دانشگاه صنعتی شریف
دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد
ریاضی کاربردی

تحلیل نظریه اطلاعاتیِ محرمانگی تفاضلی موضعی و کاربردهای آماری آن

نگارش

فیروزه ابریشمی

استاد راهنما

دکتر جواد ابراهیمی بروجنی

اسفند ۱۴۰۴

بسم الله الرحمن الرحیم

به نام خدا

دانشگاه صنعتی شریف

دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد

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

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

نگارش: فیروزه ابریشمی

کمیته ممتحنین

استاد راهنما: دکتر جواد ابراهیمی بروجنی امضاء:

استاد مشاور: استاد مشاور امضاء:

استاد مدعو: استاد ممتحن امضاء:

تاریخ:

اظهارنامه

(اصالت متن و محتوای پایان‌نامه کارشناسی ارشد)

عنوان پایان‌نامه: تحلیل نظریه اطلاعاتیِ محرمانگی تفاضلی موضعی و کاربردهای آماری آن

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

این‌جانب فیروزه ابریشمی اظهار می‌دارم:

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

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

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

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

نگارنده: فیروزه ابریشمی
تاریخ:
امضاء:

استاد راهنما: دکتر جواد ابراهیمی بروجنی
تاریخ:
امضاء:

نگارنده: فیروزه ابریشمی
تاریخ:
امضاء:

چکیده

کلیدواژه‌ها:

فهرست مطالب

فهرست جداول

فهرست تصاویر

فصل ۱

مقدمه

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

۱–۱ اهمیت داده‌ها و ضرورت حفظ حریم خصوصی

در دهه‌های اخیر، جهان شاهد رشد انفجاری در تولید و جمع‌آوری داده‌ها بوده است. پیشرفت‌های چشم‌گیر در فناوری‌های ذخیره‌سازی، محاسبات ابری و اینترنت اشیاء، منجر به انباشت حجم عظیمی از داده‌ها شده است که اغلب تحت عنوان کلان‌داده[1] شناخته می‌شوند. این داده‌ها سوخت اصلی موتورهای تصمیم‌گیری مدرن و سیستم‌های هوشمند هستند. امروزه، الگوریتم‌های یادگیری ماشین[2] و تحلیل داده[3] 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[4] افراد به وجود آورده است. داده‌های خامی که برای آموزش مدل‌های هوشمند یا استخراج آماره‌ها استفاده می‌شوند، اغلب حاوی اطلاعات حساس[5] و شخصی هستند. تاریخچه‌ی تراکنش‌های مالی، سوابق پزشکی، موقعیت‌های مکانی و حتی الگوهای جستجو در وب، همگی می‌توانند جزئیات دقیقی از زندگی

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

خصوصی افراد را فاش کنند. بنابراین، یک چالش اساسی شکل می‌گیرد: چگونه می‌توان از سودمندی[6] 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[7] (مانند نام، کد ملی و شماره تلفن) برای محافظت از هویت افراد کافی است. این فرایند که گمنام‌سازی [8] 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 [9] به شدت آسیب‌پذیر هستند. در این نوع حملات، مهاجم با استفاده از اطلاعات جانبی[10] 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[11] می‌گویند)، موفق شد پرونده پزشکی فرماندار وقت ایالت ماساچوست را شناسایی کند [۲۰].
- **مجموعه داده‌ی نتفلیکس[12]:** 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 [۱۷].
- **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 [۵].

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

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

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

۱-۱-۱ محرمانگی تفاضلی (Differential Privacy)

در پاسخ به چالش‌های امنیتی و ناکارآمدی روش‌های سنتی گمنام‌سازی، دؤرک و همکاران در سال ۲۰۰۶ مفهوم محرمانگی تفاضلی[۱۳] را معرفی کردند [۱۱]. این چارچوب ریاضی دقیق، به جای تمرکز بر ویژگی‌های ظاهری داده‌ها (مانند حذف نام‌ها)، بر فرایند تولید خروجی تمرکز دارد و تضمین می‌کند که حضور یا عدم حضور یک فرد خاص در پایگاه‌داده، تأثیر ناچیزی بر خروجی نهایی الگوریتم داشته باشد.

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

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

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

فرد به طور چشم‌گیری افزایش یابد.

برای مدل‌سازی این مفهوم، می‌توانیم از استعاره‌ی جهان‌های موازی استفاده کنیم. دو پایگاه‌داده‌ی همسایه[۱۸] (مانند $\mathcal{D}$ و $\mathcal{D}'$) را به عنوان دو جهان موازی در نظر بگیرید که در یکی، داده‌های کاربر x وجود دارد و در دیگری، این داده‌ها حذف یا تغییر یافته‌اند. هدف نهایی این است که از دیدگاه یک ناظر بیرونی (مهاجم)، این دو جهان غیرقابل تفکیک[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 ($f(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} x_i$). مهاجم می‌تواند دو پرس‌وجو انجام دهد:

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

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

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

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

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

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

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

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

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

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

این رابطه اثبات می‌کند که هیچ سطح محدودی از محرمانگی ($\varepsilon < \infty$) با توابع قطعیِ غیرثابت قابل دستیابی نیست. بنابراین، همان‌طور که در ادبیات موضوع تأکید شده است [۱۲]، برای شکستن این وابستگی قطعی و ایجاد ابهام آماری، تصادفی‌سازی[۲۷] در فرآیند مکانیزم الزامی است.

کاربردهای محرمانگی تفاضلی

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

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

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

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

- **یادگیری ماشین خصوصی**[۳۰]**:** در آموزش مدل‌های عمیق، خطر به‌خاطرسپاری[۳۱] داده‌های آموزشی وجود دارد. با استفاده از الگوریتم‌هایی نظیر DP-SGD، می‌توان مدل‌هایی آموزش داد که الگوهای کلی را یاد می‌گیرند اما قادر به بازتولید داده‌های آموزشی حساس (مانند تصاویر چهره یا متون خصوصی) نیستند.

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

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

۱-۱-۲ محرمانگی تفاضلی موضعی (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[۲۱].

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

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

گذار به مدل موضعی، حذف نیاز به اعتماد

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

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

۱–۲ کارهای پیشین و مرور ادبیات

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

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

نقطه عطف این تحول، معرفی مفهوم **محرمانگی تفاضلی**[۳۶] توسط دِوُرک و همکاران بود [۱۱]. این تعریف، برخلاف روش‌های پیشین که بر ویژگی‌های داده تمرکز داشتند، بر ویژگی‌های مکانیزم پردازش داده تمرکز دارد. در مدل استاندارد (متمرکز)، یک مکانیزم تصادفی $\mathcal{M}$ دارای شرایط ε-DP است اگر برای هر دو پایگاه‌داده همسایه $\mathcal{D}$ و $\mathcal{D}'$ (که تنها در داده یک فرد متفاوت‌اند) و برای هر زیرمجموعه از خروجی‌ها $\mathcal{S} \subseteq \mathrm{Range}(\mathcal{M})$، رابطه زیر برقرار باشد:

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

[35] Statistical Disclosure Control
[36] Differential Privacy

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

- **مقادیر کوچک ε (محرمانگی قوی):** زمانی که $\varepsilon \to ۰$، 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 [۱۶]، اما ریشه‌های عملی آن به دهه‌ها قبل باز می‌گردد.

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

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

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.*

---
[37] Privacy Budget
[38] Local Differential Privacy

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

در واقع، ساده‌ترین و نخستین نمونه از یک مکانیزم LDP، روش «پاسخ تصادفی»[۳۹] است که توسط وارنر در سال ۱۹۶۵ برای حذف سوگیری در نظرسنجی‌های حساس معرفی شد [۲۴]. 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:

$$\frac{\Pr[\mathcal{M}(x) = z]}{\Pr[\mathcal{M}(x') = z]} \leqslant \frac{p}{1 - p} \qquad (۳-۱)$$

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

$$e^{\alpha} = \frac{p}{1 - p} \quad \Rightarrow \quad \alpha = \ln\left(\frac{p}{1 - p}\right) \qquad (۴-۱)$$

این فرمول، تفسیر شهودی LDP را کامل می‌کند:

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

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

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

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

اما پاسخ تصادفی تنها راهکاری برای ایجاد محرمانگی در داده‌های دودویی است، و در کاربردهای مدرن با چالش دامنه‌ی بسیار بزرگ[41] روبروست. شرکت‌های بزرگ فناوری نیاز دارند داده‌هایی نظیر «آدرس‌های اینترنتی بازدید شده» یا «کلمات جدید تایپ‌شده» را جمع‌آوری کنند که دامنه‌ی آن‌ها($\mathcal{X}$) می‌تواند شامل میلیون‌ها حالت باشد. اعمال مستقیم RR در این حالات منجر به نویز بسیار زیاد و کاهش شدید سودمندی می‌شود. در ادامه، راهکارهای اتخاذ شده توسط بزرگ‌ترین شرکت‌های فناوری را مرور می‌کنیم:

- **گوگل و پروتکل RAPPOR:** در سال ۲۰۱۴، گوگل برای جمع‌آوری آمار تنظیمات مرورگر کروم و شناسایی بدافزارها، پروتکل RAPPOR[42] را معرفی کرد [۲۲]. چالش اصلی گوگل، جمع‌آوری رشته‌های متنی[43] بود. راه‌حل آن‌ها ترکیب پاسخ تصادفی با فیلترهای بلوم[44] 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[45] و تبدیل‌های ریاضی مانند تبدیل هادامارد[46] 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 [۲۲].

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

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

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

۱-۲-۲ چالش 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**[۴۷] داده‌ها یا همان دقت تحلیل‌های آماری است. این پدیده تحت عنوان **موازنه محرمانگی‐دقت**[۴۸] شناخته می‌شود.

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

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

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

این انباشت نویز باعث می‌شود که **نسبت سیگنال به نویز**[۴۹] در مدل موضعی بسیار پایین‌تر از مدل متمرکز باشد. به بیان دیگر، برای دستیابی به همان سطح از دقت که در مدل متمرکز وجود دارد، در مدل LDP نیازمند تعداد بسیار بیش‌تری نمونه داده هستیم.

این مسئله در کاربردهای عملی بسیار حائز اهمیت است. برای مثال، اگر هدف تخمین فراوانی یک بیماری نادر باشد، نویز اضافه شده توسط مکانیزم‌های LDP ممکن است سیگنال اصلی را کاملاً بپوشاند.

[47] Utility
[48] Privacy-Accuracy Trade-off
[49] Signal-to-Noise Ratio (SNR)

همین چالش بود که پژوهشگران را بر آن داشت تا به جای استفاده از روش‌های ساده (مثل وارنر)، به دنبال پاسخ این پرسش باشند که: «آیا می‌توان مکانیزم‌هایی طراحی کرد که با کم‌ترین میزان نویز، بیش‌ترین محرمانگی را فراهم کنند؟» و «حد نهایی این دقت کجاست؟» این پرسش‌ها زمینه را برای ورود تئوری‌های پیشرفته‌تر نظیر «تخمین مینیماکس» فراهم کرد.

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

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

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

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

$$\mathcal{E}_{LDP} \asymp \frac{۱}{\sqrt{n\alpha^{۲}}} \quad \text{در حالی که} \quad \mathcal{E}_{CDP} \asymp \frac{۱}{n\varepsilon}$$

این نتیجه که به «قانون مقیاس‌دهی کانونی»[۵۴] معروف است، حاوی دو پیام مهم است:

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

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

[50] Statistical Minimax Theory
[51] Duchi, Jordan, and Wainwright
[52] Fano's Inequality
[53] Assouad's Lemma
[54] Canonical Scaling Law

حجم داده‌ها باید با ضریب $۱/\alpha^۲$ افزایش یابد.

پس از استقرار این چارچوب نظری، تمرکز جامعه علمی بر طراحی «مکانیزم‌های بهینه ترتیب‌مقدماتی»[55] قرار گرفت که بتوانند به این کران‌های نظری دست یابند. از جمله مهم‌ترین این تلاش‌ها می‌توان به معرفی «مکانیزم‌های پله‌ای»[56] توسط کایروز و همکاران [۱۴] و توسعه پروتکل‌های پیشرفته‌ای نظیر UE (کدگذاری یگانی)[57] و OLH (هشینگ محلی بهینه)[58] توسط وانگ و همکاران [۲۲] اشاره کرد. این روش‌ها تلاش می‌کنند با بهینه‌سازی ساختار نویز و استفاده از تکنیک‌های فشرده‌سازی اطلاعات، فاصله بین عملکرد عملی و حدود نظری مینیماکس را به حداقل برسانند.

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

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

۱. **پروتکل‌های غیرتعاملی**[59]: در این حالت، تمام کاربران به صورت هم‌زمان و مستقل عمل می‌کنند. هر کاربر i مکانیزم $\mathcal{M}$ را تنها بر اساس داده‌ی خودش X_i اجرا کرده و پیام Z_i را به سرور می‌فرستد. هیچ ارتباطی بین کاربران وجود ندارد و سرور نیز هیچ بازخوردی به کاربران نمی‌دهد. به دلیل سادگی پیاده‌سازی و مقیاس‌پذیری بالا، اکثر پروتکل‌های صنعتی (مانند RAPPOR گوگل یا سیستم‌های اپل) در این دسته قرار می‌گیرند.

۲. **پروتکل‌های تعاملی**[60]: 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 $(Z_۱, \dots, Z_{i-۱})$ را از سرور دریافت کند و نویز خود را هوشمندانه‌تر تنظیم نماید. اگرچه به نظر می‌رسد این آزادی عمل باید دقت را افزایش دهد، اما دوچی و همکاران در نتایج حیرت‌انگیزی نشان دادند که برای دسته‌ی بزرگی از توابع محدب (مانند تخمین میانگین)، پروتکل‌های تعاملی هیچ مزیتی نسبت به روش‌های غیرتعاملی ندارند و نرخ مینیماکس را بهبود نمی‌بخشند [۹، ۱۴].

[55] Order-optimal Mechanisms
[56] Staircase Mechanisms
[57] Unary Encoding
[58] Optimal Local Hashing
[59] Non-interactive / Simultaneous
[60] Interactive / Sequential

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

۱-۳ بیان مسئله و اهداف پژوهش

محرمانگی تفاضلی موضعی (LDP) به عنوان یک حوزه‌ی میان‌رشته‌ای، محل تلاقی «علوم کامپیوتر» (با تمرکز بر طراحی الگوریتم و امنیت) و «آمار ریاضی» (با تمرکز بر نظریه تخمین و مینیماکس) است. این ماهیت دوگانه، اگرچه باعث غنای ادبیات موضوع شده، اما منجر به پراکندگی قابل‌توجهی در روش‌ها و ابزارهای تحلیلی گشته است.

همان‌طور که در مرور ادبیات اشاره شد، چارچوب مینیماکس که توسط دوچی و همکاران [۹] پایه‌گذاری شده، نشان می‌دهد که اعمال محدودیت محرمانگی منجر به کاهش نرخ هم‌گرایی در تخمین‌های آماری می‌شود. با این حال، اثبات این نتایج در مقالات مختلف با ابزارهای متفاوتی صورت گرفته است که منجر به نوعی «اثر برج بابل» در ادبیات علمی شده است. برای مثال:

- در برخی مسائل تخمین چگالی، پژوهشگران عمدتاً از **واگرایی کولبک-لایبلر (KL)** و نامساوی فانو استفاده کرده‌اند.
- در مسائل آزمون فرض ساده، اغلب از **فاصله‌ی تغییرات کل (TV)** و لم لوکام بهره گرفته شده است.
- در تحلیل‌های اخیرتر، **فاصله‌ی کای-دو** ($\chi^۲$) به دلیل رفتار هموارتر در همسایگی صفر و ارتباط مستقیم با واریانس، مورد توجه قرار گرفته است.

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

۱-۳-۱ رویکرد تحلیل: fـ 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 **fـ واگرایی‌ها**[۶۱] بازنویسی و تفسیر شوند. fـ واگرایی‌ها (معرفی شده توسط سیسار[۶۲])، کلاسی جامع از معیارهای فاصله بین دو توزیع احتمال P و Q هستند:

$$D_f(P\|Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right] \tag{۱-۵}$$

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

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

۱-۳-۲ اهداف و ساختار پژوهش

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

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

۲. **تبیین روش‌های کران پایین:** تشریح دقیق و گام‌به‌گام متدهای کلاسیک آماری نظیر «متد دو نقطه‌ای لوکام»[۶۵] و «نامساوی فانو»[۶۶] در بستر محرمانگی. هدف این است که نشان دهیم چگونه می‌توان

[61] f-divergences
[62] Csiszár
[63] Contraction Channel
[64] Contraction Coefficient
[65] Le Cam's Method
[66] Fano's Inequality

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

۳. **تفسیر هندسی رفتار مکانیزم‌ها:** تحلیل «رفتار انقباضی»[۶۷] مکانیزم‌ها با هدف محاسبه‌ی ضرایب انقباض برای fـ واگرایی‌های مختلف. ما نشان می‌دهیم که چگونه مکانیزم‌های محرمانگی (مانند پاسخ تصادفی) باعث کاهش فاصله بین توزیع‌ها می‌شوند و این کاهش چگونه به طور مستقیم بر خطای تخمین تأثیر می‌گذارد.

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

۱ـ۴ ساختار پایان‌نامه

ساختار ادامه‌ی این پایان‌نامه به شرح زیر سازمان‌دهی شده است:

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

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

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

[۶۷] Contraction Behavior

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

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

فصل ۲

پیش‌نیازها

۲-۱ محرمانگی تفاضلی متمرکز (CDP)

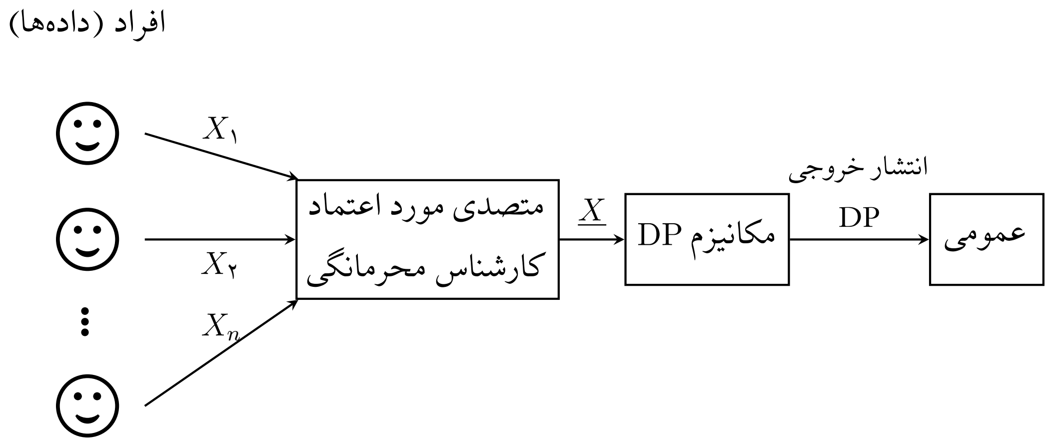
مفهوم محرمانگی تفاضلی[1] یا به اختصار DP، اولین بار توسط دِوُرک و همکاران[۱۱] معرفی شد و به سرعت به استاندارد طلایی برای حفظ حریم خصوصی در تحلیل داده‌ها تبدیل گشت. این چارچوب، یک تعریف ریاضی قوی از حریم خصوصی ارائه می‌دهد که مبتنی بر پنهان‌سازی حضور یا عدم حضور یک فرد خاص در مجموعه‌داده است.

۲-۱-۱ مدل اعتماد و تعریف رسمی

در مدل متمرکز[2]، فرض بر این است که یک متصدی مورد اعتماد[3] وجود دارد. تمام افراد داده‌های خام و حساس خود را در اختیار این متصدی قرار می‌دهند (شکل ۲-۱ را ببینید). متصدی، مجموعه‌داده‌ی کامل $\mathcal{D}$ را در اختیار دارد. وظیفه‌ی متصدی این است که با اجرای یک مکانیزم تصادفی[4] $\mathcal{M}$ بر روی مجموعه‌داده‌ی $\mathcal{D}$، نتایجی (مثلاً پاسخ به یک پرس‌وجو[5]) را به صورت عمومی منتشر کند، به طوری که اطلاعات حساس افراد فاش نشود.

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

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



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

داده‌ها»[6] می‌نامیم و آن را با $\mathcal{X}$ نمایش می‌دهیم. یک پایگاه‌داده‌ی $\mathcal{D}$ مجموعه‌ای از رکوردهاست که اعضای آن از $\mathcal{X}$ انتخاب شده‌اند. در ادبیات محرمانگی تفاضلی، پایگاه‌داده معمولاً به صورت یک بردار (هیستوگرام) $x \in \mathbb{N}^{|\mathcal{X}|}$ نمایش داده می‌شود که در آن هر مولفه x_i نشان‌دهنده‌ی تعداد تکرار عنصر iـ ام از $\mathcal{X}$ در پایگاه‌داده است.

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

$$\mathcal{D} = (۳, ۱)$$

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

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

[6] Data Universe

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

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

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

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

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

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

تعریف ۲-۴ (پایگاه‌داده‌های همسایه) *دو پایگاه‌داده‌ی $\mathcal{D}_۱$ و $\mathcal{D}_۲$ را همسایه[۷] می‌گوییم (و با $\mathcal{D}_۱ \sim \mathcal{D}_۲$ نشان می‌دهیم) اگر فاصله‌ی $\ell_۱$ آن‌ها حداکثر ۱ باشد:*

$$\|\mathcal{D}_۱ - \mathcal{D}_۲\|_۱ \leqslant ۱$$

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

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

تعریف ۲-۵ (εــ محرمانگی تفاضلی (ε-DP)) *یک مکانیزم تصادفی $\mathcal{M}$ با دامنه $\mathcal{X}^n$ و برد $\mathcal{R}$، ویژگی εــ محرمانگی تفاضلی را برآورده می‌سازد، اگر برای هر دو پایگاه‌داده‌ی همسایه‌ی $\mathcal{D}_۱$ و $\mathcal{D}_۲$ ($\mathcal{D}_۱ \sim \mathcal{D}_۲$) و برای هر زیرمجموعه از خروجی‌های ممکن $\mathcal{S} \subseteq \mathcal{R}$ (که در σــ جبر برد تعریف شده باشد)، داشته باشیم:*

$$\Pr[\mathcal{M}(\mathcal{D}_۱) \in \mathcal{S}] \leqslant \exp(\varepsilon) \cdot \Pr[\mathcal{M}(\mathcal{D}_۲) \in \mathcal{S}] \tag{۲-۲}$$

در این نامساوی، پارامتر $\varepsilon \geqslant ۰$ را ***بودجه‌ی محرمانگی***[۸] *می‌نامیم.*

[7]Neighboring / Adjacent
[8]Privacy Budget

۲-۱-۲ تفسیر پارامترهای محرمانگی

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

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

تعریف ۲-۶ (DP 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:*

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

پارامتر $\delta \in [0,1]$ را معمولاً ***احتمال شکست***[9] *یا احتمال نشت اطلاعات می‌نامند. تفسیر شهودی این است که مکانیزم با احتمال حداقل $1-\delta$، 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 ۲-۱ (هم‌ارزی نقطه‌ای) *یک مکانیزم $\mathcal{M}$ شرط ε-DP را برآورده می‌کند اگر و تنها اگر برای تمام همسایه‌های $\mathcal{D}_1 \sim \mathcal{D}_2$ شرایط زیر برقرار باشد:*

[9] Failure Probability

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

$$\frac{\Pr[\mathcal{M}(\mathcal{D}_۱) = z]}{\Pr[\mathcal{M}(\mathcal{D}_۲) = z]} \leqslant e^{\varepsilon} \tag{۲-۴}$$

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

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

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

$$\Pr[\mathcal{M}(\mathcal{D}_۱) \in \mathcal{S}] = \int_{\mathcal{S}} p(z) d\mu(z) \leqslant \int_{\mathcal{S}} e^{\varepsilon} q(z) d\mu(z) = e^{\varepsilon} \Pr[\mathcal{M}(\mathcal{D}_۲) \in \mathcal{S}]$$

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

نکته: این هم‌ارزی تنها برای $\delta = ۰$ صادق است. برای (ε, δ)-DP، بررسی نقطه‌به‌نقطه کافی نیست و شرط باید روی زیرمجموعه‌ها چک شود.

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

$$\sup_{\mathcal{D}_۱ \sim \mathcal{D}_۲} D_{\infty}(\mathcal{M}(\mathcal{D}_۱) \parallel \mathcal{M}(\mathcal{D}_۲)) \leqslant \varepsilon \tag{۲-۶}$$

در بخش‌های بعدی با تعریف دقیق واگرایی آشنا خواهیم شد.

۲-۱-۴ مکانیزم‌های پایه

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

[10] Max Divergence
[11] Noise
[12] Sensitivity

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

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

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

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

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

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

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

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

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

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

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

مثال ۲-۴ *فرض کنید f یک تابع هیستوگرام شمارشی باشد (تعداد افراد در دسته‌های مجزا). اگر مشخصات یک فرد تغییر کند، او از یک دسته خارج (تغییر -۱) و به دسته‌ی دیگر وارد (تغییر $+۱$) می‌شود. سایر دسته‌ها ثابت می‌مانند (۰). بردار تغییرات برابر است با $(۰, \dots, +۱, \dots, -۱, \dots, ۰)$. حال حساسیت‌ها را محاسبه می‌کنیم:*

[13] 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_∞:** *ماکزیمم قدرمطلق تغییرات:* $\max(|1|, |-1|, 0) = 1$.

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

انتخاب نوع حساسیت در طراحی مکانیزم، بستگی مستقیم به نوع نویز افزوده شده و ابعاد داده‌ها دارد. به طور خلاصه، حساسیت ℓ_1 برای کالیبره کردن مکانیزم لاپلاس و حساسیت ℓ_2 برای مکانیزم گوسی استفاده می‌شود. جزئیات این انتخاب و تأثیر آن بر خطای نهایی، در بخش‌های آینده و پس از معرفی این مکانیزم‌ها به تفصیل بررسی خواهد شد.

۲-۱-۵ مکانیزم‌های بنیادی محرمانگی تفاضلی

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

مکانیزم لاپلاس

ساده‌ترین و پرکاربردترین روش برای توابع عددی، افزودن نویز از توزیع لاپلاس است. توزیع لاپلاس با پارامتر مقیاس b و میانگین μ دارای تابع چگالی احتمال $h(z) = \frac{1}{2b} \exp(-\frac{|z-\mu|}{b})$ است.

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

$$\mathcal{M}_{\text{Lap}}(\mathcal{D}) = f(\mathcal{D}) + (Y_1, \dots, Y_k) \tag{۲-۱۱}$$

که در آن $Y_i \overset{i.i.d}{\sim} \text{Lap}(\frac{\Delta_1 f}{\varepsilon})$، شرط ε-DP را برآورده می‌کند.

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

$z = (z_۱, \ldots, z_k)$ بررسی می‌کنیم:

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

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

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

با اعمال این نامساوی در مجموع توان نمایی:

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

عبارت سمت راست دقیقاً برابر با نرم $\ell_۱$ تفاضل خروجی‌هاست:

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

طبق تعریف حساسیت سراسری، می‌دانیم $\|f(\mathcal{D}_۱) - f(\mathcal{D}_۲)\|_۱ \leqslant \Delta_۱ f$. بنابراین:

$$\frac{p(z|\mathcal{D}_۱)}{p(z|\mathcal{D}_۲)} \leqslant \exp\left(\frac{\Delta_۱ f}{b}\right) = \exp\left(\frac{\Delta_۱ f}{\Delta_۱ f/\varepsilon}\right) = e^{\varepsilon}$$

و حکم ثابت می‌شود. □

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 [۱۲].*

حالت تک پرس‌وجو: *حساسیت یک پرس‌وجوی شمارشی دقیقاً ۱ است ($\Delta_۱ f = ۱$)؛ زیرا افزودن یا حذف یک فرد، نتیجه‌ی شمارش را حداکثر ۱ واحد تغییر می‌دهد. بنابراین طبق قضیه ۲-۲، با افزودن نویز با مقیاس $۱/\varepsilon$ (یعنی $\mathrm{Lap}(۱/\varepsilon)$) به پاسخ واقعی، محرمانگی $(\varepsilon, 0)$-DP تضمین می‌شود. خطای مورد انتظار در این حالت $۱/\varepsilon$ است که مستقل از اندازه‌ی پایگاه‌داده می‌باشد.*

[14] Counting Queries

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

مکانیزم گوسی

زمانی که حساسیت ℓ_2 تابع بسیار کم‌تر از حساسیت ℓ_1 باشد (مثلاً در پرس‌وجوهای برداری)، مکانیزم گوسی ترجیح داده می‌شود.

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

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

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

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

$$\sigma \geqslant \sqrt{2 \ln(1.25/\delta)} \cdot \frac{\Delta_2 f}{\varepsilon} \tag{۲-۱۳}$$

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

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

[15] Privacy Loss Random Variable

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

۳. **شرط انحراف معیار:**
محاسبات نشان می‌دهد که اگر واریانس نویز ($\sigma^۲$) به اندازه کافی بزرگ باشد (رابطه صورت قضیه)، با احتمال حداقل $۱ - \delta$، نسبت چگالی‌ها کم‌تر از e^ε باقی می‌ماند.

□

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

تحلیل حساسیت:

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

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

- *مکانیزم لاپلاس باید نویزی متناسب با $d = ۱۰۰$ اضافه کند تا ε-DP حفظ شود. این حجم از نویز عملاً داده‌ها را بی‌فایده می‌کند.*
- *مکانیزم گوسی نویزی متناسب با $\sqrt{۱۰۰} = ۱۰$ اضافه می‌کند (با پذیرش یک δ ناچیز).*

*این کاهش ۱۰ برابری نویز (و در حالت کلی کاهش با ضریب $\sqrt{d}$) دلیل اصلی استفاده از مکانیزم گوسی در کاربردهای پیشرفته نظیر **یادگیری عمیق با محرمانگی تفاضلی (DP-SGD)** است که در آن گرادیان‌ها بردارهای بسیار بزرگ هستند.*

[16] Tail Bounds

مکانیزم نمایی

مکانیزم‌های قبلی برای خروجی‌های عددی بودند. اگر خروجی یک «عضو» از یک مجموعه باشد (مثلاً «بهترین» رنگ، یا «پرطرفدارترین» کاندیدا)، از مکانیزم نمایی استفاده می‌شود. این مکانیزم بر اساس یک **تابع امتیاز**[۱۷] $q(\mathcal{D}, r)$ کار می‌کند که میزان «خوبی» خروجی r را برای داده‌های $\mathcal{D}$ می‌سنجد. حساسیت این تابع به صورت $\Delta q = \max_r \max_{\mathcal{D}\sim\mathcal{D}'} |q(\mathcal{D}, r) - q(\mathcal{D}', r)|$ تعریف می‌شود.

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

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

این مکانیزم شرط ε-DP را برآورده می‌کند.

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

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

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

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

۲-۱-۶ ترکیب‌پذیری

در کاربردهای واقعی، معمولاً چندین پرس‌وجو روی یک پایگاه‌داده اجرا می‌شود. قضایای ترکیب‌پذیری[۱۸] (ساده[۱۹] و پیشرفته[۲۰]) نشان می‌دهند که چگونه بودجه‌ی محرمانگی با افزایش تعداد پرس‌وجوها انباشته

[17] Score Function
[18] Composition
[19] Basic
[20] Advanced

می‌شود.

قضیه‌ی ۲-۵ (ترکیب‌پذیری ساده) *اگر k مکانیزم $\mathcal{M}_1, \ldots, \mathcal{M}_k$ به ترتیب دارای بودجه‌های $\varepsilon_1, \ldots, \varepsilon_k$ باشند، اجرای متوالی آن‌ها روی یک پایگاه‌داده‌ی واحد، تضمین $(\sum \varepsilon_i)$-DP را فراهم می‌کند [۱۲].*

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

قضیه‌ی ۲-۶ (ترکیب‌پذیری پیشرفته) *برای هر $\delta' > 0$، اجرای k مکانیزم که هرکدام ε-DP هستند، دارای تضمین $(\varepsilon', k\delta + \delta')$-DP است که در آن:*

$$\varepsilon' \approx \varepsilon\sqrt{2k\ln(1/\delta')} + k\varepsilon(e^{\varepsilon} - 1) \tag{۲-۱۵}$$

برای مقادیر کوچک ε، جمله دوم ناچیز است و بودجه کل تقریباً با $\varepsilon\sqrt{k}$ رشد می‌کند. این نتیجه اساس بسیاری از الگوریتم‌های یادگیری ماشین خصوصی (مانند DP-SGD) است [۱].

۲-۱-۷ محرمانگی گروهی

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

قضیه‌ی ۲-۷ (محرمانگی گروهی[۲۱]) *اگر دو پایگاه‌داده $\mathcal{D}_1$ و $\mathcal{D}_2$ در k رکورد متفاوت باشند (فاصله‌ی همسایگی k)، آنگاه هر مکانیزم ε-DP برای آن‌ها تضمین $(k\varepsilon)$-DP را ارائه می‌دهد:*

$$\Pr[\mathcal{M}(\mathcal{D}_1) \in S] \leqslant e^{k\varepsilon} \Pr[\mathcal{M}(\mathcal{D}_2) \in S] \tag{۲-۱۶}$$

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

۲-۱-۸ محدودیت مدل متمرکز

با وجود تمام مزایا، مدل CDP یک نقطه‌ی ضعف اساسی دارد: نیاز به یک متصدی کاملاً مورد اعتماد. در بسیاری از سناریوهای دنیای واقعی (مانند جمع‌آوری داده از گوشی‌های هوشمند)، کاربران به سرور مرکزی

اعتماد ندارند. این عدم اعتماد، ما را به سمت مدل جایگزین، یعنی «محرمانگی تفاضلی موضعی» سوق می‌دهد. در فصل بعد با محرمانگی تفاضلی موضعی و تعاریف و قضایای اساسی آن آشنا خواهیم شد.

۲-۲ fــ واگرایی‌ها

در بخش‌های پیشین، مکانیزم‌های محرمانگی تفاضلی را ابزاری برای ایجاد «شباهت آماری» بین خروجی‌های دو پایگاه‌داده‌ی همسایه معرفی کردیم. برای کمّی‌سازی دقیق این شباهت و اثبات کران‌های پایین در فصل‌های آینده، نیازمند معیاری هستیم که فاصله میان توزیع‌های احتمالی را در یک چارچوب عمومی بسنجد. این ابزار، خانواده‌ی fــ واگرایی‌ها[۲۲] است که نخستین بار توسط سیسر [۷] و علی و سیلوِی [۲] معرفی شد.

۲-۲-۱ تعریف رسمی در فضای اندازه‌پذیر

برای ارائه تعریفی دقیق و مستقل از نوع متغیرهای تصادفی (پیوسته یا گسسته)، از زبان نظریه اندازه استفاده می‌کنیم. فرض کنید $(\Omega, \mathcal{F})$ یک فضای اندازه‌پذیر[۲۳] باشد و P و Q دو «اندازه احتمال»[۲۴] تعریف شده روی این فضا باشند.

پیش از تعریف واگرایی، مفهوم «پیوستگی مطلق» که شرط وجود چگالی نسبت است را یادآوری می‌کنیم.

تعریف ۲-۹ (پیوستگی مطلق[۲۵]) *می‌گوییم اندازه P نسبت به Q مطلقاً پیوسته است و می‌نویسیم $P \ll Q$، اگر برای هر مجموعه اندازه‌پذیر $A \in \mathcal{F}$:*

$$Q(A) = \cdot \implies P(A) = \cdot \tag{۲-۱۷}$$

این شرط تضمین می‌کند که هر رویدادی که تحت توزیع Q غیرممکن باشد، تحت P نیز غیرممکن است.

اگر $P \ll Q$ باشد، بنابر قضیه رادون‌ــ نیکودیم[۲۶]، تابعی اندازه‌پذیر و غیرمنفی روی Ω وجود دارد که «مشتق رادون‌ــ نیکودیم» P نسبت به Q نامیده می‌شود و با $\frac{dP}{dQ}$ نمایش داده می‌شود. این مشتق نقش همان نسبت درست‌نمایی[۲۷] را در حالت کلی ایفا می‌کند.

[22] f-divergences
[23] Measurable Space
[24] Probability Measures
[26] Radon-Nikodym Theorem
[27] Likelihood Ratio

تعریف ۲-۱۰ (fـ واگرایی) *فرض کنید P و Q دو اندازه احتمال روی $(\Omega, \mathcal{F})$ باشند به طوری که $P \ll Q$. برای هر تابع محدب[۲۸] $f : (0, \infty) \to \mathbb{R}$ با این شرط که $f(1) = 0$، fـ واگراییِ P از Q به صورت امید ریاضی تابع f بر روی مشتق رادون‌ـ نیکودیم (تحت اندازه Q) تعریف می‌شود:*

$$D_f(P\|Q) \triangleq \int_{\Omega} f\left(\frac{dP}{dQ}(\omega)\right) dQ(\omega) \tag{۲-۱۸}$$

یا به بیانی دیگر با استفاده از نماد امید ریاضی:

$$D_f(P\|Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right] \tag{۲-۱۹}$$

تفسیر اجزاء:

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

- **غیرمنفی بودن:** با استفاده از نامساوی ینسن[۲۹] و شرط محدب بودن f، می‌توان نشان داد که واگرایی همواره نامنفی است:

$$D_f(P\|Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right] \geqslant f\left(\mathbb{E}_Q\left[\frac{dP}{dQ}\right]\right) = f(1) = 0 \tag{۲-۲۰}$$

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

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

$$D_f(P\|Q) = \int_{\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) dx \quad \text{یا} \quad \sum_{x\in\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) \tag{۲-۲۱}$$

۲-۲-۲ نمونه‌های مهم و توابع مولد

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

---
[28] Convex
[29] Jensen's Inequality

- **فاصله تغییرات کل**[30]: این فاصله، شهودی‌ترین متریک برای سنجش تمایزپذیری دو توزیع است و بیانگر بیش‌ترین تفاوت احتمالی است که دو توزیع می‌توانند روی یک پیشامد داشته باشند. تابع مولد آن $f(t) = \frac{۱}{۲}|t - ۱|$ است.

$$\mathrm{TV}(P, Q) = \frac{۱}{۲} \int_{\mathcal{X}} |p(x) - q(x)| dx \tag{۲-۲۲}$$

$$= \sup_{A \in \mathcal{F}} |P(A) - Q(A)| \tag{۲-۲۳}$$

- **واگرایی کولبک‌ــ‌لایبلر**[31]: معروف‌ترین واگرایی در نظریه اطلاعات که آنتروپی نسبی[32] نیز نامیده می‌شود. این معیار نامتقارن است و نقش اساسی در فشرده‌سازی داده‌ها و استنتاج بیزی دارد. تابع مولد آن $f(t) = t \ln t$ است.

$$\mathrm{KL}(P \| Q) = \int_{\mathcal{X}} p(x) \ln \frac{p(x)}{q(x)} dx \tag{۲-۲۴}$$

- **اطلاعات متقابل** اگر X و V دو متغیر تصادفی باشند، اطلاعات متقابل[33] بین آن‌ها به صورت امید ریاضی واگرایی KL بین توزیع شرطی و توزیع حاشیه‌ای تعریف می‌شود:

$$I(X; V) = D_{KL}(P_{X,V} || P_X \otimes P_V) = \mathbb{E}_V \left[D_{KL}(P_{X|V} || P_X) \right] \tag{۲-۲۵}$$

این معیار نقش کلیدی در نامساوی فانو (که در بخش بعد می‌بینیم) ایفا می‌کند.

- **واگرایی کای‌ــ‌دو**[34]: تابع مولد آن $f(t) = (t - ۱)^۲$ است. این واگرایی اغلب برای تقریب‌زنی سایر فواصل (مانند KL) در همسایگی‌های کوچک استفاده می‌شود و محاسبه آن ساده‌تر است.

$$\chi^۲(P \| Q) = \int_{\mathcal{X}} \frac{(p(x) - q(x))^۲}{q(x)} dx \tag{۲-۲۶}$$

- **فاصله هلینگر‌ــ‌دو**[35]: تابع مولد آن $f(t) = (\sqrt{t} - ۱)^۲$ است. این فاصله به دلیل خواص ریاضی خوش‌رفتار (مانند متریک بودن و کران‌دار بودن)، در نظریه برآورد[36] و استخراج کران‌های مینیماکس (مانند روش Le Cam) کاربرد فراوان دارد.

$$H^۲(P, Q) = \int_{\mathcal{X}} (\sqrt{p(x)} - \sqrt{q(x)})^۲ dx \tag{۲-۲۷}$$

---

[30] Total Variation (TV)
[31] Kullback-Leibler (KL)
[32] Relative Entropy
[33] Mutual Information
[34] Chi-Squared (χ^2)
[35] Squared Hellinger
[36] Estimation Theory

- **واگرایی E_γ (یا واگرایی هاکی‌ـاستیک[۳۷]):** این واگرایی ابزاری کلیدی در تحلیل‌های مدرن حریم خصوصی و آزمون‌های فرضیه است. تابع مولد آن برای پارامتر $\gamma \geqslant ۱$ به صورت $f_\gamma(t) = [t - \gamma]_+ = \max\{۰, t - \gamma\}$ است.

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

$$E_\gamma(P\|Q) = \int_{\mathcal{X}} \max\{۰, p(x) - \gamma q(x)\}dx \tag{۲-۲۸آ}$$

$$= \sup_{A \subseteq \mathcal{X}} (P(A) - \gamma Q(A)) \tag{۲-۲۸ب}$$

$$= P(Z > \gamma) - \gamma Q(Z > \gamma) \quad (\text{که } Z = \frac{p(x)}{q(x)}) \tag{۲-۲۸ج}$$

رابطه (۲-۲۸ب) 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 «هاکی‌ـاستیک» که نخستین بار توسط ساسون و وِردو [۱۹] پیشنهاد شد، برخاسته از شکل هندسی نمودار تابع مولد $f(t)$ است (تخت بودن تا γ شبیه تیغه، و صعود خطی پس از آن شبیه دسته چوب هاکی). نمادگذاری E_γ و تدوین نقش بنیادی آن، پیش‌تر توسط پولیانسکی و همکاران [۱۸] صورت گرفته بود.

کاربرد در حریم خصوصی: شرط محرمانگی تفاضلی تقریبی ((ε, δ)-DP) دقیقاً معادل است با اینکه برای هر دو دیتابیس همسایه، واگرایی هاکی‌ـاستیک خروجی‌ها از مقدار δ تجاوز نکند: $E_{e^\varepsilon}(P\|Q) \leqslant \delta$.

- **واگرایی رنی[۳۸]:** برای پارامتر $\alpha \in (۱, \infty)$، این واگرایی به صورت زیر تعریف می‌شود:

$$D_\alpha(P\|Q) = \frac{۱}{\alpha - ۱} \ln \int_{\mathcal{X}} p(x)^\alpha q(x)^{۱-\alpha} dx \tag{۲-۲۹}$$

این واگرایی پلی میان واگرایی KL (در حد $\alpha \to ۱$) و واگرایی ماکزیمم (در حد $\alpha \to \infty$) است. اگرچه فرم لگاریتمی دارد، اما تبدیلی یک‌نوا از «واگرایی سالیس»[۳۹] است که خود یک fـواگرایی می‌باشد.

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

$$D_\infty(P\|Q) = \lim_{\alpha \to \infty} D_\alpha(P\|Q) = \sup_{x \in \mathcal{X}} \ln \frac{p(x)}{q(x)} \tag{۲-۳۰}$$

[37] Hockey-Stick Divergence
[38] Rényi Divergence
[39] Tsallis Divergence

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

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

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

نامساوی پردازش داده (DPI)

مهم‌ترین ویژگی fـ واگرایی‌ها در نظریه اطلاعات، خاصیت یک‌نواختی[۴۰] آن‌ها تحت پردازش است. این ویژگی بیان می‌کند که هیچ عملیات پردازشی روی داده‌ها (اعم از قطعی یا تصادفی) نمی‌تواند تمایزپذیری بین دو توزیع را افزایش دهد.

قضیه‌ی ۲-۸ (نامساوی پردازش داده[۴۱]) *فرض کنید P و Q دو توزیع احتمال روی فضای $\mathcal{X}$ باشند و $\kappa : \mathcal{X} \to \mathcal{Y}$ یک هسته‌ی احتمالاتی (کانال مارکوف)[۴۲] باشد که داده‌ها را از فضای $\mathcal{X}$ به $\mathcal{Y}$ نگاشت می‌کند. اگر $P\kappa$ و $Q\kappa$ توزیع‌های خروجی پس از اعمال کرنل باشند، آنگاه برای هر fـ واگرایی داریم:*

$$D_f(P\kappa\|Q\kappa) \leqslant D_f(P\|Q) \tag{۲-۳۱}$$

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[۴۳] 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 fـ واگرایی نسبت به جفت توزیع‌های ورودی خود، محدب است.

[40] Monotonicity
[42] Markov Kernel / Probability Kernel
[43] Post-processing

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

$$D_f(\lambda P_1 + (1-\lambda)P_2 \| \lambda Q_1 + (1-\lambda)Q_2) \leqslant \lambda D_f(P_1\|Q_1) + (1-\lambda)D_f(P_2\|Q_2) \quad (۲-۳۲)$$

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

روابط بین واگرایی‌ها

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

- **نامساوی پینسکر[۴۵]:** این نامساوی مشهور، ارتباط هندسه (فاصله تغییرات کل) و اطلاعات (واگرایی KL) را برقرار می‌کند و نشان می‌دهد که همگرایی در آنتروپی نسبی، همگرایی در L_1 را تضمین می‌کند:

$$\mathrm{TV}(P,Q) \leqslant \sqrt{\frac{1}{2}\mathrm{KL}(P\|Q)} \quad (۲-۳۳)$$

- **رابطه هاکی-استیک و TV:** می‌توان به سادگی دید واگرایی هاکی-استیک تعمیمی از فاصله تغییرات کل است. به طور مشخص، در نقطه $\gamma = 1$ این دو معیار بر هم منطبق می‌شوند:

$$E_1(P\|Q) = \mathrm{TV}(P,Q) \quad (۲-۳۴)$$

این تساوی پل ارتباطی مهمی بین تعاریف (ε,δ)-DP و تحلیل‌های مبتنی بر فاصله تغییرات کل فراهم می‌کند.

۲-۳ مبانی آماری و نظریه اطلاعات

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

[45] Pinsker's Inequality

۲-۳-۱ ریسک مینیماکس

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

نرخ مینیماکس[46]، کمترین خطای ممکنی است که یک تخمین‌گر در بدترین سناریو (بدترین توزیع P در کلاس $\mathcal{P}$) متحمل می‌شود.

تعریف ۲-۱۱ (نرخ مینیماکس) *برای یک کلاس از توزیع‌ها $\mathcal{P}$ و پارامتر θ، نرخ مینیماکس $\mathfrak{M}_n$ به صورت زیر تعریف می‌شود:*

$$\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho) = \inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\Phi(\rho(\hat{\theta}(Z^n), \theta(P)))] \tag{۲-۳۵}$$

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

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

۲-۴ آزمون فرض آماری و روش تقلیل

برای اثبات حدود پایین نرخ‌های مینیماکس، روش استاندارد این است که مسئله‌ی تخمین پارامتر را به یک مسئله‌ی آزمون فرض[47] تقلیل دهیم. ایده اصلی این است: اگر نتوانیم بین چند مقدار گسسته از پارامتر با دقت بالا تمایز قائل شویم، قطعاً نمی‌توانیم پارامتر را در فضای پیوسته با خطای کم تخمین بزنیم.

۲-۴-۱ آزمون فرض دودویی

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

- فرض صفر (H_0): $Z \sim P_0$

[46] Minimax Rate
[47] Hypothesis Testing

- فرض مقابل (H_1): $Z \sim P_1$

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

$$P_{err}(\psi) = \Pr_{H_0}(\psi(Z) = 1) + \Pr_{H_1}(\psi(Z) = 0) \qquad (۲-۳۶)$$

لم نیمن–پیرسون[48] نشان می‌دهد که کمترین خطای ممکن برای هر آزمون دودویی، مستقیماً با فاصله‌ی واریانس کل (d_{TV}) بین دو توزیع ارتباط دارد:

$$\inf_{\psi} P_{err}(\psi) = 1 - \|P_0 - P_1\|_{TV} \qquad (۲-۳۷)$$

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

۲-۴-۲ تقلیل تخمین به آزمون (روش بسته‌بندی)

برای استفاده از ابزارهای آزمون فرض در مسئله‌ی تخمین نرخ مینیماکس (معادله ۲-۳۵)، از تکنیک گسسته‌سازی فضای پارامتر Θ استفاده می‌کنیم. این روش شامل مراحل زیر است:

۱. **ساخت مجموعه‌ی بسته‌بندی**[49]**:** مجموعه‌ای متناهی از پارامترها $\mathcal{V} = \{\theta_1, \dots, \theta_M\} \subset \Theta$ را انتخاب می‌کنیم به طوری که از یکدیگر فاصله‌ی معناداری داشته باشند. به طور دقیق‌تر، اگر ρ متریک خطا باشد، برای هر $i \neq j$ باید داشته باشیم $\rho(\theta_i, \theta_j) \geqslant 2\delta$.

۲. **تعریف مسئله‌ی آزمون:** فرض می‌کنیم طبیعت[50] یک اندیس V را به صورت تصادفی و یکنواخت از مجموعه $\{1, \dots, M\}$ انتخاب می‌کند و داده‌ها بر اساس توزیع P_{θ_V} تولید می‌شوند. هدف، یافتن V بر اساس داده‌های مشاهده شده است.

۳. **ارتباط خطاها:** اگر یک تخمین‌گر $\hat{\theta}$ وجود داشته باشد که خطای تخمین آن با احتمال بالا کمتر از δ باشد، می‌توانیم از آن برای حل مسئله‌ی آزمون فرض استفاده کنیم (با انتخاب نزدیک‌ترین θ_i به $\hat{\theta}$).

[48] Neyman-Pearson Lemma
[49] Packing Set
[50] Nature

بنابراین، کران پایین روی خطای آزمون فرض، یک کران پایین برای خطای تخمین ایجاد می‌کند:

$$\mathfrak{M}_n(\theta(\mathcal{P})) \geqslant \Phi(\delta) \cdot \inf_{\psi} \Pr(\psi(Z^n) \neq V) \qquad (۳۸-۲)$$

۲-۴-۳ نامساوی‌های کران پایین

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

قضیه‌ی ۲-۱۰ (نامساوی لو کم[۵۱]) *این روش برای آزمون بین دو توزیع $P_۱$ و $P_۲$ استفاده می‌شود. کمینه احتمال خطا با استفاده از فاصله‌ی واریانس کل (رابطه ؟؟) کران‌دار می‌شود:*

$$\inf_{\psi} \Pr(\psi(Z^n) \neq V) \geqslant \frac{۱}{۲} (۱ - \|P_۱^n - P_۲^n\|_{TV}) \qquad (۳۹-۲)$$

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

قضیه‌ی ۲-۱۱ (نامساوی فانو[۵۲]) *زمانی که پارامتر مورد نظر متعلق به مجموعه‌ای بزرگتر $\mathcal{V}$ باشد (تعداد فرضیه‌ها $|\mathcal{V}| > ۲$)، نامساوی فانو کران پایین قوی‌تری ارائه می‌دهد که مبتنی بر اطلاعات متقابل است:*

$$\inf_{\psi} \Pr(\psi(Z^n) \neq V) \geqslant ۱ - \frac{I(Z^n; V) + \log ۲}{\log |\mathcal{V}|} \qquad (۴۰-۲)$$

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

لم ۲-۱۲ (لم اسود[۵۳]) *این لم مسئله تخمین را به چندین آزمون فرض دودویی مستقل روی مختصات یک ابرمکعب $\{-۱, ۱\}^d$ تبدیل می‌کند. نسخه دقیق‌تر آن که در [۹] استفاده شده است، کران پایین را بر اساس فاصله‌ی واریانس کل توزیع‌های مخلوط حاشیه‌ای بیان می‌کند:*

$$\mathfrak{M}_n(\theta(\mathcal{P})) \geqslant \delta \sum_{j=۱}^{d} \left[۱ - \|M_{+j}^n - M_{-j}^n\|_{TV}\right] \qquad (۴۱-۲)$$

که در آن M_{+j}^n و M_{-j}^n توزیع‌های حاشیه‌ای مخلوط روی مقادیر ۱+ و ۱− در بُعد jـ ام هستند.

فصل ۳

محرمانگی تفاضلی موضعی

۳-۱ مقدمه

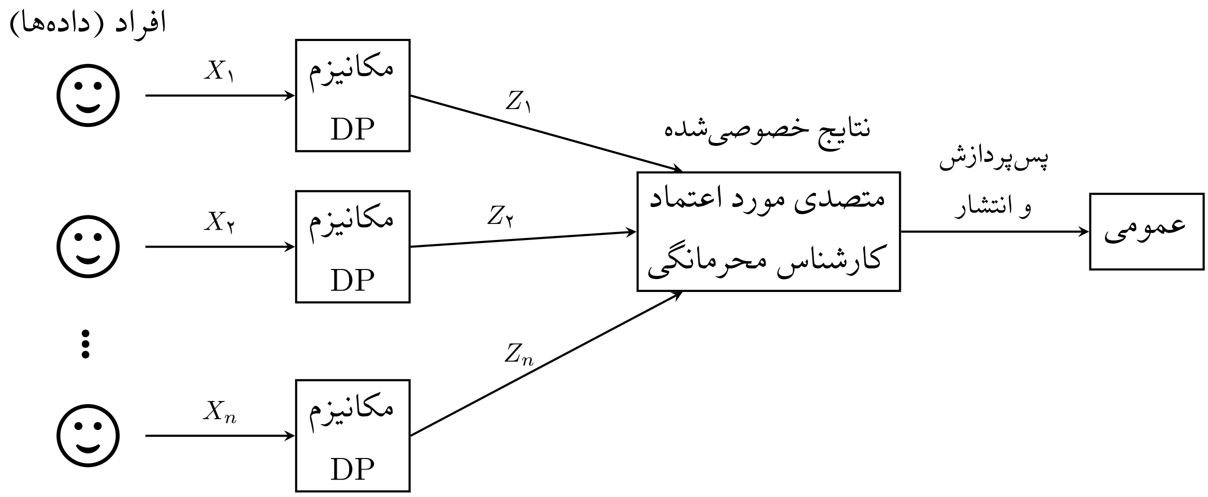
در فصل پیشین (۲)، مفاهیم بنیادی نظریه اندازه و مدل محرمانگی تفاضلی متمرکز (CDP) را بررسی کردیم. همان‌طور که در بخش ۲-۱ مشاهده شد، مدل متمرکز (CDP) بر فرض وجود یک متصدی مورد اعتماد استوار است که به داده‌های خام تمامی کاربران دسترسی دارد ($M : \mathcal{X}^n \to \mathcal{R}$). اگرچه این مدل دقت آماری بالایی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

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

۳-۲ تعاریف رسمی و مدل‌های محاسباتی

در مدل موضعی، مجموعه‌ای از n کاربر وجود دارند که هر کدام داده‌ای خصوصی $X_i \in \mathcal{X}$ در اختیار دارند. برخلاف مدل متمرکز که شرط محرمانگی روی «پایگاه‌داده‌های همسایه» تعریف می‌شد، در این‌جا

[1] Local Differential Privacy



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

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

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

۳-۲-۱ تعریف ریاضی LDP

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

تعریف ۳-۲ (α-LDP) *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-جبرِ برد) داشته باشیم:*

$$\sup_{x,x' \in \mathcal{X}} \sup_{\mathcal{S} \subseteq \mathcal{Z}} \frac{\Pr[\mathcal{M}(x) \in \mathcal{S}]}{\Pr[\mathcal{M}(x') \in \mathcal{S}]} \leqslant e^{\alpha} \quad (۳-۱)$$

نکته: در متون آماری این حوزه (مانند [۹])، معمولاً از پارامتر α برای بودجه‌ی محرمانگی موضعی استفاده می‌شود تا تمایز آن با پارامتر ε در مدل متمرکز مشخص گردد. ما نیز در این فصل و فصول بعدی از این نمادگذاری پیروی می‌کنیم.

این تعریف را می‌توان با استفاده از مفهوم «واگرایی ماکزیمم» (D_∞) که پیش‌تر معرفی شد، به صورت فشرده‌تری بیان کرد. شرط (۳-۱) دقیقاً معادل است با:

$$\sup_{x,x' \in \mathcal{X}} D_\infty(Q(\cdot|x) \parallel Q(\cdot|x')) \leqslant \alpha \tag{۳-۲}$$

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

۳-۲-۲ محرمانگی تقریبی

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

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

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

بدیهی است که اگر $\delta = ۰$ باشد، این تعریف به حالت α-LDP خالص باز می‌گردد [۲۲].

۳-۳ پروتکل‌های تعاملی و خواص ترکیب

برای تحلیل دقیق حدود مینیماکس و درک محدودیت‌های بنیادین LDP، نیازمند مدل‌سازی دقیق نحوه‌ی تعامل کاربران با سرور (یا جمع‌آورنده داده) هستیم. دوچی و همکاران [۹] پروتکل‌های موضعی را بر اساس ساختار وابستگی آماری خروجی‌ها به دو دسته‌ی کلی تقسیم می‌کنند: غیرتعاملی و تعاملی.

۳-۳-۱ پروتکل‌های غیرتعاملی

در پروتکل‌های غیرتعاملی[۴]، تمام کاربران $i = ۱, \dots, n$ مکانیزم‌های خود را به صورت هم‌زمان و مستقل از یک‌دیگر اجرا می‌کنند. اگر Z_i خروجی کاربر iـام باشد، توزیع آن تنها به داده‌ی خصوصی X_i وابسته است

[3] Likelihood Ratio
[4] Non-interactive

و هیچ وابستگی‌ای به خروجی سایر کاربران ندارد. به بیان ریاضی، توزیع مشترک خروجی‌ها به صورت حاصل‌ضرب توزیع‌های حاشیه‌ای فاکتور می‌شود:

$$\Pr(Z_۱, \dots, Z_n | X_۱, \dots, X_n) = \prod_{i=۱}^{n} Q_i(Z_i | X_i) \tag{۳-۴}$$

بسیاری از پیاده‌سازی‌های صنعتی فعلی، از جمله RAPPOR گوگل [۱۳]، در این دسته قرار می‌گیرند.

۳-۳-۲ پروتکل‌های تعاملی (ترتیبی)

در پروتکل‌های تعاملی[۵]، کاربران به نوبت داده‌های خود را ارسال می‌کنند و مکانیزم کاربر iـام می‌تواند به خروجی‌های مشاهده‌شده از کاربران پیشین $(Z_۱, \dots, Z_{i-۱})$ وابسته باشد. این مدل، آزادی عمل بیشتری را برای طراحی الگوریتم‌های تطبیقی فراهم می‌کند.

از دیدگاه آنالیز ریاضی، این فرآیند با استفاده از «کرنل‌های احتمالاتی»[۶] مدل‌سازی می‌شود. فرض کنید $Z_{۱:i-۱} = (Z_۱, \dots, Z_{i-۱})$ بردار خروجی‌های پیشین باشد که σـ فیلد $\mathcal{F}_{i-۱}$ را تولید می‌کند. مکانیزم کاربر iـام، یک کرنل احتمالاتی Q_i است که خروجی $Z_i \in \mathcal{Z}$ را مشروط بر داده‌ی خصوصی X_i و تاریخچه‌ی عمومی $Z_{۱:i-۱}$ تولید می‌کند:

$$Z_i \sim Q_i(dz_i | x_i, z_{۱:i-۱}) \tag{۳-۵}$$

شرط اساسی محرمانگی در اینجا این است که با شرطی‌سازی روی X_i و $Z_{۱:i-۱}$، متغیر Z_i باید از سایر داده‌های خصوصی $X_{j \neq i}$ مستقل باشد (شرط مارکوفی). این ساختار به پروتکل اجازه می‌دهد تا پارامترهای پرس‌وجو را به صورت پویا بر اساس اطلاعات کسب‌شده از کاربران قبلی تنظیم کند.

۳-۳-۳ قضیه ترکیب ترتیبی

یکی از ویژگی‌های بنیادین LDP، پایداری آن در برابر ترکیب است. اگر یک پروتکل شامل چندین مرحله‌ی تعاملی باشد، بودجه‌های محرمانگی با یکدیگر جمع می‌شوند. قضیه‌ی زیر، کران بالای محرمانگی را برای یک پروتکل ترتیبی بیان می‌کند [۹].

قضیه‌ی ۳-۱ (ترکیب ترتیبی[۷]) *فرض کنید در یک پروتکل تعاملی، برای هر کاربر $i \in \{۱, \dots, n\}$ و به ازای هر تاریخچه‌ی ممکن $z_{۱:i-۱} \in \mathcal{Z}^{i-۱}$، مکانیزم $Q_i(\cdot | \cdot, z_{۱:i-۱})$ دارای خاصیت α_i-LDP نسبت به*

[5] Interactive / Sequential
[6] Probability Kernels

ورودی x_i باشد. آنگاه توزیع مشترک کل خروجی‌ها $(Z_۱, \ldots, Z_n)$، دارای محرمانگی تفاضلی موضعی با بودجه‌ی مجموع است:

$$\alpha_{total} = \sum_{i=۱}^{n} \alpha_i \tag{۳-۶}$$

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 حاصل‌ضرب نسبت‌های شرطی تجزیه می‌شود:

$$\frac{P(z^n)}{P'(z^n)} = \prod_{i=۱}^{n} \frac{Q_i(z_i|x_i, z_{۱:i-۱})}{Q_i(z_i|x'_i, z_{۱:i-۱})}$$

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

۳-۴ مکانیزم‌های پایه در LDP

در این بخش، مکانیزم‌های بنیادین LDP را با رویکردی آماری تحلیل می‌کنیم. هدف اصلی در طراحی این مکانیزم‌ها، یافتن نگاشتی تصادفی است که علاوه بر ارضای شرط محرمانگی، «خطای تخمین» (که معمولاً با واریانس سنجیده می‌شود) را کمینه کند. فرض بنیادی در تمام این مکانیزم‌ها این است که برای بازیابی اطلاعات آماری (مانند هیستوگرام)، از یک «تخمین‌گر ناااریب»[۸] معکوس استفاده می‌شود.

۳-۴-۱ پاسخ تصادفی دودویی (RR)

پایه‌ای‌ترین مکانیزم α-LDP، پاسخ تصادفی[۹] (RR) برای دامنه‌ی دودویی $\mathcal{X} = \{۰, ۱\}$ 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 $۱ - p$ قرینه می‌کند:

$$\Pr[y = z|x] = \begin{cases} p & \text{if } z = x \\ ۱ - p & \text{if } z \neq x \end{cases} \tag{۳-۷}$$

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

[8] Unbiased Estimator
[9] Randomized Response

زمانی رخ می‌دهد که صورت کسر بیشترین احتمال (p) و مخرج کسر کمترین احتمال ($۱ - p$) باشد:

$$\sup_{y \in \{۰,۱\}} \frac{\Pr[y|x]}{\Pr[y|x']} = \frac{p}{۱ - p} \leqslant e^{\alpha} \tag{۳-۸}$$

با حل این نامساوی برای p (و با فرض $p > ۱/۲$ برای بی‌معنی نشدن نتیجه)، مقدار بهینه احتمال حفظ پاسخ برای بودجه‌ی α به دست می‌آید:

$$p = \frac{e^{\alpha}}{e^{\alpha} + ۱} \tag{۳-۹}$$

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

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

$$\hat{x} = \frac{y - (۱ - p)}{۲p - ۱} \tag{۳-۱۰}$$

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

$$\mathrm{Var}[\hat{x}] = \frac{e^{\alpha}}{(e^{\alpha} - ۱)^{۲}} \tag{۳-۱۱}$$

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

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

با جایگذاری در معادله تخمین‌گر:

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

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

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

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

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

□

۳-۴-۲ پاسخ تصادفی تعمیم‌یافته (GRR)

پاسخ تصادفی تعمیم‌یافته[10]، برای دامنه‌های گسسته با $k > ۲$ عنصر ($\mathcal{X} = \{۱, \dots, k\}$)، به عنوان تعمیم مستقیم RR معرفی می‌شود. این مکانیزم را می‌توان با یک «ماتریس گذار»[11] تصادفی $\mathbf{Q} \in [۰, ۱]^{k\times k}$ توصیف کرد.

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

$$\mathbf{Q} = \begin{pmatrix} p & q & \dots & q \\ q & p & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & p \end{pmatrix}_{k\times k} \tag{۳-۱۲}$$

که در آن p احتمال گزارش صادقانه و q احتمال گزارش هر یک از $k - ۱$ گزینه‌ی دیگر است.

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

$$p = \frac{e^\alpha}{e^\alpha + k - ۱}, \quad q = \frac{۱}{e^\alpha + k - ۱} \tag{۳-۱۳}$$

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

$$\mathrm{Var}_{GRR} = \frac{p(۱ - p)}{(p - q)^۲} = \frac{(e^\alpha)(k - ۱) + (k - ۱)^۲}{(e^\alpha - ۱)^۲} \approx \mathcal{O}(k) \tag{۳-۱۴}$$

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

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

برای غلبه بر مشکل کاهش دقت GRR در دامنه‌های بزرگ، خانواده‌ای از مکانیزم‌ها تحت عنوان «کدگذاری یگانی»[12] توسعه یافته‌اند. این رویکرد اساس پروتکل مشهور RAPPOR گوگل را تشکیل می‌دهد [۱۳، ۲۲].

---
[10] Generalized Randomized Response
[11] Transition Matrix
[12] Unary Encoding (UE)

تعریف ۳-۵ (کدگذاری یگانی) *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 جاها ۰ است (One-hot encoding).*

۲. ***اختلال[13] مستقل:*** *هر بیت این بردار به صورت مستقل با استفاده از یک مکانیزم باینری معکوس می‌شود. اگر v_i بیت iـ ام بردار کدگذاری شده باشد، خروجی z_i به صورت زیر تولید می‌شود:*

$$\Pr[z_i = 1] = \begin{cases} p & \textit{if } v_i = 1 \\ q & \textit{if } v_i = 0 \end{cases} \tag{۳-۱۵}$$

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

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

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:

$$\frac{\Pr[z|x]}{\Pr[z|x']} = \frac{\Pr[z_x|v_x = 1]}{\Pr[z_x|v'_x = 0]} \cdot \frac{\Pr[z_{x'}|v_{x'} = 0]}{\Pr[z_{x'}|v'_{x'} = 1]} \tag{۳-۱۶}$$

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:

$$\frac{\Pr[z|x]}{\Pr[z|x']} \leqslant \frac{p}{q} \cdot \frac{1 - q}{1 - p} = \frac{p(1 - q)}{q(1 - p)} \tag{۳-۱۷}$$

بنابراین شرط α-LDP معادل است با:

$$\alpha = \ln\left(\frac{p(1 - q)}{q(1 - p)}\right) \tag{۳-۱۸}$$

[13] Perturbation

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

$$\mathrm{Var}_{UE} = \frac{q(۱-q)}{(p-q)^۲} \tag{۳-۱۹}$$

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

دو استراتژی اصلی برای تنظیم p و q بر اساس رابطه (۳-۱۸) وجود دارد:

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

$$\mathrm{Var}_{SUE} = \frac{e^{\alpha/۲}}{(e^{\alpha/۲}-۱)^۲} \tag{۳-۲۰}$$

۲. **کدگذاری یگانی بهینه (OUE):** وانگ و همکاران [۲۲] نشان دادند که برای کمینه‌سازی واریانس در دامنه‌های بزرگ، باید اطلاعات بیت‌های ۱ (سیگنال) حفظ شود ($p=۱/۲$) و نویز روی بیت‌های ۰ (که اکثر بردار را تشکیل می‌دهند) کنترل شود ($q=\frac{۱}{e^{\alpha}+۱}$). واریانس حاصل برابر است با:

$$\mathrm{Var}_{OUE} = \frac{۴e^{\alpha}}{(e^{\alpha}-۱)^۲} \tag{۳-۲۱}$$

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 ۳-۲ (ناکارآمدی GRR در دامنه‌های بزرگ) *فرض کنید می‌خواهیم کلمات پرکاربرد را از یک لغت‌نامه با $k=۱۰۰,۰۰۰$ کلمه استخراج کنیم.*

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

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

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

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

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

این استقلال از k باعث می‌شود که خانواده UE گزینه‌ی برتر برای دامنه‌های بزرگ باشند. با این حال، OUE هزینه مخابراتی بالایی دارد (ارسال بردار با طول k). برای رفع این مشکل، نسخه بهبودیافته‌ای به نام «درهم‌سازی موضعی بهینه[۱۴]» توسط وانگ و همکاران [۲۳] 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[۱۵]، فرض می‌کنیم داده‌ها با یک تبدیل خطی به بازه‌ی $[-۱, ۱]$ نگاشت شده‌اند. در مدل موضعی، شرط محرمانگی باید برای هر جفت ورودی $x, x' \in \mathcal{X}$ برقرار باشد. بنابراین «حساسیت سراسری» (Δ) برابر با قطر دامنه است. در این صورت Δ، بیشترین مقدار ممکن را خواهد داشت:

$$\Delta = \sup_{x,x' \in [-۱,۱]} |x - x'| = |۱ - (-۱)| = ۲ \qquad (۳\text{-}۲۲)$$

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

$$z = x + \eta, \quad \eta \sim \mathrm{Lap}\left(\frac{\Delta}{\alpha}\right) = \mathrm{Lap}\left(\frac{۲}{\alpha}\right) \qquad (۳\text{-}۲۳)$$

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

$$f(z|x) = \frac{\alpha}{۴} \exp\left(-\frac{\alpha|z-x|}{۲}\right) \qquad (۳\text{-}۲۴)$$

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

$$\frac{f(z|x)}{f(z|x')} = \frac{\exp(-\frac{\alpha}{۲}|z-x|)}{\exp(-\frac{\alpha}{۲}|z-x'|)} \qquad (۳\text{-}۲۵)$$

$$= \exp\left(\frac{\alpha}{۲}(|z-x'| - |z-x|)\right) \qquad (۳\text{-}۲۶)$$

---
[14] Optimized Local Hashing (OLH)
[15] Without Loss of Generality

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

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

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

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

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

چالش ابعاد بالا (نفرین ابعاد)

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

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

$$\Delta_{\ell_۱} = \sup_{x,x' \in B_{\ell_۲}} \|x-x'\|_۱ \leqslant \sqrt{d} \cdot \sup \|x-x'\|_۲ = ۲\sqrt{d} \qquad (۳-۲۷)$$

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

- واریانس خطا در هر بُعد: $2 \times (\frac{2\sqrt{d}}{\alpha})^2 = \frac{8d}{\alpha^2}$
- خطای میانگین مربعات کل[16] برای d بُعد: $d \times \frac{8d}{\alpha^2} = \frac{8d^2}{\alpha^2}$

این نرخ رشد $\mathcal{O}(d^2)$ برای خطا، بسیار ناکارآمد است. دوچی و همکاران [9] ثابت کرده‌اند که حد پایین نظری مینیماکس برای این مسأله $\mathcal{O}(d)$ است. به همین دلیل، در ابعاد بالا از مکانیزم‌های پیشرفته‌تری مانند «توزیع برنولی چندبعدی» یا «نمونه‌برداری هایپرکیوب» استفاده می‌شود که با هندسه فضا سازگارترند.

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

$$p = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (۳-۲۸)$$

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

$$\mathrm{MSE}(\hat{p}) = \mathbb{E}\left[(\hat{p}-p)^2\right] = \mathrm{Var}[\hat{p}] + (\mathrm{Bias}[\hat{p}])^2 \qquad (۳-۲۹)$$

[16] Mean Squared Error (MSE)

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

۳-۵-۲ تحلیل در مدل متمرکز (CDP)

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

$$\hat{p}_{CDP} = \frac{۱}{n}\left(\sum_{i=۱}^{n} X_i + \eta\right), \quad \eta \sim \text{Lap}(۱/\varepsilon) \qquad (۳\text{-}۳۰)$$

خطای این تخمین‌گر برابر است با:

$$\text{MSE}_{CDP} = \text{Var}\left[\frac{\eta}{n}\right] = \frac{۱}{n^۲}\text{Var}[\eta] = \frac{۱}{n^۲}\cdot\frac{۲}{\varepsilon^۲} = \mathcal{O}\left(\frac{۱}{n^۲\varepsilon^۲}\right) \qquad (۳\text{-}۳۱)$$

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

۳-۵-۳ تحلیل در مدل موضعی (LDP)

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

$$\text{Var}[\hat{X}_i] \approx \frac{۱}{\alpha^۲} \qquad (۳\text{-}۳۲)$$

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

$$\text{MSE}_{LDP} = \text{Var}\left[\frac{۱}{n}\sum_{i=۱}^{n}\hat{X}_i\right] = \frac{۱}{n^۲}\sum_{i=۱}^{n}\text{Var}[\hat{X}_i] = \frac{۱}{n^۲}\cdot n\cdot\mathcal{O}\left(\frac{۱}{\alpha^۲}\right) = \mathcal{O}\left(\frac{۱}{n\alpha^۲}\right) \qquad (۳\text{-}۳۳)$$

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

۳-۵-۴ نتیجه‌گیری: شکاف کارایی

با مقایسه روابط (۳-۳۱) و (۳-۳۳)، تفاوت بنیادین دو مدل آشکار می‌شود (با فرض ثابت بودن بودجه‌های محرمانگی $\varepsilon \approx \alpha$):

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

این تفاوت در نرخ همگرایی پیامد بسیار مهمی در حجم نمونه[17] مورد نیاز دارد. برای رسیدن به یک دقت ثابت مشخص (مثلاً خطای τ)، تعداد کاربران مورد نیاز در هر مدل عبارت است از:

$$n_{CDP} \propto \frac{۱}{\tau}, \qquad n_{LDP} \propto \frac{۱}{\tau^۲} \tag{۳-۳۴}$$

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

$$n_{LDP} \approx \mathcal{O}(n_{CDP}^۲) \tag{۳-۳۵}$$

این رابطه که در ادبیات موضوع به «هزینه عدم اعتماد» شهرت دارد، نشان می‌دهد که مدل موضعی برای رسیدن به دقتی مشابه مدل متمرکز، نیازمند داده‌های بسیار بیشتری است. برای مثال، اگر در مدل متمرکز با ۱,۰۰۰ کاربر به دقت مطلوبی برسیم، در مدل موضعی برای همان دقت به ۱,۰۰۰,۰۰۰ کاربر نیاز خواهیم داشت [۹، ۱۵].

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؟»

[17] Sample Complexity

فصل ۴

تحلیل‌های مبتنی بر انقباض و نرخ‌های مینیماکس

۴ـ۱ مقدمه

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

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

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

۴ـ۲ محرمانگی به عنوان انقباض اطلاعاتی

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

[1] Contraction Operators
[2] Data Processing Inequality

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

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

قضیه‌ی ۴-۱ (انقباض KL در مکانیزم‌های موضعی) *فرض کنید $\mathcal{M}$ یک مکانیزم α-LDP باشد. برای هر دو ورودی $x, x' \in \mathcal{X}$، واگرایی کولبک-لایبلر بین توزیع‌های خروجی متناظر $\mathcal{M}(\cdot|x)$ و $\mathcal{M}(\cdot|x')$ با رابطه زیر محدود می‌شود:*

$$D_{KL}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) \leqslant ۴(e^{\alpha} - ۱)^{۲} \tag{۴-۱}$$

به طور دقیق‌تر، اگر $\alpha \leqslant ۱$ باشد، این کران به صورت $O(\alpha^{۲})$ رفتار می‌کند [۹].

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

$$e^{-\alpha} \leqslant \frac{q(z|x)}{q(z|x')} \leqslant e^{\alpha} \tag{۴-۲}$$

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

$$D_{KL}(P||Q) = \int p(z) \log \frac{p(z)}{q(z)} dz \tag{۴-۳}$$

$$\leqslant \int p(z) \left(\frac{p(z)}{q(z)} - ۱ + \frac{۱}{۲} \left(\frac{p(z)}{q(z)} - ۱ \right)^{۲} \right) dz \tag{۴-۴}$$

با اعمال کران‌های α-LDP بر روی نسبت p/q، جمله درجه اول صفر می‌شود و جمله درجه دوم ضریب $(e^{\alpha} - ۱)^{۲}$ را تولید می‌کند. جزئیات کامل این محاسبات در لم ۱ مقاله [۹] آمده است. نکته کلیدی این است که برای α کوچک، فاصله KL به صورت مربعی با α کاهش می‌یابد. □

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

[3] Strong Data Processing Inequality (SDPI)

۴-۲-۱ انقباض در فاصله واریانس کل

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

قضیه‌ی ۴-۲ (انقباض TV) *تحت شرایط مشابه، برای هر مکانیزم α-LDP:*

$$\|\mathcal{M}(\cdot|x) - \mathcal{M}(\cdot|x')\|_{TV} \leqslant \min\{۱, e^{\alpha} - ۱\} \cdot \|x - x'\|. \qquad (۴-۵)$$

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

۴-۳ تحلیل نرخ‌های مینیماکس با استفاده از انقباض

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

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

۲. **کران‌دار کردن اطلاعات متقابل:** استفاده از خاصیت انقباض α-LDPبرای محدود کردن اطلاعاتی که نمونه‌های مشاهده شده $Z_۱, \dots, Z_n$ درباره اندیس V می‌دهند [۶].

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

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

$$I(V; Z^n) \leqslant \sum_{i=۱}^{n} I(V; Z_i) \leqslant n \cdot \alpha^۲ \cdot C \qquad (۴-۶)$$

که در آن C ثابتی است که به هندسه مسئله بستگی دارد. این رابطه نشان می‌دهد که اطلاعات موثر با نرخ $n\alpha^۲$ رشد می‌کند، نه n. این همان دلیلی است که «اندازه نمونه موثر» در مدل موضعی برابر با $n\alpha^۲$ در نظر گرفته می‌شود.

[4] Le Cam
[5] Lower Bounds

۴-۴ مطالعه موردی: تخمین میانگین

برای نمایش قدرت این چارچوب، مسئله کلاسیک تخمین میانگین را در نظر می‌گیریم. فرض کنید هر کاربر i برداری $X_i \in [-۱, ۱]^d$ دارد و هدف تخمین میانگین جامعه $\mu = \mathbb{E}_X$ است. معیار خطا را «میانگین مربعات خطا» (MSE) در نظر می‌گیریم.

قضیه‌ی ۴-۳ (کران پایین تخمین میانگین) *برای هر مکانیزم α-LDP و هر تخمین‌گر $\hat{\mu}$، ماکسیمم خطای مورد انتظار با رابطه زیر محدود می‌شود* [۹]:

$$\inf_{\hat{\mu}, \mathcal{M}} \sup_{P} \mathbb{E}_{\|\hat{\mu}-\mu\|^۲} \geqslant \Omega\left(\frac{d}{n \min\{\alpha, \alpha^۲\}}\right) \tag{۴-۷}$$

تحلیل اثبات: برای اثبات این کران، از لم اسود استفاده می‌کنیم. فضای پارامتر را به صورت یک ابرمکعب $\{-۱, ۱\}^d$ گسسته‌سازی می‌کنیم. طبق لم اسود، خطا با مجموع فاصله‌های TV بین توزیع‌های شرطی مرتبط است. با استفاده از قضیه انقباض ۴-۱ و نامساوی پینسکر، می‌دانیم که:

$$\|\mathcal{M}(\cdot|x) - \mathcal{M}(\cdot|x')\|_{TV}^۲ \leqslant \frac{۱}{۲} D_{KL}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) \leqslant O(\alpha^۲) \tag{۴-۸}$$

بنابراین فاصله TV حداکثر از مرتبه α است. با جایگذاری این مقدار در لم اسود، کران پایین $\frac{۱}{n\alpha^۲}$ حاصل می‌شود.

این نتیجه نشان می‌دهد که برای رسیدن به خطای کم در مدل موضعی، تعداد داده‌ها باید متناسب با $۱/\alpha^۲$ افزایش یابد، که هزینه‌ی بسیار سنگین‌تری نسبت به مدل متمرکز (که متناسب با $۱/\varepsilon$ است) دارد.

۴-۵ محدودیت‌های تحلیل کلاسیک

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

همان‌طور که در رابطه (۴-۱) دیدیم، کران انقباض KL با ضریب $(e^\alpha - ۱)^۲$ رشد می‌کند. زمانی که α بزرگ باشد، این کران به سرعت به بی‌نهایت میل می‌کند و اطلاعاتی فراتر از کران بدیهی به ما نمی‌دهد. این در حالی است که به طور شهودی، حتی با α بزرگ، مکانیزم همچنان باید مقداری انقباض ایجاد کند.

این محدودیت ناشی از ذات واگرایی KL است که رفتار دنباله‌های توزیع را با حساسیت زیادی وزن‌دهی می‌کند. برای رفع این مشکل و به دست آوردن تحلیل‌های دقیق‌تر[۶] که در تمام بازه‌های α معتبر باشند، نیازمند

[6]Tight

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

فصل ۵

هم‌ارزی LDP و انقباض E_γـ واگرایی

۵-۱ مقدمه و انگیزه

در فصل پیشین، دیدیم که چگونه دوچی و همکاران [۹] از واگرایی کولبک‌ـلایبلر (KL) برای تحلیل محرمانگی تفاضلی موضعی استفاده کردند. اگرچه کران‌های آن‌ها برای رژیم‌های محرمانگی بالا (α کوچک) بسیار کارآمد هستند، اما در رژیم‌های α متوسط و بزرگ، دقت خود را از دست می‌دهند.

مشکل اصلی در آنجاست که واگرایی KL متریک «بومی» برای تعریف α-LDPنیست. تعریف α-LDP(معادله ؟؟) مبتنی بر نسبت احتمالات است، در حالی که KL مبتنی بر لگاریتم نسبت‌هاست. این ناهمخوانی باعث می‌شود که در تبدیل شرایط α-LDPبه کران‌های KL، اطلاعاتی از دست برود (lossy conversion).

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

[1]Tight

۵-۲ معرفی E_γـ واگرایی

E_γـ واگرایی یکی از اعضای کمتر شناخته‌شده‌ی خانواده fـ واگرایی‌هاست که در نظریه اطلاعات برای مقایسه نسبت درست‌نمایی توزیع‌ها کاربرد دارد.

تعریف ۵-۱ (E_γـ واگرایی) *فرض کنید P و Q دو توزیع احتمال باشند و $\gamma \geqslant 1$ یک عدد حقیقی باشد. E_γـ واگرایی بین P و Q به صورت زیر تعریف می‌شود:*

$$E_\gamma(P||Q) = \sup_{\mathcal{S}\in\sigma(\mathcal{X})} \left(P(\mathcal{S}) - \gamma Q(\mathcal{S})\right) \tag{۵-۱}$$

این تعریف را می‌توان به صورت بسته‌ی زیر نیز نوشت:

$$E_\gamma(P||Q) = \int_{\mathcal{X}} \max\{0, p(x) - \gamma q(x)\} d\mu(x) \tag{۵-۲}$$

که در آن p و q توابع چگالی احتمال هستند.

۵-۲-۱ خواص هندسی

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

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

$$E_1(P||Q) = \sup_{\mathcal{S}}(P(\mathcal{S}) - Q(\mathcal{S})) = \|P - Q\|_{TV} \tag{۵-۳}$$

  بنابراین E_γ تعمیمی از فاصله TV است.

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

۵-۳ قضیه هم‌ارزی اصلی

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

قضیه‌ی ۵-۱ (هم‌ارزی α-LDP و E_γ) *یک مکانیزم $\mathcal{M}$ در شرط α-LDP صدق می‌کند اگر و تنها اگر برای تمام جفت ورودی‌های $x, x' \in \mathcal{X}$:*

$$E_{e^\alpha}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) = ۰ \qquad (۵-۴)$$

اثبات. اثبات را در دو جهت انجام می‌دهیم.

جهت اول ($\Rightarrow$): فرض کنید $\mathcal{M}$ خاصیت α-LDP دارد. طبق تعریف ۳-۲، برای هر زیرمجموعه خروجی $\mathcal{S} \subseteq \mathcal{Z}$ و هر x, x' داریم:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leqslant e^\alpha \Pr[\mathcal{M}(x') \in \mathcal{S}] \qquad (۵-۵)$$

این نامساوی را می‌توان به صورت زیر بازنویسی کرد:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] - e^\alpha \Pr[\mathcal{M}(x') \in \mathcal{S}] \leqslant ۰ \qquad (۵-۶)$$

از آن‌جایی که این رابطه برای *تمام* $\mathcal{S}$ها برقرار است، سوپریمم آن نیز باید کوچکتر یا مساوی صفر باشد. اما طبق تعریف E_γ در معادله ۵-۱، این سوپریمم دقیقاً همان E_{e^α} است. چون E_γ نمی‌تواند منفی باشد (با انتخاب $\mathcal{S} = \emptyset$ مقدار حداقل صفر است)، پس حتماً برابر صفر است.

جهت دوم ($\Leftarrow$): فرض کنید $E_{e^\alpha}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) = ۰$. طبق تعریف سوپریمم، برای هر مجموعه دلخواه $\mathcal{S}$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] - e^\alpha \Pr[\mathcal{M}(x') \in \mathcal{S}] \leqslant ۰ \qquad (۵-۷)$$

که بلافاصله نتیجه می‌دهد:

$$\frac{\Pr[\mathcal{M}(x) \in \mathcal{S}]}{\Pr[\mathcal{M}(x') \in \mathcal{S}]} \leqslant e^\alpha \qquad (۵-۸)$$

این دقیقاً همان تعریف α-LDP است. □

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

۵-۴ بهبود کران‌های انقباض

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

$$D_{KL}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) \leqslant ۴(e^{\alpha}-۱)^{۲} \tag{۵-۹}$$

حال با استفاده از چارچوب E_{γ}، می‌توانیم کران‌های بسیار دقیق‌تری استخراج کنیم. آسوده و همکاران [۳] نشان داده‌اند که اگر شرط $E_{e^{\alpha}}=۰$ برقرار باشد، می‌توان کران‌های انقباض برای سایر fـ واگرایی‌ها را از طریق بهینه‌سازی محدب به دست آورد.

قضیه‌ی ۵-۲ (کران دقیق انقباض KL) *اگر $\mathcal{M}$ یک مکانیزم α-LDP باشد، آنگاه:*

$$D_{KL}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) \leqslant \frac{e^{\alpha}-۱}{e^{\alpha}+۱}\cdot(e^{\alpha}-۱) \tag{۵-۱۰}$$

برای مقادیر کوچک α (رژیم محرمانگی بالا)، این کران به $\alpha^{۲}/۲$ میل می‌کند که ۴ برابر کوچکتر (بهتر) از کران دوچی است.

تحلیل مقایسه‌ای: بیایید رفتار دو کران را در $\alpha\to۰$ بررسی کنیم:

- کران دوچی: $۴(e^{\alpha}-۱)^{۲}\approx ۴\alpha^{۲}$
- کران مبتنی بر E_{γ}: $\frac{\alpha^{۲}}{۲}=\frac{\alpha}{۲}\cdot\alpha \approx$ (چون $\alpha/۲ \approx \tanh(\alpha/۲)\approx \frac{e^{\alpha}-۱}{e^{\alpha}+۱}$)

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

۵-۵ تعمیم به محرمانگی تقریبی ((α,δ)-LDP)

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

$$\Pr[\mathcal{M}(x)\in\mathcal{S}] \leqslant e^{\alpha}\Pr[\mathcal{M}(x')\in\mathcal{S}]+\delta \tag{۵-۱۱}$$

با بازنویسی این رابطه داریم:

$$\sup_{\mathcal{S}}(\Pr[\mathcal{M}(x)\in\mathcal{S}]-e^{\alpha}\Pr[\mathcal{M}(x')\in\mathcal{S}]) \leqslant \delta \tag{۵-۱۲}$$

که دقیقاً معادل است با:

$$E_{e^\alpha}(\mathcal{M}(\cdot|x)||\mathcal{M}(\cdot|x')) \leqslant \delta \tag{۱۳-۵}$$

نتیجه‌ی ۵-۳ *محرمانگی تقریبی(α,δ)-LDP دقیقاً معادل محدود کردن مقدار E_{e^α}ـ واگرایی توزیع‌های خروجی به مقدار δ است. این نتیجه نشان می‌دهد که E_γـ واگرایی طبیعی‌ترین زبان برای صحبت درباره محرمانگی تفاضلی (چه خالص و چه تقریبی) است.*

۵-۶ کاربرد در تخمین توزیع گسسته

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

$$\mathfrak{M}_n \asymp \frac{k}{n(e^\alpha - ۱)^۲} \tag{۱۴-۵}$$

در حالی که استفاده از تکنیک‌های کلاسیک (دوچی)، جمله‌ای به صورت $\frac{k}{n\alpha^۲}$ را پیشنهاد می‌کرد. تفاوت این دو عبارت در رژیم α بزرگ (محرمانگی کم) آشکار می‌شود؛ جایی که $(e^\alpha - ۱)^۲$ به صورت نمایی رشد می‌کند و نشان می‌دهد که دقت می‌تواند بسیار سریع‌تر از پیش‌بینی‌های قبلی بهبود یابد.

۵-۷ انقباض قوی برای خانواده‌ی fـ واگرایی‌ها

تا اینجا دیدیم که شرط α-LDPمعادل صفر شدن E_{e^α}ـ واگرایی است. یک پرسش طبیعی و بسیار مهم این است: آیا این شرط بر روی سایر معیارهای فاصله (مثل $\chi^۲$ یا هلینجر) نیز انقباض ایجاد می‌کند؟ پاسخ مثبت است. در مقاله‌ی اخیر آسوده و ژانگ [۴]، نشان داده شده است که مکانیزم‌های موضعی خاصیت «انقباض قوی» را برای طیف وسیعی از واگرایی‌ها به ارمغان می‌آورند.

۵-۷-۱ کران دقیق برای واگرایی کای‌ـ دو ($\chi^۲$)

یکی از مهم‌ترین نتایج این پژوهش، ارائه‌ی یک ضریب انقباض دقیق برای واگرایی $\chi^۲$ است. اهمیت این واگرایی در آن است که کار با آن در محاسبات واریانس و کران‌های مینیماکس بسیار ساده‌تر از KL است.

قضیه‌ی ۵-۴ (انقباض $\chi^۲$) *فرض کنید $\mathcal{M}$ یک مکانیزم α-LDP باشد. برای هر دو توزیع ورودی P و Q، واگرایی کای‌ـ دو بین توزیع‌های خروجی با رابطه زیر محدود می‌شود:*

$$\chi^۲(\mathcal{M}P||\mathcal{M}Q) \leqslant \eta_\alpha \cdot \chi^۲(P||Q) \tag{۵-۱۵}$$

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

$$\eta_\alpha = \left(\frac{e^\alpha - ۱}{e^\alpha + ۱}\right)^۲ \tag{۵-۱۶}$$

تحلیل مجانبی: برای مقادیر کوچک α (رژیم محرمانگی بالا)، داریم:

$$\eta_\alpha \approx \left(\frac{۱ + \alpha - ۱}{۱ + \alpha + ۱}\right)^۲ \approx \left(\frac{\alpha}{۲}\right)^۲ = \frac{\alpha^۲}{۴} \tag{۵-۱۷}$$

این نتیجه بسیار قابل توجه است. یادآوری می‌کنیم که کران‌های کلاسیک دوچی (فصل ۳) ضریبی از مرتبه $O(\alpha^۲)$ داشتند، اما ضریب ۱/۴ در اینجا نشان‌دهنده‌ی یک انقباض بسیار شدیدتر است. این ضریب دقیقاً با ضریب انقباض «پاسخ تصادفی دودویی» برای واریانس مطابقت دارد و نشان می‌دهد که این کران برای کل کلاس مکانیزم‌های α-LDP «تایت» (Tight) است.

۵-۷-۲ تعمیم به سایر واگرایی‌ها

نویسندگان در [۴] نشان داده‌اند که این ضریب انقباض η_α تنها مختص $\chi^۲$ نیست، بلکه برای خانواده‌ای از واگرایی‌ها که خاصیت «تحدب مشترک» دارند (شامل فاصله هلینجر مجذور $H^۲$ و واگرایی KL) نیز صادق است.

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

$$D_{KL}(\mathcal{M}P||\mathcal{M}Q) \leqslant \eta_\alpha \cdot D_{KL}(P||Q) \tag{۵-۱۸}$$

$$H^۲(\mathcal{M}P, \mathcal{M}Q) \leqslant \eta_\alpha \cdot H^۲(P, Q) \tag{۵-۱۹}$$

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

۸-۵ نامساوی ون‌تریز خصوصی (Private van Trees Inequality)

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

نامساوی ون‌تریز (van Trees) ابزاری کلاسیک برای کران‌دار کردن خطای بیزی بر اساس «اطلاعات فیشر» است. آسوده و ژانگ [۴] نسخه‌ی خصوصی‌شده‌ی این نامساوی را ارائه کرده‌اند که ابزاری نوین در جعبه‌ابزار تحلیل محرمانگی محسوب می‌شود.

قضیه‌ی ۵-۶ (نامساوی ون‌تریز موضعی) *فرض کنید می‌خواهیم پارامتر θ را از روی مشاهدات Z^n که خروجی یک مکانیزم α-LDP هستند، تخمین بزنیم. اگر $\hat{\theta}$ هر تخمین‌گر دلخواهی باشد، آنگاه میانگین مربعات خطای بیزی[۲] دارای کران پایین زیر است:*

$$\mathbb{E}_{(\hat{\theta}-\theta)^2} \geqslant \frac{1}{\mathbb{E}_{\mathcal{I}(\theta)} + \mathcal{I}_{prior}(\pi)} \tag{۵-۲۰}$$

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

$$\mathcal{I}_{priv}(\theta) \leqslant \eta_\alpha \cdot \mathcal{I}_{orig}(\theta) \tag{۵-۲۱}$$

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

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

۹-۵ کاربردهای نوین و بهبود نرخ‌ها

استفاده از کران‌های انقباض قوی (بخش ۵-۷) و نامساوی ون‌تریز خصوصی (بخش ۵-۸) منجر به بهبود نتایج در مسائل کلاسیک می‌شود.

[۲] Bayesian Mean Square Error

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

$$R_{opt} \asymp \left(\frac{۱}{n\alpha^{۲}}\right)^{\frac{۲\beta}{۲\beta+۱}} \tag{۵-۲۲}$$

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

فصل ۶

نتیجه‌گیری

Bibliography

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 308–318, New York, NY, USA, 2016. Association for Computing Machinery.

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

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

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

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

[6] Leighton Pate Barnes, Wei Ning Chen, and Ayfer Özgür. Fisher information under local differential privacy. *IEEE Journal on Selected Areas in Information Theory*, 1:645–659, 2020.

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

[8] Paul Cuff and Lanqing Yu. Differential privacy as a mutual information constraint. In *Proceedings of the ACM Conference on Computer and Communications Security*, pages 43–54, 2016.

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

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

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

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

[13] 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.

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

[15] 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).

[16] 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.

[17] 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.

[18] 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.

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

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

[21] 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.

[22] 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.

[23] 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.

[24] 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

پرس‌وجو .... Query

پایگاه‌داده .... 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

ن

و

واگرایی .... Divergence

هـ

همسایه .... Ajacent

همسایگی .... Adjacency

ی

پیوست آ

مطالب تکمیلی

Abstract

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

Keywords: Thesis, Typesetting, Template, XePersian

Sharif University of Technology

Department of Mathematics

M.Sc. Thesis

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

By:

Firoozeh Abrishami

Supervisor:

Dr. Javad Ebrahimi Boroujeni

March 2026