقدمه ۱-۴
۶۲ محرمانگی به عنوان انقباض ۲-۴
۶۴ تفسیر رژیم‌های محرمانگی ۱-۲-۴
۶۵ نامساوی‌های پردازش داده قوی (SDPI) ۳-۴
۶۵ کران انقباض برای واگرایی کای-دو (χ^2) ۱-۳-۴
۶۶ مزیت استفاده از χ^2 نسبت به KL ۲-۳-۴
۶۷ اثبات نرخ مینی مکس برای تخمین میانگین ۴-۴
۶۷ تعریف مسئله ۱-۴-۴
۶۹ بحث و تفسیر ۲-۴-۴

۷۰

مراجع

۷۳

واژه‌نامه

۷۵

آ مطالب تکمیلی

فهرست جداول

فهرست تصاویر

- ۲-۱ مدل محرمانگی تفاضلی متمرکز (CDP) با یک متصدی مورد اعتماد. ۲۰
- ۲-۲ نمایش هندسی روش بسته‌بندی برای تقلیل مسئله‌ی تخمین به آزمون فرض. فضای پارامتر Θ با مجموعه‌ای از گوی‌های مجزا $\{\theta^1, \dots, \theta^K\}$ پوشانده شده است که تشکیل یک 2δ -بسته‌بندی می‌دهند. اگر تخمین‌گر $\hat{\theta}$ (نقطه قرمز) دارای خطای کم‌تر از δ باشد، لزوماً درون یکی از این گوی‌ها قرار می‌گیرد. از آن‌جا که گوی‌ها مجزا هستند، حداکثر یک اندیس i وجود دارد که $\hat{\theta}$ به آن نزدیک باشد؛ لذا مسئله تخمین به یافتن این اندیس (آزمون فرض چندگزینه‌ای) تقلیل می‌یابد. ۴۰
- ۳-۱ گذار از مدل متمرکز به موضعی؛ نویز به صورت موضعی روی دستگاه کاربر اضافه می‌شود. ۴۵
- ۳-۲ (الف) مدل گرافیکی نمایش‌دهنده‌ی روابط استقلال شرطی بین داده‌های خصوصی $\{X_i\}_{i=1}^n$ و متغیرهای مشاهده‌شده $\{Z_i\}_{i=1}^n$ در پروتکل‌های تعاملی؛ پیکان‌های افقی نشان‌دهنده‌ی وابستگی Z_i به تاریخچه‌ی پیشین هستند. (ب) مدل گرافیکی ساده‌تر در پروتکل‌های غیرتعاملی که در آن خروجی‌ها به شرط ورودی‌ها از یکدیگر مستقل هستند. ۴۸

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 تارکانش‌های 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 [۱۸].

- **مجموعه داده‌ی نتفلیکس^{۱۲}:** شرکت نتفلیکس مجموعه‌ای از امتیازهای کاربران به فیلم‌ها را منتشر کرد که در آن شناسه‌های کاربری با اعداد تصادفی جایگزین شده بودند. پژوهشگران نشان دادند که با استفاده از اطلاعات عمومی موجود در وب‌سایت IMDb و تطبیق الگوهای امتیازدهی، می‌توان هویت بسیاری از کاربران را با دقت بالا کشف کرد [۱۵].

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

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

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 (Differential Privacy)

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

پیش از آنکه به تعاریف صوری و ریاضی بپردازیم، ضروری است که درک عمیقی از چیستی محرمانگی و تمایز 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 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 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 قابل‌ردیابی در خروجی منجر می‌شود که بلافاصله دو جهان موازی را از هم متمایز می‌کند.

در واقع مثال میانگین را می‌توان به هر تابع قطعی $f: \mathcal{X}^n \rightarrow \mathbb{Z}$ 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-لایبلر^{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 ($\epsilon < \infty$) با توابع قطعی غیرثابت قابل دستیابی نیست. بنابراین، همان‌طور که در ادبیات موضوع تأکید شده است [۹]، برای شکستن این وابستگی قطعی و ایجاد ابهام آماری، تصادفی‌سازی^{۲۷} در فرآیند مکانیزم الزامی است.

کاربردهای محرمانگی تفاضلی

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

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

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 آموزش داد که الگوهای کلی را یاد می‌گیرند اما قادر به بازتولید داده‌های آموزشی حساس (مانند تصاویر چهره یا متون خصوصی) نیستند.

- **سیستم‌های توصیه‌گر و داده‌های مکانی**: شرکت‌های فناوری از DP برای جمع‌آوری آمارهای رفتاری (مانند پربازدیدترین وبسایت‌ها یا مکان‌های پرتردد) استفاده می‌کنند تا بدون ردیابی لحظه‌ای کاربران، کیفیت خدمات خود را بهبود بخشند (مانند مکانیزم RAPPOR در گوگل کروم).

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

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 انبوه داده‌هاست؛ به طوری که حتی پیشرفته‌ترین دیوارهای آتش^{۳۳} نیز در برابر حملات پیچیده آسیب‌پذیرند. در چنین شرایطی، اگر داده‌ها به صورت خام ذخیره شده باشند، نشت اطلاعاتی مانند آنچه در واقعه‌ی Equifax رخ داد، مکانیزم‌های محرمانگی تفاضلی مرکزی را عملاً بی‌فایده می‌کند؛ زیرا مهاجم با دور زدن مکانیزم، مستقیماً به مخزن داده‌های حساس دست می‌یابد [۱۹].

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

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 آغازگرها: از پیمایش‌های آماری تا تعریف مدرن محرمانگی

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

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

$$\Pr[\mathcal{M}(\mathcal{D}) \in S] \leq e^\epsilon \cdot \Pr[\mathcal{M}(\mathcal{D}') \in S] + \delta \quad (۱-۱)$$

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

^{۳۵}Statistical Disclosure Control

^{۳۶}Differential Privacy

^{۳۷}Privacy Budget

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

- **مقادیر کوچک ε (محرمانگی قوی):** زمانی که $\varepsilon \rightarrow 0$ ، 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 نظیر کاسی‌یسواناتان و دیگران تدوین شد [۱۳]، اما ریشه‌های عملی آن به دهه‌ها قبل باز می‌گردد.

تعریف ۱-۱ (محرمانگی تفاضلی موضعی (α, δ) -LDP): یک مکانیزم تصادفی $M : \mathcal{X} \rightarrow \mathcal{Z}$ (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 $S \subseteq \mathcal{Z}$ ، رابطه زیر برقرار باشد:

$$\Pr[M(x) \in S] \leq e^\alpha \cdot \Pr[M(x') \in S] + \delta \quad (۱-۲)$$

در این تعریف:

- α ، بودجه محرمانگی است که میزان شباهت توزیع‌های خروجی را کنترل می‌کند.

- δ ، احتمال ناچیز شکست مکانیزم در برقراری شرط محرمانگی است.

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

³⁸Local Differential Privacy

مثال‌هایی از مکانیزم و کاربرد LDP

در واقع، ساده‌ترین و نخستین نمونه از یک مکانیزم LDP، روش «پاسخ تصادفی»^{۳۹} است که توسط وارنر در سال ۱۹۶۵ برای حذف سوگیری در نظرسنجی‌های حساس معرفی شد [۲۲]. وارنر این روش را نه برای حفاظت در برابر حملات سایبری، بلکه برای تشویق پاسخ‌دهندگان به صداقت در سوالات حساس (مانند مصرف مواد مخدر یا عقاید سیاسی خاص) طراحی کرد.

سازوکار کلاسیک این روش برای یک پرسش با پاسخ «بله/خیر» به صورت زیر است: فرض کنید از کاربر 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^{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 خطا تخمین بزند. به زبان ریاضی مدرن، اگر احتمال گزارش پاسخ واقعی p باشد، نسبت احتمال خروجی‌ها برای دو ورودی متفاوت x و x' به صورت زیر محدود می‌شود:

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

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

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

این فرمول، تفسیر شهودی LDP را کامل می‌کند:

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

³⁹Randomized Response (RR)

⁴⁰Unfair

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

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

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

- **گوگل و پروتکل RAPTOR**: در سال ۲۰۱۴، گوگل برای جمع‌آوری آمار تنظیمات مرورگر کروم و شناسایی بدافزارها، پروتکل RAPTOR^{۴۲} را معرفی کرد [۲۰]. چالش اصلی گوگل، جمع‌آوری رشته‌های متنی^{۴۳} بود. راه‌حل آن‌ها ترکیب پاسخ تصادفی با فیلترهای بلوم^{۴۴} بود. در این روش، داده‌ی ورودی ابتدا به یک بردار بیتی (با استفاده از توابع درهم‌ساز) نگاشت می‌شود و سپس پاسخ تصادفی روی تک‌تک بیت‌های این فیلتر اعمال می‌گردد. این معماری به گوگل اجازه داد تا بدون دانستن ورودی دقیق هر کاربر، الگوهای پرتکرار و ناهنجاری‌ها را در مقیاس میلیونی شناسایی کند.
- **اپل و جمع‌آوری داده‌های دایره‌لغات**: شرکت اپل از LDP برای بهبود کیبورد QuickType، شناسایی ایموجی‌های پرطرفدار و داده‌های سلامت در سیستم‌عامل‌های iOS و macOS استفاده می‌کند. مسئله‌ی اپل، مخابره‌ی کارآمد داده‌ها با حفظ حریم خصوصی بود. راه‌حل اپل استفاده از تکنیک‌های مبتنی بر طرح‌ریزی^{۴۵} و تبدیل‌های ریاضی مانند تبدیل هادامارد^{۴۶} است. این تبدیل‌ها به مکانیزم اجازه می‌دهند که اطلاعات را در ابعاد پایین‌تر فشرده کند تا هم بار ارتباطی کاهش یابد و هم واریانس تخمین‌گر در دامنه‌های بزرگ کنترل شود [۲۰].

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

⁴¹High-Dimensional Domain

⁴²Randomized Aggregatable Privacy-Preserving Ordinal Response

⁴³String

⁴⁴Bloom Filters

⁴⁵Sketching

⁴⁶Hadamard Transform

استفاده کردند. تمرکز اصلی در این جا، 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.

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

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

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

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

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

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

⁴⁷Non-interactive / Simultaneous

⁴⁸Interactive / Sequential

⁴⁹Utility

⁵⁰Privacy-Accuracy Trade-off

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

- **در مدل متمرکز (CDP):** نویز تنها «یک بار» و پس از تجميع داده‌ها به نتیجه نهایی اضافه می‌شود. از آن‌جا که مجموع (یا میانگین) داده‌ها حساسیت کمی دارد، مقدار نویز معمولاً مستقل از تعداد کاربران (n) و بسیار کوچک است.

- **در مدل موضعی (LDP):** نویز باید به «تک‌تک» داده‌های ورودی اضافه شود (پیش از آنکه از دستگاه کاربر خارج شوند). وقتی تحلیل‌گر قصد دارد میانگین این داده‌ها را محاسبه کند، واریانس نویزهای n کاربر با هم جمع می‌شود.

این انباشت نویز باعث می‌شود که نسبت سیگنال به نویز^{۵۱} در مدل موضعی بسیار پایین‌تر از مدل متمرکز باشد. به بیان دیگر، برای دستیابی به همان سطح از دقت که در مدل متمرکز وجود دارد، در مدل LDP نیازمند تعداد بسیار بیش‌تری نمونه داده هستیم.

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

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

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

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

⁵¹Signal-to-Noise Ratio (SNR)

⁵²Statistical Minimax Theory

⁵³Duchi, Jordan, and Wainwright

از ابزارهایی نظیر نامساوی فانو^{۵۴} و لم اسود^{۵۵} (که در فصل سوم به تفصیل بررسی خواهند شد)، ثابت کردند که رفتار مجانبی ریسک در مدل LDP تفاوتی بنیادین با مدل متمرکز دارد. به طور مشخص، برای n کاربر و بودجه محرمانگی α ، ریسک مینی مکس (واریانس خطا) از رابطه‌ی زیر پیروی می‌کند:

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

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

۱. **کندی هم‌گرایی:** در حالی که خطای مدل متمرکز با سرعت $1/n$ کاهش می‌یابد، خطای مدل موضعی با سرعت بسیار کندتر $1/\sqrt{n}$ کم می‌شود.

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

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

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 مینی مکس) است. این

⁵⁴Fano's Inequality

⁵⁵Assouad's Lemma

⁵⁶Optimal Mechanisms

⁵⁷Staircase Mechanisms

⁵⁸Unary Encoding (UE)

⁵⁹Optimal Local Hashing (OLH)

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

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

- در برخی مسائل تخمین چگالی، پژوهشگران عمدتاً از واگرایی کولبک-لایبیلر (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.

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

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

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

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

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 «کانال انقباضی»^{۶۲} عمل می‌کند که فاصله‌ی بین توزیع‌های ورودی را فشرده می‌سازد. هدف ما در این پژوهش،

⁶⁰ f -divergences

⁶¹ Csiszár

⁶² Contraction Channel

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

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

برای غلبه بر چالش پراکندگی مفاهیم و ارائه یک تحلیل منسجم، این پایان‌نامه بر مرور و تشریح رویکرد نوینی تمرکز دارد که در سال‌های اخیر توسط E_γ -آسوده، علی اکبری و کالمون** [۲] و E_γ -آسوده و ژانگ** [۳] توسعه یافته است. این پژوهشگران نشان داده‌اند که قیود محرمانگی تفاضلی موضعی (LDP) را می‌توان به صورت دقیق و ریاضی معادل با «خواص انقباضی»^{۶۴} خانواده‌ی f -واگرایی‌ها فرمول‌بندی کرد. به طور مشخص، در این رویکرد اثبات می‌شود که شرط LDP معادل است با کران‌دار بودن ضریب انقباض برای «واگرایی‌هاکی-استیک» $(E_\gamma, \text{divergence})$ - همچنین، تحلیل‌های دقیق‌تر نشان می‌دهند که چگونه می‌توان بهترین کران‌های انقباض را برای واگرایی‌های مهمی نظیر χ^2 و KL استخراج کرد. این دیدگاه هندسی، هر مکانیزم محرمانگی را به مثابه‌ی یک کانال نویزی می‌بیند که فاصله بین توزیع‌های احتمالی ورودی را کاهش می‌دهد. در این پایان‌نامه، ما تلاش می‌کنیم تا با مطالعه‌ی دقیق این مراجع، نشان دهیم چگونه این ضرایب انقباض (η_f) محاسبه می‌شوند و چگونه می‌توان از آن‌ها برای بازتولید و بهبود کران‌های مینی مکس در مسائل تخمین و آزمون فرض استفاده کرد.

۱-۳-۳ اهداف و ساختار پژوهش

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

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

۲. بررسی ضرایب انقباض برای واگرایی‌های χ^2 و KL: مطالعه‌ی کران‌های دقیق^{۶۵} برای ضریب

⁶³Contraction Coefficient

⁶⁴Contraction Properties

⁶⁵Sharp Bounds

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

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

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

۴-۱ ساختار پایان‌نامه

ساختار ادامه‌ی این پایان‌نامه به شرح زیر سازمان‌دهی شده است:

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

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

فصل چهارم به تحلیل کران‌های پایین خطر مینی مکس^{۷۰} در مسائل تخمین آماری تحت قید محرمانگی

^{۶۶}Strong Data Processing Inequalities (SDPI)

^{۶۷}Le Cam's Lemma

^{۶۸}Fano's Lemma

^{۶۹}Randomized Response

^{۷۰}Minimax Risk

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

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

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

⁷¹Contraction Coefficient

فصل ۲

پیش‌نیازها

۱-۲ محرمانگی تفاضلی متمرکز (CDP)

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

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

¹Differential Privacy

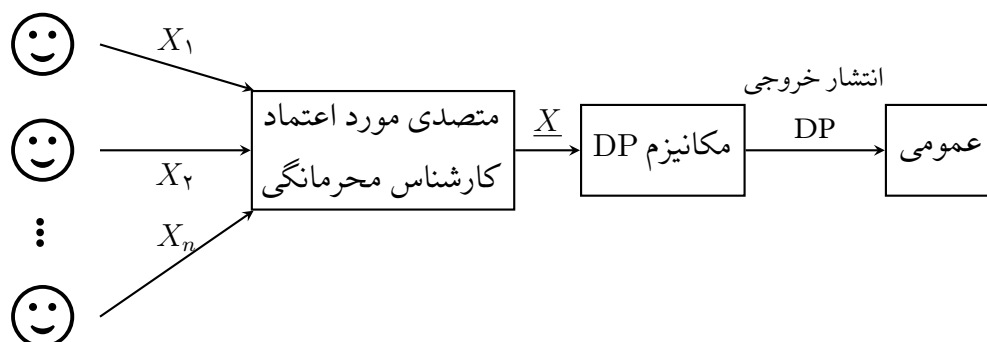
²Centralized

³Trusted Curator

⁴Randomized Mechanism

⁵Query

افراد (داده‌ها)



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

داده‌ها^۶ می‌نامیم و آن را با \mathcal{X} نمایش می‌دهیم (در ادامه‌ی پایان‌نامه نیز هرجا از \mathcal{X} استفاده شد منظور جهان داده‌هاست). یک پایگاه داده‌ی \mathcal{D} مجموعه‌ای از رکوردهاست که اعضای آن از \mathcal{X} انتخاب شده‌اند. در ادبیات محرمانگی تفاضلی، پایگاه داده معمولاً به صورت یک بردار (هیستوگرام) $x \in \mathbb{N}^{|\mathcal{X}|}$ نمایش داده می‌شود که در آن هر مولفه x_i نشان‌دهنده‌ی تعداد تکرار عنصر i - 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 $\mathcal{X} = \{\text{شاغل}, \text{بیکار}\}$. اگر در یک پایگاه داده ۳ نفر شاغل و ۱ نفر بیکار باشند، نمایش هیستوگرامی پایگاه داده \mathcal{D} به صورت بردار زیر خواهد بود:

$$\mathcal{D} = (3, 1)$$

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

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

⁶Data Universe

تعریف ۳-۲ (فاصله و نرم‌های ℓ_p) برای سنجش میزان تفاوت دو پایگاه‌داده، از مفهوم نرم ℓ_p استفاده می‌شود. در حالت کلی برای $p \geq 1$ ، فاصله‌ی ℓ_p بین دو پایگاه‌داده‌ی D_1 و D_2 (با بردارهای تکرار متناظر) به صورت زیر تعریف می‌گردد:

$$\|D_1 - D_2\|_p = \left(\sum_{i=1}^{|\mathcal{X}|} |x_{1,i} - x_{2,i}|^p \right)^{1/p} \quad (1-2)$$

در ادبیات محرمانگی تفاضلی، نرم ℓ_1 (یا فاصله‌ی منهتن) به دلیل ارتباط مستقیم آن با تعداد رکوردها، معیار اصلی محسوب می‌شود. این فاصله دقیقاً تعداد رکوردهایی را می‌شمارد که باید تغییر کنند (اضافه یا حذف شوند) تا D_1 به D_2 تبدیل شود:

$$\|D_1 - D_2\|_1 = \sum_{i=1}^{|\mathcal{X}|} |x_{1,i} - x_{2,i}|$$

مثال ۳-۲ فرض کنید $D_1 = (3, 1)$ و $D_2 = (3, 0)$ باشند (یعنی در پایگاه‌داده دوم، یک نفر بیکار حذف شده است). فاصله‌ی ℓ_1 آن‌ها برابر است با:

$$\|D_1 - D_2\|_1 = |3 - 3| + |1 - 0| = 1$$

تعریف ۴-۲ (پایگاه‌داده‌های همسایه) دو پایگاه‌داده‌ی D_1 و D_2 را همسایه^۷ می‌گوییم (و با $D_1 \sim D_2$ نشان می‌دهیم) اگر فاصله‌ی ℓ_1 آن‌ها حداکثر ۱ باشد:

$$\|D_1 - D_2\|_1 \leq 1$$

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

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

تعریف ۵-۲ (ϵ -محرمانگی تفاضلی (ϵ -DP)) یک مکانیزم تصادفی M با دامنه \mathcal{X}^n و برد R ، ویژگی ϵ -محرمانگی تفاضلی را برآورده می‌سازد، اگر برای هر دو پایگاه‌داده‌ی همسایه‌ی D_1 و D_2 ($D_1 \sim D_2$) و برای هر زیرمجموعه از خروجی‌های ممکن $S \subseteq R$ (که در σ -جبر برد تعریف شده باشد)، داشته باشیم:

$$\mathbb{P}[M(D_1) \in S] \leq \exp(\epsilon) \cdot \mathbb{P}[M(D_2) \in S] \quad (2-2)$$

در این نامساوی، پارامتر $\epsilon \geq 0$ را «بودجه‌ی محرمانگی»^۸ می‌نامیم.

⁷Neighboring / Adjacent

⁸Privacy Budget

۲-۱-۲ تفسیر پارامترهای محرمانگی

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

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

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

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

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

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 \mathcal{M} شرط ε -DP را برآورده می‌کند اگر و تنها اگر برای تمام همسایه‌های $\mathcal{D}_1 \sim \mathcal{D}_2$ شرایط زیر برقرار باشد:

^۹Failure Probability

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

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

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

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

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

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

□ (در حالت گسسته، انتگرال 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 ماکزیم:

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

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

در بخش‌های بعدی با تعریف دقیق واگرایی آشنا خواهیم شد.

۴-۱-۲ مکانیزم‌های پایه

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

¹⁰Max Divergence

¹¹Noise

¹²Sensitivity

تعریف 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 می شود، برابر است با بیشینه مقدار تغییرات خروجی تابع، به ازای تغییر تنها یک رکورد در ورودی. با استفاده از تعریف نرم ℓ_p (تعریف ۱-۲) داریم:

$$\Delta_p f = \max_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_p \quad (۷-۲)$$

که در آن ماکزیم گیری روی تمام زوج پایگاه داده های همسایه ($D_1 \sim D_2$) انجام می شود.

در میان انواع حساسیت ها، دو مورد زیر به دلیل کاربردها در مکانیزم های پایه، از اهمیت ویژه ای برخوردارند:

تعریف ۸-۲ (حساسیت ℓ_1 ، ℓ_2 و ℓ_∞) سه نوع حساسیت زیر بیش ترین کاربرد را در طراحی مکانیزم های محرمانگی دارند:

۱. حساسیت ℓ_1 ($\Delta_1 f$): این حساسیت برابر با ماکزیم فاصله منتهن بین خروجی هاست و پارامتر اصلی در مکانیزم لا پلاس می باشد:

$$\Delta_1 f = \max_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1 \quad (۸-۲)$$

۲. حساسیت ℓ_2 ($\Delta_2 f$): این حساسیت برابر با ماکزیم 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 می شود:

$$\Delta_2 f = \max_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_2 \quad (۹-۲)$$

۳. حساسیت ℓ_∞ ($\Delta_\infty f$): این حساسیت برابر با بیشینه تغییر در «تک تک مولفه های» خروجی است (نرم ماکزیم). این معیار نشان می دهد که مقدار یک درایه خاص از خروجی چقدر می تواند تغییر کند:

$$\Delta_\infty f = \max_{D_1 \sim D_2} \|f(D_1) - f(D_2)\|_\infty \quad (۱۰-۲)$$

مثال ۴-۲ فرض کنید f یک تابع هیستوگرام شمارشی باشد (تعداد افراد در دسته های مجزا). اگر مشخصات یک فرد تغییر کند، او از یک دسته خارج (تغییر -۱) و به دسته دیگری وارد (تغییر $+۱$) می شود. سایر دسته ها ثابت می مانند (۰). بردار تغییرات برابر است با $(۰, \dots, +۱, \dots, -۱, \dots, ۰)$. حال حساسیت ها را محاسبه می کنیم:

^{۱۳} ℓ_p -Global Sensitivity

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

این مثال به وضوح رابطه $\Delta_1 f \leq \Delta_2 f \leq \Delta_\infty f$ 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 کالبره کردن مکانیزم لاپلاس و حساسیت l_2 برای مکانیزم گوسی استفاده می‌شود.

۵-۱-۲ مکانیزم‌های بنیادی محرمانگی تفاضلی

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

مکانیزم لاپلاس

ساده‌ترین و پرکاربردترین روش برای توابع عددی، افزودن نویز از توزیع لاپلاس است. توزیع لاپلاس با پارامتر مقیاس b و میانگین μ دارای تابع چگالی احتمال $h(z) = \frac{1}{b} \exp\left(-\frac{|z-\mu|}{b}\right)$ است.

قضیه‌ی ۲-۲ (محرمانگی 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 $Y_i \stackrel{i.i.d}{\sim} \text{Lap}\left(\frac{\Delta_1 f}{\epsilon}\right)$ ، شرط ϵ -DP 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 $b = \frac{\Delta_1 f}{\epsilon}$ ، نسبت چگالی احتمال را برای یک بردار خروجی دلخواه

$z = (z_1, \dots, z_k)$ 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 نمای:

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

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 $\Delta_1 f \leq \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1$. 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) پرس‌وجوهای شمارشی^{۱۴}، پرس‌وجوهایی به فرم «چه تعداد از اعضای پایگاه‌داده ویژگی P را دارند؟» هستند. این نوع توابع بلوک‌های سازنده‌ی بسیاری از تحلیل‌های آماری و داده‌کاوی (مانند هیستوگرام‌ها) هستند [۹].

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

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

¹⁴Counting Queries

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

مکانیزم گوسی

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

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

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

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

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

اثبات. [طرح کلی] اثبات دقیق این قضیه نیازمند تحلیل «متغیر تصادفی زیان محرمانگی»^{۱۵} است که جزئیات کامل آن در [9, Appendix A] موجود است. ایده اصلی این است که برخلاف توزیع لاپلاس، در توزیع گوسی نسبت $\frac{p(z)}{q(z)}$ کران دار نیست، اما احتمال رخداد مقادیری که این نسبت را بزرگ می‌کنند (نواحی دمی توزیع)، با δ محدود می‌شود. \square

مثال ۲-۶ (انتشار آماره‌های چندگانه و اثر ابعاد) فرض کنید یک بیمارستان می‌خواهد میانگین d ویژگی حیاتی مختلف را منتشر کند. داده‌های هر بیمار را می‌توان به صورت یک بردار $v \in [0, 1]^d$ در نظر گرفت. اگر یک بیمار پرونده‌اش را تغییر دهد، در بدترین حالت تمام d ویژگی او می‌تواند از ۰ به ۱ تغییر کنند.

تحلیل حساسیت:

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

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

¹⁵Privacy Loss Random Variable

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

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

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

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

مکانیزم نمایی

مکانیزم‌های قبلی برای خروجی‌های عددی بودند. اگر خروجی یک «عضو» از یک مجموعه باشد، از مکانیزم نمایی استفاده می‌شود. این مکانیزم بر اساس یک تابع امتیاز^{۱۶} $q(\mathcal{D}, r)$ کار می‌کند که میزان «خوبی» خروجی r را می‌سنجد. حساسیت این تابع به صورت $\Delta q = \max_r \max_{\mathcal{D} \sim \mathcal{D}'} |q(\mathcal{D}, r) - q(\mathcal{D}', r)|$ تعریف می‌شود.

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

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

این مکانیزم شرط ε -DP را برآورده می‌کند.

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

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

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

¹⁶Score Function

۶-۱-۲ ترکیب پذیری

در کاربردهای واقعی، معمولاً چندین پرس و جو روی یک پایگاه داده اجرا می شود. قضایای ترکیب پذیری^{۱۷} نشان می دهند که چگونه بودجه ی محرمانگی انباشته می شود.

قضیه ی ۵-۲ (ترکیب پذیری ساده) اگر k مکانیزم $\mathcal{M}_1, \dots, \mathcal{M}_k$ به ترتیب دارای بودجه های $\varepsilon_1, \dots, \varepsilon_k$ باشند، اجرای متوالی آن ها روی یک پایگاه داده ی واحد، تضمین $DP - (\sum \varepsilon_i)$ را فراهم می کند [۹].

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

قضیه ی ۶-۲ (ترکیب پذیری پیشرفته) برای هر $\delta' > 0$ ، اجرای k مکانیزم که هر کدام $\varepsilon - DP$ 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 $\varepsilon\sqrt{k}$ رشد می کند.

۷-۱-۲ محرمانگی گروهی

محرمانگی تفاضلی نه تنها از یک فرد، بلکه به صورت خودکار از گروه های کوچک نیز محافظت می کند [۹].

قضیه ی ۷-۲ (محرمانگی گروهی^{۱۸}) اگر دو پایگاه داده \mathcal{D}_1 و \mathcal{D}_2 در k رکورد متفاوت باشند (فاصله ی همسایگی k)، آنگاه هر مکانیزم $\varepsilon - DP$ برای آن ها تضمین $DP - (k\varepsilon)$ را ارائه می دهد:

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

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

¹⁷Composition

۸-۱-۲ 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 فصل بعد با محرمانگی تفاضلی موضعی و تعاریف و قضایای اساسی آن آشنا خواهیم شد.

۲-۲ f -واگرایی‌ها

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

۱-۲-۲ تعریف رسمی در فضای اندازه‌پذیر

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

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

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

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

۳. نسبت به اجتماع شمارا بسته باشد: اگر A_1, A_2, \dots دنباله‌ای از مجموعه‌ها در \mathcal{F} باشند، آنگاه

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

¹⁹ f -divergences

²⁰ Measure Theory

²² Field

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

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

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \quad (17-2)$$

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

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

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

$$\mu(A) = 0 \implies P(A) = 0. \quad (18-2)$$

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

قضیه ۸-۲ (رادون-نیکودیم^{۲۷}) اگر $\mu \ll P$ باشد (و μ یک اندازه σ -متناهی باشد)، آنگاه یک تابع منحصر به فرد (تقریباً همه‌جا) غیرمنفی و اندازه‌پذیر $p: \Omega \rightarrow [0, \infty)$ وجود دارد که برای هر $A \in \mathcal{F}$:

$$P(A) = \int_A p d\mu. \quad (19-2)$$

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

۲-۲-۲ تعریف f -واگرایی

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

²³Measurable Space

²⁵Countable Additivity

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

$$D_f(P||Q) \triangleq \int_{\Omega} f\left(\frac{p(x)}{q(x)}\right) q(x) d\mu(x). \quad (20-2)$$

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

$$0 \cdot f\left(\frac{0}{0}\right) = 0 \cdot$$

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

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

تفسیر اجزاء:

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

- غیرمنفی بودن: با استفاده از نامساوی یسن^{۲۹} برای تابع محدب f ، می‌توان نشان داد که واگرایی همواره نامنفی است. از آنجا که $\mathbb{E}_Q\left[\frac{dP}{dQ}\right] = \int p(x) d\mu = 1$ داریم:

$$D_f(P||Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right] \geq f\left(\mathbb{E}_Q\left[\frac{dP}{dQ}\right]\right) = f(1) = 0. \quad (21-2)$$

بنابراین همواره $D_f(P||Q) \geq 0$. همچنین اگر f در نقطه $t = 1$ اکیداً محدب^{۳۰} باشد، تساوی $D_f(P||Q) = 0$ برقرار است اگر و تنها اگر $P = Q$ (تقریباً همه‌جا).

در حالت‌های خاص که فضای نمونه Ω گسسته یا پیوسته (اقلیدسی) باشد و چگالی‌های p و q نسبت به یک اندازه پایه (مانند شمارشی یا لبگ) وجود داشته باشند، مشتق رادون-نیکودیم به نسبت معمولی

²⁹Jensen's Inequality

³⁰Strictly Convex

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

$$D_f(P\|Q) = \int_{\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) dx \quad \text{یا} \quad \sum_{x \in \mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) \quad (22-2)$$

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

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

۳-۲-۲ نمونه‌های مهم و توابع مولد

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

تعریف ۲-۱۳ (تغییرات کل) فاصله تغییرات کل^{۳۴}، شهودی‌ترین متریک برای سنجش تمایزپذیری دو توزیع است و بیان‌گر بیش‌ترین تفاوت احتمالی است که دو توزیع می‌توانند روی یک پیشامد داشته باشند. تابع مولد آن $f(t) = \frac{1}{2}|t-1|$ است.

$$TV(P, Q) = \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)| dx \quad (23-2)$$

$$= \sup_{A \subseteq \mathcal{X}} |P(A) - Q(A)| \quad (24-2)$$

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

³¹Distance

³²Symmetry

³³Triangle Inequality

³⁴Total Variation (TV)

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

$$KL(P||Q) = \int_{\mathcal{X}} p(x) \ln \left(\frac{p(x)}{q(x)} \right) dx \quad (25-2)$$

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

$$I(X; V) = KL(P_{X,V} || P_X \otimes P_V) = \mathbb{E}_V [KL(P_{X|V} || P_X)] \quad (26-2)$$

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

تعریف ۱۶-۲ (کای-دو) واگرایی کای-دو^{۳۸}، اغلب برای تقریب‌زنی سایر فواصل (مانند KL) در همسایگی‌های کوچک استفاده می‌شود و به دلیل فرم مربعی، محاسبه آن معمولاً ساده‌تر است. تابع مولد آن $f(t) = (t - 1)^2$ است.

$$\chi^2(P||Q) = \int_{\mathcal{X}} \frac{(p(x) - q(x))^2}{q(x)} dx \quad (27-2)$$

تعریف ۱۷-۲ (هلینگر-دو) فاصله هلینگر-دو^{۳۹}، تابع مولد آن $f(t) = (\sqrt{t} - 1)^2$ است. این فاصله به دلیل خواص ریاضی خوش‌رفتار (مانند متریک و کران‌دار بودن بین ۰ و ۲)، در نظریه برآورد^{۴۰} و استخراج کران‌های مینی‌مکس (مانند روش $Le\ Cam$) کاربرد فراوان دارد.

$$H^{\chi}(P, Q) = \int_{\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx \quad (28-2)$$

تعریف ۱۸-۲ (E_{γ} یا هاکی-استیک) واگرایی E_{γ} یا واگرایی هاکی-استیک^{۴۱}، ابزاری کلیدی در تحلیل‌های مدرن حریم خصوصی و آزمون‌های فرضیه است. تابع مولد آن برای پارامتر $\gamma \geq 1$ به صورت $f_{\gamma}(t) = [t - \gamma]_+ = \max\{0, t - \gamma\}$ است.

³⁵Kullback-Leibler (KL)

³⁶Relative Entropy

³⁷Mutual Information

³⁸Chi-Squared (χ^2)

³⁹Squared Hellinger

⁴⁰Estimation Theory

⁴¹Hockey-Stick Divergence

تعریف و صورت‌های معادل: تعریف اصلی بر اساس انتگرال جرم اضافی نسبت درست‌نمایی است، اما صورت‌های معادل زیر بینش عملیاتی‌تری ارائه می‌دهند:

$$E_\gamma(P\|Q) = \int_{\mathcal{X}} \max\{0, p(x) - \gamma q(x)\} dx \quad (2-29)$$

$$= \sup_{A \subseteq \mathcal{X}} (P(A) - \gamma Q(A)) \quad (2-29 \text{ ب})$$

$$= P(Z > \gamma) - \gamma Q(Z > \gamma) \quad \left(Z = \frac{p(x)}{q(x)} \text{ که} \right) \quad (2-29 \text{ ج})$$

رابطه (2-29 ب) نشان می‌دهد که این واگرایی بیان‌گر بیشینه‌ی تفاضل وزن‌دار احتمالات است که مستقیماً با موازنه خطای نوع اول و دوم در آزمون فرضیه مرتبط است.

نکات تحلیلی و تاریخی:

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

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

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

$$E_{e^\epsilon}(P\|Q) \leq \delta \quad (2-30)$$

تعریف ۱۹-۲ (رنی) واگرایی رنی^{۴۲}، با پارامتر $\alpha \in (1, \infty)$ ، به صورت زیر تعریف می‌شود:

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \ln \left(\int_{\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} dx \right) \quad (2-31)$$

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

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

$$D_\infty(P\|Q) = \lim_{\alpha \rightarrow \infty} D_\alpha(P\|Q) = \sup_{x \in \mathcal{X}} \ln \left(\frac{p(x)}{q(x)} \right) \quad (2-32)$$

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

⁴²Rényi Divergence

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

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

نامساوی پردازش داده (DPI)

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

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

$$D_f(P_K \| Q_K) \leq D_f(P \| Q) \quad (۳۳-۲)$$

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

تحدب مشترک

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

قضیه‌ی ۱۰-۲ (تحدب مشترک^{۴۷}) نگاشت $(P, Q) \mapsto D_f(P \| Q)$ یک تابع محدب مشترک است. یعنی برای هر $\lambda \in [0, 1]$ و توزیع‌های P_1, P_2, Q_1, Q_2 :

$$D_f(\lambda P_1 + (1 - \lambda)P_2 \| \lambda Q_1 + (1 - \lambda)Q_2) \leq \lambda D_f(P_1 \| Q_1) + (1 - \lambda)D_f(P_2 \| Q_2) \quad (۳۴-۲)$$

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

⁴³Monotonicity

⁴⁵Markov Kernel / Probability Kernel

⁴⁶Post-processing

روابط بین واگرایی‌ها

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

- **نامساوی پینسکر^{۴۸}:** این نامساوی مشهور، ارتباط هندسه (فاصله تغییرات کل) و اطلاعات (واگرایی KL) را برقرار می‌کند و نشان می‌دهد که همگرایی در آنتروپی نسبی، همگرایی در نرم L_1 را تضمین می‌کند:

$$\|P - Q\|_{TV} \leq \sqrt{\frac{1}{4} \text{KL}(P\|Q)} \quad (۳۵-۲)$$

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

$$E_1(P\|Q) = \|P - Q\|_{TV} = \frac{1}{2} \|P - Q\|_1 \quad (۳۶-۲)$$

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

- **نمایش انتگرالی f -واگرایی‌ها^{۴۹}:** یکی از عمیق‌ترین نتایج نظری در مقاله ساسون و وردو [۱۷] (قضیه ۱۱)، بیان می‌کند که E_γ به عنوان یک «مؤلفه‌ی سازنده» یا «پایه» برای تمامی f -واگرایی‌ها عمل می‌کند. هر f -واگرایی محدب D_f (برای تابع f دوبار مشتق‌پذیر با $f(1) = 0$) را می‌توان به صورت ترکیب انتگرالی خطی از واگرایی‌های E_γ بازنویسی کرد:

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

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

⁴⁸Pinsker's Inequality

⁴⁹Integral Representation of f-Divergences

۳-۲ مبانی آماری و کران‌های اطلاعاتی

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

۱-۳-۲ نظریه تصمیم و ریسک مینی‌مکس

در چارچوب «نظریه تصمیم آماری»^{۵۰}، مسأله‌ی تخمین را می‌توان به عنوان یک بازی بین «طبیعت» (که پارامتر θ را انتخاب می‌کند) و «آمارگر» (که تخمین‌گر $\hat{\theta}$ را انتخاب می‌کند) مدل‌سازی کرد.

فرض کنید فضای نمونه \mathcal{X}^n و کلاس مدل‌های آماری $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ داده شده‌اند. داده‌های مشاهده‌شده $X^n = (X_1, \dots, X_n)$ متغیرهای تصادفی مستقلی هستند که از توزیع P_θ تولید شده‌اند. برای سنجش کیفیت یک تخمین‌گر $\hat{\theta}$ ، به یک معیار فاصله‌ی متریک (یا شبه‌متریک) $\rho : \Theta \times \Theta \rightarrow \mathbb{R}_{\geq 0}$ روی فضای پارامتر نیاز داریم.

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

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

که در آن $\Phi : [0, \infty) \rightarrow [0, \infty)$ تابعی غیرنزولی با شرط $\Phi(0) = 0$ است. انتخاب‌های رایج شامل $\Phi(t) = t^2$ (زیان مربعات) و $\Phi(t) = |t|$ (زیان قدرمطلق) است.

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

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

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

⁵⁰Statistical Decision Theory

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

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

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

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

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

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

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

تعریف ۲-۲۴ (مجموعه بسته‌بندی) مجموعه‌ی متناهی $\mathcal{V} = \{\theta_1, \dots, \theta_M\} \subset \Theta$ یک 2δ -بسته‌بندی^{۵۴} برای فضای (Θ, ρ) نامیده می‌شود اگر اعضای آن حداقل به اندازه 2δ از یک‌دیگر فاصله داشته باشند:

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

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

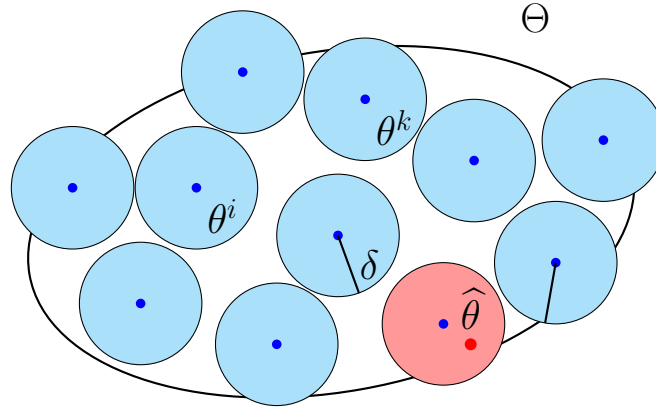
قضیه‌ی ۲-۱۱ (کران پایین عمومی) فرض کنید $\{\theta_1, \dots, \theta_M\}$ یک 2δ -بسته‌بندی برای Θ باشد. آنگاه ریسک مینی‌مکس به خطای آزمون فرض محدود می‌شود:

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

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

⁵³Minimax Approach

⁵⁴Packing



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

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

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

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

$$\begin{aligned} \mathfrak{M}_n(\Theta) &= \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{P_{\theta}} \left[\Phi(\rho(\hat{\theta}, \theta)) \right] \\ &\geq \inf_{\hat{\theta}} \sup_{\theta \in \mathcal{V}} \mathbb{E}_{P_{\theta}} \left[\Phi(\rho(\hat{\theta}, \theta)) \right] \\ &\geq \inf_{\hat{\theta}} \frac{1}{M} \sum_{v=1}^M \mathbb{E}_{P_{\theta_v}} \left[\Phi(\rho(\hat{\theta}, \theta_v)) \right] \end{aligned} \quad (2-44)$$

۲. استفاده از نامساوی مارکوف: از آن‌جا که Φ تابعی صعودی و غیرمنفی است، می‌توانیم از نامساوی

مارکوف استفاده کنیم. برای هر θ ثابت داریم:

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

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

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

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

$$\rho(\hat{\theta}, \theta_k) \leq \rho(\hat{\theta}, \theta_v) \quad (۴۷-۲)$$

از طرفی، چون v یک 2δ -بسته‌بندی است، طبق نامساوی مثلث داریم:

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

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

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

و در نتیجه برای احتمالات داریم:

$$\Pr_{P_{\theta_v}}(\psi(X^n) \neq v) \leq \Pr_{P_{\theta_v}}(\rho(\hat{\theta}, \theta_v) \geq \delta) \quad (۵۰-۲)$$

۵. جمع‌بندی: با جایگذاری ۵۰-۲ در نامساوی مارکوف (۴۵-۲) و میانگین‌گیری روی تمام v ها

(مرحله ۱)، داریم:

$$\begin{aligned}\mathfrak{M}_n(\Theta) &\geq \frac{1}{M} \sum_{v=1}^M \Phi(\delta) \cdot \Pr(\rho(\hat{\theta}, \theta_v) \geq \delta) \\ &\geq \Phi(\delta) \cdot \frac{1}{M} \sum_{v=1}^M \Pr(\psi(X^n) \neq v) \\ &= \Phi(\delta) \cdot \bar{P}_{err}(\psi)\end{aligned}\quad (51-2)$$

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

۳-۳-۲ نامساوی‌های کران پایین

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

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

$$\inf_{\psi} \bar{P}_{err}(\psi) \geq \frac{1}{4} (1 - \|P_1^n - P_2^n\|_{TV}) \quad (52-2)$$

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

قضیه ۱۳-۲ (نامساوی فانو^{۵۶}) زمانی که پارامتر مورد نظر متعلق به مجموعه‌ای بزرگ‌تر \mathcal{V} باشد ($M = |\mathcal{V}| > 2$)، نامساوی فانو کران پایین قوی‌تری ارائه می‌دهد:

$$\inf_{\psi} \bar{P}_{err}(\psi) \geq 1 - \frac{I(X^n; \mathcal{V}) + \log 2}{\log M} \quad (53-2)$$

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

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

لم ۱۴-۲ (لم اسود^{۵۷}) این لم ابزاری قدرتمند برای فضاهای پارامتر با ابعاد بالا (مانند $\{-1, 1\}^d$) است. قدرت لم اسود در شکستن یک مسئله d -بعدی دشوار به d مسئله ۱-بعدی مستقل است.

$$\mathfrak{M}_n(\Theta) \geq \frac{\delta}{2} \sum_{j=1}^d \left[1 - \|M_{+j}^n - M_{-j}^n\|_{TV} \right] \quad (۵۴-۲)$$

که در آن M_{+j}^n و M_{-j}^n توزیع‌های مخلوط حاشیه‌ای هستند (میانگین توزیع‌هایی که بیت j -ام آن‌ها به ترتیب $+1$ و -1 است). اگر در هر بُعد تمایز قائل شدن سخت باشد، مجموع خطاها با تعداد ابعاد d جمع شده و کران دقیقی^{۵۸} می‌سازد.

⁵⁸Tight Bound

فصل ۳

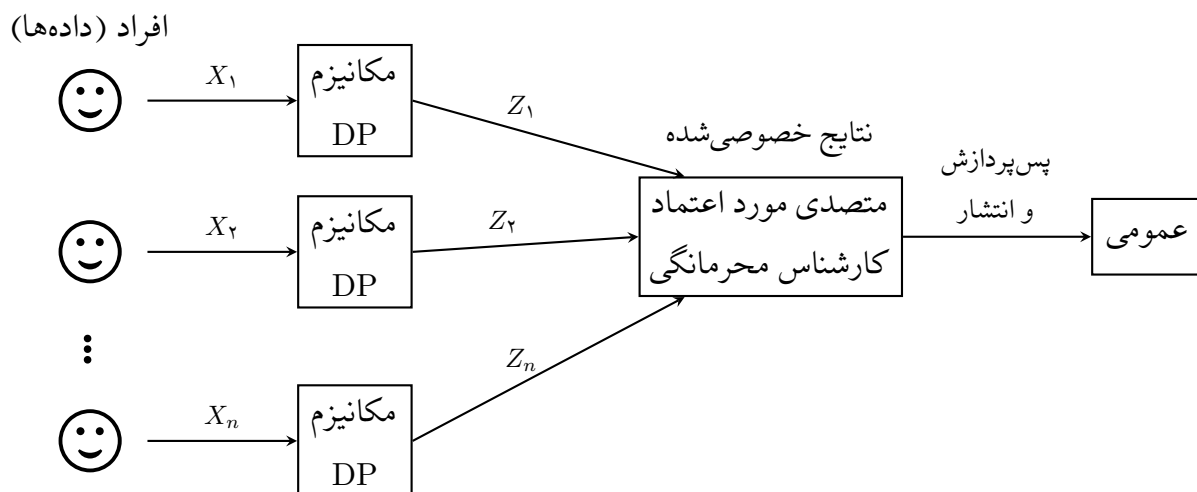
محرمانگی تفاضلی موضعی

۱-۳ مقدمه

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

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

^۱Local Differential Privacy



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

۳-۲ تعاریف رسمی و مدل های محاسباتی

در مدل محرمانگی موضعی، فرض می کنیم n کاربر در سیستم حضور دارند. داده های خصوصی این کاربران را می توان به صورت دنباله ای از متغیرهای تصادفی مستقل و هم توزیع^۲ X_1, \dots, X_n مدل سازی کرد که هر یک از توزیع احتمال ناشناخته ای P روی دامنه ای \mathcal{X} نمونه برداری شده اند:

$$X_i \stackrel{i.i.d.}{\sim} P, \quad X_i \in \mathcal{X}, \quad i = 1, \dots, n. \quad (3-1)$$

هدف نهایی تحلیل گر در این بستر، معمولاً تخمین ویژگی های آماری این توزیع P (مانند میانگین یا چگالی احتمال) بر اساس خروجی های نویزدار است.

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

تعریف ۳-۱ (تصادفی ساز موضعی^۳) یک مکانیزم تصادفی $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{Z}$ را یک تصادفی ساز موضعی می نامیم که توسط یک هسته ای انتقال احتمال^۴ یا توزیع شرطی Q مشخص می شود. به بیان دقیق ریاضی، $Q = \{Q(\cdot|x) : x \in \mathcal{X}\}$ خانواده ای از توزیع های احتمال روی فضای خروجی \mathcal{Z} است که برای هر ورودی

²Independent and Identically Distributed (i.i.d.)

⁴Probability Transition Kernel

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

$$Q(z|x) \triangleq \Pr[\mathcal{M}(x) = z], \quad \forall z \in \mathcal{Z}. \quad (2-3)$$

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

مثال ۱-۳ (کانال دودویی متقارن به عنوان تصادفی‌ساز) ساده‌ترین نمونه از یک تصادفی‌ساز موضعی، مدلی است که در آن فضای ورودی و خروجی هر دو دودویی هستند ($\mathcal{X} = \mathcal{Z} = \{0, 1\}$). فرض کنید مکانیزم \mathcal{M} ، بیت ورودی را با احتمال p حفظ کرده و با احتمال $1 - p$ تغییر می‌دهد (فلیپ می‌کند). در این صورت ماتریس انتقال Q به شکل زیر خواهد بود:

$$Q = \begin{pmatrix} Q(0|0) & Q(1|0) \\ Q(0|1) & Q(1|1) \end{pmatrix} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}. \quad (3-3)$$

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

۱-۲-۳ تعریف ریاضی LDP

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

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

$$\sup_{x, x' \in \mathcal{X}} \sup_{\mathcal{S} \subseteq \mathcal{Z}} \frac{\mathbb{P}[\mathcal{M}(x) \in \mathcal{S}]}{\mathbb{P}[\mathcal{M}(x') \in \mathcal{S}]} \leq e^\alpha \quad (4-3)$$

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

^۵Row-stochastic Matrix

^۶Randomized Response

این تعریف را می‌توان با استفاده از مفهوم «واگرایی ماکزیمم» (D_∞) که پیش‌تر معرفی شد، به صورت فشرده‌تری بیان کرد. شرط (۳-۴) دقیقاً معادل است با:

$$\sup_{x, x' \in \mathcal{X}} D_\infty(Q(\cdot|x) \parallel Q(\cdot|x')) \leq \alpha \quad (۳-۵)$$

این رابطه نشان می‌دهد که α -LDP محدودیتی سخت‌گیرانه بر روی «نسبت درست‌نمایی»^۷ توزیع‌های خروجی اعمال می‌کند و تضمین می‌دهد که مشاهده‌ی خروجی z ، اطلاعات اندکی درباره‌ی ورودی x افشا می‌کند.

۲-۲-۳ محرمانگی تقریبی

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

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

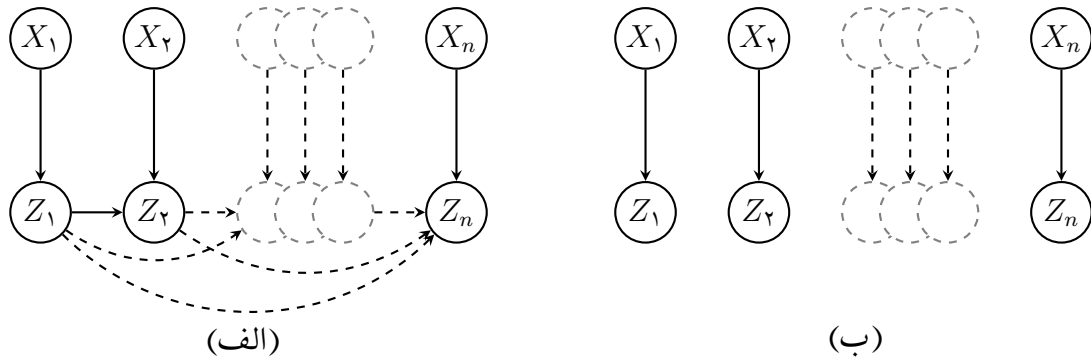
$$\mathbb{P}[\mathcal{M}(x) \in \mathcal{S}] \leq e^\alpha \cdot \mathbb{P}[\mathcal{M}(x') \in \mathcal{S}] + \delta \quad (۳-۶)$$

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

۳-۳ پروتکل‌های تعاملی و خواص ترکیب

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

⁷Likelihood Ratio



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

۳-۳-۱ پروتکل‌های غیرتعاملی

در پروتکل‌های غیرتعاملی^۸، تمام کاربران $i = 1, \dots, n$ مکانیزم‌های خود را به صورت هم‌زمان و مستقل از یک‌دیگر اجرا می‌کنند. اگر Z_i خروجی کاربر i -ام باشد، توزیع آن تنها به داده‌ی خصوصی X_i وابسته است و هیچ وابستگی‌ای به خروجی سایر کاربران ندارد. به بیان ریاضی، توزیع مشترک خروجی‌ها به صورت حاصل ضرب توزیع‌های حاشیه‌ای فاکتور می‌شود:

$$\mathbb{P}(Z_1, \dots, Z_n | X_1, \dots, X_n) = \prod_{i=1}^n Q_i(Z_i | X_i) \quad (۷-۳)$$

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

۳-۳-۲ پروتکل‌های تعاملی (ترتیبی)

در پروتکل‌های تعاملی^۹، کاربران به نوبت داده‌های خود را ارسال می‌کنند و مکانیزم کاربر i -ام می‌تواند به خروجی‌های مشاهده‌شده از کاربران پیشین (Z_1, \dots, Z_{i-1}) وابسته باشد. این مدل، آزادی عمل بیشتری را برای طراحی الگوریتم‌های تطبیقی فراهم می‌کند.

از دیدگاه آنالیز ریاضی، این فرآیند با استفاده از «هسته‌های احتمالاتی»^{۱۰} مدل‌سازی می‌شود. فرض کنید $Z_{1:i-1} = (Z_1, \dots, Z_{i-1})$ بردار خروجی‌های پیشین باشد که σ -فیلد \mathcal{F}_{i-1} را تولید می‌کند. مکانیزم کاربر

⁸Non-interactive

⁹Interactive / Sequential

¹⁰Probability Kernels

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

$$Z_i \sim Q_i(dz_i | x_i, z_{1:i-1}) \quad (۸-۳)$$

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

۳-۳-۳ قضیه ترکیب ترتیبی

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

قضیه‌ی ۱-۳ (ترکیب ترتیبی^{۱۱}) فرض کنید در یک پروتکل تعاملی، برای هر کاربر $i \in \{1, \dots, n\}$ و به ازای هر تاریخچه‌ی ممکن $z_{1:i-1} \in \mathcal{Z}^{i-1}$ ، مکانیزم $Q_i(\cdot | \cdot, z_{1:i-1})$ دارای خاصیت α_i -LDP نسبت به ورودی x_i باشد. آنگاه توزیع مشترک کل خروجی‌ها (Z_1, \dots, Z_n) ، دارای محرمانگی تفاضلی موضعی با بودجه‌ی مجموع است:

$$\alpha_{total} = \sum_{i=1}^n \alpha_i \quad (۹-۳)$$

اثبات. اثبات بر پایه خاصیت زنجیره‌ای واگرایی ماکزیمم (D_∞) یا تجزیه‌ی نسبت‌های درست‌نمایی استوار است. اگر P و P' دو توزیع احتمال روی دنباله‌ی خروجی‌ها Z^n باشند که ناشی از دو دنباله ورودی x^n و x'^n هستند، نسبت احتمال توأم به حاصل ضرب نسبت‌های شرطی تجزیه می‌شود:

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

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

۴-۳ مکانیزم‌های پایه در LDP

در این بخش، مکانیزم‌های بنیادین LDP را با رویکردی آماری تحلیل می‌کنیم. هدف اصلی در طراحی این مکانیزم‌ها، یافتن نگاهی تصادفی است که علاوه بر ارضای شرط محرمانگی، «خطای تخمین» (که معمولاً

با واریانس سنجیده می‌شود) را کمینه کند. فرض بنیادی در تمام این مکانیزم‌ها این است که برای بازیابی اطلاعات آماری (مانند هیستوگرام)، از یک «تخمین‌گر نااریب»^{۱۲} معکوس استفاده می‌شود.

۳-۴-۱ پاسخ تصادفی دودویی (RR)

پایه‌ای‌ترین مکانیزم α -LDP، پاسخ تصادفی^{۱۳} (RR) برای دامنه‌ی دودویی $\mathcal{X} = \{0, 1\}$ است. این مکانیزم را می‌توان به صورت یک کانال متقارن دودویی مدل‌سازی کرد که ورودی x را با احتمال p حفظ کرده و با احتمال $1 - p$ قرینه می‌کند:

$$\mathbb{P}[y = z | x] = \begin{cases} p & \text{if } z = x \\ 1 - p & \text{if } z \neq x \end{cases} \quad (10-3)$$

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

$$\sup_{y \in \{0, 1\}} \frac{\mathbb{P}[y | x]}{\mathbb{P}[y | x']} = \frac{p}{1 - p} \leq e^\alpha \quad (11-3)$$

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

$$p = \frac{e^\alpha}{e^\alpha + 1} \quad (12-3)$$

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

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

$$\hat{x} = \frac{y - (1 - p)}{2p - 1} \quad (13-3)$$

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

$$\text{Var}[\hat{x}] = \frac{e^\alpha}{(e^\alpha - 1)^2} \quad (14-3)$$

¹²Unbiased Estimator

¹³Randomized Response

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

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

با جایگذاری در معادله تخمین‌گر:

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

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

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

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

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

□

۳-۴-۲ پاسخ تصادفی تعمیم‌یافته (GRR)

پاسخ تصادفی تعمیم‌یافته^{۱۴}، برای دامنه‌های گسسته با $k > 2$ عنصر ($\mathcal{X} = \{1, \dots, k\}$)، به عنوان تعمیم مستقیم RR معرفی می‌شود. این مکانیزم را می‌توان با یک «ماتریس گذار»^{۱۵} تصادفی $\mathbf{Q} \in [0, 1]^{k \times k}$ توصیف کرد.

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

$$\mathbf{Q} = \begin{pmatrix} p & q & \dots & q \\ q & p & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p \end{pmatrix}_{k \times k} \quad (15-3)$$

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

¹⁴Generalized Randomized Response

¹⁵Transition Matrix

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

$$p = \frac{e^\alpha}{e^\alpha + k - 1}, \quad q = \frac{1}{e^\alpha + k - 1} \quad (16-3)$$

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

$$\text{Var}_{GRR} = \frac{p(1-p)}{(p-q)^2} = \frac{(e^\alpha)(k-1) + (k-1)^2}{(e^\alpha - 1)^2} \approx \mathcal{O}(k) \quad (17-3)$$

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

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

برای غلبه بر مشکل کاهش دقت GRR در دامنه‌های بزرگ، خانواده‌ای از مکانیزم‌ها تحت عنوان «کدگذاری یکانی»^{۱۶} توسعه یافته‌اند. این رویکرد اساس پروتکل مشهور RAPPOR گوگل را تشکیل می‌دهد [۱۰، ۲۰].

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

۱. **کدگذاری قطعی:** ورودی $x \in \{1, \dots, k\}$ به یک بردار بیتی $v \in \{0, 1\}^k$ تبدیل می‌شود که تنها در موقعیت x برابر با ۱ و در سایر مکان‌ها ۰ است (کدگذاری وان-هات^{۱۷}).

۲. **اختلال^{۱۸} مستقل:** 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:

$$\mathbb{P}[z_i = 1] = \begin{cases} p & \text{if } v_i = 1 \\ q & \text{if } v_i = 0 \end{cases} \quad (18-3)$$

مثال ۳-۲ فرض کنید دامنه شامل ۴ آیت باشد ($\mathcal{X} = \{1, 2, 3, 4\}$) و ورودی کاربر $x = 2$ باشد.

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

¹⁶Unary Encoding (UE)

¹⁷One-Hot Encoding

¹⁸Perturbation

۲. اعمال نویز: هر بیت مستقل پرتاب می‌شود. ممکن است خروجی نهایی $z = [0, 1, 1, 0]$ شود (بیت سوم از ۰ به ۱ تغییر کرده است).

اثبات α -LDP: برای بررسی شرط α -LDP، نسبت احتمال خروجی برداری $z = (z_1, \dots, z_k)$ را برای دو ورودی متمایز x و x' محاسبه می‌کنیم. بردارهای متناظر v و v' تنها در دو موقعیت تفاوت دارند: موقعیت x (که $v_x = 1, v'_x = 0$) و موقعیت x' (که $v_{x'} = 0, v'_{x'} = 1$). در سایر موقعیت‌ها ($j \neq x, x'$) بیت‌ها یکسان و برابر صفر هستند و در نسبت احتمالات ساده می‌شوند. از آن‌جا که بیت‌ها مستقل هستند:

$$\frac{\mathbb{P}[z|x]}{\mathbb{P}[z|x']} = \frac{\mathbb{P}[z_x|v_x=1]}{\mathbb{P}[z_x|v'_x=0]} \cdot \frac{\mathbb{P}[z_{x'}|v_{x'}=0]}{\mathbb{P}[z_{x'}|v'_{x'}=1]} \quad (19-3)$$

بیشینه‌ی این کسر زمانی رخ می‌دهد که صورت کسر ماکزیمم و مخرج مینیمم شود؛ یعنی زمانی که $z_x = 1$ (حفظ بیت ۱) و $z_{x'} = 0$ (حفظ بیت ۰) باشد. در این حالت:

$$\frac{\mathbb{P}[z|x]}{\mathbb{P}[z|x']} \leq \frac{p}{q} \cdot \frac{1-q}{1-p} = \frac{p(1-q)}{q(1-p)} \quad (20-3)$$

بنابراین شرط α -LDP معادل است با:

$$\alpha = \ln \left(\frac{p(1-q)}{q(1-p)} \right) \quad (21-3)$$

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

$$\text{Var}_{UE} = \frac{q(1-q)}{(p-q)^2} \quad (22-3)$$

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

دو استراتژی اصلی برای تنظیم p و q بر اساس رابطه (۲۱-۳) وجود دارد:

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

$$\text{Var}_{SUE} = \frac{e^{\alpha/2}}{(e^{\alpha/2}-1)^2} \quad (23-3)$$

۲. کدگذاری یکانی بهینه (OUE): وانگ و همکاران [۲۰] نشان دادند که برای کمینه‌سازی واریانس در دامنه‌های بزرگ، باید اطلاعات بیت‌های ۱ (سیگنال) حفظ شود ($p = 1/2$) و نویز روی بیت‌های ۰ (که اکثر بردار را تشکیل می‌دهند) کنترل شود ($q = \frac{1}{e^{\alpha}+1}$). واریانس حاصل برابر است با:

$$\text{Var}_{OUE} = \frac{4e^{\alpha}}{(e^{\alpha}-1)^2} \quad (24-3)$$

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

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

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

- در مکانیزم GRR، طبق رابطه (۳-۱۷)، واریانس تقریباً با k رشد می‌کند:

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

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

- در مکانیزم OUE، واریانس مستقل از k است:

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

این استقلال از k باعث می‌شود که خانواده UE گزینه‌ی برتر برای دامنه‌های بزرگ باشند. با این حال، OUE هزینه مخابراتی بالایی دارد (ارسال بردار با طول k). برای رفع این مشکل، نسخه بهبودیافته‌ای به نام «درهم‌سازی موضعی بهینه»^{۱۹} توسط وانگ و همکاران [۲۱] معرفی شده است. این روش با استفاده از توابع درهم‌ساز، ورودی را فشرده کرده و بدون افزایش واریانس، هزینه مخابراتی را کاهش می‌دهد.

۳-۴-۵ مکانیزم لاپلاس موضعی

برای داده‌های عددی پیوسته، رویکرد استاندارد تعمیم مکانیزم لاپلاس از مدل متمرکز به مدل موضعی است. فرض کنید دامنه ورودی یک بازه‌ی کران‌دار $\mathcal{X} \subset \mathbb{R}$ باشد. بدون کاستن از کلیت^{۲۰}، فرض می‌کنیم داده‌ها با یک تبدیل خطی به بازه‌ی $[-1, 1]$ نگاشت شده‌اند. در مدل موضعی، شرط محرمانگی باید برای هر جفت ورودی $x, x' \in \mathcal{X}$ برقرار باشد. بنابراین «حساسیت سراسری» (Δ) برابر با قطر دامنه است. در این صورت Δ ، بیشترین مقدار ممکن را خواهد داشت:

$$\Delta = \sup_{x, x' \in [-1, 1]} |x - x'| = |1 - (-1)| = 2 \quad (۳-۲۵)$$

¹⁹Optimized Local Hashing (OLH)

²⁰Without Loss of Generality

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

$$z = x + \eta, \quad \eta \sim \text{Lap}\left(\frac{\Delta}{\alpha}\right) = \text{Lap}\left(\frac{2}{\alpha}\right) \quad (26-3)$$

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

$$f(z|x) = \frac{\alpha}{4} \exp\left(-\frac{\alpha|z-x|}{2}\right) \quad (27-3)$$

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

$$\frac{f(z|x)}{f(z|x')} = \frac{\exp\left(-\frac{\alpha}{4}|z-x|\right)}{\exp\left(-\frac{\alpha}{4}|z-x'|\right)} \quad (28-3)$$

$$= \exp\left(\frac{\alpha}{4}(|z-x'| - |z-x|)\right) \quad (29-3)$$

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

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

بنابراین نسبت احتمال با $e^\alpha = \exp\left(\frac{\alpha}{4} \cdot 2\right)$ کران دار می‌شود.

مثال ۳-۴ فرض کنید می‌خواهیم دمای بدن یک بیمار را گزارش کنیم. اگر دامنه تغییرات دما $[35, 42]$ درجه باشد، طول بازه $42 - 35 = 7$ است. اگر بدون نرمال‌سازی از لاپلاس استفاده کنیم، باید نویزی متناسب با $\frac{2}{\alpha}$ اضافه کنیم. اما در روش استاندارد، ابتدا دما را به $[-1, 1]$ نگاشت می‌کنیم (که حساسیت ۲ شود)، نویز با مقیاس $\frac{2}{\alpha}$ اضافه می‌کنیم و در سمت سرور مجدداً نتیجه را به مقیاس اصلی برمی‌گردانیم. حال اگر بخواهیم دمای بدن یک بیمار را که به بازه $[-1, 1]$ نرمالایز شده است، با بودجه $\alpha = 1$ منتشر کنیم. اگر مقدار واقعی $x = 0.5$ باشد:

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

- یک نمونه تصادفی ممکن است $z = 0.5 + (-1/2) = -0.7$ باشد.

- واریانس خطا برابر است با $8 = 2b^2$. این واریانس برای یک مقدار در بازه $[-1, 1]$ بسیار زیاد است و نشان می‌دهد که LDP برای داده‌های عددی تک بعدی خطای زیادی تحمیل می‌کند مگر اینکه n (تعداد کاربران) بسیار زیاد باشد.

۳-۴-۶ مکانیزم تخمین میانگین دوجی (۱-بیتی)

همان‌طور که در مثال قبل دیدیم، مکانیزم لاپلاس به دلیل دامن‌های نامتناهی خروجی، واریانس قابل توجهی را تحمیل می‌کند. برای حل این مشکل در داده‌های کران‌دار (مثلاً $[-1, 1]$)، دوجی و همکاران [۷] یک مکانیزم بهینه ارائه کردند که خروجی را به شدت کوانتیده (باینری) می‌کند.

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

الگوریتم ۱ مکانیزم ۱-بیتی دوجی

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

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

$$u = \begin{cases} 1 & \text{با احتمال } \frac{1+x}{2} \\ -1 & \text{با احتمال } \frac{1-x}{2} \end{cases} \quad (30-3)$$

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

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

$$z = z' \cdot \frac{e^\alpha + 1}{e^\alpha - 1} \quad (31-3)$$

۵: خروجی: مقدار z .

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

$$\text{Var}_{Duchi} \leq \left(\frac{e^\alpha + 1}{e^\alpha - 1} \right)^2 \approx \mathcal{O} \left(\frac{1}{\alpha^2} \right) \quad (32-3)$$

مزیت اصلی این روش نسبت به لاپلاس، کران‌دار بودنِ قطعیِ خروجی است که در تحلیل‌های بدترین حالت^{۲۱} اهمیت زیادی دارد.

²¹Worst-case

چالش ابعاد بالا (نفرین ابعاد)

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

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

$$\Delta_{\ell_1} = \sup_{x, x' \in B_{\ell_2}} \|x - x'\|_1 \leq \sqrt{d} \cdot \sup_{x, x' \in B_{\ell_2}} \|x - x'\|_2 = \sqrt{d} \quad (3-33)$$

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

$$\bullet \text{ واریانس خطا در هر بُعد: } \frac{\Lambda d}{\alpha^2} = \left(\frac{\sqrt{d}}{\alpha} \right)^2$$

$$\bullet \text{ خطای میانگین مربعات کل}^{22} \text{ برای } d \text{ بُعد: } d \times \frac{\Lambda d}{\alpha^2} = \frac{\Lambda d^2}{\alpha^2}$$

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

مقایسه با مدل متمرکز: در مثال ۲-۶ فصل قبل دیدیم که در مدل متمرکز، مکانیزم لاپلاس با افزایش ابعاد (d) دچار افت کارایی می‌شود و نویز با ضریب d رشد می‌کند (که با استفاده از مکانیزم گوسی به \sqrt{d} کاهش می‌یابد). اما در مدل موضعی، این پدیده بسیار شدیدتر است. در اینجا نه تنها نویز با d رشد می‌کند، بلکه به دلیل عدم تمرکز و جمع شدن خطاهای تک‌تک کاربران، خطای نهایی (MSE) با d^2 افزایش می‌یابد. به همین دلیل، راهکارهای ساده‌ی مدل متمرکز (مانند افزودن نویز به هر بُعد) در مدل موضعی تقریباً بلااستفاده هستند.

۳-۵ چالش سودمندی و هزینه عدم اعتماد

همان‌طور که دیدیم، مدل LDP گلوگاه اعتماد به سرور مرکزی را حذف می‌کند. اما این افزایش امنیت بدون هزینه نیست. در این بخش، با یک تحلیل دقیق ریاضی نشان می‌دهیم که حذف متصدی مورد اعتماد منجر

²²Mean Squared Error (MSE)

به کاهش شدید دقت آماری (سودمندی) می‌شود. برای این منظور، ساده‌ترین مسئله آماری یعنی «تخمین میانگین جامعه» را در دو مدل متمرکز و موضعی مقایسه می‌کنیم.

۳-۵-۱ تعریف مسئله: تخمین میانگین دودویی

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

$$p = \frac{1}{n} \sum_{i=1}^n X_i \quad (3-34)$$

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

$$\text{MSE}(\hat{p}) = \mathbb{E}[(\hat{p} - p)^2] = \text{Var}[\hat{p}] + (\text{Bias}[\hat{p}])^2 \quad (3-35)$$

ما در هر دو مدل از تخمین‌گرهای نااریب ($\text{Bias} = 0$) استفاده می‌کنیم، بنابراین خطا صرفاً ناشی از واریانس نویز تزریق شده است.

۳-۵-۲ تحلیل در مدل متمرکز (CDP)

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

$$\hat{p}_{CDP} = \frac{1}{n} \left(\sum_{i=1}^n X_i + \eta \right), \quad \eta \sim \text{Lap}(1/\epsilon) \quad (3-36)$$

خطای این تخمین‌گر برابر است با:

$$\text{MSE}_{CDP} = \text{Var} \left[\frac{\eta}{n} \right] = \frac{1}{n^2} \text{Var}[\eta] = \frac{1}{n^2} \cdot \frac{2}{\epsilon^2} = \mathcal{O} \left(\frac{1}{n^2 \epsilon^2} \right) \quad (3-37)$$

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

۳-۵-۳ تحلیل در مدل موضعی (LDP)

در مدل موضعی با محدودیت LDP، هیچکس به X_i های خام دسترسی ندارد. هر کاربر به صورت مستقل مکانیزم پاسخ تصادفی (RR) را روی داده خود اجرا می کند و \hat{X}_i را گزارش می دهد. طبق نتایج بخش ۳-۴-۱، واریانس تخمین گر هر کاربر برای α های کوچک ($\alpha < 1$) تقریباً برابر است با:

$$\text{Var}[\hat{X}_i] \approx \frac{1}{\alpha^2} \quad (38-3)$$

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

$$\text{MSE}_{LDP} = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n \hat{X}_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[\hat{X}_i] = \frac{1}{n^2} \cdot n \cdot \mathcal{O} \left(\frac{1}{\alpha^2} \right) = \mathcal{O} \left(\frac{1}{n\alpha^2} \right) \quad (39-3)$$

در اینجا خطا با سرعت $1/\sqrt{n}$ به سمت صفر میل می کند.

۴-۵-۳ نتیجه گیری: شکاف کارایی

با مقایسه روابط (۳۷-۳) و (۳۹-۳)، تفاوت بنیادین دو مدل آشکار می شود (با فرض ثابت بودن بودجه های محرمانگی $\epsilon \approx \alpha$):

• مدل متمرکز: نرخ همگرایی خطا $\mathcal{O}(1/n)$ است.

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

این تفاوت در نرخ همگرایی پیامد بسیار مهمی در حجم نمونه^{۲۳} مورد نیاز دارد. برای رسیدن به یک دقت ثابت مشخص (مثلاً خطای τ)، تعداد کاربران مورد نیاز در هر مدل عبارت است از:

$$n_{CDP} \propto \frac{1}{\tau}, \quad n_{LDP} \propto \frac{1}{\tau^2} \quad (40-3)$$

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

$$n_{LDP} \approx \mathcal{O}(n_{CDP}^2) \quad (41-3)$$

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

²³Sample Complexity

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

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

فصل ۴

تحلیل مینی مکس و هندسه اطلاعاتی در LDP

۴-۱ مقدمه

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

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

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

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

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

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

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

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

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

۲-۴ محرمانگی به عنوان انقباض

همان‌طور که در فصل پیش‌نیازها (۲-۲) بحث شد، f -واگرایی‌ها ابزاری قدرتمند برای سنجش تمایز بین توزیع‌های احتمالی هستند. ایده مرکزی در تحلیل مینی‌مکس تحت LDP این است که مکانیزم‌های محرمانگی به عنوان «عملگرهای انقباضی»^۳ عمل می‌کنند. به عبارت دیگر، شرط α -LDP باعث می‌شود که توزیع‌های خروجی مکانیزم برای هر دو ورودی دلخواه، به یک‌دیگر بسیار نزدیک شوند و در نتیجه واگرایی بین آن‌ها محدود شود.

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

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

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

^۱Privacy as Contraction

^۲Data Processing Inequalities

^۳Contraction Operators

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

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

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

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

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

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

که در آن $q(\cdot|x)$ تابع چگالی (یا جرم) احتمال مکانیزم است. عبارت سمت چپ نامساوی (۴-۱) را بسط می‌دهیم (که همان واگرایی متقارن یا واگرایی جفریز است):

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

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

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

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

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

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

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

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

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

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

۱-۲-۴ تفسیر رژیم‌های محرمانگی

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

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

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

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

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

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

۳-۴ نامساوی‌های پردازش داده قوی (SDPI)

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

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

$$\eta_f(\mathcal{M}) = \sup_{P \neq Q} \frac{D_f(\mathcal{M}(P) \parallel \mathcal{M}(Q))}{D_f(P \parallel Q)} \quad (۱۰-۴)$$

اگر $\eta_f(\mathcal{M}) < ۱$ باشد، می‌گوییم مکانیزم دارای خاصیت SDPI است.

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

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

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

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

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

$$D_{\chi^2}(M_1 \parallel M_2) \leq (e^\alpha + 1)(e^\alpha - 1)^2 \|P_1 - P_2\|_{TV}^2 \quad (۱۱-۴)$$

^۴Strong Data Processing Inequality (SDPI)

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

۲-۳-۴ مزیت استفاده از χ^2 نسبت به KL

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

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

۲. سادگی در مخلوط‌ها: در روش‌هایی مانند لم اسود یا فانو، ما نیاز به محاسبه واگرایی بین «مخلوطی از توزیع‌ها» ($\bar{P} = \mathbb{E}_{P_\theta}$) داریم. به دلیل ساختار انتگرالی $D_{\chi^2}(P||Q) = \int \frac{P^2}{Q} - 1$ ، محاسبه کران برای مخلوط‌ها ساده‌تر از KL است که لگاریتم در داخل انتگرال دارد (نامساوی ینسن در KL جهت نامطلوبی دارد، اما در χ^2 می‌توان راحت‌تر کران بالا پیدا کرد).

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

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

۴-۴ اثبات نرخ مینی مکس برای تخمین میانگین

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

۱-۴-۴ تعریف مسئله

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

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

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

قضیه ۴-۳ (کران پایین مینی مکس برای تخمین میانگین) برای هر مکانیزم α -LDP با $\alpha \in (0, 1]$ ، خطای مینی مکس در تخمین میانگین یک توزیع روی مکعب $\{-1, 1\}^d$ حداقل از مرتبه زیر است:

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

که $c > 0$ یک ثابت عددی مطلق است.

اثبات. برای اثبات این قضیه، از روش «لم اسود»^۵ (فصل ۲) استفاده می‌کنیم. استراتژی کلی این است که یک زیرمجموعه گسسته از فضای پارامتر (یک ابرمکعب) بسازیم و نشان دهیم که تشخیص رأس‌های این مکعب تحت محدودیت α -LDP دشوار است.

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

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

^۵Assouad's Lemma

که $\Delta \in (0, 1]$ پارامتری است که بعداً مقدار دقیق آن را تعیین می‌کنیم. به عبارت دیگر، بردار میانگین متناظر با v برابر است با $\theta_v = \Delta \cdot v$.

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

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

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

۳. **استفاده از خاصیت انقباض LDP:** اکنون باید فاصله تغییرات کل $\|M_v^n - M_{v'}^n\|_{TV}$ را کران‌دار کنیم. دو همسایه v و v' را در نظر بگیرید که تنها در یک مؤلفه (مثلاً مؤلفه j) تفاوت دارند. طبق نامساوی پینسکر و خاصیت تانسوری واگرایی KL برای نمونه‌های مستقل (Z_1, \dots, Z_n) :

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

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

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

چون P_v و $P_{v'}$ توزیع‌های برنولی ضربی هستند که تنها در مؤلفه j تفاوت دارند (با میانگین‌های $\Delta + 1$ و $-\Delta$)، فاصله تغییرات کل آن‌ها دقیقاً برابر است با تفاوت میانگین‌ها تقسیم بر دامنه (در این جا ساده‌سازی شده):

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

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

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

بنابراین:

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

۴. تنظیم پارامتر Δ : برای این که کران پایین در رابطه (۴-۱۵) غیر صفر و بزرگ باشد، باید عبارت داخل پرانتز مثبت باشد. ما Δ را چنان انتخاب می کنیم که فاصله TV برابر یک مقدار ثابت کوچک (مثلاً $1/2$) شود:

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

با جایگذاری این مقدار Δ در رابطه لم اسود:

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

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

۲-۴-۴ بحث و تفسیر

نتیجه به دست آمده در قضیه ۴-۳ دارای پیام های مهمی برای طراحی سیستم های خصوصی است:

۱. هزینه محرمانگی: در مقایسه با تخمین میانگین در حالت غیرخصوصی (یا متمرکز) که نرخ خطا $\frac{d}{n}$ است، در حالت α -LDP خطا با ضریب $\frac{1}{\alpha^2}$ افزایش می یابد. این یعنی برای جبران نویز اضافه شده توسط محرمانگی، حجم داده ها (n) باید متناسب با مربع بودجه محرمانگی افزایش یابد.

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

۳. بهینگی مکانیزم ها: در ۳-۴-۶ دیدیم که مکانیزم تخمین میانگین دوجی دقیقاً واریانسی از مرتبه $\frac{d \log d}{n\alpha^2}$ (برای روش های تنک) یا $\frac{d}{n\alpha^2}$ تولید می کند. تطابق کران پایین ثابت شده در این جا با کران بالای آن مکانیزم ها، ثابت می کند که الگوریتم های موجود نه تنها خوب، بلکه «بهینه مینی مکس» هستند و بهبود قابل توجهی در نرخ همگرایی آن ها ممکن نیست.

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

Bibliography

- [1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):131–142, 1966.
- [2] Shahab Asoodeh, Maryam Aliakbarpour, and Flavio P. Calmon. Local differential privacy is equivalent to contraction of an f -divergence. In *2021 IEEE International Symposium on Information Theory (ISIT)*, page 545–550. IEEE Press, 2021.
- [3] Shahab Asoodeh and Huanyu Zhang. Contraction of locally differentially private mechanisms. *IEEE Journal on Selected Areas in Information Theory*, 5:385–395, 2024.
- [4] Michael Barbaro and Tom Zeller. A face is exposed for aol searcher no. 4417749. *New York Times*, 01 2006.
- [5] Imre Csiszár. Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten. *A Magyar Tudományok Akadémia Matematikai Kutató Intézetének Közleményei*, 8(1-2):85–108, 1963.
- [6] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438, 2013.
- [7] John C Duchi, Michael I Jordan, and Martin J Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.
- [8] Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.
- [9] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*, 9(3–4):211–407, 2014.

- [10] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14*, page 1054–1067, New York, NY, USA, 2014. Association for Computing Machinery.
- [11] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.*, 17(1):492–542, January 2016.
- [12] Gautam Kamath. CS860: Algorithms for private data analysis – lecture 17: Local differential privacy. Lecture Notes, University of Waterloo, 2020. <http://www.gautamkamath.com/CS860notes/lec17.pdf> (Accessed: 2026-02-14).
- [13] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 531–540, 2008.
- [14] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE, 2017.
- [15] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125, 2008.
- [16] Yury Polyanskiy, H. Vincent Poor, and Sergio Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theor.*, 56(5):2307–2359, May 2010.
- [17] Igal Sason and Sergio Verdu. f -divergence inequalities. *IEEE Trans. Inf. Theor.*, 62(11):5973–6006, November 2016.
- [18] Latanya Sweeney. k -anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.
- [19] U.S. House of Representatives Committee on Oversight and Government Reform. The equifax data breach. Majority staff report, U.S. House of Representatives, December 2018.
- [20] Teng Wang, Xuefeng Zhang, Jingyu Feng, and Xinyu Yang. A comprehensive survey on local differential privacy toward data statistics and analysis. *Sensors*, 20:1–48, 2020.
- [21] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In *Proceedings of the 26th*

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

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

واژه‌نامه

الف	
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