نظریه اطلاعات شبکه، ترم بهار ۱۳۹۴-۱۳۹۳

مدرس: امین زاده گوهری

جلسه ۲

مرور بر مفاهیم اصلی نظریه اطلاعات نقطه به نقطه

در نظریه اطلاعات به روابطی بر میخوردیم که عبارات آنتروپی در آنها ظاهر میشود. برخی اوقات این عبارات شامل ماکزیمم گیری روی مجموعه ای از توزیع ها میشود. مثلا ظرفیت کانال نقطه به نقطه به شکل $\max_{p(x)} I(X;Y)$ میباشد که در اینجا کانال $p(y|x)$ ثابت بوده و توزیع ورودی $p(x)$ تغییر میکند. برعکس در پاسخ به برخی سؤالات ممکن است $p(x)$ ثابت باشد و کانال $p(y|x)$ تغییر کند. بنابراین فهمیدن نحوه رفتار عبارات آنتروپیک زمانی که بخشی از توزیع مشترک میان متغیرها ثابت فرض شده و بخشی دیگر متغیر باشد میتواند مفید باشد.

جهت انگیزه دادن به بحث این جلسه مساله زیر را در نظر بگیرید: فرض کنید که توزیع ثابت $p(a,b,x)$ را روی سه متغیر تصادفی A,B,X داشته باشیم. کانالی را متصور شوید که X را به عنوان ورودی گرفته و مستقل از مقادیر A و B متغیر Y را تولید میکند. یعنی توزیع مشترک متغیرهای A,B,X,Y بشکل

$$p(a,b,x,y) = p(a,b,x)p(y|x)$$

بوده و یا زنجیره مارکف

$$A,B - X - Y$$

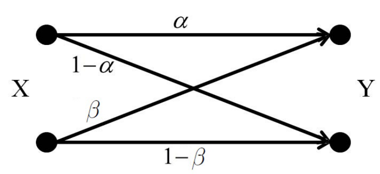
برقرار میباشد. فرض کنید که توزیع $p(a,b,x)$ ثابت بوده اما کانال $p(y|x)$ متغیر است. در این صورت سؤال این است که عبارت زیر را چگونه محاسبه میکنیم؟

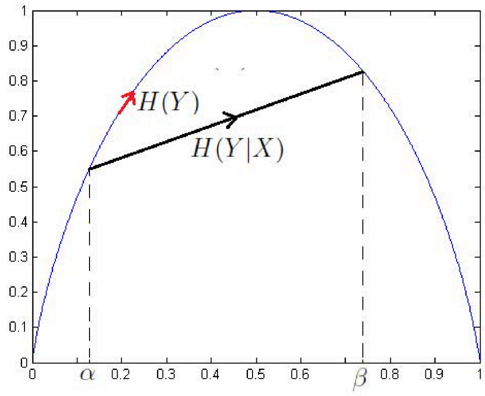
$$\min_{p(y|x)} I(X;A|Y) - I(X;B|Y)$$

این کار از نظر محاسباتی آسان به نظر نمیرسد زیرا نیازمند مینیمم گیری روی تمامی کانال ها میباشد. در این جلسه ابزار های لازم برای حمله به این مساله را بدست خواهیم آورد.

۱ یک مثال

عبارت $\max_{p(x)} I(X;Y)$ مربوط به ظرفیت کانال زیر را در نظر بگیرید:





شکل ۱: تغییرات $H(Y|X)$ و $H(Y)$ بر حسب توزیع ورودی برای کانال دودویی

در این صورت

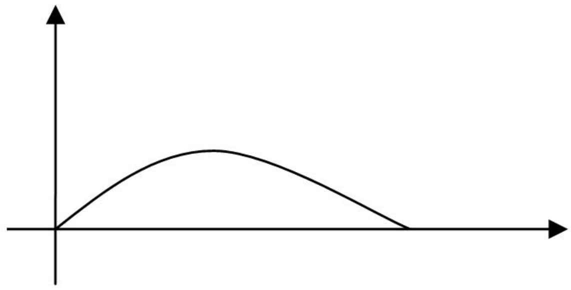
$$\begin{aligned} H(Y|X) &= (1-p)H(|X=0) + pH(Y|X=1) \\ &= (1-p)h(\alpha) + ph(\beta). \\ H(Y) &= h\big((1-p)\alpha + p\beta\big). \end{aligned}$$

فرض کنید که p از صفر تا یک تغییر داده شود. تغییرات $H(Y|X)$ و $H(Y)$ در شکل ۱ نشان داده شده است. در حالت $p=0$ مقادیر $H(Y)=H(Y|X)=h(\alpha)$ بوده و در حالت $p=1$ مقادیر $H(Y)=H(Y|X)=h(\beta)$ میباشد. با تغییر p از صفر تا یک $H(Y|X)$ روی یک خط عبور می نماید. با تغییر p مقدار $H(Y)$ روی نمودار تابع آنتروپی باینری حرکت می کند.

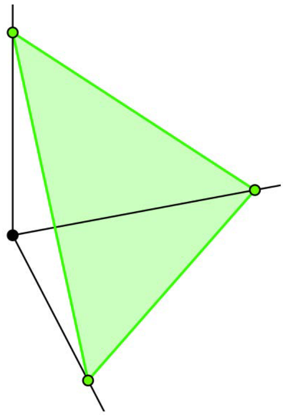
در ازای p داده شده در بازه $[0,1]$، در نقطه $(1-p)\alpha+p\beta$ روی محور افقی نمودار قرار خواهیم داشت و مقدار $H(Y|X)=(1-p)h(\alpha)+pH(\beta)$ از این طریق بدست می آید که به مقدار تابع آنتروپی باینری در نقاط مرزی $p=0,1$ نگاه کرده و با یک خط آنها را به هم وصل میکنیم. سپس به مقدار این خط در نقطه $(1-p)\alpha+p\beta$ نگاه میکنیم و از روی آن $H(Y|X)$ را بدست می آوریم.

میدانیم که $H(Y) \geq H(Y|X)$ همواره درست است. بنابراین رابطه $H(Y) \geq H(Y|X)$ مقعر بودن $H(Y)$ را نتیجه میدهد. با تغییر p از صفر تا یک نمودار $I(X;Y)=H(Y)-H(Y|X)$ مانند شکل ۲ میباشد. چون این نمودار تفاضل یک تابع مقعر و خطی است پس حتما مقعر است. دقت کنید که بر خلاف تابع آنتروپی باینری مشتق آن در $p=0$ و $p=1$ لزوما بینهایت نیست.

در حالت کلی تر غیردودویی هم $I(X;Y)$ بر حسب توزیع ورودی همواره یک تابع مقعر است.



شکل ۲: تغییرات $I(X;Y)$ بر حسب توزیع ورودی برای کانال دودویی



شکل ۳: سیمپلکس احتمالاتی در سه بعد (منبع: ویکیپدیا)

۲ تصور هندسی از توزیع های احتمالاتی

فرض کنید که متغیر مقادیر $\mathcal{X} = \{0, 1, 2, \cdots, n\}$ را اختیار میکند. در این صورت بردار

$$[p(X = 0), p(X = 1), \cdots, p(X = n)]$$

که از کنار هم گذاشتن احتمالات حاصل میشود، برداری در فضای $n + 1$ بعدی خواهد بود. مولفه های این بردار نامنفی بوده و جمع آنها برابر یک است. پس به زیرفضای زیر از $\mathbb{R}^{n+1}$ تعلق دارد:

$$\mathcal{P}_n = \{(p_0, p_1, p_2, \cdots, p_n) \in \mathbb{R}^{n+1} \mid \Sigma_{i=0}^{n} p_i = 1 \text{ و } \quad p_i \geq 0 \qquad \forall i\}$$

ناحیه $\mathcal{P}_n$ یک چند وجهی در فضای $\mathbb{R}^{n+1}$ است اما بعد آن از بعد فضا یکی کمتر است (یعنی یک چندوجهی n بعدی است). به این چندوجهی سیمپلکس احتمالاتی[1] گفته میشود. سیمپلکس احتمالاتی در سه بعد در شکل ۳ نشان داده شده است. سیمپلکس احتمالاتی در سه بعد یک مثلث است؛ در دو بعد یک خط است؛ در چهار بعد یک هرم است و الی آخر. رئوس این چندوجهی نقاط زیر هستند:

[1] Probability Simplex

$$v_1 = (1, 0, 0, \cdots, 0),$$
$$v_2 = (0, 1, 0, \cdots, 0),$$
$$\vdots$$
$$v_{n+1} = (0, 0, 0, \cdots, 1).$$

دلیل اینکه به این چندوجهی سیمپلکس گفته میشود این است که ساده ترین شکل ممکن n بعدی (از لحاظ داشتن حداقل تعداد رئوس) میباشد. در صفحه دو بعدی مثلث با سه راس ساده ترین شکل است؛ در یک فضای یک بعدی، پاره خط با دو راس ساده ترین شکل است و الی آخر.

پس هر توزیع احتمالاتی روی مجموعه $\mathcal{X} = \{0, 1, 2, \cdots, n\}$ نقطه ای در سیمپلکس احتمالاتی است. تغییر دادن این توزیع به معنی تغییر دادن نقطه کار روی سیمپلکس احتمالاتی است.

در صورتی که دو توزیع احتمالی π_1 و π_2 در سیمپلکس احتمالی داشته باشیم، فاصله میان آنها معمولا با توجه به

$$D(\pi_1 \| \pi_2) = \sum_x \pi_1(x) \log \frac{\pi_1(x)}{\pi_2(x)}$$

و یا متر فاصله مجموع[۲]

$$\|\pi_1 - \pi_2\| = \sum_x |\pi_1(x) - \pi_2(x)|$$

محاسبه میشود.

۱.۲ توابع روی سیمپلکس احتمالاتی

تابع $f(p(x))$ که یک توزیع احتمال دلخواه $p(x)$ روی $\mathcal{X}$ را به عددی حقیقی تبدیل میکند را میتوان متصور شد.

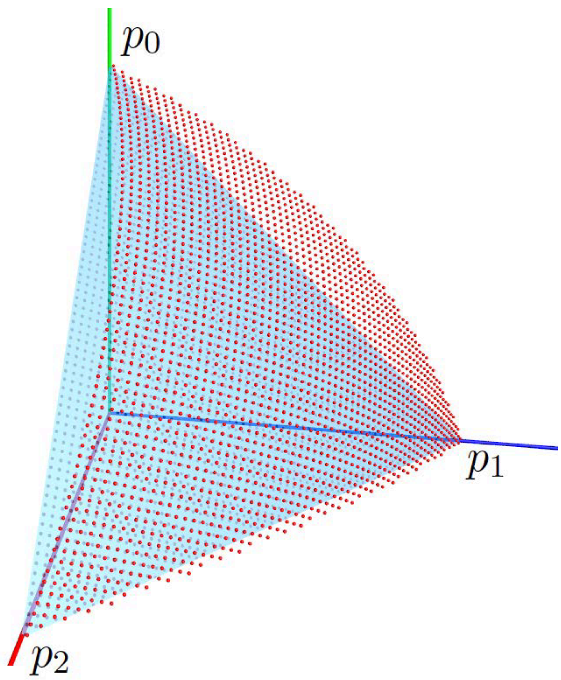
$$f : \mathcal{P} \mapsto \mathbb{R}.$$

یک تابع مهم از این خانواده تابع آنتروپی $f(p(x)) = H(X)$ است که در رئوس ناحیه سیمپلکس احتمالاتی برابر صفر بوده و در توزیع یکنواخت به مقدار بیشینه اش میرسد. این تابع روی سیمپلکس احتمالاتی تابعی مقعر است. نمودار این تابع در شکل ۴ آمده است.

اما توابع دیگری را هم میتوان متصور شد. در ازای کانال ثابت $p(a|x)$، با داشتن توزیع $p(x)$ توزیع $p(a)$ را میتوان محاسبه کرد و از روی آن آنتروپی این توزیع را محاسبه کرد:

$$f(p(x)) = H(A).$$

[۲] Total variation distance



شکل ۴: نمودار تابع آنتروپی. Image courtesy of John Williamson 2006.

تابع فوق نیز تابعی مقعر روی سیمپلکس احتمالاتی $p(x)$ است.

همچنین در صورتی که کانال ثابت $p(a,b|x)$، با داشتن توزیع $p(x)$ توزیع $p(a,b)$ را میتوان محاسبه کرد و از روی آن آنتروپی این این دو متغیر را محاسبه کرد. در این صورت میتوان تابع زیر را نیز در نظر گرفت:

$$f(p(x)) = H(A) - H(B).$$

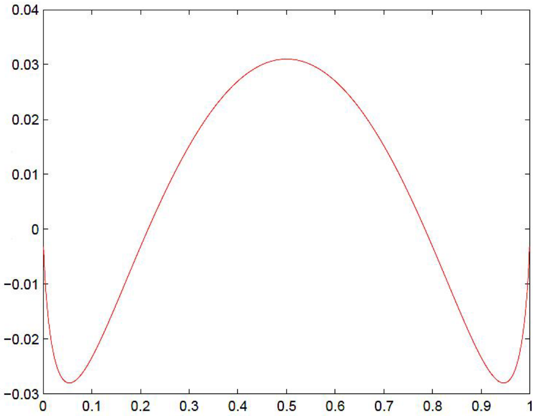
این تابع تفاضل دو تابع مقعر است و حاصل آن در حالت کلی نه محدب است و نه مقعر. مثال دیگر میتواند تابع

$$f(p(x)) = I(X;A) - I(X;B)$$

باشد. نمودار این تابع در شکل ۵ در صورتی که فرض کنیم که X دودویی باشد، و کانال از X به A کانال BSC با پارامتر $\frac{1}{10}$، و کانال از A به C کانالی BEC با پارامتر $\frac{1}{2}$ باشد، آمده است. در اینجا نمودار بر حسب $p = p(X=0)$ رسم شده است.

۳ عبور یک توزیع از یک کانال به معنی پخش شدگی است

فرض کنید که توزیع $p(x)$ از متغیر تصادفی X ثابت است و آن را از کانال متغیری $p(y|x)$ گذر داده ایم. هدف از این بخش فراهم کردن تصویری هندسی از اثر کانال روی یک توزیع ثابت و مشخص است. برای شروع فرض کنید که $\mathcal{X} = \{0,1\}$ و X توزیع ثابت $p(x)$ روی این الفبا داشته باشد. فرض کنید که گیرنده متغیر X را از طریق کانال $p(y|x)$



شکل ۵: نمودار تابع $f(p(x)) = I(X;A) - I(X;B)$ بر حسب $p = p(X=0)$ در صورتی که X دودویی باشد، و کانال از X به A کانال BSC با پارامتر $\frac{1}{10}$، و کانال از A به C کانالی BEC با پارامتر $\frac{1}{2}$ باشد.

مشاهده کرده باشد. برای سادگی فرض کنید که الفبای Y دودویی باشد. در صورتی که گیرنده مقدار $Y=0$ را دریافت کند، توزیع X از نقطه نظر او برابر خواهد بود با

$$\pi_0 = \begin{pmatrix} p(X=0|Y=0) \\ \\ p(X=1|Y=0) \end{pmatrix},$$

و در صورتی که گیرنده مقدار $Y=1$ را دریافت کند، توزیع X از نقطه نظر او برابر خواهد بود با

$$\pi_1 = \begin{pmatrix} p(X=0|Y=1) \\ \\ p(X=1|Y=1) \end{pmatrix}.$$

هر دوی نقاط π_0 و π_1 روی سیمپلکس احتمالاتی هستند. اما موقعیت آنها نسبت به

$$\pi = \begin{pmatrix} p(X=0) \\ \\ p(X=1) \end{pmatrix}$$

چگونه است؟ با توجه به خاصیت تابع توزیع احتمال داریم:

$$p(X=0) = p(X=0|Y=0)p(Y=0) + p(X=0|Y=1)p(Y=1)$$

$$p(X=1) = p(X=1|Y=0)p(Y=0) + p(X=1|Y=1)p(Y=1)$$

می توان معادلات بالا را بصورت زیر نوشت:

$$\begin{pmatrix} p(X=0) \\ \\ p(X=1) \end{pmatrix} = \begin{pmatrix} p(X=0|Y=0) \\ \\ p(X=1|Y=0) \end{pmatrix} p(Y=0) + \begin{pmatrix} p(X=0|Y=1) \\ \\ p(X=1|Y=1) \end{pmatrix} p(Y=1).$$

در نتیجه π بر روی پاره خطی که نقاط π_0 و π_1 را به هم وصل میکند قرار دارد. همچنین π مرکز ثقل این نقاط است اگر روی نقطه π_0 وزنه ای با وزن $p(Y=0)$ و روی نقطه π_1 وزنه ای با وزن $p(Y=1)$ قرار دهیم. در صورتی که Y بجای دو مقدار سه مقدار میگرفت، آنوقت باز هم میتوانستیم نشان دهیم

$$\pi = p(Y=0)\pi_0 + p(Y=1)\pi_1 + p(Y=2)\pi_2.$$

در حالت کلی

مفهوم عبور دادن X از یک کانال یعنی انتخاب چند نقطه از ناحیه و اختصاص وزن به این نقاط، بطوریکه مرکز ثقل این نقاط همان نقطه مربوط به توزیع مفروض X باشد.

مثال ۱ نحوه پخش شدگی π به موقعیت نقطه π بستگی دارد. حتی با فرض ثابت بودن کانال $p(y|x)$ با تغییر π نحوه پخش شدگی میتواند تغییر کند. اگر π یکی از رئوس سیمپلکس احتمالاتی باشد، آنوقت عبور از هر کانال دلخواهی باعث پخش شدگی آن نمیشود. چون π در نقطه گوشه ای قرار گرفته است، مرکز ثقل مجموعه ای از نقاط متمایز در سیمپلکس احتمالاتی نمیتواند باشد. در نتیجه $\pi_i = \pi$ برای هر i باید برقرار باشد. این را بصورت مستقیم هم میتوان تحقیق کرد. اگر π یکی از رئوس سیمپلکس احتمالاتی باشد آنوقت X یک عدد ثابت را با احتمال یک میگیرد. در نتیجه توزیع X به شرط $Y=y$ هم با احتمال یک برابر همان عدد ثابت خواهد بود.

قضیه ۲ فرض کنید که π و مجموعه ای از نقاط $\pi_0, \pi_1, \cdots, \pi_k$ به ما داده شده باشد بطوریکه بتوان وزن های نامنفی $\omega_0, \omega_1, \cdots, \omega_k$ با مجموع یک را یافت که

$$\pi = \omega_0\pi_0 + \omega_1\pi_1 + \cdots + \omega_k\pi_k$$

در این صورت اگر متغیر X توزیع π را داشته باشد آنگاه میتوان کانال $p(y|x)$ را بگونه ای یافت که

$$p(Y=i) = \omega_i$$
$$p(X|Y=i) = \pi_i$$

۱.۳ میزان پخش شدگی و اطلاعات متقابل

فرض کنید که X توزیع ثابت و مفروض $p(x)$ را داشته باشد. فرض کنید که گیرنده متغیر X را از طریق کانال $p(y|x)$ مشاهده کرده باشد. ابتدا کانال خاص زیر را در نظر بگیرید که در آن الفبای $\mathcal{Y} = \{0\}$ تک عضوی است. یا به عبارت دیگر

$$p_{Y|X}(0|x) = 1, \qquad \forall x.$$

در این صورت گیرنده همواره مقدار صفر را دریافت کرده و هیچ اطلاعات جدیدی در مورد X بدست نمی آورد

$$I(X;Y) = 0.$$

توزیع X از نقطه نظر او برابر $p(x)$ خواهد بود. در اینجا کانال هیچگونه پخش شدگی ایجاد نمیکند و نقطه $p(x)$ را به همان نقطه تبدیل میکند.

حالت افراطی دیگر این است که Y اطلاعات کاملی راجع به X بدهد (یعنی $Y = X$). در این صورت توزیع X به شرط Y همواره یکی از رئوس سیمپلکس احتمالاتی خواهد بود و میزان پخش شدگی حداکثر مقدار ممکن است. مشاهده میکنیم که میان اطلاعاتی که کانال منتقل میکند و میزان پخش شدگی ارتباطی وجود دارد. این رابطه به شکل زیر قابل بیان است:

$$I(X;Y) = \sum_y p(y) D\big(p(x|y) \| p(x)\big)$$

که در آن $D\big(p(x|y)\|p(x)\big)$ فاصله میان نقطه جدید در سیمپلکس احتمالاتی (پس از مشاهده y) و نقطه اولیه میباشد. به عبارت دیگر اگر فاصله میان نقاط پخش شده و نقطه اولیه را با توجه به تابع وزن مربوط به نقاط پخش شده میانگین گیری کنیم، به اطلاعات متقابل میرسیم.

نکته ۳ در مثال ۱ دیدیم که نحوه پخش شدگی توسط یک کانال ثابت $p(y|x)$ به $p(x)$ ربط دارد. با توجه به بحث های قبلی مفهوم ظرفیت کانال $\max_{p(x)} I(X;Y)$ را میتوان اینگونه تفسیر کرد: برای کانال داده شده موقعیت $p(x)$ را بگونه ای بیابید بطوریکه متوسط پخش شدگی توزیع توسط کانال به حداکثر مقدار ممکنش برسد.

۴ مقدار متوسط توابع آنتروپیک روی سیمپلکس احتمالی پس از گذر از کانال

تابع $f(p(x))$ که یک توزیع احتمال دلخواه $p(x)$ روی $\mathcal{X}$ را به عددی حقیقی تبدیل میکند را در نظر بگیرید. فرض کنید که X توزیع ثابت و مفروض $p(x)$ را داشته باشد. فرض کنید که گیرنده متغیر X را از طریق کانال $p(y|x)$ مشاهده کرده باشد. در این صورت توزیع جدید X از نقطه نظر گیرنده پس از مشاهده $Y = y$ بصورت $p(x|Y = y)$ خواهد بود. و مقدار تابع f در این نقطه روی سیمپلکس احتمالاتی برابر $f\big(p(x|Y = y)\big)$ خواهد بود. گیرنده با احتمال $p(Y = y)$ مشاهده $Y = y$ را خواهد داشت، پس مقدار متوسط تابع f از نقطه نظر گیرنده برابر است با

$$\sum_y p(Y = y) f\big(p(x|Y = y)\big).$$

از لحاظ هندسی رابطه

$$\sum_y p(Y = y) f\big(p(x|Y = y)\big).$$

به معنی زیر است. نقطه $\pi = p(x)$ توسط کانال به نقاط $\pi_y = p(x|Y = y)$ پخش میشود. میانگین نقاط $\sum_y p(y)\pi_y$ همان نقطه π است. ما با استفاده از همین وزن ها مقدار تابع f را در این نقاط ترکیب میکنیم.

مثال ۴ فرض کنید که $f(p(x)) = H(X)$. در این صورت مقدار متوسط تابع از نقطه نظر گیرنده برابر است با

$$\sum_y p(Y = y) H(X|Y = y) = H(X|Y).$$

فرض کنید که $f(p(x)) = H(X)$

مثال ۵ کانال ثابتی مانند $p(a|x)$ را در نظر بگیرید. تابع f را اینگونه تعریف میکنیم:

$$f(p(x)) = H(A, X).$$

عبارت

$$\sum_y p(Y=y) f\big(p(x|Y=y)\big)$$

را در نظر بگیرید. از آنجایی که برای محاسبه $f\big(p(x|Y=y)\big)$ برای هر مقدار y، طبق تعریف f باید از کانال ثابت $p(a|x)$ استفاده کنیم. پس اگر بخواهیم صحبت از توزیع مشترک A, X, Y بکنیم، باید $p(a|xy) = p(a|x)$ و یا زنجیره مارکف $A - X - Y$ را داشته باشیم. در نتیجه برای متغیرهای A, X, Y با این توزیع مشترک داریم:

$$\sum_y p(Y=y) f\big(p(x|Y=y)\big) = \sum_y p(Y=y) H(A, X|Y=y) = H(A, X|Y).$$

۱.۴ پوش مقعر و پوش محدب گیری از نمودار یک تابع

برای تابع داده شده f، پوش محدب آن "نزدیکترین" تقریب محدب از تابع f است و آن را با $\mathbb{C}[f]$ نشان میدهیم. خاصیت اصلی $\mathbb{C}[f]$ این است که اولا $\mathbb{C}[f] \leq f$ یعنی این تابع کران پایینی برای f میباشد (نمودارش زیر نمودار f قرار میگیرد). خاصیت دوم این است که برای هر تابع محدب g که نمودارش زیر نمودار f قرار بگیرد $g \leq f$ داریم: $g \leq \mathbb{C}[f]$. به عبارت دیگر $\mathbb{C}[f]$ بزرگترین تابع محدبی است که نمودارش زیر نمودار f قرار میگیرد.

پوش مقعر یک تابع که آن را با $\mathbb{K}[f]$ نشان میدهیم بشکل مشابهی تعریف میشود: $\mathbb{K}[f]$ کوچکترین تابع مقعری است که نمودارش بالای نمودار f قرار میگیرد. پوش محدب یک تابع محدب با خودش برابر است، و پوش مقعر یک تابع مقعر با خودش برابر است.

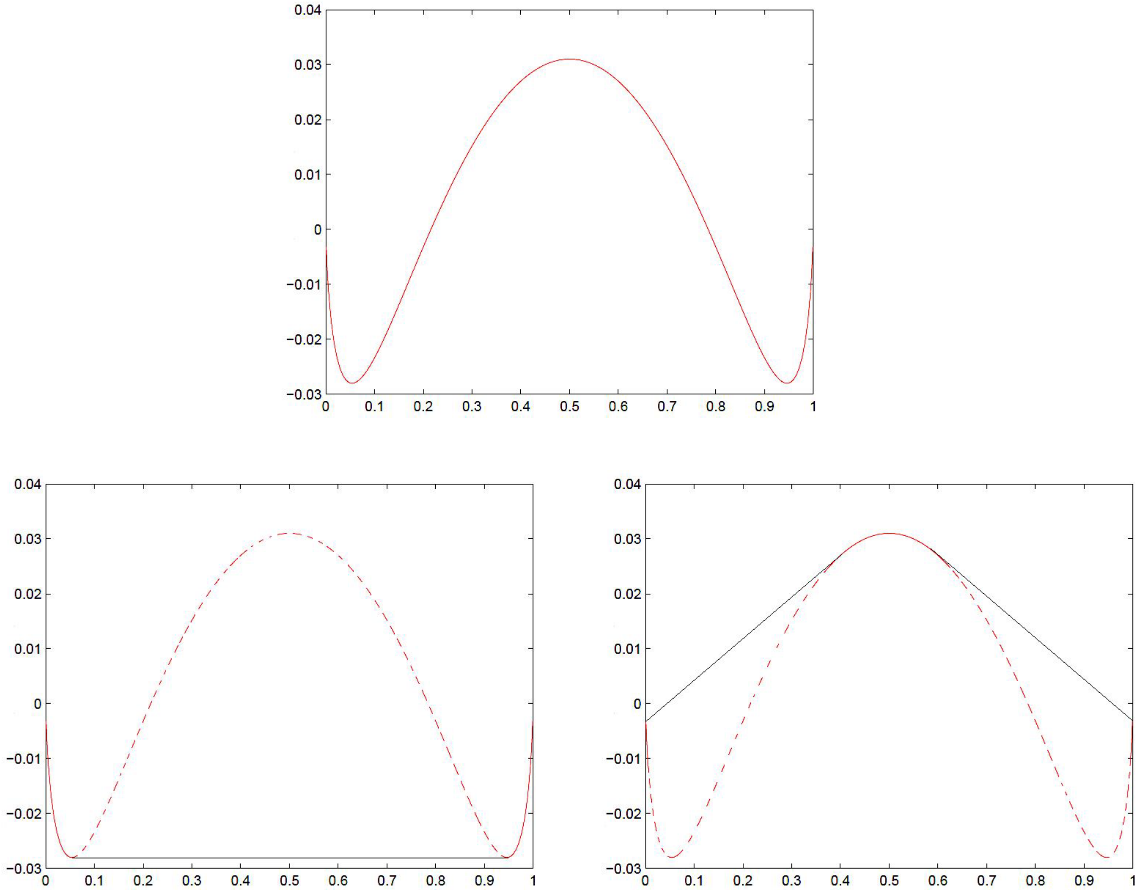
نمودارهای شکل ۶ پوش مقعر و محدب یک تابع را نمایش میدهد. جهت دستیابی به پوش محدب یک تابع باید خطوط مماس بر منحنی را یافت. این خطوط مماس با وصل کردن نقاطی از منحنی بدست می آیند. همانگونه که در نمودار شکل ۷ نشان داده شده جهت یافتن پوش مقعر در نقطه π کافی است که خط گذرا از نقاط π_1 و π_2 را در نظر بگیریم. اگر نقطه π را به دو نقطه π_1 و π_2 "پخش" کنیم، و ترکیب خطی مناسب از مقادیر تابع در نقاط π_1 و π_2 را حساب کنیم، میتوانیم به ارتفاع این خط مماس برسیم. بصورت مشخص اگر نقطه π ترکیب خطی محدب نقاط π_1 و π_2 با وزن های ω_1 و ω_2 باشد، $\pi = \omega_1\pi_1 + \omega_2\pi_2$ ، آنوقت با محاسبه ترکیب خطی میتوان به $\omega_1 f(\pi_1) + \omega_2 f(\pi_2)$ در نقطه π رسید. اما خطی که نقاط $(\pi_1, f(\pi_1))$ و $(\pi_2, f(\pi_2))$ را به هم وصل میکند دارای معادله خط

$$\frac{f(\pi_1) - f(\pi_2)}{\pi_1 - \pi_2}(x - \pi_1) + f(\pi_1)$$

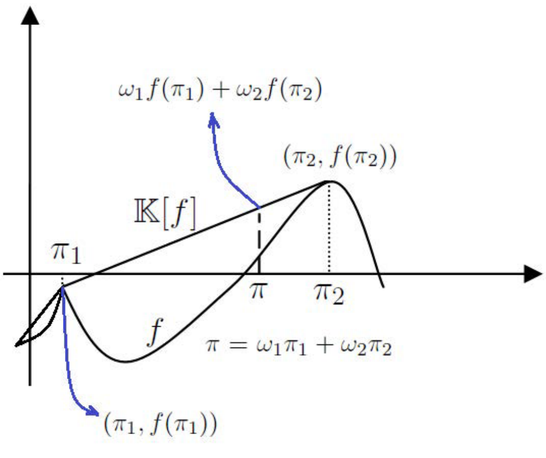
است که اگر مقدار آن را در $x = \pi = \omega_1\pi_1 + \omega_2\pi_2$ حساب کنیم به همان عبارت $\omega_1 f(\pi_1) + \omega_2 f(\pi_2)$ میرسیم:

$$\frac{f(\pi_1) - f(\pi_2)}{\pi_1 - \pi_2}(\omega_1\pi_1 + \omega_2\pi_2 - \pi_1) + f(\pi_1) = \omega_1 f(\pi_1) + \omega_2 f(\pi_2).$$

برای یافتن خط مماس میتوانیم تمامی حالات ممکن پخش شدگی را در نظر گرفته و از میان آنها بالاترین خط را بیابیم. این جستجو در قضیه زیر بشکلی ریاضی بیان شده است:



شکل ۶: (بالا) نمودار یک تابع. (پایین راست) پوش مقعر تابع. (پایین چپ) پوش محدب تابع.



شکل ۷: یک تابع و پوش مقعر آن. پوش مقعر متشکل از خطوط مماس و بخش های بیرون زده منحنی است.

قضیه ۶ روابط زیر در مورد پوش محدب و مقعر یک تابع f که روی سیمپلکس احتمالاتی تعریف شده برقرار است:

$$\mathbb{C}[f](p(x)) = \min_{p(y|x)} \sum_y p(Y=y) f\big(p(x|Y=y)\big).$$

$$\mathbb{K}[f](p(x)) = \max_{p(y|x)} \sum_y p(Y=y) f\big(p(x|Y=y)\big).$$

نکته ۷ این روابط نشان میدهند که پوش محدب با محاسبه میانگین مقادیر تابع برای پخش شدگی های مختلف بدست و انتخاب پخش شدگی ای که کمترین مقدار را بدست میدهد، قابل محاسبه است. و همینطور در مورد پوش مقعر.

اثبات: تنها رابطه را در مورد پوش محدب ثابت میکنیم. اثبات برای پوش مقعر مشابه است. کافی است که دو خاصیت پوش محدب را بررسی کنیم. اولا

$$\min_{p(y|x)} \sum_y p(Y=y) f\big(p(x|Y=y)\big) \leq f\big(p(x)\big)$$

زیرا در سمت چپ روی تمامی کانال ها مینیمم میگیریم و یک انتخاب $p(y|x)$ در سمت چپ کانال بدیهی است که هیچگونه پخش شدگی انجام نمیدهد: کانالی که در آن الفبای $\mathcal{Y} = \{0\}$ تک عضوی است. یا به عبارت دیگر

$$p_{Y|X}(0|x) = 1, \qquad \forall x.$$

جهت بررسی خاصیت دوم، فرض کنید که تابع محدب g نمودارش زیر نمودار f قرار بگیرد، یعنی $g \leq f$. در این صورت برای هر کانال دلخواه $q(y|x)$ داریم:

$$g\big(\pi_y\big) \leq f\big(\pi_y\big), \qquad \forall y,$$

و یا

$$g\big(p(x|Y=y)\big) \leq f\big(p(x|Y=y)\big), \qquad \forall y.$$

در نتیجه

$$\sum_y p(Y=y) g\big(p(x|Y=y)\big) \leq \sum_y p(Y=y) f\big(p(x|Y=y)\big)$$

اما چون تابع g محدب است داریم:

$$\sum_y p(Y=y) g\big(p(x|Y=y)\big) \geq g\big(\sum_y p(Y=y) p(x|Y=y)\big) = g\big(p(x)\big)$$

در نتیجه

$$g\big(p(x)\big) \leq \sum_y p(Y=y) f\big(p(x|Y=y)\big).$$

اما چون رابطه برای هر کانال دلخواه $q(y|x)$ برقرار است، داریم:

$$g\big(p(x)\big) \leq \min_{q(y|x)} \sum_y p(Y=y) f\big(p(x|Y=y)\big).$$

پس خاصیت دوم هم برقرار است. پس

$$\mathbb{C}[f](p(x)) = \min_{p(y|x)} \sum_y p(Y=y) f\big(p(x|Y=y)\big).$$

$\square$

مثال ۸ برای یک کانال ثابت $p(a|x)$ تابع

$$f\big(p(x)\big) = I(X;A)$$

را در نظر بگیرید. این تابع مقعر است چون اطلاعات متقابل برای یک کانال ثابت بر حسب توزیع ورودی مقعر است. پس باید با پوش مقعرش برابر باشد:

$$\begin{aligned}
\mathbb{C}[f](p(x)) &= \min_{p(y|x)} \sum_y p(Y=y) f\big(p(x|Y=y)\big) \\
&= \min_{p(y|x)} \sum_y p(Y=y) I(X;A|Y=y) \\
&= \min_{p(y|x)} I(X;A|Y).
\end{aligned}$$

اما برای هر کانال دلخواه $p(y|x)$ داریم:

$$I(X;A|Y) \leq I(XY;A) = I(X;A) + I(Y;A|X) = I(X;A).$$

پس

$$\begin{aligned}
\mathbb{C}[f](p(x)) &= \min_{p(y|x)} I(X;A|Y) \\
&= I(X;A) \\
&= f\big(p(x)\big).
\end{aligned}$$

نکته ۹ در صورتی که تابع f بر حسب جملات از نوع آنتروپی تعریف شده باشد، آنوقت عبارت

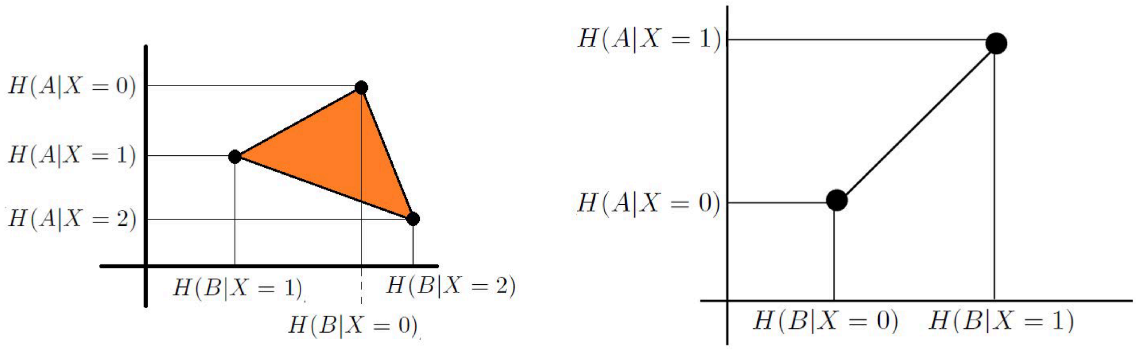
$$\sum_y p(Y=y) f\big(p(x|Y=y)\big)$$

معادل با مشروط کردن کل عبارت تابع f به Y است. به عنوان مثال اگر $f\big(p(x)\big) = H(X)$ آنوقت

$$\sum_y p(Y=y) f\big(p(x|Y=y)\big) = \sum_y p(Y=y) H(X|Y=y) = H(X|Y).$$

و یا یک مثال دیگر: در صورتی که کانال ثابت $p(a,b|x)$ را داشته باشیم میتوان تابع زیر را در نظر گرفت:

$$f(p(x)) = I(X;A) - I(X;B)$$



شکل ۸: تغییرات $[H(A|X), H(B|X)]$ بر حسب توزیع ورودی برای کانال با ورودی دودویی (شکل سمت راست) و ورودی سه تایی (شکل سمت چپ)

آنوقت

$$\begin{aligned}\sum_y p(Y=y) f\big(p(x|Y=y)\big) &= \sum_y p(Y=y) I(X;A|Y=y) - I(X;B|Y=y) \\ &= I(X;A|Y) - I(X;B|Y).\end{aligned}$$

این موضوع معنی ای برای عملگر مشروط کردن به دست میدهد.

از بحث بالا این نتیجه را میگیریم که برای محاسبه

$$\min_{p(y|x)} I(X;A|Y) - I(X;B|Y)$$

کافی است که تابع f را رسم کرده و پوش محدب آن را حساب کنیم. این کار از نظر محاسباتی به مراتب آسان تر از مینیمم گیری روی تمامی کانال ها میباشد. خصوصا اگر X دودویی باشد، فضای سیمپلکس احتمالاتی یک بعدی خواهد بود و محاسبه پوش محدب و مقعر کار آسانی خواهد بود.

۵ مشروط کردن و پوش محدب گیری

در بالا ارتباط میان مشروط کردن و پوش محدب گیری تا حدی مشخص شد. با یک مثال متفاوت این بحث را به پایان میرسانیم. فرض کنید که X متغیر تصادفی دودویی است. کانال ثابت $p(a,b|x)$ را در نظر بگیرید. در این صورت اگر p احتمال یک بودن X باشد، نقطه زیر در صفحه مختصات با تغییر p چگونه تغییر مینماید؟

$$\begin{pmatrix} H(A|X) \\ \\ H(B|X) \end{pmatrix}$$

توجه کنید که

$$\begin{pmatrix} H(A|X) \\ \\ H(B|X) \end{pmatrix} = \begin{pmatrix} H(A|X=0)(1-p) + H(A|X=1)p \\ \\ H(B|X=0)(1-p) + H(B|X=1)p \end{pmatrix}$$
$$= \begin{pmatrix} H(A|X=0) \\ \\ H(B|X=0) \end{pmatrix} (1-p) + \begin{pmatrix} H(A|X=1) \\ \\ H(B|X=1) \end{pmatrix} p$$

ملاحظه می شود با تغییر p از صفر تا یک مکان هندسی به یک خط منجر می گردد. این موضوع در شکل ۸ نشان داده شده است.

اما اگر X بجای دو مقدار سه مقدار میگرفت مکان هندسی چطور می بود؟ در این صورت مکان هندسی بجای یک خط یک مثلث می بود.

در حالت کلی جواب پوش محدب[۳] مجموعه نقاط زیر میباشد:

$$\begin{pmatrix} H(A|X=0) \\ \\ H(B|X=0) \end{pmatrix}, \begin{pmatrix} H(A|X=1) \\ \\ H(B|X=1) \end{pmatrix}, \begin{pmatrix} H(A|X=2) \\ \\ H(B|X=2) \end{pmatrix}, \cdots$$

میبینیم که مشروط کردن در واقع عملگری است که ترکیب خطی محدب را محاسبه میکند.

[۳] Convex Hull