همبستگی ماکسیمال

یاشار طالبی‌راد و حسین نادری

چکیده

در این مقاله ابتدا به بررسی علت مهم بودن همبستگی ماکسیمال [۱] و یکی از کاربردهای آن در تئوری اطلاعات اشاره می‌کنیم و به تعریف آن می‌پردازیم. سپس خواص ابتدایی آن را اثبات می‌کنیم و در انتها چند نامساوی پرکاربرد را بیان و اثبات می‌کنیم.

۱ مقدمه

یکی از مباحث مهم در تئوری اطلاعات، حفظ حریم داده[۲] است. مسأله‌هایی که در این مبحث مطرح می‌شوند را معمولا به این شکل شبیه‌سازی می‌کنند که دو نفر (مثل آلیس و باب) که از دو منبع اطلاعاتی مختلف (مثل X و Y) که توزیع توأم آن‌ها داده شده اطلاعات دریافت می کنند، آیا می‌توانند دو متغیر تصادفی A و B را که توزیع آن‌ها نیز داده شده است شبیه‌سازی کنند یا خیر؛ و اگر پاسخ بله است چگونه این کار را انجام دهند. معیارهای وابستگی[۳] مانند همبستگی ماکسیمال در پاسخ‌دادن به چنین سؤال‌هایی کمک می‌کنند.

۲ مدل‌سازی ریاضی

با توجه به این که شبیه‌سازی صد درصدی معمولا در کاربردهای واقعی غیرممکن است، در اکثر موارد کافیست فقط به مقدار لازم به توزیع A و B نزدیک شویم. یعنی آلیس و باب با انجام عملیات موضعی روی داده‌ها، متغیرهای تصادفی جدیدی می‌سازند که به A و B نزدیک باشند. عملیات موضعی روی متغیر تصادفی‌ای مثل X به معنی اعمال تابعی مانند f روی X و به‌دست آوردن $X' = f(X)$ است. حال فرض کنید آلیس و باب به‌جای این که فقط یک نمونه از X و Y داشته باشند، بتوانند به تعداد دلخواهی از آن نمونه تولید کنند؛ یعنی X^n و Y^n را داشته باشند. یک مدل پرکاربرد برای این مسأله، شبیه‌سازی غیرتعاملی[۴] نام دارد. شبیه‌سازی غیرتعاملی وقتی اتفاق می افتد که به ازای هر $\epsilon > ۰$، توابع قطعی f, g وجود داشته باشند که $d_{\mathrm{TV}}\Big((f(X^n), g(Y^n)), (A, B)\Big) < \epsilon$، که در تعریف بالا d_{TV} برابر است با

$$d_{\mathrm{TV}}(A, B) = \sup_{T \in \mathcal{F}} \{|\mathcal{P}_A(T) - \mathcal{P}_B(T)|\}$$

قضیه ۱. *شبیه‌سازی غیرتعاملی امکان‌پذیر است اگر و فقط اگر (A, B) در بستار[۵] مجموعه‌ی زیر باشد:*

$$\Big\{(A', B') \mid A' \to X^n \to Y^n \to B' \text{ زنجیر مارکوف باشد}\Big\}$$

اما بررسی شرط فوق نیز آسان نیست چون نمی‌توان اعضای

[1] Maximal Correlation
[2] Data Privacy
[3] Measures of Dependence
[4] Non-interactive Simulation
[5] Closure

مجموعه‌ی بالا را به سادگی مشخص کرد. اما به‌جای مشخص‌کردن اعضا، می‌توان کران‌هایی پیدا کرد که اعضای آن مجموعه باید در آن کران‌ها صدق کنند. به این ترتیب اگر A و B و X و Y داده‌شده ای در آن کران‌ها صدق نکرد، می‌توان مطمئن بود که (A,B) در آن مجموعه نیست و بنابراین شبیه‌سازی غیرتعاملی امکان‌پذیر نیست. معیارهای وابستگی در یافتن این نوع کران‌ها کمک می‌کنند.

۳ معیارهای وابستگی

تاکنون معیارهای متعددی برای اندازه گیری میزان وابستگی دو متغیر تصادفی تعریف شده است. برای مثال یکی از پرکاربردترین آن‌ها، ضریب همبستگی پیرسون [۶]

$$\rho(X;Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

است. هر معیار وابستگی خوب است چند ویژگی زیر را داشته باشد:

۱. تحت عملیات موضعی عوض نشود.

۲. میزان وابستگی ۰ است اگر و فقط اگر دو متغیر مستقل باشند.

۳. میزان وابستگی ۱ است اگر و فقط اگر رابطه‌ی قطعی‌ای بین دو متغیر تصادفی وجود داشته باشد.

بنابراین خاصیت ۱ بیان می‌کند که وابستگی دو متغیر تصادفی در یک زنجیر مارکوف، یا گذر از یک کانال، یا انجام عملیات موضعی زیاد نشود. به طور دقیق‌تر برای هر معیار وابستگی C و هر زنجیر مارکوف $X\longrightarrow Y\longrightarrow Z$ باید داشته باشیم:

$$C(X;Z)\leq C(X;Y)$$

این نامساوی به نامساوی پردازش اطلاعات[۷] معروف است. با استفاده از این نامساوی، می‌توان به کران‌هایی در قضیه ۱ دست یافت. اما هر معیار وابستگی این ویژگی را ندارد. برای مثال ضریب همبستگی پیرسون ویژگی های ۱ و ۲ را ندارد.

یکی دیگر از این معیار های وابستگی، اطلاعات متقابل[۸] است. با استفاده از نامساوی پردازش اطلاعات برای اطلاعات متقابل در قضیه ۱ می‌توان گفت

$$I(A,B)\leq I(X^n;Y^n)=nI(X;Y)$$

اما سمت راست نامساوی وقتی $n\longrightarrow\infty$ برابر صفر یا بینهایت می‌شود. پس این معیار کمک چندانی نمی‌کند.

در [۳] بیان شده که یک خاصیت خوب دیگر می‌تواند خاصیت تانسوری [۹] باشد. گفته می‌شود معیار وابستگی C خاصیت تانسوری دارد وقتی $C(X^n;Y^n)=C(X;Y)$. به این ترتیب اطلاعات متقابل به علت جمعی[۱۰] بودن، تانسوری نیست. یک معیار مشابه که این مشکل را حل می‌کند، همبستگی ماکسیمال نام دارد که اثباتی از تانسوری بودن آن در [۵] ارائه شده است. تازگی ثابت شده است که مفهومی به نام نوار ابرانقباض پذیری[۱۱] کران‌های بهتری از همبستگی ماکسیمال نتیجه می‌دهد. به تعدادی از این کران‌ها در [۵] اشاره شده است.

۴ تعاریف اولیه

تعریف ۱. *همبستگی ماکسیمال بین دو متغیر تصادفی X,Y عبارت است از بیشینه‌ی ضریب همبستگی پیرسون بین $f(X)$ و $g(Y)$ روی همه توابع f و g.*

[6] Pearson Correlation Coefficient
[7] Data-Processing Inequality
[8] Mutual Information
[9] Tensorization
[10] Additive
[11] Hypercontractivity Band

به عبارت دیگر

$$\rho_m(X;Y) = \max_{f,g}\left|\rho(f(X);g(Y))\right|$$
$$= \max_{f,g}\frac{\left|\mathrm{Cov}(f(X),g(Y))\right|}{\sigma_{f(X)}\sigma_{g(Y)}}$$

که در آن f و g توابع ثابت نیستند. وقتی یکی از X یا Y به طور قریب‌به‌یقین [12] ثابت باشد، همبستگی ماکسیمال ۰ تعریف می‌شود. همچنین Cov به معنی کوواریانس و σ_X همان انحراف معیار X است که به‌ترتیب عبارتند از

$$\mathrm{Cov}(X,Y) = \mathrm{E}\Big[(X-\mathrm{E}[X])\overline{(Y-\mathrm{E}[Y])}\Big]$$
$$= \mathrm{E}[X\overline{Y}] - \mathrm{E}[X]\mathrm{E}[\overline{Y}]$$
$$\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathrm{Cov}(X,X)} = \sqrt{\mathrm{E}\big[|X|^۲\big] - \big|\mathrm{E}[X]\big|^۲}$$

دقت کنید تعریف‌های بالا تعمیم حالت متغیرتصادفی حقیقی به حالت مختلط هستند.

۵ خواص

گزاره ۱. *برای هر دو متغیر تصادفی X,Y داریم*

$$۰ \leq \rho_m(X;Y) \leq ۱$$

برهان. طبق نامساوی کوشی‌ـ‌شوارتز [13] داریم

$$|\mathrm{Cov}(Z,W)|^۲ \leq \mathrm{Var}(Z)\mathrm{Var}(W)$$
$$\Rightarrow |\mathrm{Cov}(Z,W)| \leq \sqrt{\mathrm{Var}(Z)\mathrm{Var}(W)}$$
$$\Rightarrow \frac{|\mathrm{Cov}(Z,W)|}{\sqrt{\mathrm{Var}(Z)}\sqrt{\mathrm{Var}(W)}} \leq ۱ \Rightarrow |\rho(Z;W)| \leq ۱$$

با قرار دادن $Z=f(X), W=g(Y)$ طرف راست نامساوی ثابت می‌شود. طرف چپ بنابر تعریف بدیهی است. ■

برای بررسی چند خاصیت ابتدایی درباره‌ی استقلال، نیاز به تعریف تابع مشخصه[14] داریم. تابع مشخصه‌ی متغیر تصادفی X عبارت است از

$$\varphi_X(s) = \mathrm{E}[e^{isX}]$$

همچنین برای بردار تصادفی (X,Y) به این صورت تعریف می‌شود

$$\varphi_{X,Y}(s,t) = \mathrm{E}[e^{isX} \times e^{itY}]$$

تابع مشخصه مانند تابع توزیع احتمال، رفتار متغیر تصادفی را به طور یکتا مشخص می‌کند. به این معنی که $X_۱$ و $X_۲$ هم‌توزیع‌اند اگر و فقط اگر $\varphi_{X_۱} = \varphi_{X_۲}$.

لم ۱. *متغیر های تصادفی X و Y مستقل‌اند اگر و فقط اگر به‌ازای هر s و t داشته باشیم $\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t)$.*

برهان. ابتدا فرض کنید X و Y مستقل باشند. چون هر تابع‌هایی از متغیرهای مستقل خود نیز مستقل اند، برای هر t و s دو متغیر e^{itY} و e^{isX} نیز مستقل‌اند. در نتیجه

$$\mathrm{E}[e^{isX}e^{itY}] = \mathrm{E}[e^{isX}]\mathrm{E}[e^{itY}]$$

برای طرف دیگر ابتدا توجه کنید ضرب توزیع متغیرهای تصادفی $X \sim \mathcal{P}_X$ و $Y \sim \mathcal{P}_Y$ یعنی $P_X \times P_Y$ تابع مشخصه‌ای برابر $E[e^{isX}]E[e^{itY}]$ دارد که طبق فرض همان تابع مشخصه بردار تصادفی (X,Y) است. چون تبدیل فوریه [15] وارون‌پذیر است لذا تابع مشخصه یکتاست و $\mathcal{P}_X \times \mathcal{P}_Y = \mathcal{P}_{(X,Y)}$ که یعنی X و Y مستقل‌اند و حکم ثابت می‌شود. ■

قضیه ۲. *متغیر های تصادفی حقیقی X و Y مستقل‌اند اگر و فقط اگر $\rho_m(X;Y) = ۰$.*

برهان. ابتدا فرض کنید X و Y مستقل باشند. چون هر تابع‌هایی از متغیرهای مستقل، خود نیز مستقل اند، برای هر f و g داریم $\mathrm{E}[f(X)\overline{g(Y)}] = \mathrm{E}[f(X)]\mathrm{E}[\overline{g(Y)}]$ و در نتیجه

$$\forall f,g: \qquad \rho(f(X);g(Y)) = \frac{\mathrm{Cov}(f(X),g(Y))}{\sigma_X\sigma_Y} = ۰$$

[12] Almost Surely
[13] Cauchy-Schwarz Inequality
[14] Characteristic Function
[15] Fourier Transform

که یعنی همبستگی ماکسیمال برابر ۰ است.

برای طرف دیگر قضیه، اگر $\rho_m(X;Y) = ۰$ باشد، برای s و t ثابت قرار دهید $f(X) = e^{isX}$ و $g(Y) = e^{itY}$. توجه کنید که هرتابعی از e^{isX} تابعی از X نیز هست، و هر تابعی از e^{itY} تابعی از Y نیز هست پس بنابرتعریف

$$\rho_m(e^{isX}; e^{itY}) \le \rho_m(X;Y)$$

ولی $\rho_m(e^{isX}; e^{itY}) \ge ۰$ و $\rho_m(X;Y) = ۰$ پس

$$\begin{aligned}\rho_m(e^{isX}; e^{itY}) = ۰ &\Rightarrow \mathrm{Cov}(e^{isX}, e^{itY}) = ۰\\ &\Rightarrow \mathrm{E}[e^{isX}\overline{e^{itY}}] = \mathrm{E}[e^{isX}]\mathrm{E}[\overline{e^{itY}}]\end{aligned}$$

اما $\overline{e^{itY}} = e^{-itY}$ پس با تغییرمتغیر $t \leftarrow -t$ و طبق لم ۱، X و Y مستقل‌اند. ■

لم ۲. *برای هر دو تابع f و g می‌توان دو تابع f' و g' یافت به طوری که*

$$\mathrm{E}[f'(X)] = \mathrm{E}[g'(Y)] = ۰$$
$$\|f'(X)\| = \|g'(Y)\| = ۱$$
$$\rho\Big(f'(X); g'(Y)\Big) = \rho\Big(f(X); g(Y)\Big)$$

که در این جا $\|Z\| = \|Z\|_۲ = \sqrt{\mathrm{E}[|Z|^۲]}$.

برهان. ابتدا فرض کنید $f' = f - \mathrm{E}[f(X)]$. در این صورت داریم

$$\rho(f'(X); g(Y)) = \frac{\mathrm{Cov}(f(X) - \mathrm{E}[f(X)], g(Y))}{\sigma_{f(X)-\mathrm{E}[f(X)]}\sigma_{g(Y)}}$$

اما می‌دانیم واریانس و کوواریانس با اضافه شدن عددی ثابت به متغیرتصادفی تغییر نمی‌کنند. یعنی

$$\mathrm{Var}(Z+c) = \mathrm{Var}(Z), \qquad \mathrm{Cov}(Z+c, W) = \mathrm{Cov}(Z, W)$$

پس $\mathrm{E}[f'(X)] = ۰$ و

$$\rho(f'(X); g(Y)) = \frac{\mathrm{Cov}(f(X), g(Y))}{\sigma_{f(X)}\sigma_{g(Y)}} = \rho(f(X); g(Y))$$

و به همین ترتیب می‌توان $\mathrm{E}[g'(Y)]$ را نیز بدون تغییر همبستگی برابر ۰ کرد. به این کار سنترینگ [16] می‌گویند. حال با فرض این‌که امیدریاضی $f(X)$ و $g(Y)$ صفر است قراردهید $f' = f/\|f(X)\|$. در این صورت

$$\begin{aligned}\rho(f'(X); g(Y)) &= \frac{\mathrm{Cov}(f(X)/\|f(X)\|, g(Y))}{\sigma_{f(X)/\|f(X)\|}\sigma_{g(Y)}}\\ &= \rho(f(X); g(Y))\end{aligned}$$

زیرا برای هر عدد حقیقی c داریم

$$\mathrm{Var}(cZ) = c^۲\,\mathrm{Var}(Z), \qquad \mathrm{Cov}(cZ, W) = c\,\mathrm{Cov}(Z, W)$$

همچنین

$$\|f'(X)\| = \Big\| f(X)/\|f(X)\| \Big\| = \|f(X)\|/\|f(X)\| = ۱$$

و به همین ترتیب می‌توان $\|g'(Y)\|$ را نیز بدون تغییر همبستگی برابر ۱ کرد. به این کار اسکیلینگ [17] می گویند. به این ترتیب لم ثابت می‌شود.

■

با توجه به لم فوق، برای پیداکردن همبستگی ماکسیمال بین دو متغیر تصادفی، کافیست توجهمان را به توابع با امید ۰ و نرم ۱ و واریانس مثبت معطوف کنیم. برای سادگی به جای $f(X)$ از نماد f و به جای $g(Y)$ از g استفاده می‌کنیم. توجه کنید

$$\begin{aligned}\rho(f(X); g(Y)) &= \frac{\mathrm{E}[(f - \mathrm{E}[f])(g - \mathrm{E}[g])]}{\sqrt{(\mathrm{E}[f^۲] - \mathrm{E}[f]^۲)(\mathrm{E}[g^۲] - \mathrm{E}[g]^۲)}}\\ &= \frac{\mathrm{E}[f(X)g(Y)]}{\sqrt{\|f\|^۲\|g\|^۲}} = \mathrm{E}[f(X)g(Y)]\end{aligned}$$

پس کافیست $\mathrm{E}[f(X)g(Y)]$ را بیشینه کنیم. همچنین داریم

$$\begin{aligned}\mathrm{E}[f(X)g(Y)] &= \mathrm{E}\Big[\mathrm{E}\big[f(X)g(Y)\big|Y\big]\Big]\\ &= \mathrm{E}\Big[g(Y)\mathrm{E}[f(X)|Y]\Big] \le \|g(Y)\| \times \|\mathrm{E}[f(X)|Y]\|\end{aligned}$$

که در گام آخر از نامساوی کوشی‌ـ شوارتز استفاده شده‌است. حال دقت کنید $\|g(Y)\| = ۱$ و اگر f را داده‌شده فرض کنیم، برای بیشینه‌کردن همبستگی باید بین تمامی g ها آن تابعی را بیابیم که نامساوی بالا را به تساوی تبدیل کند. می‌دانیم حالت تساوی

[16] Centering
[17] Scaling

نامساوی کوشی‌ـ‌شوارتز وقتی اتفاق می‌افتد که دو متغیر تصادفی در راستای هم، یعنی ضریبی از هم باشند، یعنی

$$g(Y) = c\mathrm{E}[f(X)|Y]$$

اما باید $\|g\| = ۱$ و $\mathrm{E}[g] = ۰$ باشد. پس

$$g(Y) = \frac{\mathrm{E}[f(X)|Y]}{\|\mathrm{E}[f(X)|Y]\|}$$

و برای این f و g بهینه طبق نامساوی بالا داریم

$$\begin{aligned}\rho_m(X;Y) &= \max_{f:\mathrm{E}[f]=۰,\|f\|=۱} \|\mathrm{E}[f(X)|Y]\| \\ &= \max_{f:\mathrm{E}[f]=۰,\|f\|=۱} \sqrt{\mathrm{E}\Big[\mathrm{E}[f(X)|Y]^۲\Big]} \qquad \text{(I)}\end{aligned}$$

به همین ترتیب می‌توان برای $f(X)$ نیز رابطه مشابهی پیدا کرد

$$f(X) = \frac{\mathrm{E}[g^*(Y)|X]}{\|\mathrm{E}[g^*(Y)|X]\|}$$

که در این جا g^* جواب بهینه‌سازی زیر است

$$\rho_m(X;Y) = \max_{g:\mathrm{E}[g]=۰,\|g\|=۱} \|\mathrm{E}[g(Y)|X]\|$$

پس f و g بهینه که در همبستگی ماکسیمال صدق می‌کنند، باید در این روابط نیز صدق کنند، که به آن‌ها معادلات نقطه ثابت[۱۸] می‌گویند. با استفاده از این روابط و الگوریتم امیدریاضی شرطی متناوب [۱۹] می‌توان از یک حدس اولیه شروع کرد و به طور تکراری[۲۰] به f و g بهینه میل کرد. منبع [۱] به این الگوریتم و کاربردهای آن به تفصیل پرداخته است.

برای حالت‌های خاصی از X و Y، روش‌های ساده‌ای برای محاسبه‌ی ρ_m ارائه شده است. برای مثال در [۴] ثابت شده اگر حداقل یکی از X و Y الفبای دوحرفی داشته باشد، داریم

$$\rho_m^۲(X;Y) = \left(\sum_{x,y} \frac{p(x,y)^۲}{p(x)p(y)}\right) - ۱$$

۶ کاربردها

همان‌طور که در ابتدای متن اشاره کردیم، یکی از کاربردهای همبستگی ماکسیمال در محاسبه کران‌هایی برای بررسی شبیه‌سازی غیرتعاملی است. چند کران پرکاربرد را در این قسمت بیان و اثبات می‌کنیم.

لم ۳. *فرض کنید $X \longrightarrow Y \longrightarrow Z$ یک زنجیر مارکوف باشد. چون X و Z به شرط Y مستقل‌اند، برای هر دو تابع f و g داریم*

$$\begin{aligned}\mathrm{E}[f(X)|Y]\mathrm{E}[g(Z)|Y] = \mathrm{E}[f(X)g(Z)|Y] \Rightarrow \\ \mathrm{E}[f(X)g(Z)] = \mathrm{E}[\mathrm{E}[f(X)g(Z)|Y]] \\ = \mathrm{E}[\mathrm{E}[f(X)|Y]\mathrm{E}[g(Z)|Y]]\end{aligned}$$

لم ۴. *اگر $X \longrightarrow Y \longrightarrow Z$ زنجیر مارکوف باشد، $\mathrm{E}[f(X)g(Z)] \le \rho_m(X;Y)\rho_m(Z;Y)$.*

برهان. طبق لم بالا داریم

$$\begin{aligned}&\mathrm{E}[f(X)g(Z)] = \mathrm{E}[\mathrm{E}[f(X)|Y]\mathrm{E}[g(Z)|Y]] \\ &\overset{(a)}{\le} \|\mathrm{E}[f(X)|Y]\| \times \|\mathrm{E}[g(Z)|Y]\| \overset{\text{طبق}(I)}{\le} \rho_m(X;Y)\rho_m(Z;Y)\end{aligned}$$

که در آن (a) از نامساوی کوشی‌ـ‌شوارتز نتیجه شده است. ■

یک نتیجه‌ی مهم این لم این است که

$$\begin{aligned}\rho_m(X;Z) &= \underset{\text{بهینه } f,g}{E}[f(X)g(Z)] \\ &\le \rho_m(X;Y)\rho_m(Z;Y) \le \rho_m(X;Y)\end{aligned}$$

که نامساوی آخر از گزاره ۱ نتیجه شده است. این نتیجه همان نامساوی پردازش اطلاعات برای همبستگی ماکسیمال است. از آن‌جایی که توزیع توأم X و Y را می‌توان به شکل یک کانال مخابراتی و یک توزیع ورودی برای آن نیز بیان کرد، گاهی به‌جای محاسبه‌ی همبستگی X و Y، همبستگی توزیع کانال و توزیع ورودی را بررسی می‌کنند. مثال‌هایی از این نوع بررسی در [۵] آمده

[18] Fixed-Point Equations
[19] Alternating Conditional Expectation algorithm
[20] Iterative

است. یک مسأله‌ی پرکاربرد در این خصوص، تکرار یک کانال است. یعنی $X \longrightarrow Y \longrightarrow Z$ که در آن (X,Y) با (Z,Y) هم‌توزیع باشد. یک نامساوی مهم برای این حالت را که از لم بالا نتیجه می‌شود این‌جا آورده‌ایم.

نتیجه ۱. *اگر $X \longrightarrow Y \longrightarrow Z$ زنجیر مارکوف باشد و (X,Y) با (Z,Y) هم‌توزیع باشد، داریم*

$$\rho_m^{۲}(X;Y) = \max_{f:\mathrm{E}[f]=۰,\|f\|=۱} \mathrm{E}[f(X)f(Z)]$$

برهان. چون (X,Y) با (Z,Y) هم‌توزیع است داریم

$$\begin{aligned} &\mathrm{E}[f(X)|Y] = \mathrm{E}[f(Z)|Y] \\ &\Rightarrow \mathrm{E}[f(X)|Y]^{۲} = \mathrm{E}[f(X)|Y]\mathrm{E}[f(Z)|Y] \\ &\Rightarrow \mathrm{E}\Big[\mathrm{E}[f(X)|Y]^{۲}\Big] = \mathrm{E}\Big[\mathrm{E}[f(X)|Y]\mathrm{E}[f(Z)|Y]\Big] \\ &= \mathrm{E}\Big[\mathrm{E}[f(X)f(Z)|Y]\Big] = \mathrm{E}[f(X)f(Z)] \end{aligned}$$

اما طبق (I) داریم

$$\begin{aligned} \rho_m^{۲}(X;Y) &= \max_{f:\mathrm{E}[f]=۰,\|f\|=۱} \mathrm{E}[\mathrm{E}[f(X)|Y]^{۲}] \\ &= \max_{f:\mathrm{E}[f]=۰,\|f\|=۱} \mathrm{E}[f(X)f(Z)] \end{aligned}$$

و حکم ثابت می‌شود.

دقت کنید چون

$$\begin{aligned} &\max_{f:\mathrm{E}[f]=۰,\|f\|=۱} \mathrm{E}[f(X)f(Z)] \\ &\leq \max_{f,g:\mathrm{E}[f]=\mathrm{E}[g]=۰,\|f\|=\|g\|=۱} \mathrm{E}[f(X)g(Z)] \\ &= \rho_m(X;Z) \end{aligned}$$

می‌توان نتیجه گرفت $\rho_m^{۲}(X;Y) \leq \rho_m(X;Z)$. همچنین طبق نتیجه‌ی لم قبل داریم

$$\rho_m(X;Z) \leq \rho_m(X;Y)\rho_m(Z;Y) = \rho_m^{۲}(X;Y)$$

که نتیجه می‌دهد $\rho_m(X;Z) = \rho_m^{۲}(X;Y)$.

نتیجه ۲.

$$\sup_{\substack{X\longrightarrow Y\longrightarrow Z \\ \rho_m(Y;Z)\neq ۰}} \frac{\rho_m(X;Z)}{\rho_m(Z;Y)} = \rho_m(X;Y)$$

توجه کنید طبق نتیجه‌ی لم ۴

$$\frac{\rho_m(X;Z)}{\rho_m(Z;Y)} \leq \rho_m(X;Y)$$

اثباتی برای رخ‌دادن تساوی در نامساوی بالا در [۴] ارائه شده‌است. یکی از توزیع‌های پرکاربرد برای X و Y، توزیع نرمال دومتغیره است. در [۲] همبستگی ماکسیمال برای این توزیع

$$\rho_m(X;Y) = |\rho(X;Y)|$$

محاسبه شده‌است. همچنین قضیه‌ی زیر را داریم

$$\rho_m^{۲}(X;Y) \leq ۱ - ۲^{-۲I(X;Y)} \leq (۲\ln ۲)I(X;Y)$$

اثباتی از این قضیه در [۴] آمده است. نکته‌ی جالب راجع به این نامساوی این است که همبستگی ماکسیمال را به اطلاعات متقابل ربط می‌دهد.

۷ نتیجه گیری

همان‌طور که دیدیم، همبستگی ماکسیمال در تئوری اطلاعات و حفظ حریم داده کاربردهای بسیاری دارد و کمک می‌کند بفهمیم چه موقع یک شبیه‌سازی امکان‌پذیر نیست.

۸ تشکر

در پایان از استادان گران‌قدر دکتر سلمان ابوالفتح بیگی و دکتر پویا شریعت‌پناهی که بدون رهنمودهای ایشان نگارش این نوشته ناممکن بود کمال سپاس‌گزاری را داریم.

مراجع

[1] J. Friedman L. Beriman. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association, 1985.

[2] H. O. Lancaster. Some properties of the bivariate normal distribution considered in the form of the contingency table. Biometrika, Volume 44, Issue 1-2, page 289–292, 1957.

[3] A. Renyi. On measures of dependence. Acta Mathematica Hungarica.

[4] T. Linder S. Asoodeh, F. Alajaji. On maximal correlation, mutual information and data privacy. 2015 IEEE 14th Canadian Workshop on Information Theory (CWIT).

[5] Y. Park Z. Yin. Hypercontractivity, maximal correlation and non-interactive simulation.